

# Final Report to



---

## OPTIMIZING DEVELOPMENT STRATEGIES TO INCREASE RESERVES IN UNCONVENTIONAL GAS RESERVOIRS 07122-35.Final

---

September 30, 2011

Duane A. McVay  
Texas Engineering Experiment Station  
3000 TAMU, College Station, TX 77843-3000  
Phone: 979-862-8466  
[mcvay@pe.tamu.edu](mailto:mcvay@pe.tamu.edu)

J. Eric Bickel  
The University of Texas at Austin  
1 University Station, C2200, Austin, TX 78712-0292  
Phone: 512-232-8316  
[ebickel@mail.utexas.edu](mailto:ebickel@mail.utexas.edu)

## LEGAL NOTICE

This report was prepared by the Texas Engineering Experiment Station and the University of Texas at Austin as an account of work sponsored by the Research Partnership to Secure Energy for America, RPSEA. Neither RPSEA members of RPSEA, the National Energy Technology Laboratory, the U.S. Department of Energy, nor any person acting on behalf of any of the entities:

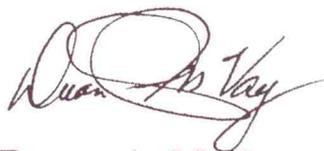
- a. MAKES ANY WARRANTY OR REPRESENTATION, EXPRESS OR IMPLIED WITH RESPECT TO ACCURACY, COMPLETENESS, OR USEFULNESS OF THE INFORMATION CONTAINED IN THIS DOCUMENT, OR THAT THE USE OF ANY INFORMATION, APPARATUS, METHOD, OR PROCESS DISCLOSED IN THIS DOCUMENT MAY NOT INFRINGE PRIVATELY OWNED RIGHTS, OR
- b. ASSUMES ANY LIABILITY WITH RESPECT TO THE USE OF, OR FOR ANY AND ALL DAMAGES RESULTING FROM THE USE OF, ANY INFORMATION, APPARATUS, METHOD, OR PROCESS DISCLOSED IN THIS DOCUMENT.

THIS IS A FINAL REPORT. THE DATA, CALCULATIONS, INFORMATION, CONCLUSIONS, AND/OR RECOMMENDATIONS REPORTED HEREIN ARE THE PROPERTY OF THE U.S. DEPARTMENT OF ENERGY.

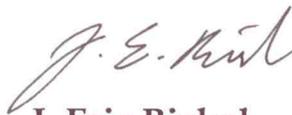
REFERENCE TO TRADE NAMES OR SPECIFIC COMMERCIAL PRODUCTS, COMMODITIES, OR SERVICES IN THIS REPORT DOES NOT REPRESENT OR CONSTITUTE AN ENDORSEMENT, RECOMMENDATION, OR FAVORING BY RPSEA OR ITS CONTRACTORS OF THE SPECIFIC COMMERCIAL PRODUCT, COMMODITY, OR SERVICE.

**This page intentionally left blank**

Submitted January 27, 2012, by

A handwritten signature in dark ink, appearing to read "Duane A. McVay". The signature is fluid and cursive, with the first name being the most prominent.

Duane A. McVay

A handwritten signature in dark ink, appearing to read "J. Eric Bickel". The signature is cursive and somewhat stylized, with the first name being the most prominent.

J. Eric Bickel

## Abstract

While the U.S. unconventional gas resource base is large, developing it in an economic and environmentally friendly way is challenging, due in part to the large uncertainty inherent in unconventional reservoirs. The objective of this research was to develop technology and tools to help operators determine optimal well spacing and completion strategies in highly uncertain and risky unconventional gas reservoirs as quickly as possible.

Probabilistic reservoir models integrated with Bayesian decision models were developed to optimize development strategies in unconventional reservoirs. The integrated reservoir and decision models were applied in two case studies, one in Deep Basin tight gas sands in the Berland River area, Alberta, and one in the northern Barnett shale of Montague, Cooke, and Wise counties, Texas.

In the first case study, probabilistic reservoir simulation models were configured to investigate the effect of downspacing in the Gething D formation, and to determine the optimal primary spacing and secondary downspacing for the reservoir. Two different reservoir simulation models were developed and integrated with Bayesian decision models – a single-well model and a multi-well model. Monte Carlo simulation was utilized to predict production profiles for a wide variety of reservoir properties and different development scenarios. Geostatistical techniques were used to model dependencies in reservoir properties and production responses among wells. Under illustrative economic assumptions, we determined with both models that the optimal primary spacing is 160 acres with no downspacing in the second stage.

In the second case study, we considered the northern Barnett shale formation located in the oil window within Montague, Cooke, and Wise counties. Probabilistic decline curves were used as the reservoir model. We correlated decline curve parameters with decision parameters and reservoir parameters using linear regression. We analyzed two different data sets of well and production data and found a clear correlation between decline curve parameters and perforated horizontal interval. Correlation of decline curve parameters to well spacing is less certain. While the correlations are low in these models, reflecting the large uncertainty inherent in shale reservoirs, we present an illustrative economic example to demonstrate how these integrated reservoir/decision models can aid operator decision making.

The integrated reservoir/decision models developed in this study can provide decision makers with an understanding of the uncertainty inherent in unconventional gas reservoir development and allow them to evaluate tradeoffs between these risks and other factors, such as increased capital spending, to optimize well spacing and completion strategies. The methodologies developed in this project can be applied to other unconventional reservoirs with minor changes.

## Table of Contents

A	Introduction.....	14
A.1	Statement and Significance of the Problem .....	14
A.2	Background and Existing Technologies/Methodologies.....	16
A.3	Objective of the project .....	18
B	Case study 1: Deep Basin TIGHT sands.....	19
B.1	Introduction of Deep Basin tight sands .....	19
B.2	Reservoir Model.....	21
B.2.1	Objective of Reservoir Model.....	21
B.2.2	Monte Carlo Simulation with Reservoir Simulation .....	21
B.2.3	Probabilistic Reservoir Model and Decision Context.....	26
B.2.4	Geostatistical Model .....	29
B.2.5	Reservoir Simulation Results.....	32
B.3	Decision Model .....	36
B.3.1	Overview.....	36
B.3.2	Illustrative Example .....	38
B.4	Multi-Well Reservoir Simulation Model .....	41
B.4.1	Overview.....	41
B.4.2	Geostatistics Model.....	41
B.4.3	Comparison with Single Well Model .....	54
B.4.4	Results.....	58
B.5	Conclusions .....	61
C	Case Study 2: Northern Barnett Shale play .....	62
C.1	Introduction .....	62
C.2	Decline Curve Based Reservoir Model.....	64
C.2.1	Decline Curve Parameters.....	64
C.2.2	Decision Parameters.....	66
C.2.3	Linear Regression Models .....	67
C.2.4	GOR Model.....	71
C.2.5	Monte Carlo Simulation of Decline Curve Model.....	74
C.3	Decision Analysis.....	76
C.4	Updated Model Using a Larger Data Set .....	78
C.4.1	Introduction.....	78
C.4.2	Decline Curve Based Reservoir Model.....	78
C.4.3	GOR Model.....	83
C.4.4	Illustrative Economic Example Using the Second Regression Model .....	85
C.4.5	Discussion.....	99
C.5	Conclusions .....	101
D	Impact to Producers .....	101
E	Technology Transfer Efforts.....	103
F	Conclusions.....	104
G	Recommendations.....	104
H	References.....	105
I	Bibliography of technology transfer efforts.....	109
J	List of Acronyms and Abbreviations .....	110

K	Appendix A: SAS output .....	114
L	AppendIX B: Discretization and simulation.....	115
M	AppendIX C: Discretization and the value of information.....	151

## List of Figures

Fig. 1—Location map showing Deep Basin of Alberta (Smith et al. 1984).....	20
Fig. 2—Stratigraphic column of the Lower Cretaceous in the Deep Basin area (modified from Hietala and Connolly 1984).....	20
Fig. 3— Schematic illustration of the integrated reservoir and decision modeling tools. ....	21
Fig. 4— Block diagram representation of the applications and tools used in this study. ....	23
Fig. 5 – Histograms and probability distribution functions of uncertain input variables used in the modeling of the Deep Basin (Gething D) reservoir: (a) net pay, (b) porosity, (c) formation depth, (d) reservoir size. ....	23
Fig. 6 – (a) Probability distribution plot (PDF), (b) cumulative distribution plot (CDF) of permeability. ....	24
Fig. 7 – Porosity-permeability cloud. ....	25
Fig. 8 – Gridblocks model based on an aspect ratio of 7: (a) x-direction centroids, (b) y-direction centroids. ....	26
Fig. 9 – Schematic illustration showing (a) possible well locations, (b) representative well locations.....	27
Fig. 10 – Possible two-stage downspacing combinations. Numbers on top of arrows indicate Stage 1 wells and additional Stage 2 wells. Wells 1 through 4 are the representative wells. ....	28
Fig. 11 – Porosity map of the Deep Basin (Gething D) reservoir (provided by UGR).....	30
Fig. 12 – Variogram analysis for porosity: (a) Variogram map (distances are in meters), (b) major and (c) minor direction empirical variograms, (d) major and (e) minor direction theoretical variograms with covariance and correlation functions (data from Deep Basin (Gething D) provided by UGR). ....	31
Fig. 13 – Crossplots of natural log of individual-well, discounted stage production (MMscf) downspacing from 640 to 80 ac after a 1-yr Stage 1, (a) Stage 2 vs Stage 1 of the original well, and (b) the infill well in Stage 2 vs the original well in Stage 1. ....	34
Fig. 14 – Schematic decision tree for optimal spacing decision.....	37
Fig. 15 – Optimal policy and ENPVs for our illustrative example with 1-year Stage 1. Monetary values in \$MM/section.....	40
Fig. 16 – Optimal policy and ENPVs for our illustrative example with 3-year and 5-year Stage 1. Monetary values in \$MM/section.....	40
Fig. 17 – Variogram map for gas porosity. The white line shows the major direction of anisotropy. ....	43
Fig. 18 – Variogram map for net-to-gross ratio. The white line shows the major direction of anisotropy.....	44
Fig. 19 – Variogram map for permeability. The white line shows the major direction of anisotropy. ....	45
Fig. 20 – An exponential variogram model for gas porosity. a) is the variogram in the major direction of anisotropy and b) is the variogram in the minor direction.....	46
Fig. 21 – A Gaussian variogram model for NTG. a) is the variogram in the major direction of anisotropy and b) is the variogram in the minor direction.....	47
Fig. 22 – A spherical variogram model for permeability. a) is the variogram in the major direction of anisotropy and b) is the variogram in the minor direction.....	48
Fig. 23 – An example of a log including facies for a well of the area under study. The facies log is presented on Track 4. Sand facies are represented by the dark yellow color and flood plain deposits by the brown color. Depths are in meters. ....	51
Fig. 24 – A realization of porosity for the Gething D formation obtained through geostatistical procedures. The boxes in light purple represent UGR’s lease areas. The shadowed gray box represents the section selected to extract the property maps to be used in the multi-well reservoir model for this project.....	52
Fig. 25 – A realization of net-to-gross for the Gething D formation obtained through geostatistical procedures. The boxes in light purple represent UGR’s lease areas. The shadowed gray box represents the section selected to extract the property maps to be used in the multi-well reservoir model for this project.....	52
Fig. 26 – A realization of permeability for the Gething D formation obtained through geostatistical procedures. The boxes in light purple represent UGR’s lease areas. The shadowed gray box represents the section selected to extract the property maps to be used in the multi-well reservoir model for this project.....	53
Fig. 27 – An example realization of reservoir property maps for the section of the Gething D interval, Berland River Area, showing the properties in individual layers. ....	53
Fig. 28 – Simulation grid of the study area showing a top (a) and bottom (b) view of the three-dimensional cube of net-to-gross ratio (NTG).....	54

Fig. 29 — a) The multi-well reservoir model yielded a distribution of production that matches the actual distribution of Gething production and single-well reservoir model results quite well, b) revised production distribution for multi-well reservoir model where the lowest 10% of production values was removed from Stage-1 results. A better match with the actual distribution of Gething production and single-well reservoir model is observed. .... 56

Fig. 30 — a) The multi-well reservoir model yielded a distribution of best-month gas production that matches the actual distribution of Gething production. b) Lowest 10% of production values removed from Stage 1. A better match with the actual distribution of Gething production is observed. .... 57

Fig. 31 – Barnett shale thermal maturity map. Northern Barnett shale study area (yellow) is located in the oil window. (Modified from Montgomery et al., 2005)..... 63

Fig. 32 – Overview of monthly production history and drilling activity in the northern Barnett study area shows continuous increase of oil and gas production and drilling activity through the end of 2010..... 64

Fig. 33 – Decline curve fit historical production from 2<sup>nd</sup> month (peak oil in first year) to 43<sup>rd</sup> month (end of historical production). .... 65

Fig. 34 – Definition and example of shorter well spacing (WS1), longer well spacing (WS2), and average well spacing (WSA)..... 67

Fig. 35 – Initial production rate is positively related to length of perforated interval with  $R^2 = 0.2881$ . .... 70

Fig. 36 – CP6to1 is positively related to shorter well spacing (WS1); average CP6to1 is 3.22 when shorter well spacing is greater than 2000 ft..... 70

Fig. 37 – CP6to1 is positively related to fluid per perforated interval with  $R^2 = 0.1895$ . .... 71

Fig. 38 – GOR type curve fitted to average GOR of 64 wells ..... 72

Fig. 39 – Initial oil percentage is negatively related to vitrinite reflectance ( $R_o$ ). .... 73

Fig. 40 – Gas production forecast generated by multiplying oil production forecast and GOR curve..... 73

Fig. 41 – Tornado diagrams for cumulative oil and gas production at 25 years. .... 75

Fig. 42 – Decline curve regression model fits actual study-area oil and gas production profiles well. .... 76

Fig. 43 –  $\ln(q_i)$  is positively related to perforated interval (2<sup>nd</sup> data set) ..... 79

Fig. 44 – CP6to1 vs. perforated interval. Seven outliers for wells with less than 1500-ft PI are observed. .... 80

Fig. 45 – CP6to1 is negatively correlated to perforated interval, excluding 7 outliers. .... 81

Fig. 46 – CP6to1\_ex peaks around 340 tighter well spacing ..... 81

Fig. 47 – Comparison between first and second data sets in CP6to1 vs. tighter well spacing ..... 83

Fig. 48 – The average GOR was fit by a type curve. .... 84

Fig. 49 – The peak three months’ oil percentage were correlated with vitrinite reflectance ( $R_o$ ). .... 84

Fig. 50 Combine oil production forecast with GOR curve to generate gas production forecast ..... 85

Fig. 51 – Occupied area by a single well in normalized area ..... 88

Fig. 52 – Well count per 640 acres vs. perforated interval..... 89

Fig. 53 – NPV(excluding capital) vs. perforated interval. Normalized to 640 acres. .... 89

Fig. 54 – NPV/Capital vs. perforated interval. Normalized to 640 acres. .... 90

Fig. 55 – NPV/Capital vs. perforated interval (mean and P50). Normalized to 640 acres..... 90

Fig. 56 – NPV-Capital vs. perforated interval. Normalized to 640 acres..... 91

Fig. 57 – NPV-Capital vs. perforated interval (mean and P50). Normalized to 640 acres. .... 91

Fig. 58 – Well count per 640 acres vs. tighter well spacing..... 92

Fig. 59 –NPV(excluding capital) vs. tighter well spacing. Normalized to 640 acres. .... 93

Fig. 60 – NPV/Capital vs. tighter well spacing. Normalized to 640 acres..... 93

Fig. 61 – NPV/Capital vs. tighter well spacing (mean and P50). Normalized to 640 acres..... 94

Fig. 62 – NPV-Capital vs. tighter well spacing. Normalized to 640 acres. .... 94

Fig. 63 – NPV-Capital vs. tighter well spacing (mean and P50). Normalized to 640 acres. .... 95

Fig. 64 – NPV (excluding capital) per section vs. perforated interval and tighter well spacing. .... 96

Fig. 65 –Oil production per section vs. perforated interval and tighter well spacing. .... 96

Fig. 66 –NPV/Capital per section vs. perforated interval and tighter well spacing. .... 97

Fig. 67 –NPV-Capital per section vs. perforated interval and tighter well spacing. .... 97

Fig. 68 – Comparison between 1<sup>st</sup> and 2<sup>nd</sup> data sets vs. perforated interval. Values exclude capital costs and are normalized to 3500-ft x 1000-ft area. .... 98

Fig. 69 – Comparison between 1 <sup>st</sup> and 2 <sup>nd</sup> data sets vs. tighter well spacing. Values exclude capital costs and are normalized to 3500-ft x 1000-ft area. ....	99
Fig. 70 – Probabilistic single-well production forecasts for 5500 ft perforated interval and 340 ft well spacing. ....	100
Figure B1– Gaussian quadrature for the normal distribution. ....	149
Figure B2 – Illustration of Equal Areas discretization method. ....	149
Figure B3: Error in Swanson’s Mean with direct application to lognormal distribution. Compare to Megill (1984, Fig. B.1).....	1500
Figure B4: Example of decision situation that may not lend itself well to Monte Carlo methods. ....	1500
Figure C1: Decision tree used to value information under three-point discrete approximation. ....	1566
Figure C2: Comparison of approximate EVOIs to exact EVOI ( $\mu = \$10$ million and $\sigma = \$20$ million).....	1577
Figure C3: Ratio of approximate to exact EVOI ( $c = 0.5$ ). ....	1588
Figure C4: Sensitivity of the ratio of the approximate to exact EVOI to $c$ . $c = 0.25$ (panel a) and $c = 0.75$ (panel b). ....	15959
Figure C5: Equivalent number of Monte Carlo samples ( $c = 0.5$ , $\tau = 0.95$ ). Values above 100,000 are not shown. ....	1622
Figure C6: Sensitivity of the equivalent number of Monte Carlo samples ( $\tau = 0.95$ ) to $c$ . $c = 0.25$ (panel a) and $c = 0.75$ (panel b). Values above 100,000 are not shown.....	1633

## List of Tables

Table 1 — Uncertain input variables involved in the modeling of Deep Basin (Gething D) reservoir (data from UGR) .....	24
Table 2 — Input parameters employed in the Deep Basin (Gething D) reservoir model (data from UGR) .....	26
Table 3 — Possible downspacing combinations evaluated in the final reservoir model .....	27
Table 4 — Correlation matrix for porosity.....	32
Table 5 — Correlation matrix for net pay.....	32
Table 6 — Correlation matrix for permeability .....	32
Table 7—Statistics for natural log of individual-well, discounted stage production (MMscf) for original and infill wells for a 1-year Stage 1.....	35
Table 8 — Statistics for natural log of individual-well, discounted stage production (MMscf) for original and infill wells for a 3-year Stage 1.....	35
Table 9 — Statistics for natural log of individual-well, discounted stage production (MMscf) for original and infill wells for a 5-year Stage 1.....	36
Table 10 — Theoretical variogram parameters. ....	42
Table 11 — Optimal development strategy results for Stage-1 length of 1 year. Monetary values are per section. ....	59
Table 12 — Optimal development strategy results for Stage-1 length of 3 years. Monetary values are per section. ....	60
Table 13 — Optimal development strategy results for Stage-1 length of 5 year. Monetary values are per section. ....	61
Table 14 – $R^2$ Between Decline Curve Parameters and Decision/Reservoir Parameters .....	65
Table 15 – Reservoir Model Input Uncertainty.....	74
Table 16 – Cumulative Oil and Gas Uncertainty Range.....	76
Table 17 – Impact of perforated interval on oil and gas production at 25 years.....	78
Table 18 – Impact of well spacing on oil and gas production at 25 years.....	78
Table 19 – Impact of stimulation fluid volume on oil and gas production at 25 years .....	78
Table 20 – CP6to1 and CP6to1_ex average used for Fig. 47.....	82
Table 21 – Overall economics and capital cost per well .....	85
Table 22 – Oil, gas and condensate prices.....	86
Table 23 – Monthly OPEX per well .....	86
Table 24 – Single-well results using 5500-ft perforated interval and 340-ft well spacing. ....	100
Table 25 – Results normalized to 640 acres using 5500-ft perforated interval and 340-ft well spacing. ....	100

Table 26 – Range of reservoir models developed .....	102
Table B1: Two-, Three-, and Four-Point Gaussian Quadrature Formulae for Common Distributions .....	145
Table B2: Three-Point Discretization Shortcuts .....	145
Table B3: Three-Point Moment Matching Weights for P90, P50, P10 Values .....	145
Table B4: Three-Point Moment Matching Values for 25-50-25 Weights (same pdfs as Table 3) .....	146
Table B5: Three-Point Moment Matching Values for 30-40-30 Weights (same pdfs as Table 3) .....	146
Table B6: Representative Four-Point Discretizations for Lognormal .....	146
Table B7: 95% S-Equivalences for Discretization Shortcuts .....	146
Table B8: Approximation Error in Mean for Illustrative Example .....	148
Table B9: Approximation Error in Variance for Illustrative Example .....	148
Table C1: Three-Point Discretization Shortcuts .....	153
Table C2: 95% S-Equivalences for the Illustrative VOI Example ( $c = 0.5$ and $\rho = 0.6$ ).....	161

## Executive Summary

Unconventional's share of U.S. oil and gas production has increased dramatically in the last 15 years. While the resource base is large, 847 Tcf technically recoverable according to the Annual Energy Outlook 2011 (EIA 2011), developing it in an economic and environmentally friendly way is challenging, due in part to the large uncertainty inherent in unconventional reservoirs. The objective of this research is to develop technology and tools to help operators determine optimal well spacing and completion strategies in highly uncertain and risky unconventional gas reservoirs as quickly as possible.

Probabilistic reservoir models integrated with Bayesian decision models were developed to optimize development strategies in unconventional reservoirs. The integrated reservoir and decision models were applied in two case studies, one in Deep Basin tight gas sands in the Berland River area, Alberta, and one in the northern Barnett shale of Montague, Cooke, and Wise counties, Texas.

In the first case study, we considered the Gething D formation in the Deep Basin tight gas sands. The reservoir was developed with vertical wells with well spacing ranging from 40 to 640 acres. Probabilistic reservoir simulation models were configured to investigate the effect of downspacing in the Gething D formation, and to determine the optimal primary spacing and secondary downspacing for the reservoir. Two different reservoir simulation models were developed and integrated with decision models – a single-well model and a multi-well model.

The first model was a single-well, single-layer, single-phase reservoir simulation model. Monte Carlo simulation with a 1000 realizations was utilized to predict production profiles for a wide variety of reservoir properties and different development scenarios. Geostatistical techniques were used to model dependencies in reservoir properties and production responses among wells. Under illustrative economic assumptions, we determined the optimal primary spacing is 160 acres with no downspacing in the second stage.

The second model built was a multi-well reservoir simulation model, which better modeled correlation and interference of production between wells. Sequential Gaussian Simulation (SGS) was used to generate 1,000 reservoir property maps. Those 1,000 reservoir property maps were then included in reservoir simulation models to generate probabilistic production forecasts. The multi-well generated a better history match of actual production than the single-well model, while the optimal strategy remained the same as determined by the single-well model.

In the second case study, we considered the northern Barnett shale formation was located in the oil window within Montague, Cooke, and Wise counties. The formation was developed using horizontal wells with multiple hydraulic fractures.

Probabilistic decline curves were used as the reservoir model. We correlated decline curve parameters (logarithm of initial oil rate,  $\ln(q_i)$ , and the ratio of cumulative oil production at 6 months to 1 month, CP6to1) with decision parameters (primarily perforated interval and well spacing) and reservoir parameters using linear regression techniques. GOR for individual wells was modeled by combining a GOR type curve based on historical data with a linear regression model of initial GOR to thermal maturity. The oil production forecasts were multiplied by GOR to generate the gas production forecast.

Two different data sets of well and production data were used for this integrated reservoir/decision modeling case study. In the first data set, 42 horizontal wells or leases with well spacing data available and at least 6 months production through December 2010 were analyzed. In the regression models,  $\ln(q_i)$  is positively correlated with perforated interval with  $R^2$  equal to 29% and CP6to1 is positively related with well spacing and fracture fluid volume per perforated interval with  $R^2$  equal to 32%. Individual-well production increases as perforated interval and well spacing increase, while production normalized to a common area decreases as perforated interval and well spacing increase.

In the second model, 84 wells with well spacing data available and at least 6 months production through January 2011 were analyzed. In the regression models,  $\ln(q_i)$  is positively correlated with perforated interval with  $R^2$  equal to 21%; CP6to1 is negatively correlated with perforated interval and is correlated in a stepwise fashion with well spacing ( $R^2$  between CP6to1 and the two decision parameters totals 20%). Under illustrative economic assumptions, maximum average NPV/Capital for a single well equals 1.42, with P90, P50, and P10 values of 0.22, 0.87 and 3.07, respectively, at well spacing of approximately 340 ft and perforated interval of approximately 5500 ft.

The different integrated reservoir and decision modeling systems developed in this project can help operators in unconventional gas reservoirs increase reserves and accelerate production, while protecting the environment, by determining the optimal well spacing and completion strategy as early in reservoir life as possible. These models provide decision makers with an understanding of the uncertainty inherent in unconventional gas reservoir development and allow them to evaluate tradeoffs between these risks and other factors such as increased capital spending. While specific strategies were determined for the two case studies presented, these results are specific to the reservoir properties and economic assumptions used in these applications and cannot be generalized. The methodologies developed in this project can be applied to other unconventional reservoirs with minor changes.

## A INTRODUCTION

### A.1 Statement and Significance of the Problem

According to the National Petroleum Council, North America has over 5000 trillion cubic feet (TCF) of natural gas resources in shale or tight sand formations. While this resource base is large, developing it in an economic and environmentally sensitive manner is challenging. For example, in the lower 48 United States, around 847 TCF of the shale gas and tight gas is estimated to be recoverable using existing technology. In specific development areas such as the Barnett Shale, current recovery per well averages just 7% of gas in place--far below the 20% that many believe is achievable. In addition to low recovery rates, operators in unconventional reservoirs must invest large amounts of capital and face significant risks.

Reliance on unconventional gas is projected to increase with gas shales, tight gas sands, and coalbed methane playing a significant role. Despite our increased experience, unconventional plays remain risky. While geologic risk and volumetric uncertainty are less important than in conventional plays, complex heterogeneities and reservoir properties relating to producibility are key. In addition, drivers of profitability such as completion and stimulation efficiency are highly uncertain. In the face of this risk, operators must make sound development decisions such as optimal well spacings and completion practices. These decisions must balance the need to conserve capital and protect the environment by avoiding over drilling, with the objective of profitably maximizing production by achieving the optimal well spacing as quickly as possible. Previous unconventional gas developments, such as the Carthage Field (Cotton Valley), have taken decades to optimize well spacing (McKinney et al. 2002). In cases like this it was estimated, based on a hypothetical 500-well tight gas field, that suboptimal development could reduce the value of the unconventional asset by 50%, on the order of \$100 million. This value loss is likely to be even greater in the current supply environment, which is likely to persist for the next several decades. Therefore, we must create more rapid development plans. For this reason, the National Petroleum Council recently recommended a "major increase in research funding" to "optimize drilling and completion methods" via integrated reservoir

characterization of geologic, seismic, petrophysical, and engineering data (NPC, 2007, pp. 196-7).

Historically, optimal well spacing has been determined by comparing the performance of wells drilled at different spacings using some form of statistical comparison. When data from multiple infill programs spanning 10-40 years are available, the risk of over drilling is fairly low. However, in emerging plays such as the non-core Barnett Shale, the Fayetteville Shale, and many emerging tight gas sand plays, historical infill programs are not available to evaluate optimal spacing with traditional methods and we do not have the luxury of developing these fields over the next 40 years.

Determining optimal completion and stimulation strategies can also be quite difficult, due both to technical challenges and to the difficulty of evaluating effectiveness of different completion and stimulation practices in these complex, highly heterogeneous reservoirs. As an example, stimulation treatments in the Barnett Shale have evolved from massive hydraulic fracture treatments in the 1980's and 1990's to the light-sand waterfracs being used today (Coulter et al. 2004). Reeves et al. (2000) illustrate the difficulty of determining stimulation effectiveness in unconventional tight gas sands.

In summary, the nation must speed development of its tight sand and shale gas reservoirs. Historic development patterns are too slow to meet future demand and will further increase our reliance on non-domestic sources of gas—particularly gas from outside of North America. Operators would like to reach the optimal spacing and determine the optimal completion method as quickly as possible. However, the significant uncertainty present in unconventional reservoirs makes rapid development risky. Without a way to effectively manage this uncertainty, operators may under invest in the development of their assets and/or develop them more slowly.

## A.2 Background and Existing Technologies/Methodologies

### Reservoir Modeling

The most accurate way to optimize development of unconventional gas reservoirs is to conduct a complete integrated reservoir evaluation, including geological, geophysical, petrophysical, and reservoir engineering analyses and interpretations. This method includes developing a geological model of the study reservoir, estimating distributions of static reservoir properties such as porosity and permeability, constructing and calibrating a reservoir simulation model of the reservoir and wellbores, and then using the reservoir model to predict and optimize development of the reservoir. This is the method advocated by Anadarko (Newsham and Rushing 2001; Rushing and Newsham 2001). While it may be accurate, this approach can be prohibitively time-consuming and expensive. Given the marginal nature of unconventional gas resources, this approach is not often employed, particularly by small- to medium-sized operators with lean staffs.

Various authors have proposed alternate approaches. For example, Guan et al. (2002) described a new technique for large-scale infill evaluations in tight gas reservoirs. The technique, referred to as Mosaic or moving window technology, was an extension of the method described by Voneiff and Cipolla (1996). The method consists of a multitude of local analyses, each in an areal window centered around an existing well. A model-based linear 4D regression equation is applied within each window:

$$BY = f(VBY, G_p/A, A)$$

where

$BY$  = best year, the best 12 consecutive months of production divided by 12.  $BY$  has been demonstrated to correlate well with long-term production and is used as a proxy for ultimate recovery.

$VBY$  = virgin best year, the  $BY$  of a well at virgin conditions. Depletion effects are removed by computing the  $BY$  of a local area at a time before depletion using a 2D regression of  $BY$  versus well start date.  $VBY$  is used as a proxy for  $kh$ , or reservoir quality.

$G_p/A$  = cumulative production divided by well spacing.

$A =$  well spacing, area of Voronoi polygon around each well based on well locations.

This is used as a proxy for drainage area.

Once the regression equation coefficients are determined for each window, performance can be estimated for infill wells by substituting the appropriate values for candidate infill well conditions (well spacing,  $G_p/A$ ,  $VBY$ ). The primary advantages of the moving window technique are its speed and its reliance upon only well location and production data. It is used to conduct infill screening studies of projects consisting of 1000's of wells and can be used to evaluate an entire basin in a few man-days.

Others have used simulation-based regression and automated prediction methods for selecting infill candidates (Chenget al. 2006; Cheng et al. 2006; Gao and McVay 2004). A reservoir simulator serves as the forward model, which calculates well production responses from reservoir description data. Sensitivity coefficients are calculated internally and used in the inversion of historical production data to estimate the porosity and permeability fields. Using the estimated property fields and forward model, expected performance of potential infill wells are determined using automated prediction methods. Since the method is simulation-based, all the data required to initialize a reservoir simulator are required for application. However, since the goal is rapid, approximate estimation of infill potential, the authors do not conduct a detailed reservoir characterization study. Instead, in an initial application, they simply use average properties from whatever data are available. The result is an approximate, but very fast, method for assessing infill performance in unconventional gas reservoirs. Gao and McVay (2004) demonstrated that this method is more accurate than moving window methods, with comparable data requirements and computation time.

Teufel et al. (2004), in a US DOE-sponsored project, developed an approximate tool called the Infill Well Locator Calculator (IWLC) to assist operators in the drilling of infill wells in low-permeability gas reservoirs. This tool was applied to evaluate infill potential in the Mesaverde Group in San Juan Basin. The tool is versatile, but only provides a qualitative evaluation of infill performance since it ignores heterogeneity within the test area. In addition, it provides only a deterministic assessment and does not quantitatively consider the large uncertainty inherent in the assessment.

## Decision Modeling

Decision analysis (DA) and uncertainty quantification (UQ) methodologies were introduced to the oil and gas industry almost 50 years ago (Grayson 1960). Most large companies (the Majors) have decision and risk analysis groups that analyze the risk and profitability of major assets. However, mid-size and small operators still rely primarily on deterministic estimates (Bickel and Bratvold 2007). While the Majors have applied DA/UQ extensively, the vast majority of these applications have been in the context of conventional hydrocarbon accumulations. Unconventional resources pose special challenges (Haskett and Brown 2005; Stabell 2005). For example, volumetric uncertainty is the primary risk in conventional plays, while risks related to producibility are key in unconventional reservoirs. In addition to the lack of experience in unconventional plays, the industry as a whole does not rigorously estimate the value of its information gathering and pilot programs (Bratvold et al. 2007). The optimal design of such programs in unconventional reservoirs, where the risks are high, are likely to pay large dividends. For example, optimal exploration programs can turn uneconomic reservoirs into profitable plays (Bickel and Smith 2006; Bickel et al. 2011).

### **A.3 Objective of the project**

A major long-term goal of our research program is the development of technology and tools to optimize exploitation of unconventional resources in the U.S. and worldwide for the public benefit. Given the considerable uncertainty associated with these resources, a second long-term goal is to develop and disseminate the tools and methodologies that enable oil and gas operators to make the best decisions in the face of significant uncertainty and risk. Previous work in this area has focused on the optimal use of 3D and 4D seismic (Bickel et al. 2006; Gibson et al. 2007), the development of optimal sequential exploration programs (Bickel and Smith 2006; Bickel et al. 2011), and probabilistic methods for assessing unconventional reserves and potential resources (Cheng et al. 2007; Salazar et al. 2007).

The specific objectives of this project were to (1) develop integrated reservoir and decision models to help operators in unconventional gas reservoirs increase reserves and accelerate production, while protecting the environment, by determining the optimal well

spacing and completion strategy as quickly as possible, and (2) apply these integrated reservoir and decision models to two unconventional plays, Deep Basin tight sands and Northern Barnett shale.

## **B CASE STUDY 1: DEEP BASIN TIGHT SANDS**

### **B.1 Introduction of Deep Basin tight sands**

The Deep Basin is located in western Alberta, Canada (**Fig. 1**). Discovered in 1976, the Deep Basin is considered as one of North America's giant gas fields, with gas reserves ranging between 50 to 150 Tcf (Masters 1979). **Fig. 2** depicts the stratigraphic column of the lower Cretaceous in the Deep Basin area. The depositional environments in the red highlighted interval are: (1) Cadomin - alluvial fan and plain, (2) Gething - fluvial deposits and flood plain, (3) Bluesky - shoreface sand and shales (Smith et al. 1984). The formation of interest in this study is the Gething.

The Gething formation consists primarily of interbedded very-fine- to medium-grained sandstones, siltstones, mudstones and coal sediments. The sequence is terrestrial and is described as a low-relief interior drainage plain. Sandstones are fining upward or thin-bedded. Trough and planar cross-bedded, ripple-bedded and parallel-laminated sandstones are common. The representative drainage pattern of the Gething sandstones is a series of active channel systems that generally trended northeast at the time of deposition. The implication of this depositional environment is that Gething reservoirs often consist of layered, discontinuous, heterogeneous sands that are challenging to describe and develop.

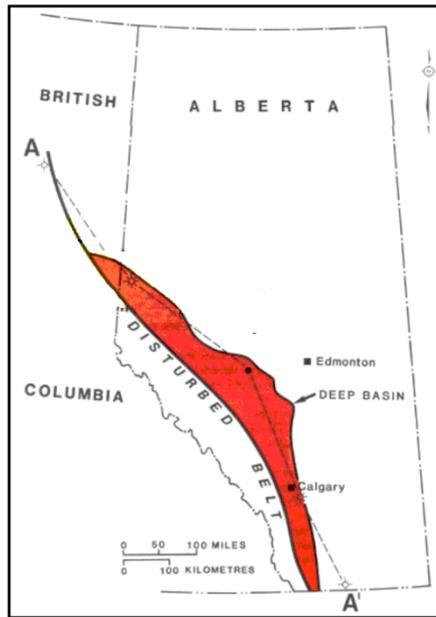


Fig. 1—Location map showing Deep Basin of Alberta (Smith et al. 1984).

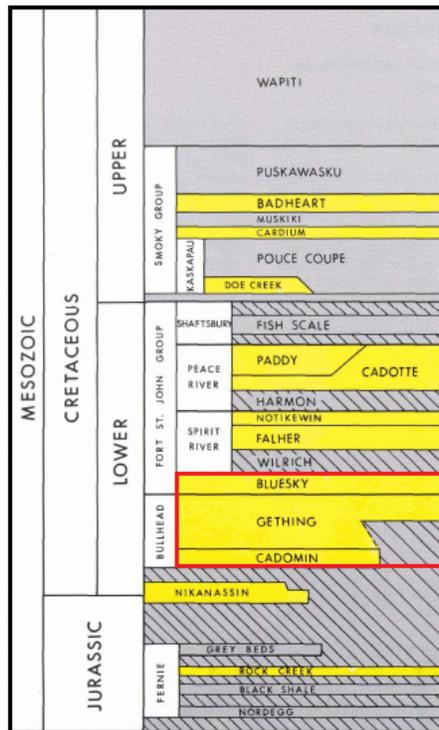
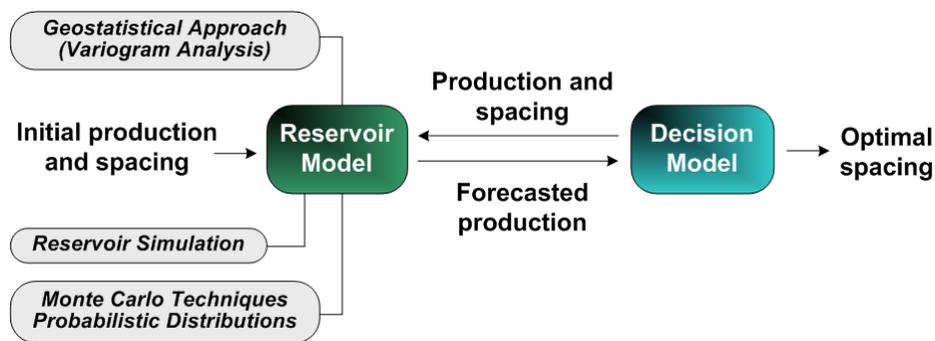


Fig. 2—Stratigraphic column of the Lower Cretaceous in the Deep Basin area (modified from Hietala and Connolly 1984).

## B.2 Reservoir Model

### B.2.1 Objective of Reservoir Model

The core of the technology developed in this work is a reservoir model that involves explicit modeling of subsurface uncertainty and a decision model that incorporates uncertainty to optimally manage risk (**Fig. 3**). The reservoir model uses Monte Carlo methods and single-well reservoir simulation to predict production profiles for a wide variety of reservoir properties and different development scenarios. Geostatistical techniques are used to model dependencies in reservoir properties and production responses among wells. To integrate all the different techniques used in constructing the reservoir model and automate the performing of thousands of simulations, Microsoft Visual Basic® for Applications (VBA) code was developed. In this way, the reservoir model can provide a full spectrum of production profiles under different downspacing scenarios.



**Fig. 3— Schematic illustration of the integrated reservoir and decision modeling tools.**

### B.2.2 Monte Carlo Simulation with Reservoir Simulation

Most tight gas reservoirs are characterized by complex geological, geophysical and petrophysical properties and a high level of heterogeneity and, hence, involve considerable uncertainty. Because of the large number of unknowns in the spatial distribution of reservoir properties, a deterministic approach remains inadequate to quantify the range of variability and assess the risk in outcomes such as cumulative gas production or net present value. Consequently, to model production uncertainties and quantify the risk involved in the development decisions, a stochastic modeling tool, @RISK from the Palisade Corporation, was coupled to the reservoir simulator IMEX from Computer Modeling Group. This allowed the

generation of thousands of simulations to evaluate combinations of unknown parameters within their ranges of uncertainty.

A single-well, one-layer, single-phase reservoir simulation model was built for modeling the Deep Basin (Gething) reservoir. Although the Gething reservoir is a commingled, multilayer system, it is dominated by the Gething D interval when present. We modeled only the Gething D sand because it can be represented by a single simulation layer for simplicity and, in particular, for speed, since the simulations are done in a Monte Carlo context. The reservoir grid was constructed to model a hydraulically fractured well located centrally in a rectangular drainage area.

A VBA code was generated in Excel® to perform thousands of simulations automatically. An IMEX input file template was created in Excel and uncertain parameters were defined by @RISK distribution functions. **Fig. 4** is a block diagram representation of the applications and tools used in the construction of the Gething reservoir model. The uncertain parameters and their associated distributions used in modeling the Deep Basin (Gething D) reservoir are presented in **Table 1**. Histograms and cumulative probability distribution functions of these uncertain input variables are illustrated in **Fig. 5**. In addition to the parameters defined in Table 1, the average reservoir pressure is also uncertain since it is computed in the model by multiplying a constant pressure gradient by the formation depth, which itself is an uncertain variable.

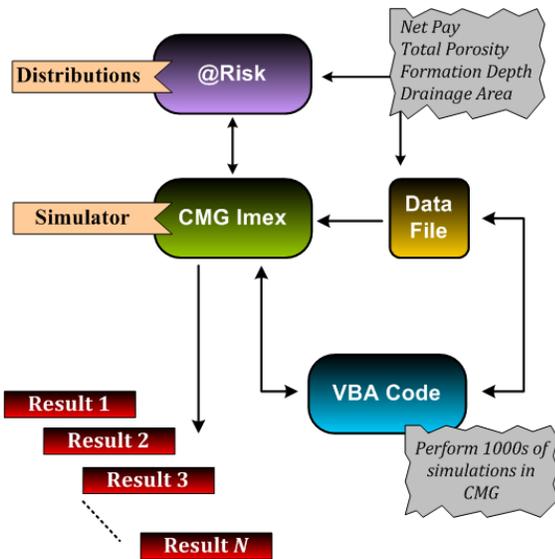


Fig. 4— Block diagram representation of the applications and tools used in this study.

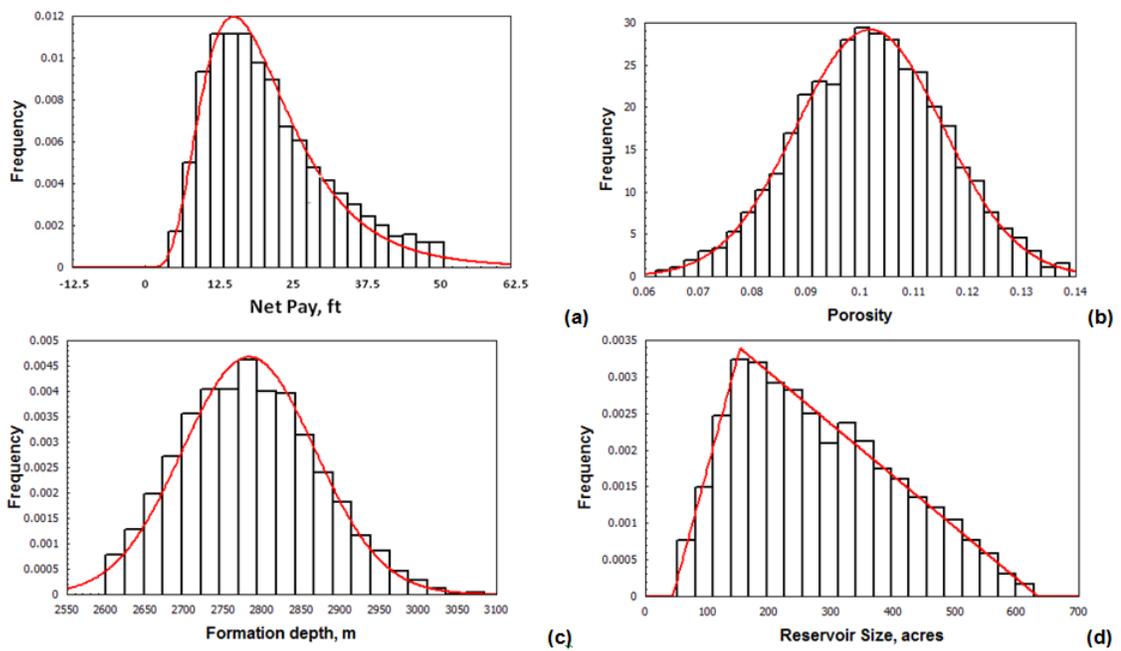


Fig. 5 – Histograms and probability distribution functions of uncertain input variables used in the modeling of the Deep Basin (Gething D) reservoir: (a) net pay, (b) porosity, (c) formation depth, (d) reservoir size.

**Table 1 — Uncertain input variables involved in the modeling of Deep Basin (Gething D) reservoir (data from UGR)**

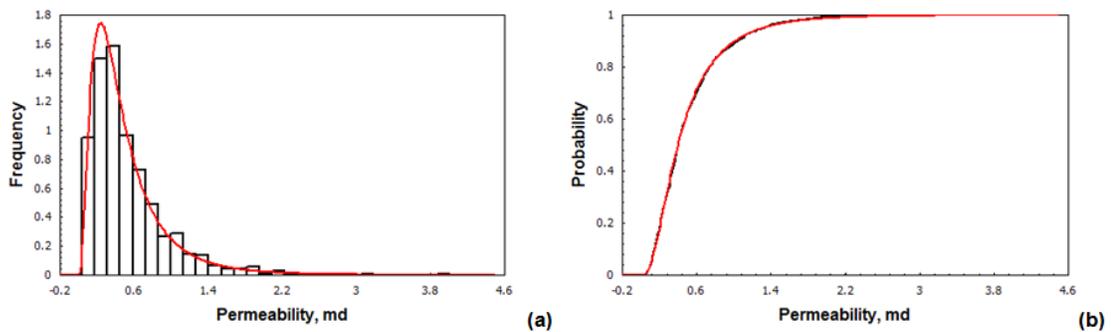
Variable Parameters	Distribution	Minimum	Maximum	Mean	Most Likely	Std dev.	Shift
	Type						
Net Pay, m	Log-normal	0.4	15	6.5	-	4.45	0.6
Total Porosity, %	Normal	6	14	10.2	-	1.4	-
Formation Depth, m	Normal	2600	3100	2780	-	91	-
Reservoir Size, acres	Triangular	40	640	-	160	-	-

The permeability model is based on a porosity-permeability relationship conditioned to core data from the field and well performance data. The correlation is given by

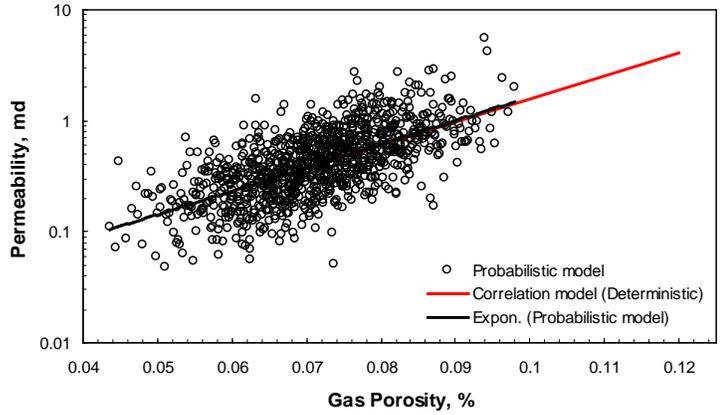
$$k_{gas}(md) = 1.92 * \exp(47.517 * \text{Gas Porosity} + \ln 0.0071) \dots\dots\dots (1)$$

Uncertainty is incorporated into the permeability model as well by making it a random variable. The expression inside the exponential function is assigned as the mean of a normal distribution with a standard deviation of 0.55. This model results in a log-normal distribution for permeability (**Fig. 6**) and a cloud on a semi-log porosity-permeability plot (**Fig. 7**).

Reservoirs that originate in non-marine, fluvial channel environments such as the Deep Basin (Gething) reservoir are often characterized by laterally discontinuous sand bodies. The fluvial channel geometry is generally defined by an elongated length/width (aspect) ratio. An aspect ratio of 7 was used in the modeling of the Deep Basin (Gething) reservoir.



**Fig. 6 — (a) Probability distribution plot (PDF), (b) cumulative distribution plot (CDF) of permeability.**



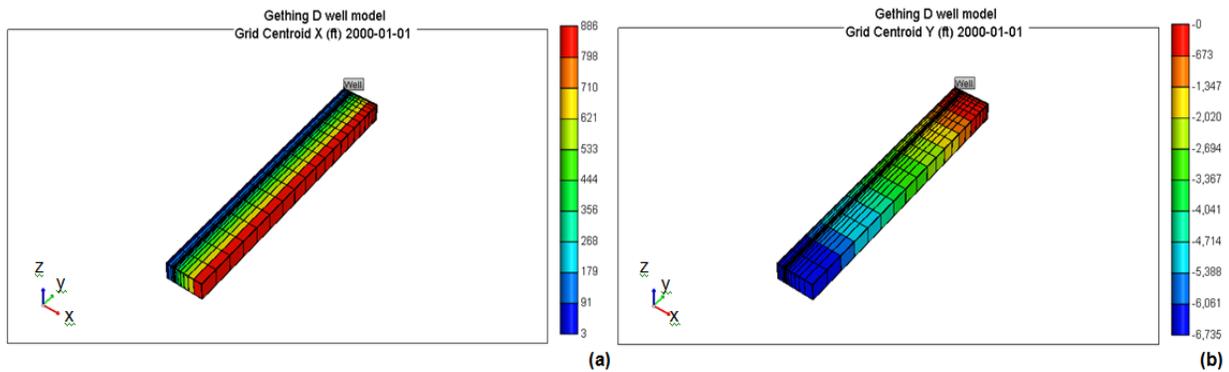
**Fig. 7 — Porosity-permeability cloud.**

We model a quarter of the drainage area, as is typically done for single-hydraulically-fractured-well models. Fig. 8 displays  $x$ - and  $y$ -direction centroids for the grid. The grid includes a hydraulic fracture near the wellbore, along the  $x$ -axis. To accurately model gas flow in the hydraulically fractured well, the hydraulic fracture is explicitly modeled by refined grids along the fracture and around the wellbore (**Fig. 8**).

Other input parameters in the Gething D reservoir model are listed in **Table 2**. The hydraulic fracture has a half-length of 200 ft, a simulated width of 0.04 ft, and a dimensionless fracture conductivity ( $F_{cD}$ ) of 1.3. The fracture permeability is computed as:

$$k_f (md) = \frac{F_{cD} * L_f * k}{w} \dots\dots\dots(2)$$

Since permeability is a random variable, unrealistically large values of fracture permeability can result from using a fixed  $F_{cD}$ . Thus, the fracture conductivity,  $wk_f$ , was limited to a maximum value of 150 md-ft, which means that the fracture permeability,  $k_f$ , cannot exceed 3750 md.



**Fig. 8 — Gridblocks model based on an aspect ratio of 7: (a) x-direction centroids, (b) y-direction centroids.**

**Table 2 — Input parameters employed in the Deep Basin (Gething D) reservoir model (data from UGR)**

Pressure Gradient, psi/ft	0.28
Reservoir Temperature, °C	90
Aspect Ratio	7
Gas Gravity	0.71
Water Saturation, %	30
Fracture Length, ft	200
Fracture Width, ft	0.04
Dimensionless Fracture Conductivity	1.3

### B.2.3 Probabilistic Reservoir Model and Decision Context.

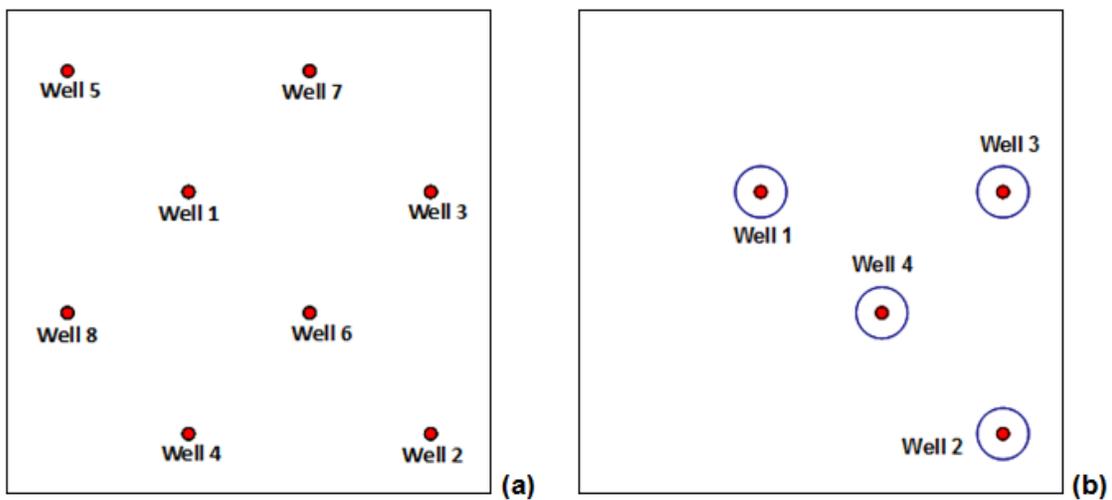
The Gething probabilistic reservoir model was configured to illustrate the effect of downspacing in the Deep Basin area, and to determine the optimal development strategy for the reservoir. Specifically, the decision context modeled consists of two development decisions: the primary and secondary well spacings. The intention is to start development on a particular well spacing (640, 320, 160 or 80 acres) and then decide on whether to downspace or not based on results of production at the primary spacing for a particular period of time. Thus, the model evaluates all possible two-stage downspacing combinations between 640, 320, 160 and 80 acres (Table 3) and simulation results quantify best-month production, stage-end average pressure, discounted stage production and 20-year discounted cumulative production for each well.

The reservoir model is established assuming the drilling of up to 8 wells in a section (Fig. 9a). The wells are numbered consistent with the order in which they would be drilled if spacing was progressively reduced from 640-ac spacing to 320-ac, 160-ac, and then 80-ac

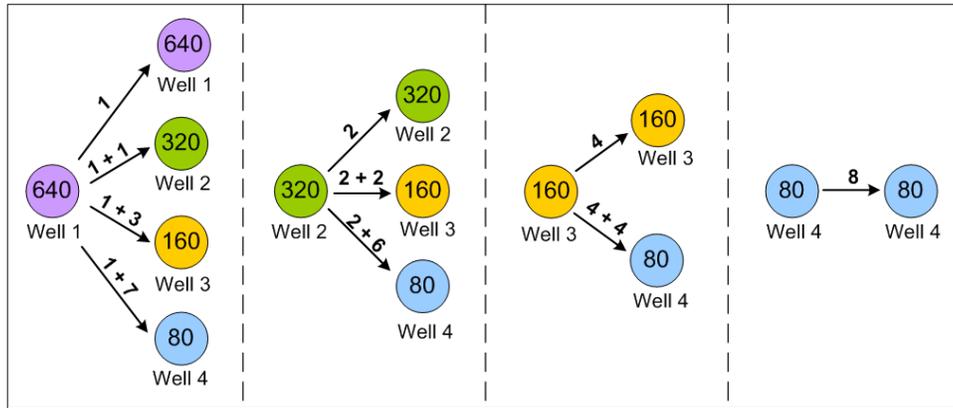
spacing. For simplicity and shorter computation times, only four of the eight wells (**Fig. 9b**) are modeled explicitly. Wells 1 through 4 correspond to well spacings of 640, 320, 160 and 80 acres, respectively. For each downspacing combination, only one well is simulated representing all the wells that will be drilled at each stage (**Fig. 10**). For instance, when downspacing from 320 acres (Stage 1) to 80 acres (Stage 2), 2 wells are produced in Stage 1 and 6 more are drilled in Stage 2. In this case, Well 2 and Well 4 will be simulated to represent all Stage 1 and Stage 2 wells, respectively. The production values from the one well simulated in Stage 1 are multiplied by 2 and the production values from the one well simulated in Stage 2 are multiplied by 6. To account for areal variability in reservoir properties at different well locations, all four wells are modeled individually, but sampled from the same probabilistic distributions. In this way, the reservoir heterogeneity is modeled in an approximate fashion even though single-well modeling is used.

**Table 3 — Possible downspacing combinations evaluated in the final reservoir model**

Stage-2 \ Stage-1	640	320	160	80
640	X	X	X	X
320	-	X	X	X
160	-	-	X	X
80	-	-	-	X



**Fig. 9 — Schematic illustration showing (a) possible well locations, (b) representative well locations.**



**Fig. 10 — Possible two-stage downspacing combinations. Numbers on top of arrows indicate Stage 1 wells and additional Stage 2 wells. Wells 1 through 4 are the representative wells.**

The initial reservoir pressure of the first-stage wells is computed using the initial pore pressure gradient of 0.28 psi/ft. Assignment of initial reservoir pressure for second-stage wells is complicated by the possibility of pressure interference and was modeled probabilistically. Pressure interference between first- and second-stage wells depends on the well spacing, i.e., number of wells drilled in a section and areas that are being drained by these wells. One would expect a greater chance of pressure communication between wells when the field is developed on a smaller spacing; as more wells are drilled, the probability of pressure interference increases. Also, the greater the reservoir (sand-body) size of each well, the greater the chance of communication between adjacent wells. We estimated the probability of pressure communication between wells for the various combinations of first- and second-stage well spacing and reservoir size and applied these discrete probabilities in the reservoir modeling. Based on the probability for a particular case, the Stage 2 well is either in pressure communication with the Stage 1 well (in which case the average reservoir pressure at the end of Stage 1 is designated as the initial pressure of the Stage 2 well) or it is not in communication (in which case the initial pressure of the Stage 2 well is set at the time-zero initial reservoir pressure).

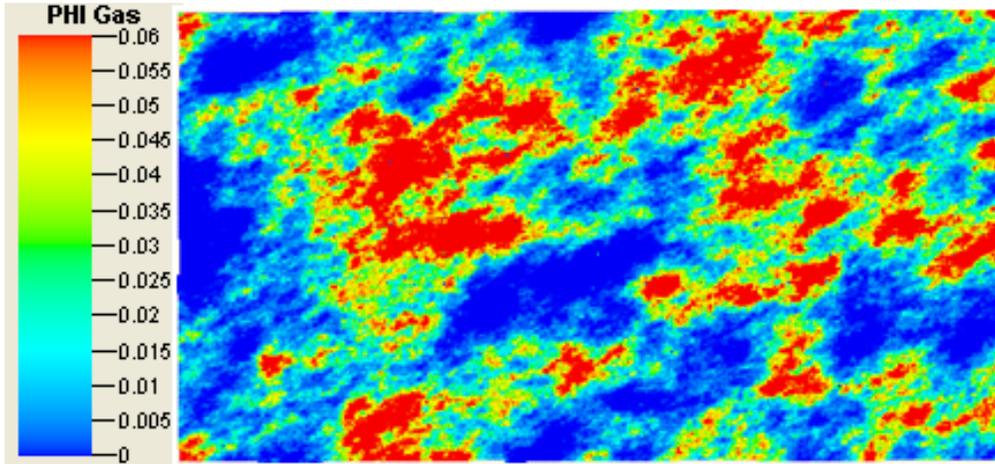
Completion efficiency risk of Deep Basin (Gething) wells was also modeled probabilistically. Based on information provided by UGR, a stimulation failure rate of 10% was assumed and incorporated into the reservoir model. Thus, there was a 10% probability of

stimulation failure, in which case the fracture permeability  $k_f$  is set equal to the formation permeability. Otherwise, the fracture permeability is defined by Eq. 2.

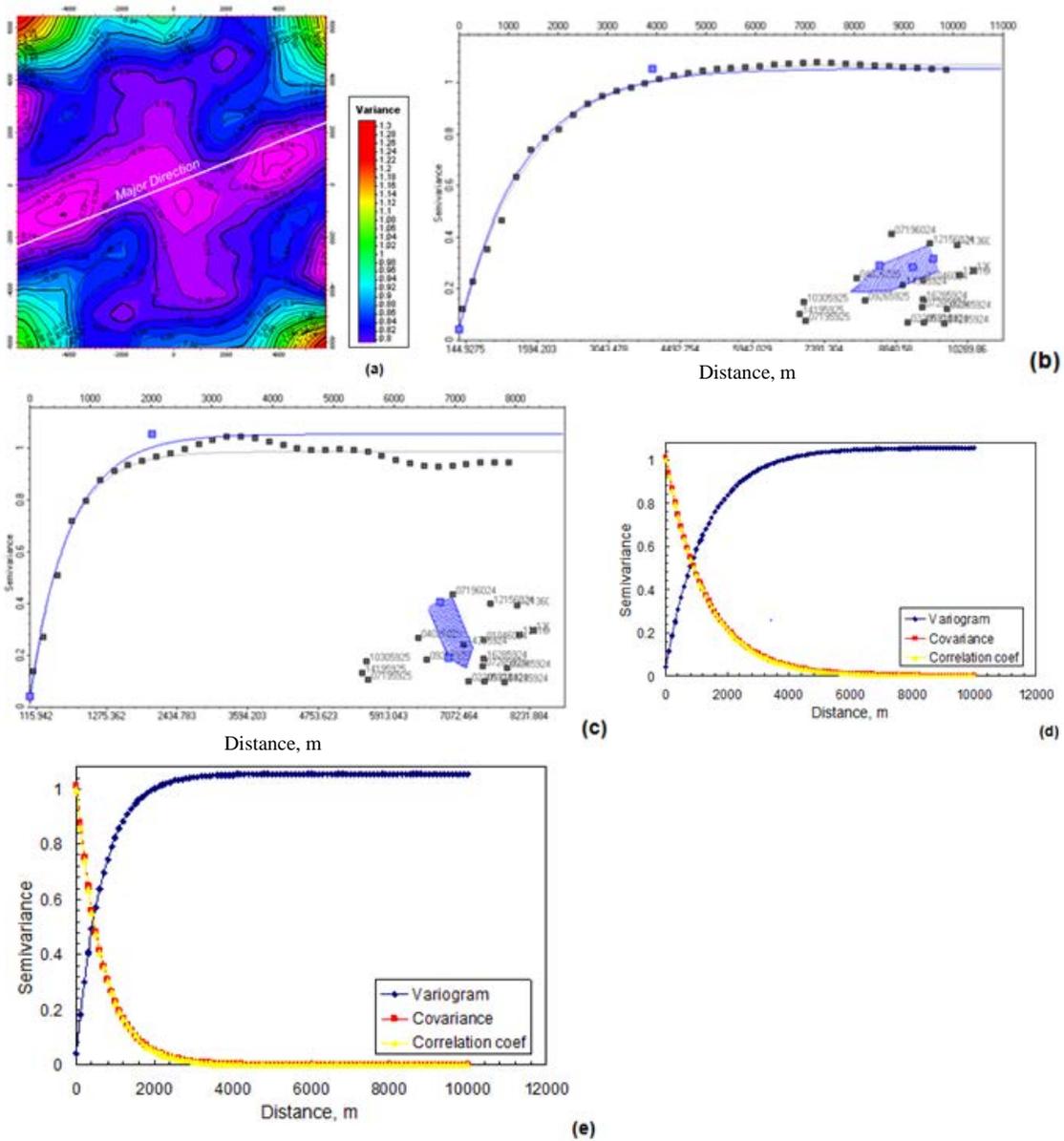
#### **B.2.4 Geostatistical Model**

Spatial correlation of reservoir properties affects dependencies in production responses between wells, which are needed in decision modeling. We developed spatial correlations of reservoir properties in the Deep Basin (Gething D) reservoir from geostatistical data included in an integrated reservoir study provided by UGR. From the spatial correlation data, we determined the correlation coefficients for reservoir properties between well locations separated by specific distances.

Porosity and porosity variogram maps of a 13x7-section study area are shown in **Fig. 11** and **Fig. 12a**, respectively. Based on the variogram map, the major and minor continuity directions for porosity are determined to be  $68^\circ$  and  $338^\circ$ , respectively. The corresponding major and minor direction empirical variograms and the theoretical models fitted to these experimental variograms are shown in **Fig. 12b** and **c**. These empirical variograms are of exponential type with a sill of 1.055 and nugget of 0.034. The major range is established as 3913 m and the minor range as 2015 m. Fitting an exponential model to the empirical porosity variograms, the theoretical variograms were calculated for both major and minor directions of continuity, along with their associated covariance and correlation functions (**Fig. 12d** and **e**). Similar analyses were performed for net pay and permeability. The correlation plots were used to compute the correlation coefficients for reservoir properties based on the distances between planned well locations in a section. **Table 4** – **Table 6** present the correlation matrices displaying the correlation coefficients between all well pairs for porosity, net pay thickness and permeability. These correlation coefficients were then incorporated into the probabilistic reservoir model to account for dependencies in reservoir properties between wells.



**Fig. 11 — Porosity map of the Deep Basin (Gething D) reservoir (provided by UGR).**



**Fig. 12 – Variogram analysis for porosity: (a) Variogram map (distances are in meters), (b) major and (c) minor direction empirical variograms, (d) major and (e) minor direction theoretical variograms with covariance and correlation functions (data from Deep Basin (Gething D) provided by UGR).**

**Table 4 — Correlation matrix for porosity**

Well no.	1	2	3	4
1	1			
2	0.18	1		
3	0.30	0.30	1	
4	0.43	0.43	0.43	1

**Table 5 — Correlation matrix for net pay**

Well no.	1	2	3	4
1	1			
2	0.56	1		
3	0.75	0.75	1	
4	0.87	0.87	0.87	1

**Table 6 — Correlation matrix for permeability**

Well no.	1	2	3	4
1	1			
2	0.37	1		
3	0.54	0.54	1	
4	0.67	0.67	0.67	1

### **B.2.5 Reservoir Simulation Results**

Reservoir simulation combined with Monte Carlo analysis provided a practical and quick method to generate a multitude of production forecasts with various reservoir properties under different development scenarios. These production profiles were incorporated into a flexible decision model, which allowed calculation of the expected net present value for each scenario and selection of the optimal development strategy. To facilitate the integration of the reservoir and decision models, discounted stage cumulative production values were used in the decision model. The discounted stage cumulative production for each drilling location was summarized by its mean, standard deviation and pair-wise correlation coefficients between it and all other well locations. This information was used to specify a joint probability distribution for each development scenario.

The reservoir model was run with 1-, 3- and 5-year Stage-1 durations to understand the trade-off between Stage-1 duration and well spacing. We refer to the wells drilled during Stage 1 (S1) as “Package 1” (P1) and the new wells drilled during Stage 2 (S2) as “Package 2” (P2). Only P1 wells produce during S1, while both P1 and P2 wells produce during S2. The wells in each stage will be drilled on spacings of 640, 320, 160, or 80 acres. Thus, each stage decision is

defined by four numbers: stage, package, spacing, and duration. For example, S1-P1-320-1 signifies that wells were drilled on 320-acre spacing during Stage 1 and produced for 1 year.

The reservoir model forecasts production vs. time, from which we calculate discounted-cumulative-production (DCP) at a discount rate of 10%. We take the natural logarithm of this quantity and refer to it as ln-DCP. Fig. 13 display scatter plots of S2 ln-DCP versus S1 ln-DCP for P1 and P2. In particular, both figures are for the case where S1 was drilled on 640-acre spacing and produced for one year (S1-P1-640-1). **Fig. 13a** presents the S2 production of P1 if we downspace to 80 acres (i.e., how the original P1 well performs over the remaining 19 years in the presence of the seven new wells). Fig. 13b relates the 19-year ln-DCP from an 80-acre S2-P2 (S2-P2-80-19) to S1 production. The correlations between Stages 1 and 2 for P1 are 0.83 when going from 640s to 80s and 0.60 for P2 when going from 640s to 80s. Because of this relationship, we can use the S1 production to forecast mean S2 production and the accompanying uncertainty. To do so, we assume that the uncertainty between stages is jointly lognormally distributed (i.e., ln-DCP is jointly normal between stages). Under this assumption, the posterior mean  $\mu_{2|1}$  and standard deviation  $\sigma_{2|1}$  of S2 ln-DCP given S1 ln-DCP ( $\ln Q_1$ ) is

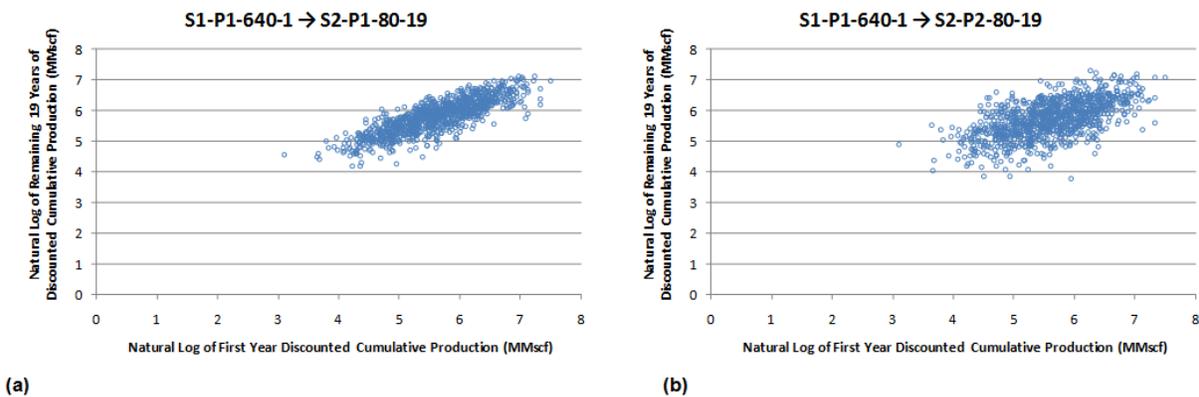
$$\mu_{2|1} = \mu_2 + \rho\sigma_2 \frac{\ln Q_1 - \mu_1}{\sigma_1} \dots\dots\dots(3)$$

$$\sigma_{2|1} = \sigma_2(1 - \rho^2)^{\frac{1}{2}} \dots\dots\dots(4)$$

where  $\mu_2$  and  $\sigma_2$  are the prior mean and standard deviation, respectively, of S2 ln-DCP. Likewise,  $\mu_1$  and  $\sigma_1$  are the prior mean and standard deviation for S1 ln-DCP. The correlation between S1 and S2 is  $\rho$ . Although we do not explicitly show this in our notation, the means and standard deviations for each stage are functions of the package, spacing, and stage length.

**Table 7** presents the mean production, standard deviation, and correlation for each possible S1-S2 spacing combination with a 1-year S1 pilot phase. We see that S1-P1 production tells us more (has a higher correlation coefficient) about how P1 will perform in Stage 2 than how P2 will perform. This is understandable since P2 is a completely new set of wells. However, the correlation between S2-P2 is still greater than 0.50, which is quite robust. We also see that correlations generally increase with increased downspacing, although they begin to

decrease beyond 160-acre spacing for P1. We believe this is related to the assumptions regarding reservoir size, which has a triangular distribution with a most-likely reservoir size of 160 acres (Table 1). **Table 8** and **Table 9** present these same statistics for a 3-year S1 and a 5-year S1, respectively. We notice the same trends as for a 1-year S1. However, by comparing between the tables we see that increasing the length of S1 decreases the correlation with S2 performance, which may be counter-intuitive. We believe this is related to the onset of depletion effects in Stage 1. Permeabilities in the Gething are relatively high for tight gas reservoirs, which results in boundary-dominated flow being reached within a couple of years in many cases (and sometimes months for certain combinations of permeability and reservoir size).



(a) (b)  
**Fig. 13 — Crossplots of natural log of individual-well, discounted stage production (MMscf) downspacing from 640 to 80 ac after a 1-yr Stage 1, (a) Stage 2 vs Stage 1 of the original well, and (b) the infill well in Stage 2 vs the original well in Stage 1.**

**Table 7—Statistics for natural log of individual-well, discounted stage production (MMscf) for original and infill wells for a 1-year Stage 1**

1-Year Pilot Phase									
	S1-P1	S2-P1				S2-P2			
spacing	<b>640</b>	<b>640</b>	<b>320</b>	<b>160</b>	<b>80</b>	<b>640</b>	<b>320</b>	<b>160</b>	<b>80</b>
mean, $\mu$	5.61	6.29	6.27	6.24	5.81		6.44	6.08	5.73
std dev, $\sigma$	0.71	0.79	0.76	0.59	0.54		0.74	0.65	0.60
corr, $\rho$	1.00	0.76	0.77	0.86	0.83		0.32	0.49	0.60
spacing	<b>320</b>	<b>640</b>	<b>320</b>	<b>160</b>	<b>80</b>	<b>640</b>	<b>320</b>	<b>160</b>	<b>80</b>
mean, $\mu$	5.60		6.26	6.18	5.80			6.14	5.78
std dev, $\sigma$	0.73		0.77	0.62	0.56			0.65	0.61
corr, $\rho$	1.00		0.78	0.85	0.84			0.52	0.60
spacing	<b>160</b>	<b>640</b>	<b>320</b>	<b>160</b>	<b>80</b>	<b>640</b>	<b>320</b>	<b>160</b>	<b>80</b>
mean, $\mu$	5.10			6.10	5.76				5.85
std dev, $\sigma$	0.75			0.60	0.54				0.57
corr, $\rho$	1.00			0.89	0.87				0.64
spacing	<b>80</b>	<b>640</b>	<b>320</b>	<b>160</b>	<b>80</b>	<b>640</b>	<b>320</b>	<b>160</b>	<b>80</b>
mean, $\mu$	4.99				5.62				
std dev, $\sigma$	0.73				0.53				
corr, $\rho$	1.00				0.82				

**Table 8— Statistics for natural log of individual-well, discounted stage production (MMscf) for original and infill wells for a 3-year Stage 1**

3-Year Pilot Phase									
	S1-P1	S2-P1				S2-P2			
spacing	<b>640</b>	<b>640</b>	<b>320</b>	<b>160</b>	<b>80</b>	<b>640</b>	<b>320</b>	<b>160</b>	<b>80</b>
mean, $\mu$	6.22	5.70	5.67	5.63	5.23		5.87	5.54	5.24
std dev, $\sigma$	0.72	0.97	0.94	0.73	0.68		0.94	0.84	0.79
corr, $\rho$	1.00	0.73	0.73	0.78	0.74		0.36	0.46	0.51
spacing	<b>320</b>								
mean, $\mu$	6.20		5.65	5.58	5.21			5.68	5.35
std dev, $\sigma$	0.72		0.94	0.75	0.68			0.84	0.79
corr, $\rho$	1.00		0.72	0.77	0.74			0.41	0.53
spacing	<b>160</b>								
mean, $\mu$	5.82			5.57	5.25				5.49
std dev, $\sigma$	0.71			0.59	0.54				0.65
corr, $\rho$	1.00			0.78	0.75				0.50
spacing	<b>80</b>								
mean, $\mu$	5.63				4.94				
std dev, $\sigma$	0.68				0.54				
corr, $\rho$	1.00				0.64				

**Table 9 — Statistics for natural log of individual-well, discounted stage production (MMscf) for original and infill wells for a 5-year Stage 1**

5-Year Pilot Phase									
	S1-P1	S2-P1				S2-P2			
spacing	<b>640</b>	<b>640</b>	<b>320</b>	<b>160</b>	<b>80</b>	<b>640</b>	<b>320</b>	<b>160</b>	<b>80</b>
mean, $\mu$	6.43	5.19	5.16	5.12	4.74		5.39	5.08	4.80
std dev, $\sigma$	0.72	1.05	1.02	0.81	0.76		1.07	1.00	0.92
corr, $\rho$	1.00	0.69	0.69	0.71	0.67		0.30	0.39	0.45
spacing	<b>320</b>								
mean, $\mu$	6.42		5.13	5.06	4.71			5.27	4.95
std dev, $\sigma$	0.71		1.02	0.83	0.77			0.98	0.92
corr, $\rho$	1.00		0.69	0.71	0.68			0.37	0.42
spacing	<b>160</b>								
mean, $\mu$	6.08			5.09	4.80				5.18
std dev, $\sigma$	0.68			0.59	0.55				0.70
corr, $\rho$	1.00			0.71	0.66				0.42
spacing	<b>80</b>								
mean, $\mu$	5.84				4.34				
std dev, $\sigma$	0.62				0.58				
corr, $\rho$	1.00				0.50				

### B.3 Decision Model

#### B.3.1 Overview

We can use the reservoir simulation results described above to choose the optimal Stage 1 and Stage 2 spacing strategy. To do so we use the schematic decision tree summarized in Fig. 14. We begin by choosing an initial spacing. We then observe S1 production results. Given these results, we make a S2 spacing decision and finally observe Stage 2 production. We assume that our objective is to maximize expected net present value (ENPV). Though not explicitly shown in this simple decision tree, our model does not allow the operator to choose a Stage 2 spacing that is greater than Stage 1. Likewise, selecting the same spacing in Stages 2 and 1 is simply a decision not to downspace.

These trees are solved starting at the end (right) and “rolling-back.” Thus, we cannot make the optimal S1 decision without first determining the optimal S2 spacing given every possible S1 outcome. To simplify this procedure, we discretize the production uncertainties with discrete probability mass functions containing five branches. Specifically, we weight the P99.8, P91.2, P50, P8.8, and P0.2 of the excess distribution function (i.e., the probability of exceeding a particular production level) by 0.011, 0.222, 0.533, 0.222, and 0.011, respectively



### B.3.2 Illustrative Example

We applied the integrated reservoir and decision models described above to a hypothetical case based on the Deep Basin (Gething D) formation. We made the following illustrative economic assumptions:

Gas Price: \$5.50 per Mcf

Marginal Cost: \$1.00 per Mcf

Fixed Cost: \$50,000 per year per well

Drilling Cost: \$1.5 MM per well

Discount Rate: 10%

No royalty

In this paper, we do not consider uncertainty in these economic parameters in order to focus attention on subsurface uncertainties. However, there is nothing in our method that would preclude this and any actual application should include market uncertainties.

A summary of our results for a 1-year Stage 1 appears in Fig. 15. Economics are computed on a section basis. Suppose we begin with a spacing of 640 acres. If we observe low production during the first year (P99.8), the optimal S2 spacing is to stay with 640 acres. The NPV of this scenario is \$0.43 MM. If we instead observe higher production (P91.2), the optimal S2 decision is to downspace to 320 acres, which has an NPV of \$1.22 MM. We would also downspace to 320 acres if we observed P50 production. Observing high production (P8.8) would make it economic to downspace to 160 acres, with an associated NPV of \$8.04 MM. Under very high S1 production (P0.2), we would downspace all the way to 80 acres and earn an NPV \$19.49 MM. Weighting each of these production scenarios by their probability of occurrence, we obtain an ENPV for an initial 640-acre spacing of \$4.05 MM.

This first example clearly demonstrates the value of an optimal dynamic strategy, which takes advantage of information gained during the first stage. For example, if one had to commit to 640-acre spacing from the outset and could not downspace, we calculate (not shown) that the

ENPV would be \$2.92 MM, or over \$1 MM (25%) less. This \$1 MM difference is the value of the information that is gained via the staged strategy. Thus, we see the value that can be gained through better use of information and better decision making. This level of improvement is typical (see Bickel and Smith (2006) and Bickel et al. (2008) for another example).

**Fig. 15** includes a similar analysis for each S1 spacing followed by optimal S2 spacing. For example, under an initial spacing of 320 acres, we would downspace to 160 acres if we observed P50 or greater production. Under either 160- or 80-acre initial spacings, we would not downspace under any S1 production scenario. The ENPVs for initial spacings of 640, 320, 160, and 80 acres are \$4.05 MM, \$5.93 MM, \$6.75 MM, and \$5.09 MM, respectively. Thus, the optimal S1 spacing is 160 acres, in which case we will not downspace during S2. This does not imply that downspacing is never beneficial. These results are specific to the particular reservoir model and economic assumptions used in this illustrative example.

**Fig. 16** presents our results for a 3-year and 5-year S1. The best initial spacing in each of these cases is also 160 acres. The ENPV for the 3-year Stage 1 is \$6.22 MM, while it is \$5.90 MM for a 5-year Stage 1. Thus, there is no economic benefit to extending the S1 duration beyond 1 year to either 3 years or 5 years, which is due in part to the decreasing correlation of S2 to S1 performance with increased S1 duration as discussed earlier. The complete optimal policy is a 1-year Stage 1 on 160 acres followed by no downspacing. Again, these results are specific to the reservoir properties and economic assumptions used in this example and cannot be generalized.

1-Year Stage One							
Stage 1			Stage 2		Stage 1	Stage 2	Total
spacing	prob	prod	opt space	S1 E[NPV]	S2 NPV	NPV	
640	0.011	P99.8	640	\$ 0.04	\$ 0.39	\$ 0.43	
\$ 4.05	0.222	P91.2	320	\$ 0.04	\$ 1.18	\$ 1.22	
	0.533	P50	320	\$ 0.04	\$ 3.26	\$ 3.31	
	0.222	P8.8	160	\$ 0.04	\$ 8.00	\$ 8.04	
	0.011	P0.2	80	\$ 0.04	\$ 19.45	\$ 19.49	
320	0.011	P99.8	320	\$ 0.10	\$ 0.76	\$ 0.86	
\$ 5.93	0.222	P91.2	320	\$ 0.10	\$ 2.00	\$ 2.10	
	0.533	P50	160	\$ 0.10	\$ 4.67	\$ 4.77	
	0.222	P8.8	160	\$ 0.10	\$ 11.63	\$ 11.73	
	0.011	P0.2	160	\$ 0.10	\$ 26.69	\$ 26.78	
160	0.011	P99.8	160	\$ (2.26)	\$ 1.58	\$ (0.67)	
\$ 6.75	0.222	P91.2	160	\$ (2.26)	\$ 3.72	\$ 1.46	
	0.533	P50	160	\$ (2.26)	\$ 7.81	\$ 5.55	
	0.222	P8.8	160	\$ (2.26)	\$ 16.18	\$ 13.92	
	0.011	P0.2	160	\$ (2.26)	\$ 36.01	\$ 33.75	
80	0.011	P99.8	80	\$ (5.46)	\$ 2.54	\$ (2.92)	
\$ 5.09	0.222	P91.2	80	\$ (5.46)	\$ 5.18	\$ (0.28)	
	0.533	P50	80	\$ (5.46)	\$ 9.58	\$ 4.12	
	0.222	P8.8	80	\$ (5.46)	\$ 17.47	\$ 12.01	
	0.011	P0.2	80	\$ (5.46)	\$ 33.71	\$ 28.25	

**Fig. 15 — Optimal policy and ENPVs for our illustrative example with 1-year Stage 1. Monetary values in \$MM/section.**

3-Year Stage One							5-Year Stage One														
Stage 1			Stage 2		Stage 1	Stage 2	Total	Stage 1			Stage 2		Stage 1	Stage 2	Total						
spacing	prob	prod	opt space	S1 E[NPV]	S2 NPV	NPV	spacing	prob	prod	opt space	S1 E[NPV]	S2 NPV	NPV	spacing	prob	prod	opt space	S1 E[NPV]	S2 NPV	NPV	
640	0.011	P99.8	640	\$ (0.06)	\$ 0.06	\$ (0.01)	640	0.011	P99.8	640	\$ 1.99	\$ (0.09)	\$ 1.90	640	0.011	P99.8	640	\$ 1.99	\$ (0.09)	\$ 1.90	
\$ 1.68	0.222	P91.2	640	\$ (0.06)	\$ 0.39	\$ 0.33	\$ 2.85	0.222	P91.2	640	\$ 1.99	\$ 0.11	\$ 2.10	0.222	P91.2	640	\$ 1.99	\$ 0.11	\$ 2.10	\$ 2.10	
	0.533	P50	640	\$ (0.06)	\$ 1.22	\$ 1.16		0.533	P50	640	\$ 1.99	\$ 0.62	\$ 2.61		0.533	P50	640	\$ 1.99	\$ 0.62	\$ 2.61	\$ 2.61
	0.222	P8.8	320	\$ (0.06)	\$ 3.92	\$ 3.86		0.222	P8.8	640	\$ 1.99	\$ 1.94	\$ 3.93		0.222	P8.8	640	\$ 1.99	\$ 1.94	\$ 3.93	\$ 3.93
	0.011	P0.2	320	\$ (0.06)	\$ 11.57	\$ 11.51		0.011	P0.2	320	\$ 1.99	\$ 6.43	\$ 8.42		0.011	P0.2	320	\$ 1.99	\$ 6.43	\$ 8.42	\$ 8.42
320	0.011	P99.8	320	\$ 2.59	\$ 0.12	\$ 2.71	320	0.011	P99.8	320	\$ 3.90	\$ (0.18)	\$ 3.73		0.011	P99.8	320	\$ 3.90	\$ (0.18)	\$ 3.73	\$ 3.73
\$ 5.56	0.222	P91.2	320	\$ 2.59	\$ 0.77	\$ 3.35	\$ 5.48	0.222	P91.2	320	\$ 3.90	\$ 0.21	\$ 4.11		0.222	P91.2	320	\$ 3.90	\$ 0.21	\$ 4.11	\$ 4.11
	0.533	P50	320	\$ 2.59	\$ 2.31	\$ 4.89		0.533	P50	320	\$ 3.90	\$ 1.15	\$ 5.05		0.533	P50	320	\$ 3.90	\$ 1.15	\$ 5.05	\$ 5.05
	0.222	P8.8	320	\$ 2.59	\$ 6.17	\$ 8.76		0.222	P8.8	320	\$ 3.90	\$ 3.59	\$ 7.49		0.222	P8.8	320	\$ 3.90	\$ 3.59	\$ 7.49	\$ 7.49
	0.011	P0.2	320	\$ 2.59	\$ 17.57	\$ 20.16		0.011	P0.2	320	\$ 3.90	\$ 11.05	\$ 14.96		0.011	P0.2	320	\$ 3.90	\$ 11.05	\$ 14.96	\$ 14.96
160	0.011	P99.8	160	\$ 1.45	\$ 0.76	\$ 2.22	160	0.011	P99.8	160	\$ 3.45	\$ 0.13	\$ 3.58		0.011	P99.8	160	\$ 3.45	\$ 0.13	\$ 3.58	\$ 3.58
\$ 6.22	0.222	P91.2	160	\$ 1.45	\$ 2.03	\$ 3.48	\$ 5.90	0.222	P91.2	160	\$ 3.45	\$ 0.90	\$ 4.35		0.222	P91.2	160	\$ 3.45	\$ 0.90	\$ 4.35	\$ 4.35
	0.533	P50	160	\$ 1.45	\$ 4.23	\$ 5.68		0.533	P50	160	\$ 3.45	\$ 2.18	\$ 5.63		0.533	P50	160	\$ 3.45	\$ 2.18	\$ 5.63	\$ 5.63
	0.222	P8.8	160	\$ 1.45	\$ 8.35	\$ 9.80		0.222	P8.8	160	\$ 3.45	\$ 4.43	\$ 7.88		0.222	P8.8	160	\$ 3.45	\$ 4.43	\$ 7.88	\$ 7.88
	0.011	P0.2	160	\$ 1.45	\$ 17.22	\$ 18.67		0.011	P0.2	160	\$ 3.45	\$ 8.98	\$ 12.44		0.011	P0.2	160	\$ 3.45	\$ 8.98	\$ 12.44	\$ 12.44
80	0.011	P99.8	80	\$ (0.15)	\$ 0.87	\$ 0.72	80	0.011	P99.8	80	\$ 2.08	\$ (0.30)	\$ 1.77		0.011	P99.8	80	\$ 2.08	\$ (0.30)	\$ 1.77	\$ 1.77
\$ 4.19	0.222	P91.2	80	\$ (0.15)	\$ 2.14	\$ 1.99	\$ 3.45	0.222	P91.2	80	\$ 2.08	\$ 0.35	\$ 2.43		0.222	P91.2	80	\$ 2.08	\$ 0.35	\$ 2.43	\$ 2.43
	0.533	P50	80	\$ (0.15)	\$ 4.02	\$ 3.88		0.533	P50	80	\$ 2.08	\$ 1.25	\$ 3.33		0.533	P50	80	\$ 2.08	\$ 1.25	\$ 3.33	\$ 3.33
	0.222	P8.8	80	\$ (0.15)	\$ 7.04	\$ 6.89		0.222	P8.8	80	\$ 2.08	\$ 2.58	\$ 4.66		0.222	P8.8	80	\$ 2.08	\$ 2.58	\$ 4.66	\$ 4.66
	0.011	P0.2	80	\$ (0.15)	\$ 12.54	\$ 12.39		0.011	P0.2	80	\$ 2.08	\$ 4.81	\$ 6.88		0.011	P0.2	80	\$ 2.08	\$ 4.81	\$ 6.88	\$ 6.88

**Fig. 16 — Optimal policy and ENPVs for our illustrative example with 3-year and 5-year Stage 1. Monetary values in \$MM/section.**

## B.4 Multi-Well Reservoir Simulation Model

### B.4.1 Overview

We implemented in this part a probabilistic, multi-well reservoir model based on reservoir simulation techniques coupled with stochastic methods and geostatistical characterization. Two major extensions over the previous probabilistic single-well model described above (and in Turkarslan et al. 2010) are: (1) key reservoir parameters are incorporated in the form of areal reservoir property maps obtained through geostatistical procedures and (2) it incorporates multiple wells in the model rather than using just one well.

The reservoir model simulates one section of the reservoir (640 acres). A data input file for CMG's IMEX simulator is created by the VBA code. Key reservoir parameters (porosity, permeability and *NTG*) are included as geostatistical maps and initial pressure is defined by a @RISK distribution function. The single-section multi-well reservoir model includes up to 8 wells in the section. The wells are numbered consistent with the order in which they would be drilled if spacing was progressively reduced from 640-ac spacing to 320-ac, 160-ac, and then 80-ac spacing. The reservoir model is formulated for the analysis of hydraulically fractured gas wells.

### B.4.2 Geostatistics Model

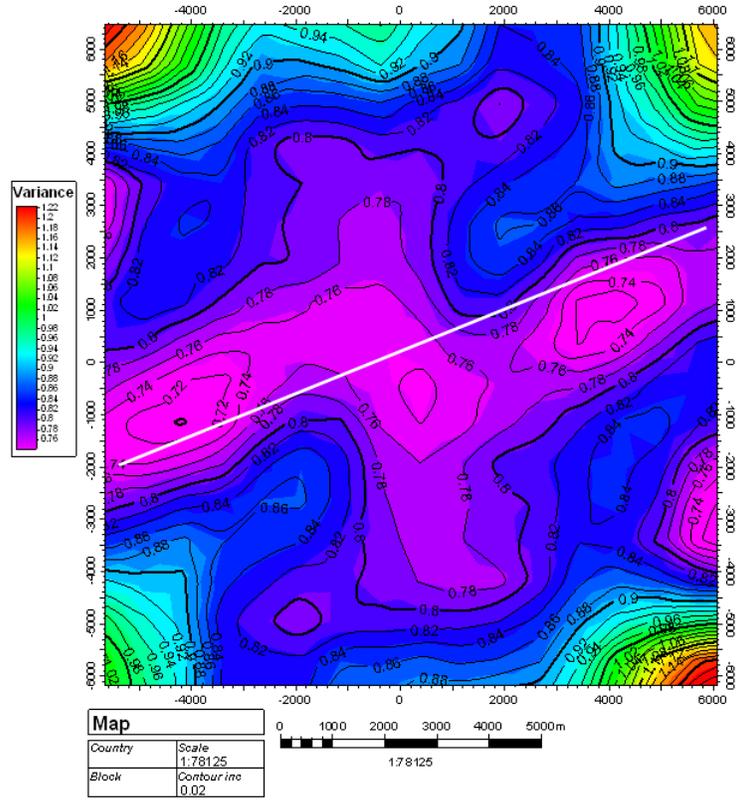
The procedure developed in this research includes the use of geostatistical characterization to generate multiple possible representations of the primary reservoir parameters to populate the reservoir model. To evaluate the uncertainty in the reservoir model, 1,000 maps of key reservoir parameters, including net-to-gross ratio (*NTG*), gas porosity ( $\phi_{\text{gas}}$ ) and gas permeability (*k<sub>gas</sub>*), were obtained. The wells with available log data that can be used for interpretation from the UGR's integrated field study of the Berland River area were used as input to generate the maps.

Basic reservoir parameters were upscaled. As the simulation grid cells often are much larger than the sample density for well logs, well log data must be scaled up before they can be entered into the grid. Using the available upscaled logs at the wells location I first constructed variogram maps of the main properties to evaluate the direction of anisotropy and determine in which direction there is enough stable data to construct empirical variograms to be used for the

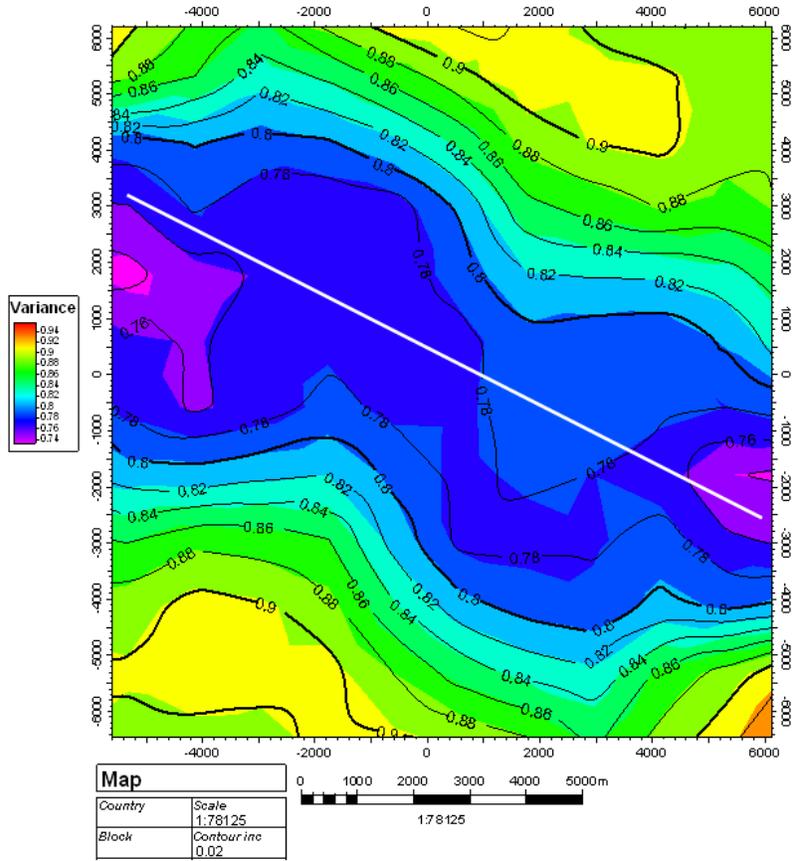
property modeling. The variogram map is obtained by calculating variograms in all possible azimuths and plotting them in a Cartesian plane. This gives the appearance of a contour map, where the contour maps are lines connecting points of equal variance. If the data have any anisotropy, it will be reflected in this variogram map in the form of elliptical contours. The direction of the elliptical closure will indicate the anisotropy direction. Once the anisotropy is identified, then the variograms are obtained. PETREL’s data analysis module was used to obtain the variogram maps and then the empirical variograms for the main properties in the major and minor directions of anisotropy. The major direction defines the direction in which the sample points have the strongest correlation (the main angle is specified as the clockwise angle from the north for the main search directions), the minimum direction is perpendicular to the major direction. **Fig. 17** is the variogram map for gas porosity; from this map we observed the major direction of anisotropy for gas porosity to be 68 degrees. Similarly, **Fig. 18** and **Fig. 19** show the variogram maps for NTG and permeability. The major directions of anisotropy are 297 degrees for NTG and 264 degrees for permeability. The empirical variograms were fitted by variogram models to obtain the theoretical variograms. **Table 10** shows the type of variogram models used and the variogram parameters. **Fig. 20** to **Fig. 22** show the fitted variograms for each property in both the major and minor directions.

**Table 10 — Theoretical variogram parameters.**

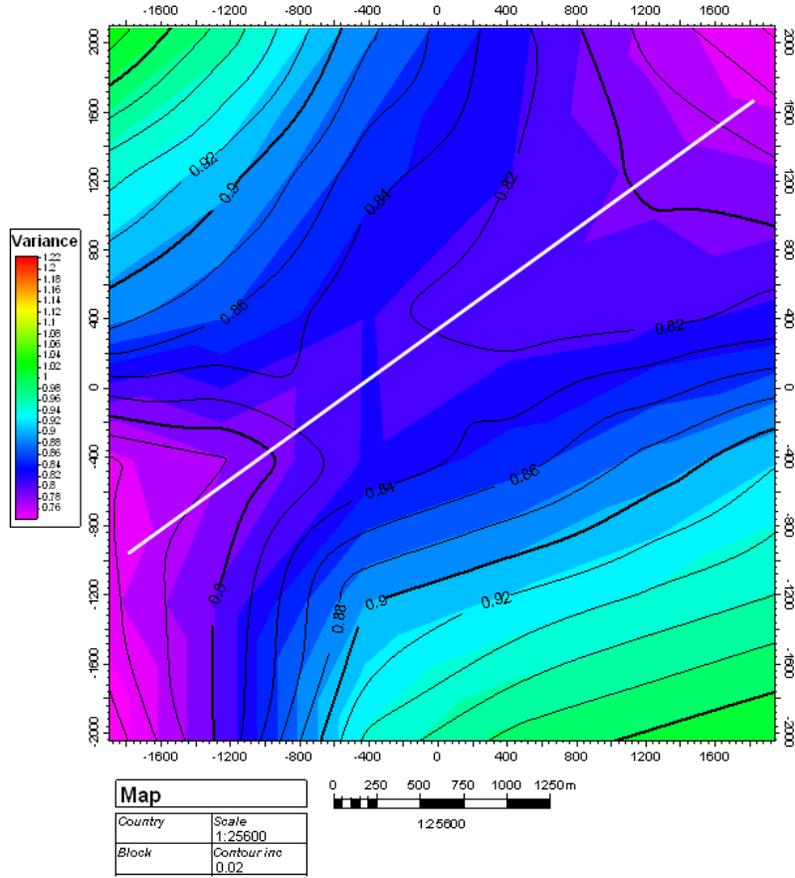
Property	Variogram Model	Angle of Major Direction	Range Major Direction (m)	Range Minor Direction (m)	Sill	Nugget
Gas Porosity	Exponential	68	3910	2015	1.054	0.034
NTG	Gaussian	247	2570	2470	0.952	0.116
Permeability	Spherical	264	2520	2430	0.997	0.050



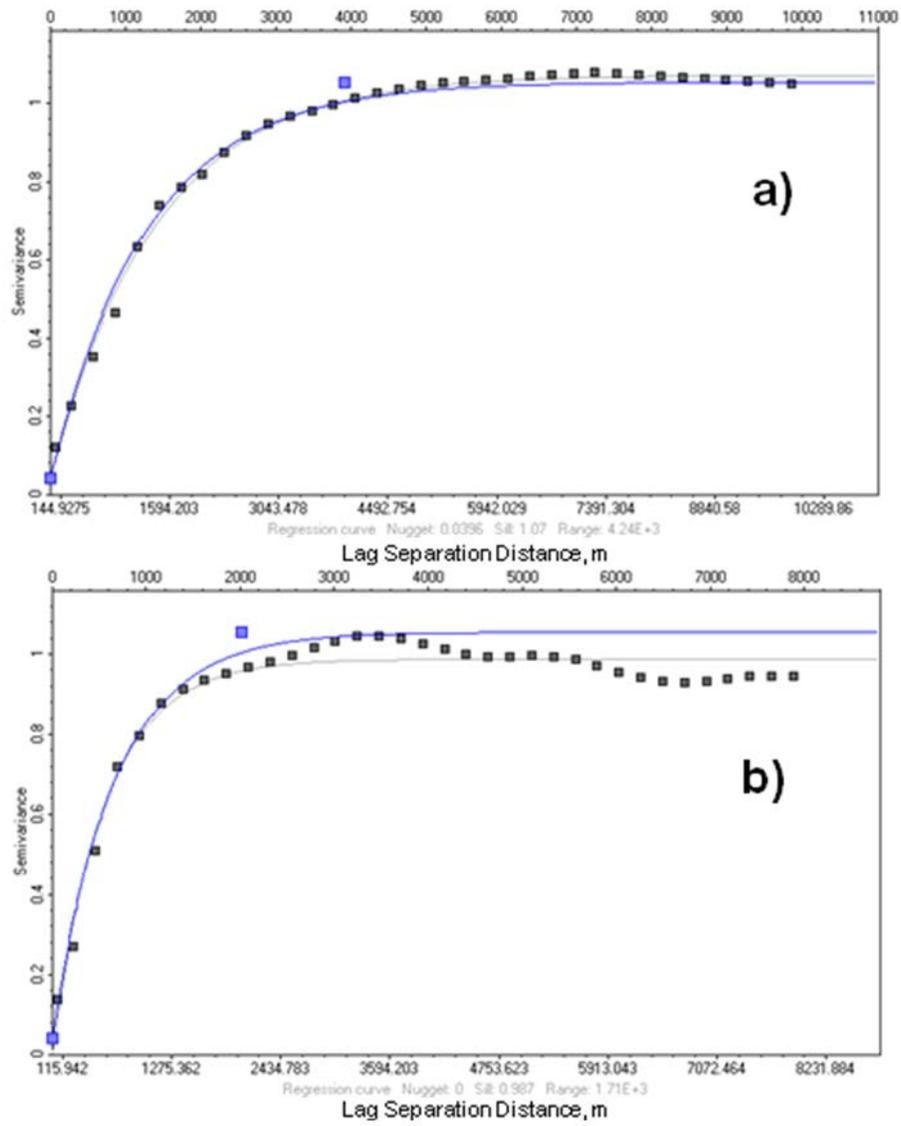
**Fig. 17 — Variogram map for gas porosity. The white line shows the major direction of anisotropy.**



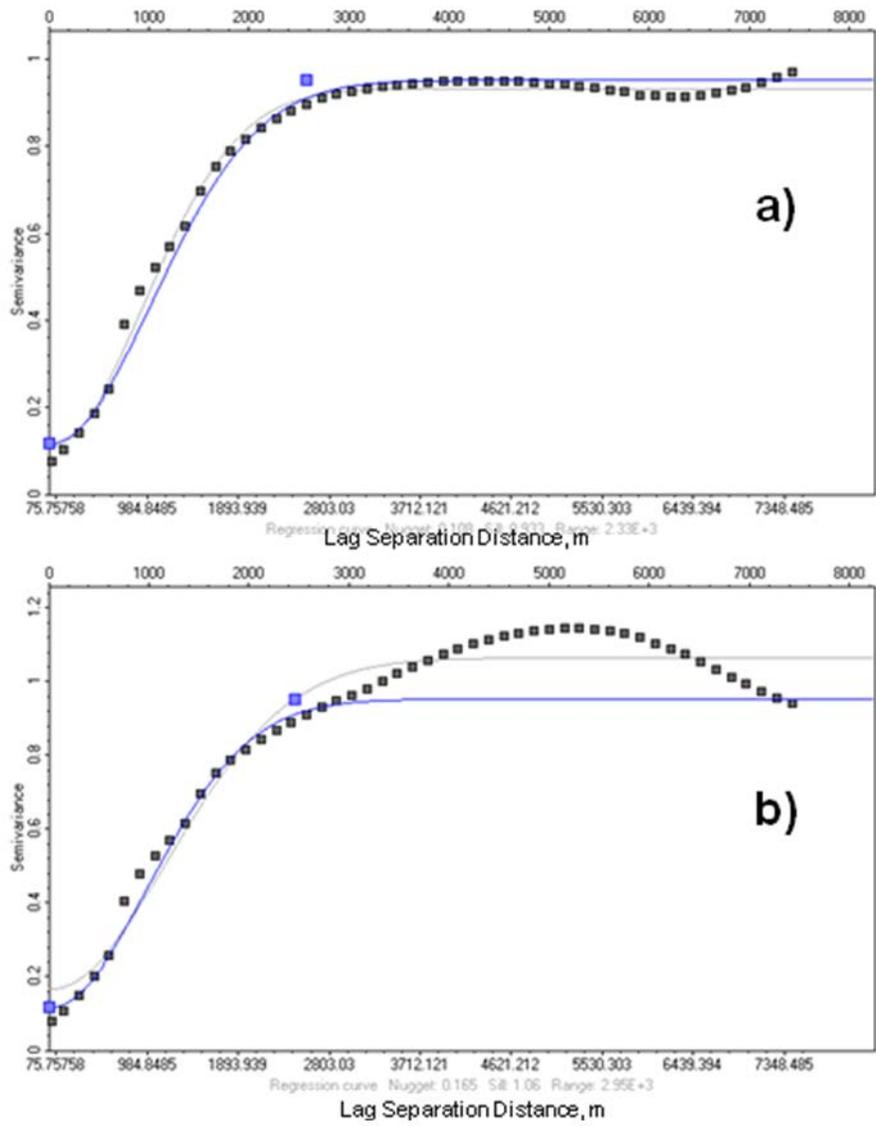
**Fig. 18 — Variogram map for net-to-gross ratio. The white line shows the major direction of anisotropy.**



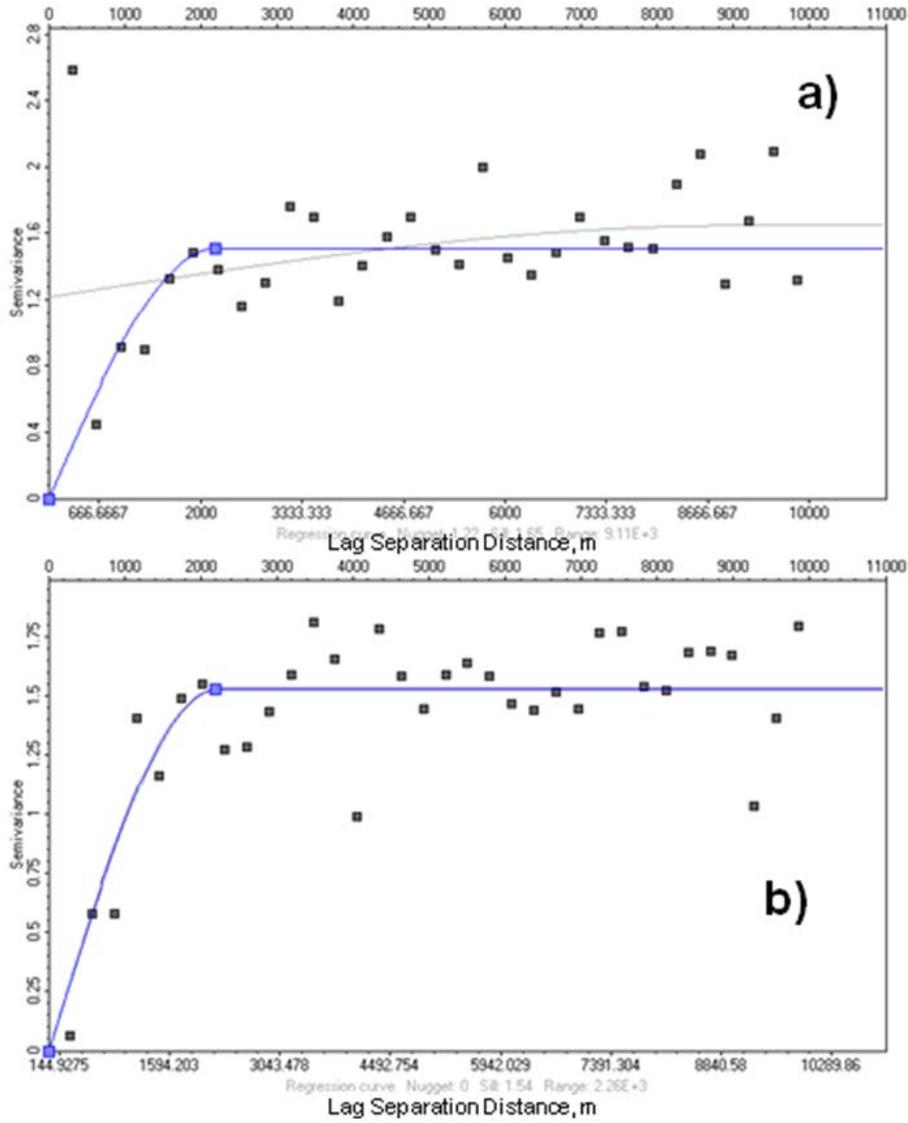
**Fig. 19 — Variogram map for permeability. The white line shows the major direction of anisotropy.**



**Fig. 20 — An exponential variogram model for gas porosity. a) is the variogram in the major direction of anisotropy and b) is the variogram in the minor direction.**



**Fig. 21 — A Gaussian variogram model for NTG. a) is the variogram in the major direction of anisotropy and b) is the variogram in the minor direction.**



**Fig. 22 — A spherical variogram model for permeability. a) is the variogram in the major direction of anisotropy and b) is the variogram in the minor direction.**

We used a Sequential Gaussian Simulation (SGS) algorithm from PETREL's petrophysical modeling process which is a stochastic method of interpolation based on kriging, which honors the well data, input distributions, variograms and trends, to generate the 1,000 property maps. Upscaled logs, variograms and input distributions were used during the petrophysical modeling. All cells in the simulation grid were given values. The upscaled parameters are fixed at the wellbore of existing wells and are honored by the algorithm; the

model then was populated with the three pertinent parameters away from wellbores. Ordinary kriging was used as part of the SGS process to generate 1,000 reservoir properties maps. Facies were defined, one to represent sands and one to represent the flood plain deposits. Logs of these facies were generated for the available wells and the facies modeling was done in conjunction with property modeling. **Fig. 23** shows an example of a facies log for one of the wells in the area.

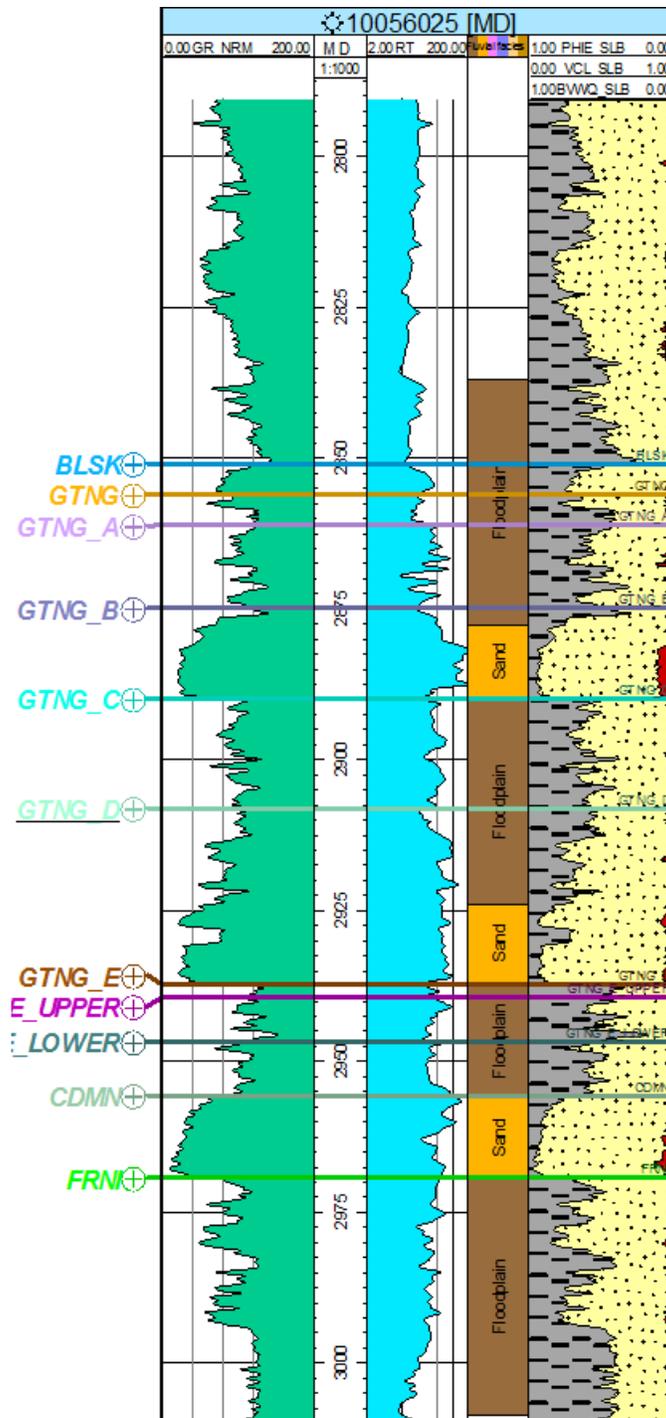
For permeability, a co-kriging option was used to steer the simulation using the spatial distribution of a second variable together with a correlation coefficient to calculate the contribution of the secondary variable at each point. The pore volume normalized by the arithmetic mean was used as the secondary variable and a correlation coefficient of 0.68 determined by PETREL was used. The correlation coefficient was estimated based on the upscaled log values and it is given for the normal score transformed data of both primary and secondary variables. The correlation coefficient is obtained from a cross-plot of the two variables in the normal score space and basically summarizes the relationship between the two variables.

**Fig. 24** to **Fig. 26** show one of the individual realizations of the porosity, net-to-gross ratio and permeability maps of the Gething D formation for the entire Berland River Area obtained through the geostatistical procedure described above. Gas porosity ranges from 0 to 0.1, net-to-gross ratio ranges from 0 to 1, and gas permeability ranges from 0.001 to 2.3 md. These property maps show the heterogeneous nature of the reservoir. The gray area on these figures represents the section from where the 1,000 property maps were obtained for the multi-well reservoir model for this project.

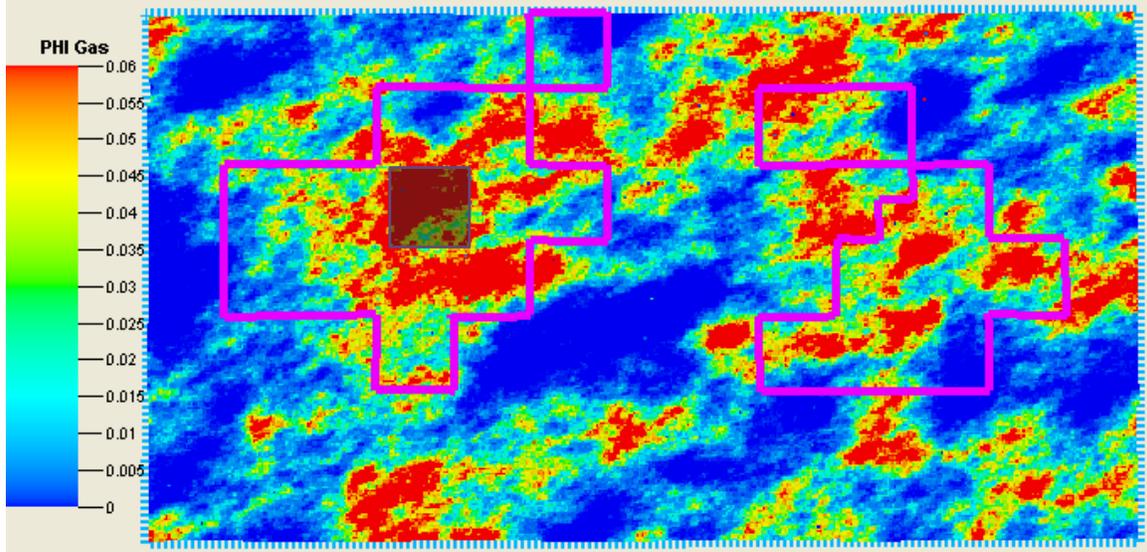
An example of one realization of reservoir properties map for the selected section of the Berland River Area is presented in **Fig. 27**. **Fig. 28** also shows one realization of the property *NTG* on a 3-D display of the simulation grid. Note the high degree of heterogeneity seen through

the model both laterally and vertically in the study area. The high degree of heterogeneity both vertically and laterally displayed on Figs. 21 and 22 highlights the need for using a multi-well model to simulate the tight gas reservoir.

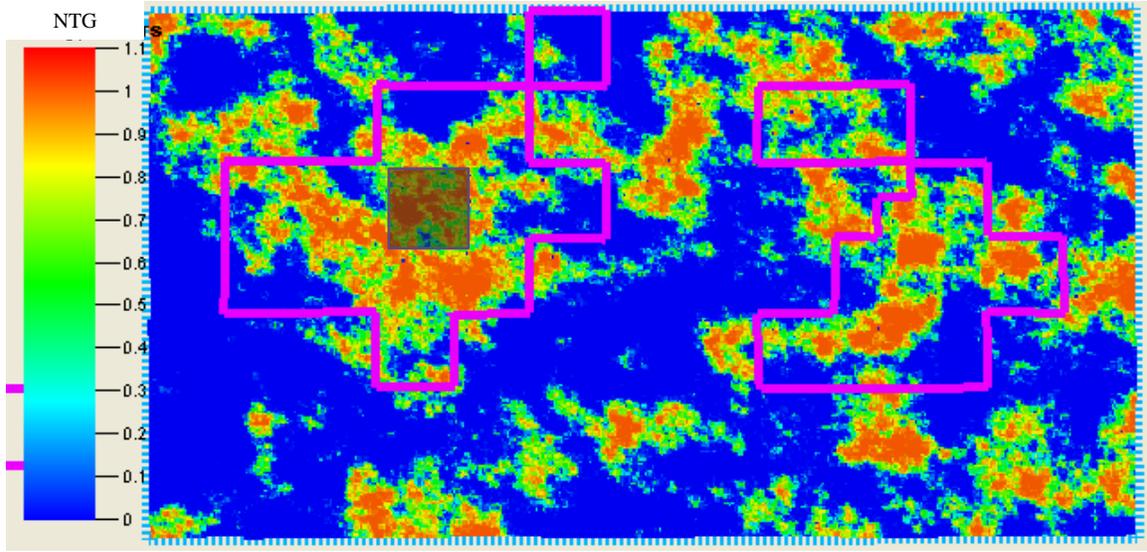
After the reservoir simulation using the 1,000 realizations of properties maps was initiated, some issues were found with the maps that needed correction before the simulation could be continued. The most significant was related to adjustment required to the NTG value for certain cells that need to change its value from 0% to 0.5% to avoid undesired stops in the simulation.



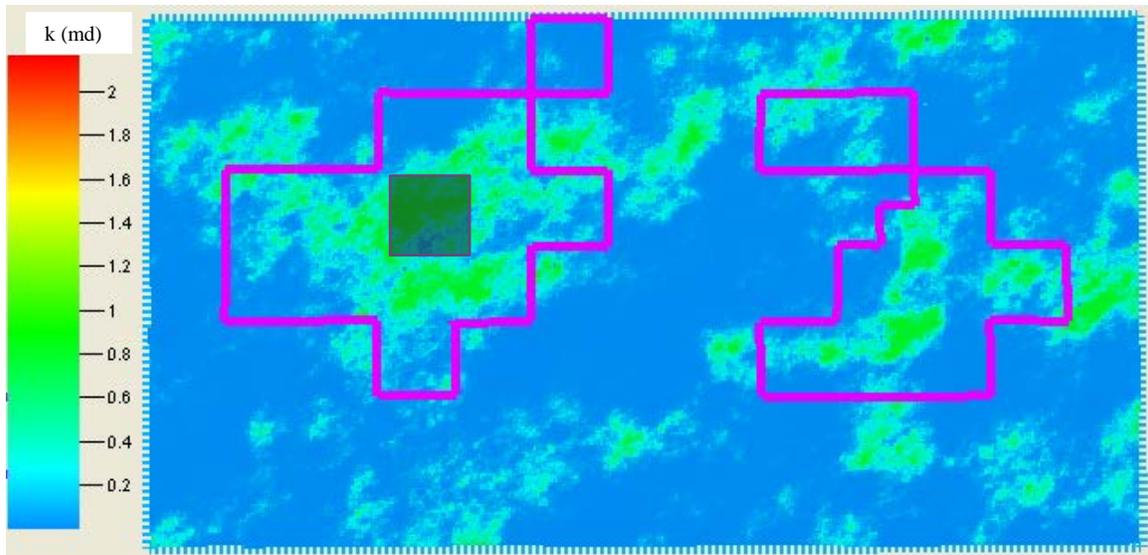
**Fig. 23 — An example of a log including facies for a well of the area under study. The facies log is presented on Track 4. Sand facies are represented by the dark yellow color and flood plain deposits by the brown color. Depths are in meters.**



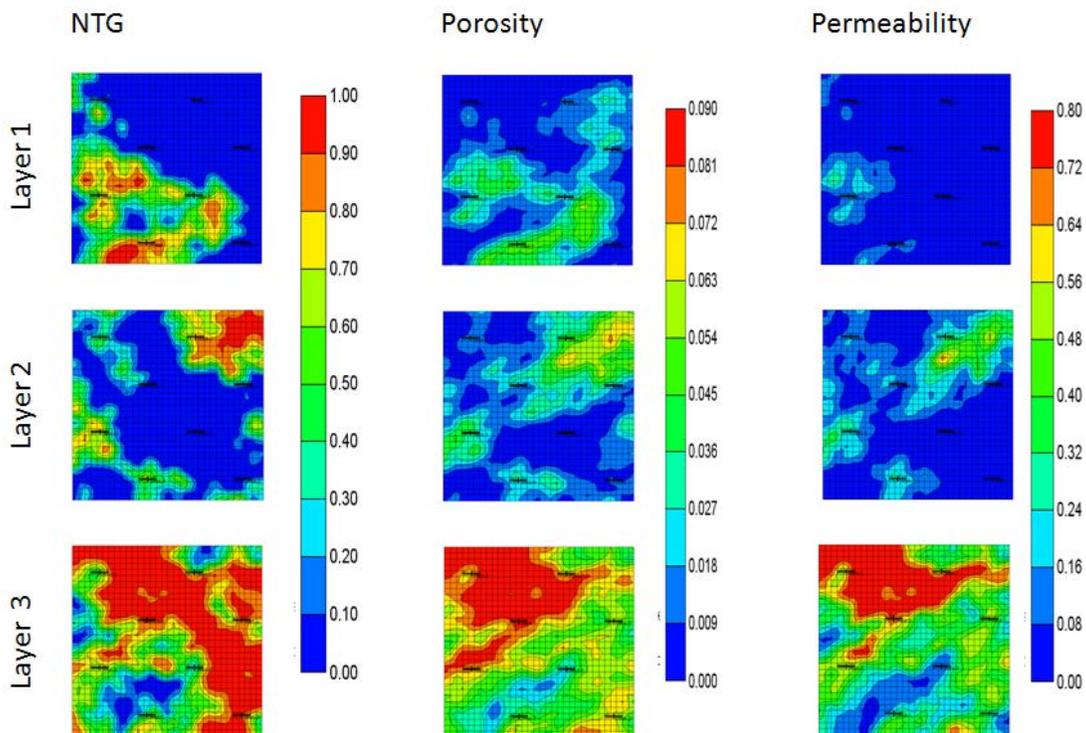
**Fig. 24** — A realization of porosity for the Gething D formation obtained through geostatistical procedures. The boxes in light purple represent UGR’s lease areas. The shadowed gray box represents the section selected to extract the property maps to be used in the multi-well reservoir model for this project.



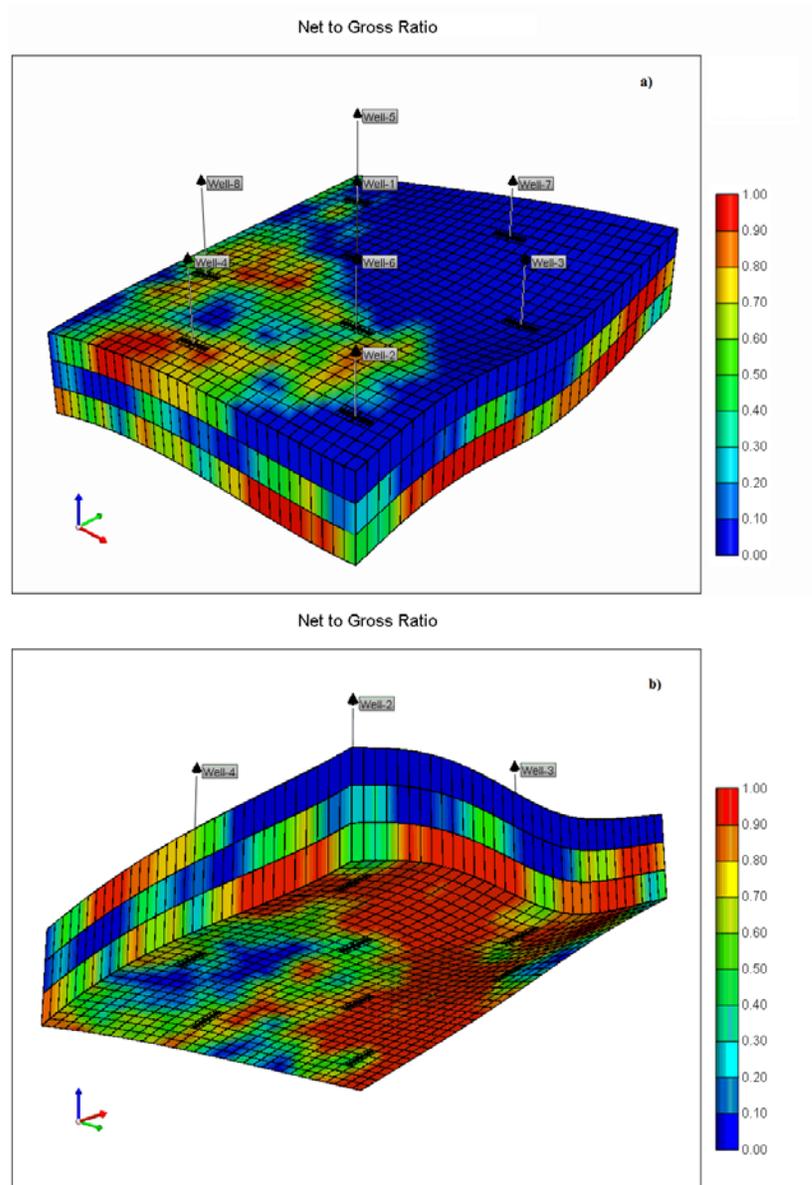
**Fig. 25** — A realization of net-to-gross for the Gething D formation obtained through geostatistical procedures. The boxes in light purple represent UGR’s lease areas. The shadowed gray box represents the section selected to extract the property maps to be used in the multi-well reservoir model for this project.



**Fig. 26** — A realization of permeability for the Gething D formation obtained through geostatistical procedures. The boxes in light purple represent UGR's lease areas. The shadowed gray box represents the section selected to extract the property maps to be used in the multi-well reservoir model for this project.



**Fig. 27** — An example realization of reservoir property maps for the section of the Gething D interval, Berland River Area, showing the properties in individual layers. Permeability is in md.



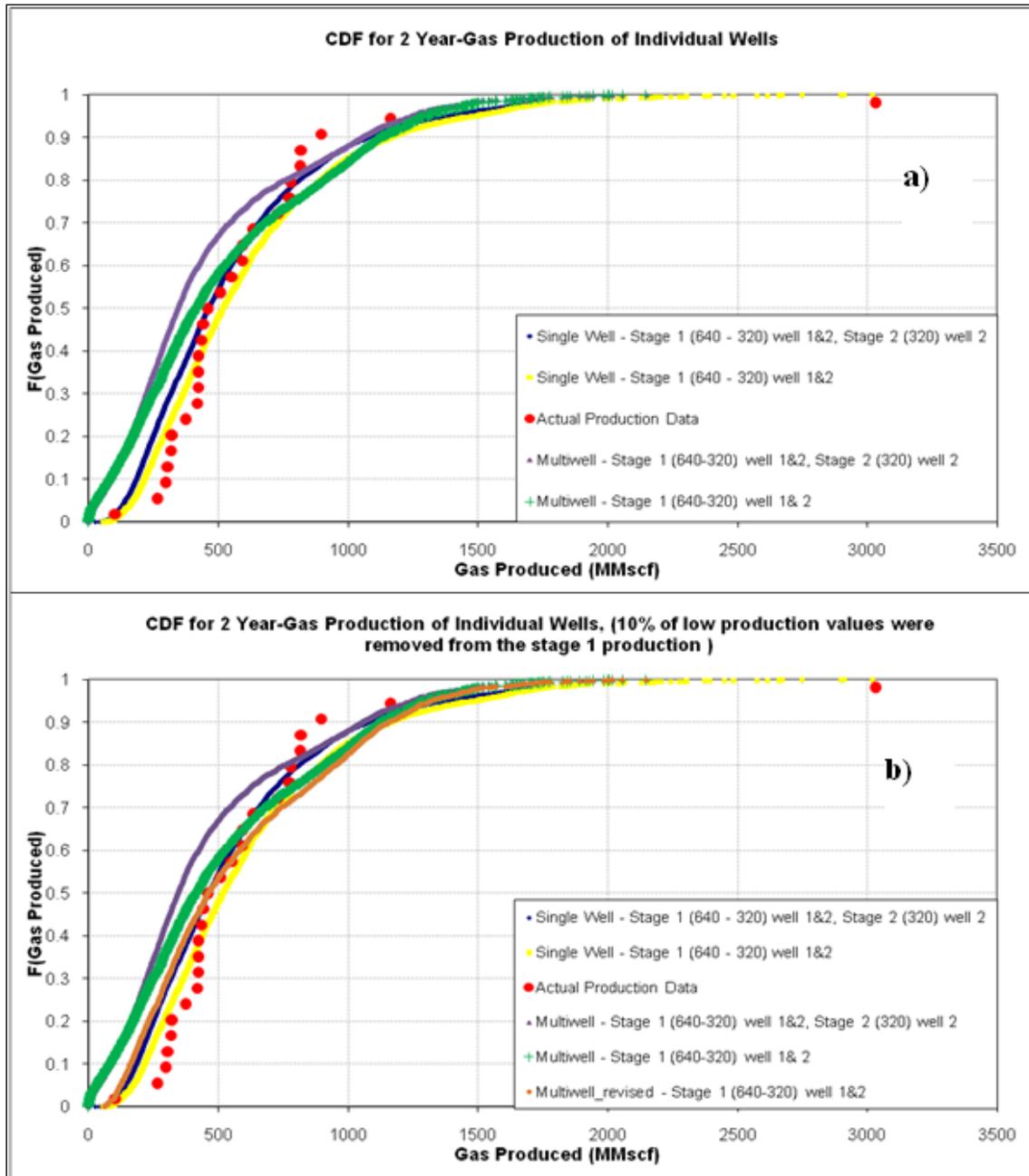
**Fig. 28** — Simulation grid of the study area showing a top (a) and bottom (b) view of the three-dimensional cube of net-to-gross ratio (NTG).

### B.4.3 Comparison with Single Well Model

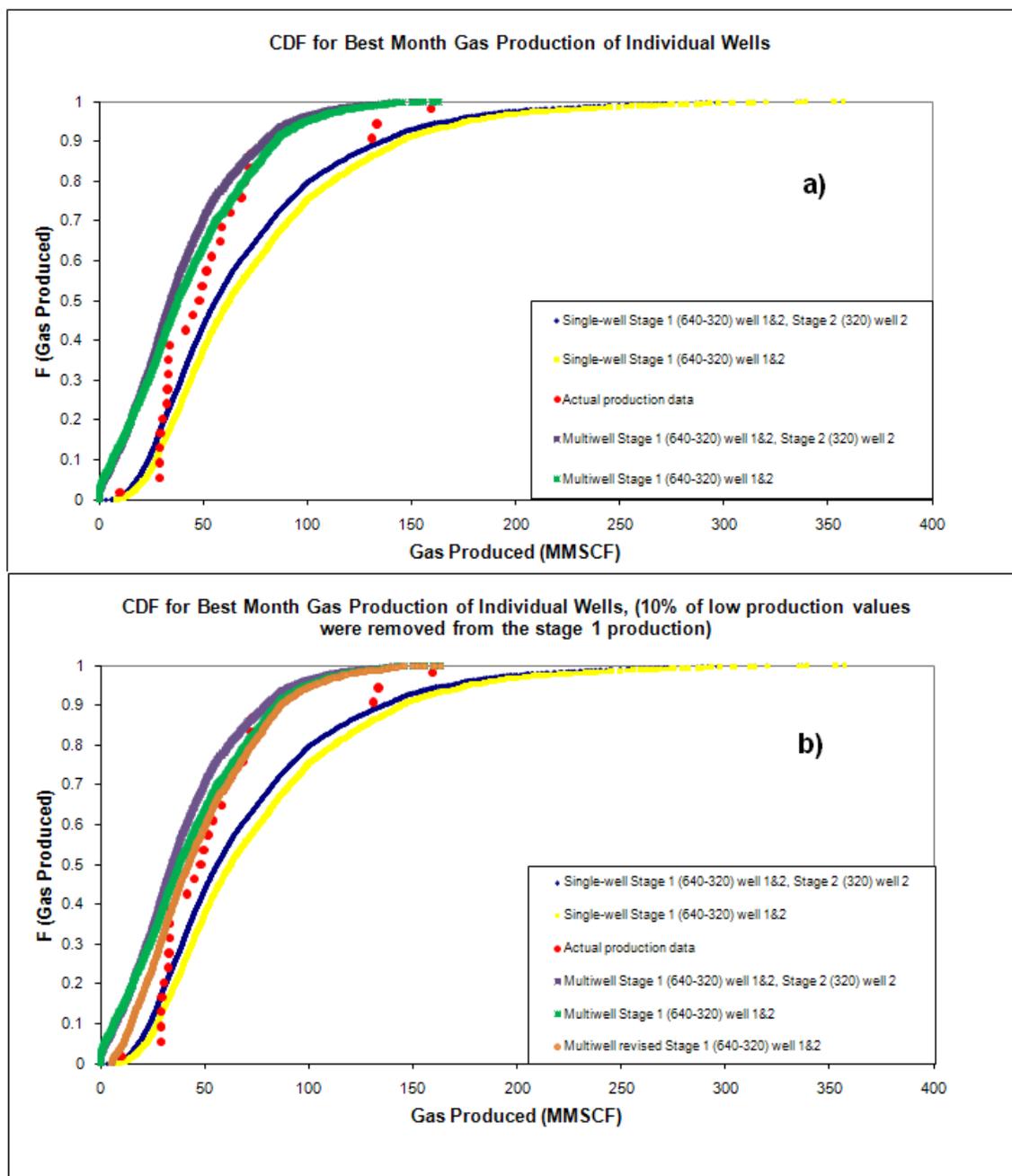
The new reservoir model yielded a distribution of production that matches the actual distribution of Gething production and the single-well reservoir model production results quite well (**Fig. 29a**). Since around 10% of the maps had problems with wells in zones of very low

pay, the lowest 10% of production values were removed from the Stage-1 production results for 640- and 320-acre spacing for Wells 1 & 2 and the cumulative distribution functions (CDF) was constructed again (**Fig. 29b**). A better comparison between actual data, single-well model results and multi-well model results is obtained in this case.

Similarly, comparison was done for best-month production on individual wells (Wells 1 & 2). The multi-well model produced a distribution that matches much better the actual production results from the Berland River area than the single-well model (**Fig. 30a**). Similarly, 10% of the lowest production values were removed and the cumulative distribution plot was constructed again and a better comparison between the multi-well reservoir results and the actual production data is observed (**Fig. 30b**).



**Fig. 29 — a) The multi-well reservoir model yielded a distribution of production that matches the actual distribution of Gething production and single-well reservoir model results quite well, b) revised production distribution for multi-well reservoir model where the lowest 10% of production values was removed from Stage-1 results. A better match with the actual distribution of Gething production and single-well reservoir model is observed.**



**Fig. 30 — a) The multi-well reservoir model yielded a distribution of best-month gas production that matches the actual distribution of Gething production. b) Lowest 10% of production values removed from Stage 1. A better match with the actual distribution of Gething production is observed.**

#### B.4.4 Results

The same illustrative economic assumptions used in the single-well reservoir simulation model were used here:

Gas Price: \$5.50 per Mcf

Marginal Cost: \$1.00 per Mcf

Fixed Cost: \$50,000 per year per well

Drilling Cost: \$1.5 MM per well

Discount Rate: 10%

No royalty

In **Table 11** to **Table 13** the initial spacing and the recommended downspacing for Stage 2 are determined. The results for Stage-1 length of 1 year (Table 11) shows that if we start with a 640-acre spacing and the initial production is high ( $P_{90}$ ), the downspace alternative that maximizes the NPV is to continue on 640-acre spacing; in this case no additional wells will be drilled in Stage 2. The total NPV of this scenario, including Stage 1, is \$15.88 MM. If we observe low production ( $P_{10}$ ), the optimal Stage-2 decision is to downspace to 160 acres, in which case the NPV is \$2.21 MM. We also downspace to 160 acres if  $P_{50}$  production is observed; for this case the NPV is \$4.45 MM. Weighting each of these production scenarios with their probability of occurrence, we obtain an expected NPV for an initial 640-acre spacing of \$6.67 MM. Similar analyses were done for initial spacing of 320, 160 and 80 acres and their estimated NPVs are \$6.04 MM, \$7.10 MM and \$6.25 MM, respectively. Thus, the optimal Stage-1 spacing is 160 acres, in which no further downspacing is required. This example demonstrates the value when a dynamic strategy is used. A dynamic strategy takes advantage of the information gained during the initial stage to develop the optimal downspacing program.

Table 12 and Table 13 present the results for 3-year and 5-year Stage-1 lengths, respectively. The best initial spacing for the 3-year Stage-1 length is 160 acres and its estimated NPV is \$6.63 MM. For 5-year Stage-1 length, the optimum is to start with 640-acre spacing and not downspace, in which case the expected NPV is \$6.39MM. We observe from these results that extending the duration of Stage 1 beyond 1 year does not represent an economic benefit, which is consistent with and results from the decreasing correlation coefficients between Stage 2 and Stage 1 performance with increased Stage 1 duration previously discussed. The optimal development policy is a 1-year Stage 1 of 160 acres followed by no downspacing. These results are specific to the reservoir model and economic assumptions used in this work and cannot be generalized.

**Table 11 — Optimal development strategy results for Stage-1 length of 1 year. Monetary values are per section.**

<u>Stage 1</u> Spacing, acres	Probability	Observed DCP	<u>Stage 2</u> Optimal Spacing acres	<u>Stage 1</u> E[NPV], \$MM	<u>Stage 2</u> NPV, \$MM	<u>Total</u> NPV, \$MM	Estimated NPV, Optimal Policy, \$MM
640	0.25	P90	640	0.94	14.64	15.58	6.67
	0.50	P50	160	0.94	3.51	4.45	
	0.25	P10	160	0.94	1.27	2.21	
320	0.25	P90	320	0.39	10.31	10.70	6.04
	0.50	P50	320	0.39	4.97	5.36	
	0.25	P10	320	0.39	2.37	2.76	
160	0.25	P90	160	0.43	10.55	10.98	7.10
	0.50	P50	160	0.43	6.24	6.67	
	0.25	P10	160	0.43	3.66	4.09	
80	0.25	P90	80	0.68	7.89	8.57	6.25
	0.50	P50	80	0.68	5.38	6.06	
	0.25	P10	80	0.68	3.63	4.31	

**Table 12 — Optimal development strategy results for Stage-1 length of 3 years. Monetary values are per section.**

<u>Stage 1</u> Spacing, acres	Probability	Observed DCP	<u>Stage 2</u> Optimal Spacing, acres	<u>Stage 1</u> E[NPV], \$MM	<u>Stage 2</u> NPV, \$MM	<u>Total</u> NPV, \$MM	Estimated NPV, Optimal Policy, \$MM
640	0.25	P90	640	3.19	8.74	11.94	6.46
	0.50	P50	640	3.19	1.98	5.17	
	0.25	P10	320	3.19	0.38	3.57	
320	0.25	P90	320	2.75	5.62	8.37	5.85
	0.50	P50	320	2.75	2.75	5.50	
	0.25	P10	320	2.75	1.28	4.03	
160	0.25	P90	160	3.79	4.47	8.26	6.63
	0.50	P50	160	3.79	2.68	6.47	
	0.25	P10	160	3.79	1.54	5.33	
80	0.25	P90	80	4.28	1.56	5.85	5.24
	0.50	P50	80	4.28	0.92	5.20	
	0.25	P10	80	4.28	0.44	4.72	

**Table 13 — Optimal development strategy results for Stage-1 length of 5 year. Monetary values are per section.**

<u>Stage 1</u> Spacing acres	Probability	Observed DCP	<u>Stage 2</u> Optimal Spacing, acres	<u>Stage 1</u> E[NPV], \$MM	<u>Stage 2</u> NPV, \$MM	<u>Total</u> NPV, \$MM	Estimated NPV, Optimal Policy, \$MM
640	0.25	P90	640	4.42	5.27	9.69	6.39
	0.50	P50	640	4.42	1.21	5.63	
	0.25	P10	640	4.42	0.17	4.59	
320	0.25	P90	320	4.00	3.05	7.06	5.67
	0.50	P50	320	4.00	1.50	5.50	
	0.25	P10	320	4.00	0.65	4.65	
160	0.25	P90	160	5.20	1.85	7.04	6.28
	0.50	P50	160	5.20	1.02	6.22	
	0.25	P10	160	5.20	0.45	5.65	
80	0.25	P90	80	5.18	(0.43)	4.75	4.53
	0.50	P50	80	5.18	(0.66)	4.52	
	0.25	P10	80	5.18	(0.85)	4.33	

## B.5 Conclusions

- Two integrated reservoir models (single-well and multi-well simulation based) and decision models were built to quantify the uncertainty and identify optimal well spacing in Deep Basin (Gething) tight sands.
- The optimal well spacing for a \$5.5/Mscf gas price for both models is 160 acres without any downspacing recommended. This result is specific to the reservoir properties and economic assumptions used in this illustrative example and cannot be generalized.
- The multi-well reservoir model generated a better match with historical production data; however, it is more difficult and time consuming to apply.
- The methods and tools developed in this part of the project can be applied, with some modification, to other unconventional plays.

## C CASE STUDY 2: NORTHERN BARNETT SHALE PLAY

### C.1 Introduction

Discovered in 1980, the Barnett shale play located in the Fort Worth basin was one of the first successful shale gas plays developed using horizontal wells and slick-water fracturing techniques. Our study area is in the northern part of the Barnett shale play, and consists of those parts of Montague, Cooke and Wise counties that are southwest of the Muenster Arch and have latitudes greater than  $33.4^\circ$  (yellow area in **Fig. 31**). The northern Barnett play is located inside the oil window (vitrinite reflectance,  $R_o$ ,  $< 1.2$ ). The average producing GOR of the northern Barnett shale play is 10 Mscf/STB through the end of 2010 (**Fig. 32**).

The initial oil production rate of existing producing wells ranges from less than 1 STB/D to 1500 STB/D. This large uncertainty in shale well performance makes it difficult for operators to identify the optimal development strategy in the early stages of development. Our well database included 64 horizontal wells in the study area, selected based on the following criteria:

- Inside the northern Barnett shale study area,
- Horizontal well with known completion data and well spacing,
- Produced for at least six months with an established decline, and
- The least-squares fitted decline curve is a reasonable representation of the production profile.

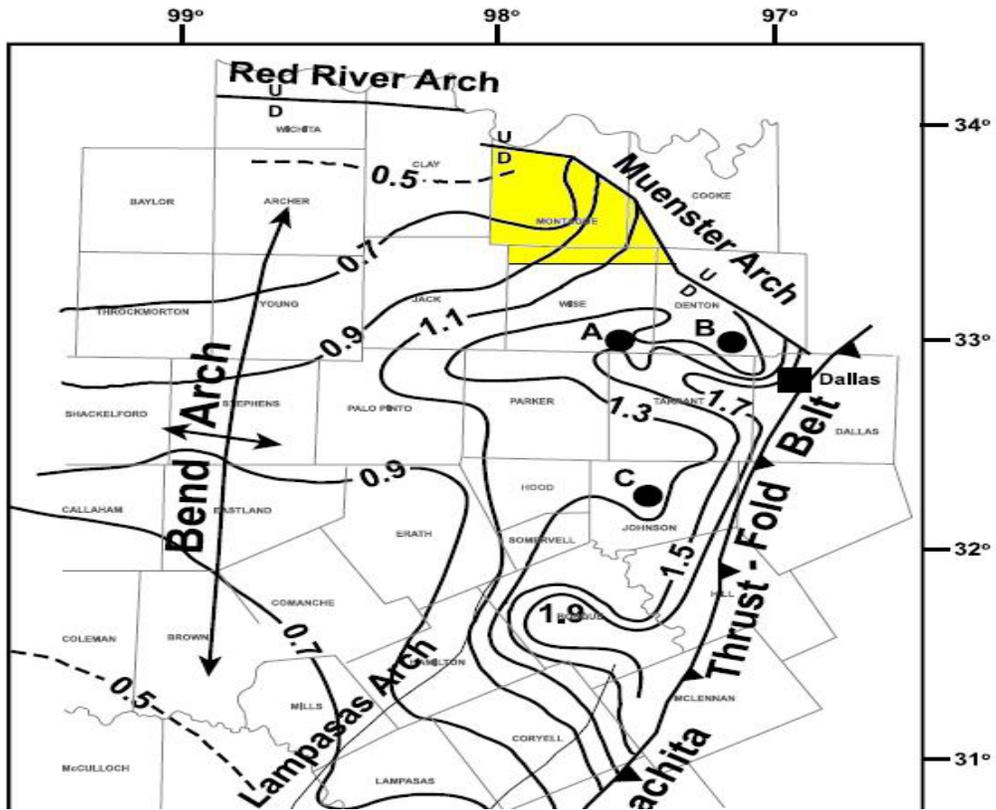
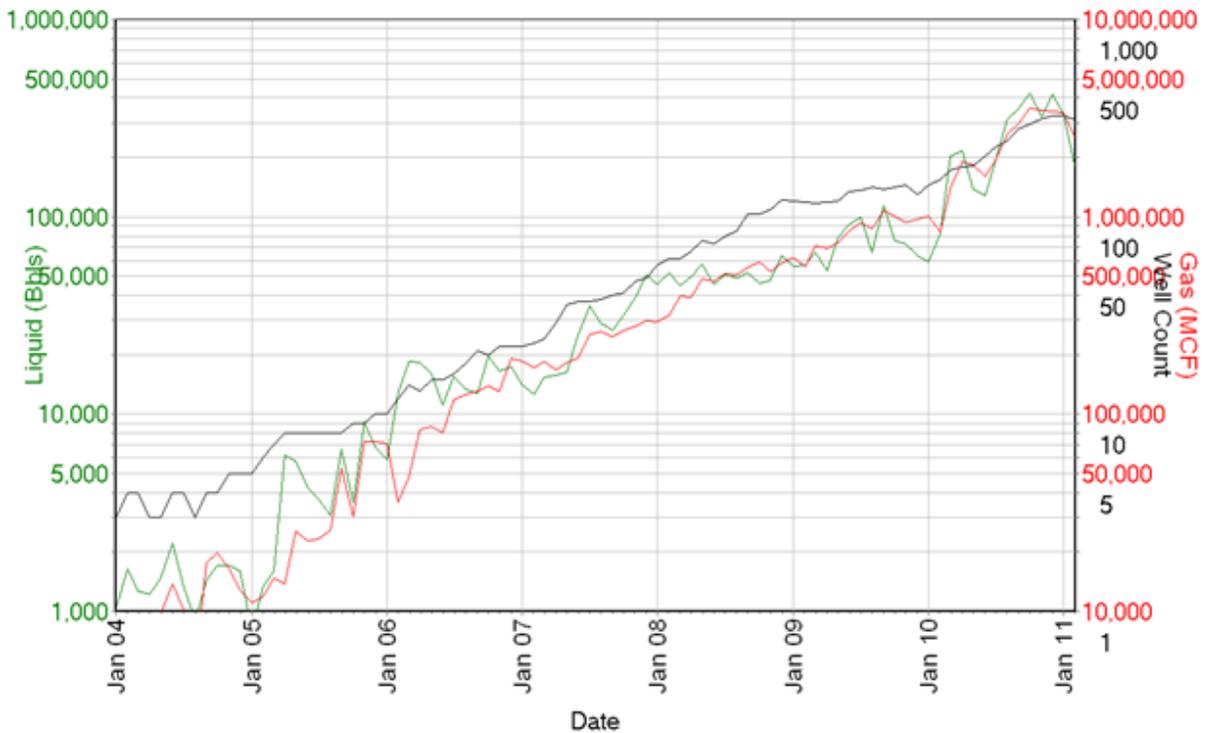


Fig. 31 – Barnett shale thermal maturity map. Northern Barnett shale study area (yellow) is located in the oil window. (Modified from Montgomery et al., 2005)



**Fig. 32 – Overview of monthly production history and drilling activity in the northern Barnett study area shows continuous increase of oil and gas production and drilling activity through the end of 2010.**

## C.2 Decline Curve Based Reservoir Model

### C.2.1 Decline Curve Parameters

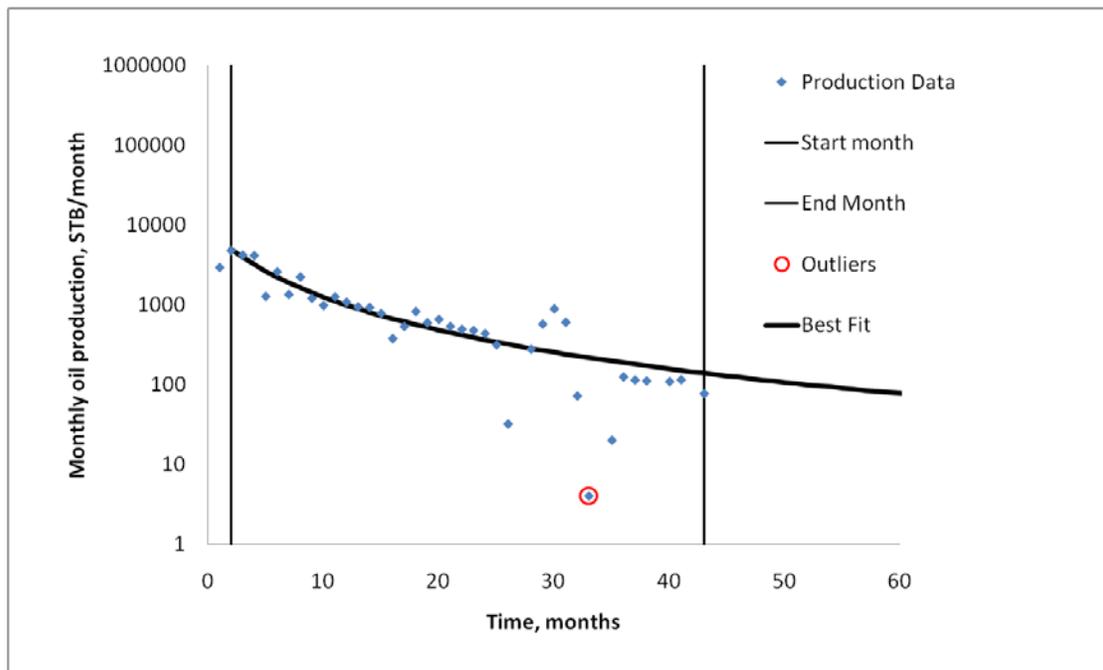
The monthly oil and gas production of 64 horizontal wells in the study area were extracted from a public database. Since the study area is inside the oil window and oil is more economic than natural gas under current conditions, the monthly oil production and monthly equivalent oil production (6 Mscf = 1 STB) were fitted with Arps' decline curves. Because the  $R^2$  values of the oil production decline curve parameters were greater than the  $R^2$  values of equivalent oil production decline curve parameters when correlated with decision/reservoir parameters (**Table 14**), monthly oil production is used as the primary production phase in the rest of the paper and monthly gas production is calculated by multiplying monthly oil production by GOR. Some of the 64 wells have been restimulated; however, decline curve analysis assumes the operating condition of the producing well remains the same. Thus, for restimulated wells we truncated the production data at the time of restimulation in the decline

curve analysis. The production data were fitted with hyperbolic decline curves using least-squares regression of the logarithm of production versus time. Influential outliers were ignored (Fig. 33). The decline exponent  $b$  was constrained between 0 and 2, while the initial decline rate was constrained to between 0 and 50 1/year.

**Table 14 – R<sup>2</sup> Between Decline Curve Parameters and Decision/Reservoir Parameters**

	Fluid	Prop	Stages	Interval	WS1	WS2	WSA	Ro	Thick	Depth	Fluid_Interval	Prop_Interval
Oil_qi	0.01	0.08	0.02	0.15	0.03	0.08	0.06	0.00	0.17	0.05	0.07	0.04
Oil_In(qi)	0.04	0.20	0.05	0.29	0.11	0.17	0.19	0.03	0.06	0.01	0.13	0.14
Oil_Di	0.00	0.03	0.03	0.03	0.07	0.08	0.05	0.07	0.00	0.01	0.14	0.03
Oil_b	0.02	0.02	0.00	0.01	0.03	0.00	0.00	0.03	0.00	0.02	0.02	0.01
Oil_CP6to1	0.01	0.21	0.06	0.18	0.20	0.13	0.13	0.00	0.01	0.00	0.19	0.18
Eq_qi	0.05	0.11	0.03	0.21	0.05	0.13	0.10	0.01	0.13	0.07	0.06	0.07
Eq_In(qi)	0.11	0.25	0.08	0.36	0.13	0.22	0.23	0.00	0.03	0.02	0.11	0.18
Eq_Di	0.03	0.01	0.00	0.02	0.04	0.01	0.00	0.00	0.07	0.15	0.01	0.00
Eq_b	0.02	0.04	0.00	0.03	0.00	0.01	0.01	0.02	0.14	0.08	0.00	0.02
Eq_CP6to1	0.01	0.03	0.02	0.02	0.10	0.04	0.02	0.00	0.00	0.02	0.06	0.03

**Note: Black numbers indicate positive relationship, red numbers indicate negative relationship, and yellow highlighted numbers were selected for use in the linear regression models.**



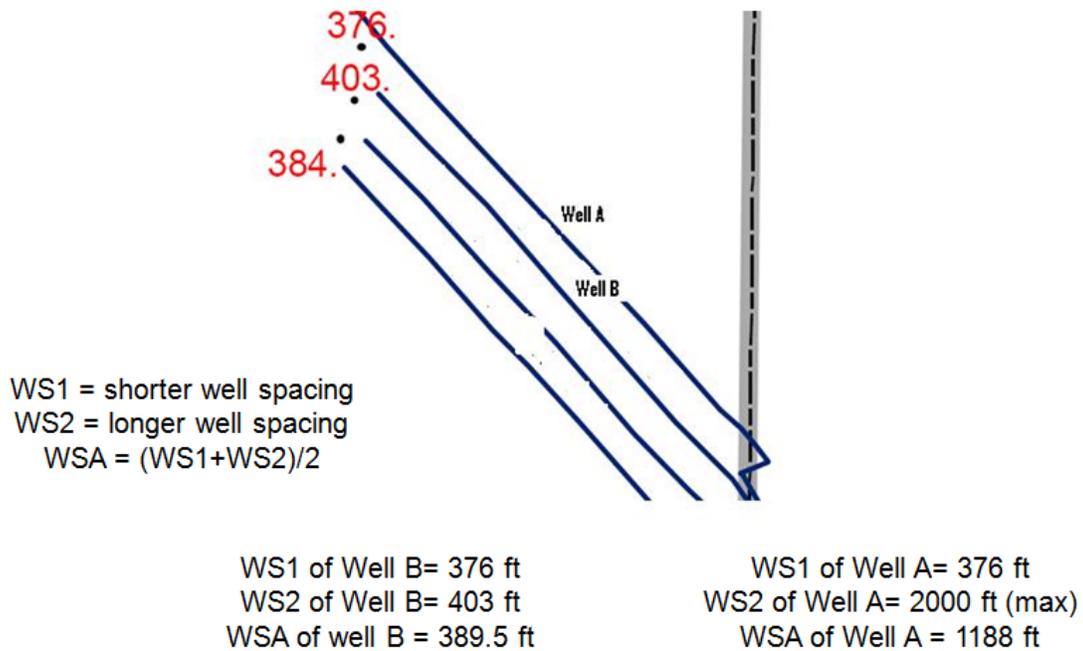
**Fig. 33 – Decline curve fit historical production from 2<sup>nd</sup> month (peak oil in first year) to 43<sup>rd</sup> month (end of historical production).**

### C.2.2 Decision Parameters

Most shale gas wells currently are drilled horizontally and stimulated with multiple fractures. The decisions to be made with regard to drilling, completion and stimulation of these wells include length of perforated interval, horizontal well spacing, fracturing fluid volume, proppant volume, number of stages, and others. We call these parameters “decision parameters.” Reservoir properties that affect production performance but cannot be controlled by the operator, such as formation thickness, formation depth, thermal maturity and others, are termed “reservoir parameters.”

Many decision parameters, such as perforated interval, amount of fracturing proppant, and volume of fracturing fluid, can be extracted from public databases for existing wells. However, reservoir parameters are not typically available because data such as core and log data are not available from public data sets. Some reservoir parameters used in this study, such as formation thickness, were estimated from reservoir property maps presented in the literature, e.g., Montgomery et al. (2005). We digitized the reservoir property maps and interpolated reservoir parameters at each well location.

A well parameter that has significant impact on production performance is well spacing. In trying to quantify this impact, we considered three of well spacing parameters: the shorter well spacing, the longer well spacing, and average well spacing. The shorter well spacing (WS1) is the smaller of the two spacings between the well and the two neighbor wells on either side, the longer well spacing (WS2) is the greater of the two spacings, while the average well spacing (WSA) is the average of WS1 and WS2 (**Fig. 34**). For well spacings greater than 2000 ft, 2000 ft was assigned as the well spacing because we believe well spacing greater than 2000 ft will not make any difference on production.



**Fig. 34 – Definition and example of shorter well spacing (WS1), longer well spacing (WS2), and average well spacing (WSA)**

### C.2.3 Linear Regression Models

The conventional decline curve parameters initial production rate ( $q_i$ ), initial decline rate ( $D_i$ ), and decline exponent  $b$  were used as the dependent variables for linear regression initially. However, the  $R^2$  values between  $q_i$  and  $D_i$  and the decision/reservoir parameters were low. Since the initial rate  $q_i$  has a wide range and the linear regression model was dominated by the high-initial-rate wells, we substituted  $q_i$  with  $\ln(q_i)$  to obtain a better correlation with the decision parameters. The ratio of cumulative production at 6 months to 1 month ( $CP6to1$ ) was introduced as a proxy for  $D_i$ , the instantaneous decline rate at time zero (Eq. 5). Because  $D_i$  can range widely and have extremely high values, it is not a good parameter for linear regression. Because  $CP6to1$  is related to instantaneous decline rate but is averaged over a longer time period, the  $R^2$  between  $CP6to1$  and decision parameters is higher than the  $R^2$  between  $D_i$  and decision parameters.

$$CP6to1 = \frac{\left(1 - \left(1 + 6 \frac{D_i}{12} b\right)^{1-\frac{1}{b}}\right)}{\left(1 - \left(1 + \frac{D_i}{12} b\right)^{1-\frac{1}{b}}\right)} \dots \dots \dots (5)$$

Linear regression was used to correlate decision and reservoir parameters with decline curve parameters. Decline curve parameters  $\ln(q_i)$ ,  $CP6to1$  and  $b$  were modeled as linear functions of decision and reservoir parameters with errors (Eqs. 6, 7, and 8). The errors between the linear regression models and the least-squares fitted decline curve parameters are assumed to have a joint multi-normal distribution (Eq. 9).

$$\ln(q_i) = 3.782191508 + 0.000346087 \times \text{Perforated interval} \dots \dots \dots (6)$$

$$CP6to1 = \begin{cases} -0.75656431 + 0.001495744 \times 2000 + 0.022245455 \times \frac{\text{Fluid}}{\text{Interval}} + \varepsilon_{CP6to1}(WS1 \geq 2000) \\ 1.08676577 + 0.001495744 \times WS1 + 0.022245455 \times \frac{\text{Fluid}}{\text{Interval}} + \varepsilon_{CP6to1}(WS1 < 2000) \end{cases} \dots (7)$$

$$b = 0.9715 + \varepsilon_b \dots \dots \dots (8)$$

$$\begin{pmatrix} \varepsilon_{\ln(q_i)} \\ \varepsilon_{CP6to1} \\ \varepsilon_b \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.74618 & -0.27053 & -0.08429 \\ -0.27053 & 0.5073 & 0.13709 \\ -0.08429 & 0.13709 & 0.31539 \end{pmatrix} \right) \dots \dots \dots (9)$$

Table 14 shows the  $R^2$  values of each pair of the decision/reservoir parameters with decline curve parameters resulting from fits of either oil production data or equivalent oil (oil plus gas) production data. The linear relationships between decline curve parameters and decision/reservoir parameters selected for use in the linear regression models are highlighted in yellow. The crossplots between decline curve parameters and decision/reservoir parameters in Eqs. 2, 3, 4, and 5 are shown in **Fig. 35**, **Fig. 36**, and **Fig. 37**. Statistics like  $R^2$ ,  $p$ -value,  $F$ -value, and  $t$ -value for every parameter included in the linear models (Eqs. 2, 3, and 4) can be found in the SAS output tables in Appendix A.

In the regression model the parameter  $\ln(q_i)$  is related only to the length of perforated interval, with a  $R^2 = 0.2887$  (Fig. 35).  $CP6to1$  is related to the shorter well spacing and fluid volume per perforated interval. Based on the linear regression model, when the shorter well

spacing and the volume per perforated interval are greater, the CP6to1 will be greater; i.e., the average decline rate over the first 6 months will be smaller (Fig. 36 and Fig. 37).

In Table 14, all of the  $R^2$  values are relatively low. We typically hope to see  $R^2$  values for physical and engineering relationships that exceed 0.8. There are several reasons for the relatively low  $R^2$  values for this linear regression model. First, the relationships between decline curve parameters and decision parameters are not as straightforward as many physical and engineering relationships. Second, the uncertainty of horizontal drilling in shale plays is large, to which any operator will attest. When the dependent parameters (decline curve parameters) are functions of multiple factors (decision/reservoir parameters and more), the  $R^2$  for each parameter with the dependent parameters will be low. Although the  $R^2$  values are relatively low, another statistical measure, the  $p$ -value, shows the linear model is statistically significant. The  $p$ -value represents the probability that the corresponding independent variable (decision/reservoir parameters) and dependent variables (decline curve parameters) are not linearly related. According to the statistical output tables in Appendix A, every decision/reservoir parameter included in the proposed model has a  $p$ -value less than 5%, which means there is a 95% chance that the linear relationship between each decision/reservoir parameter and selected decline curve parameter holds.

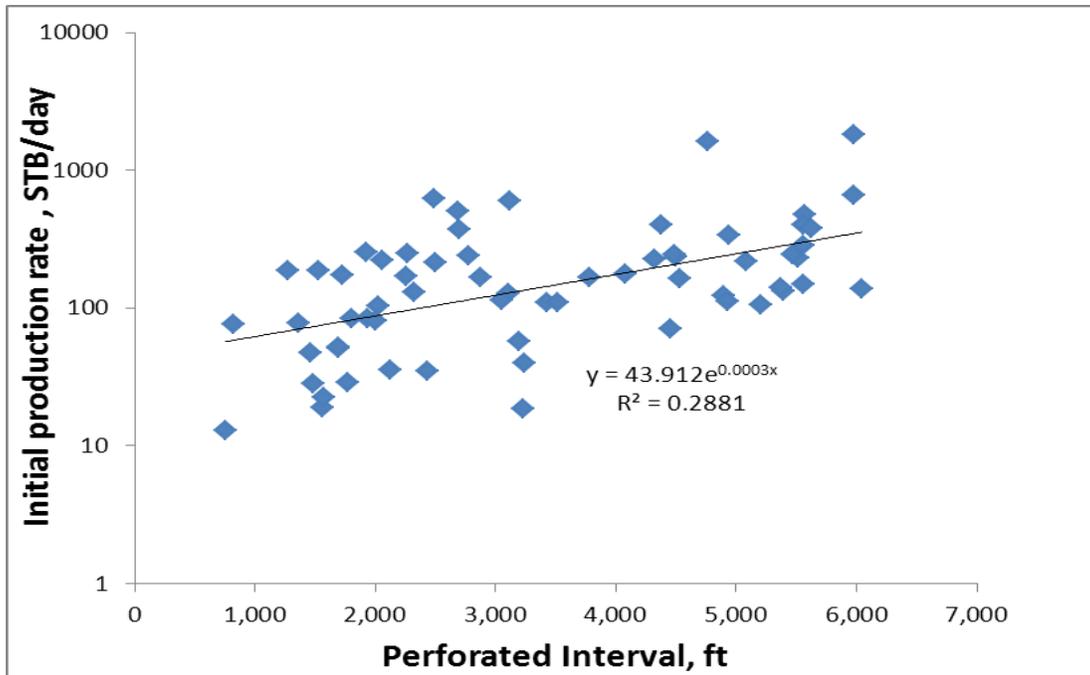


Fig. 35 – Initial production rate is positively related to length of perforated interval with  $R^2 = 0.2881$ .

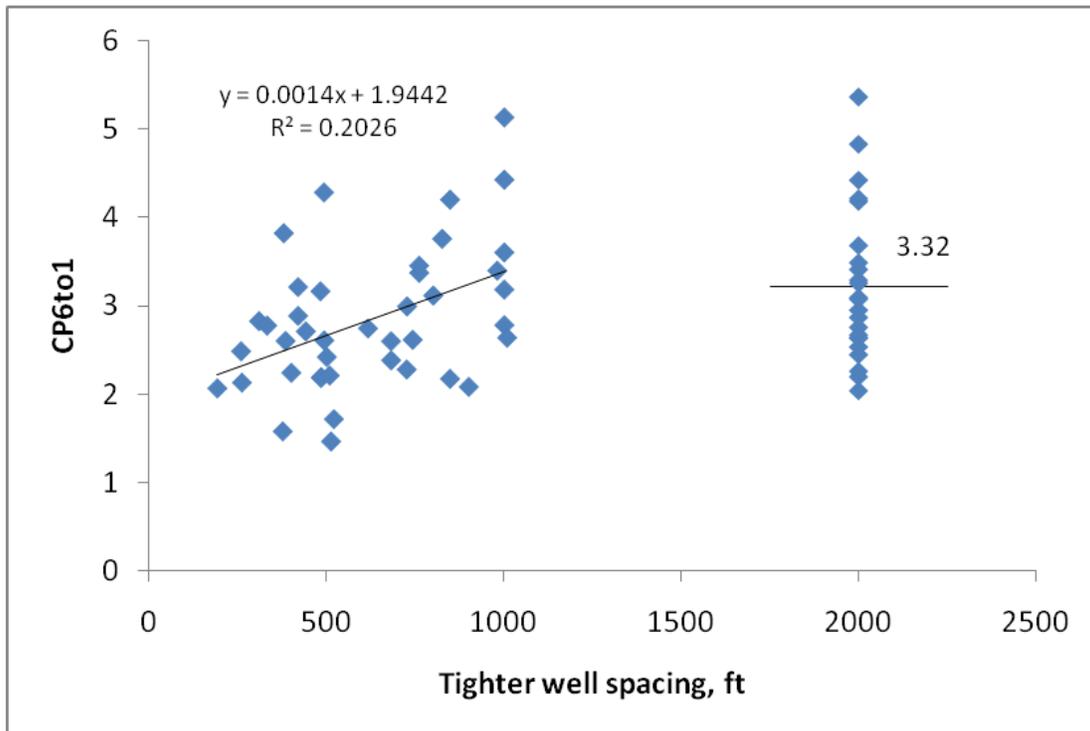
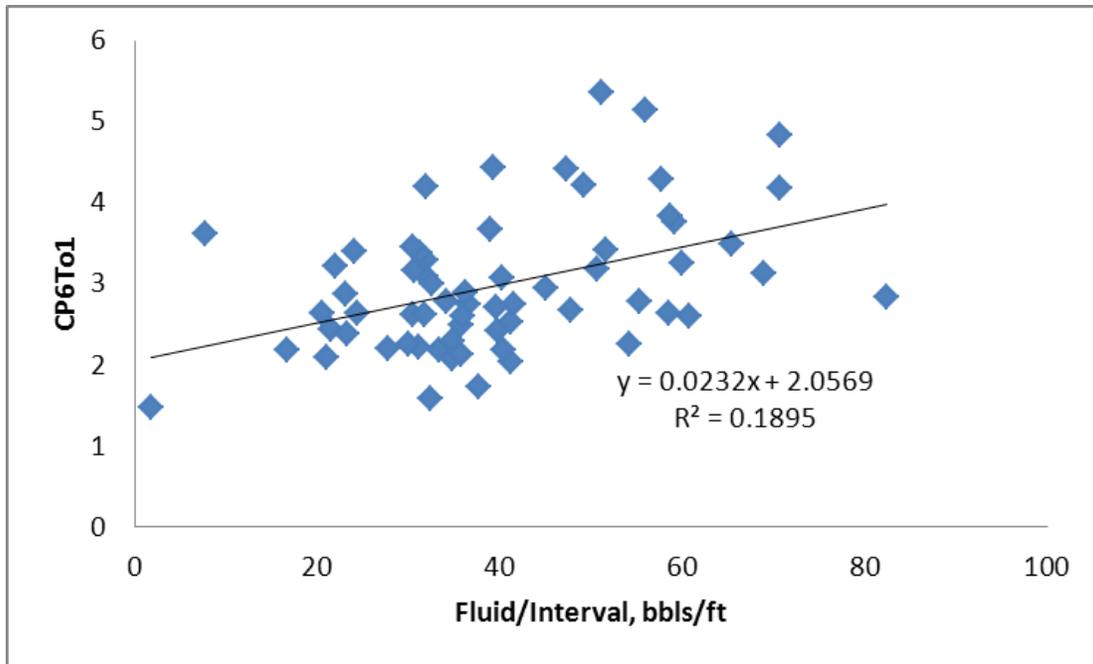


Fig. 36 – CP6to1 is positively related to shorter well spacing (WS1); average CP6to1 is 3.22 when shorter well spacing is greater than 2000 ft.



**Fig. 37 – CP6to1 is positively related to fluid per perforated interval with  $R^2 = 0.1895$ .**

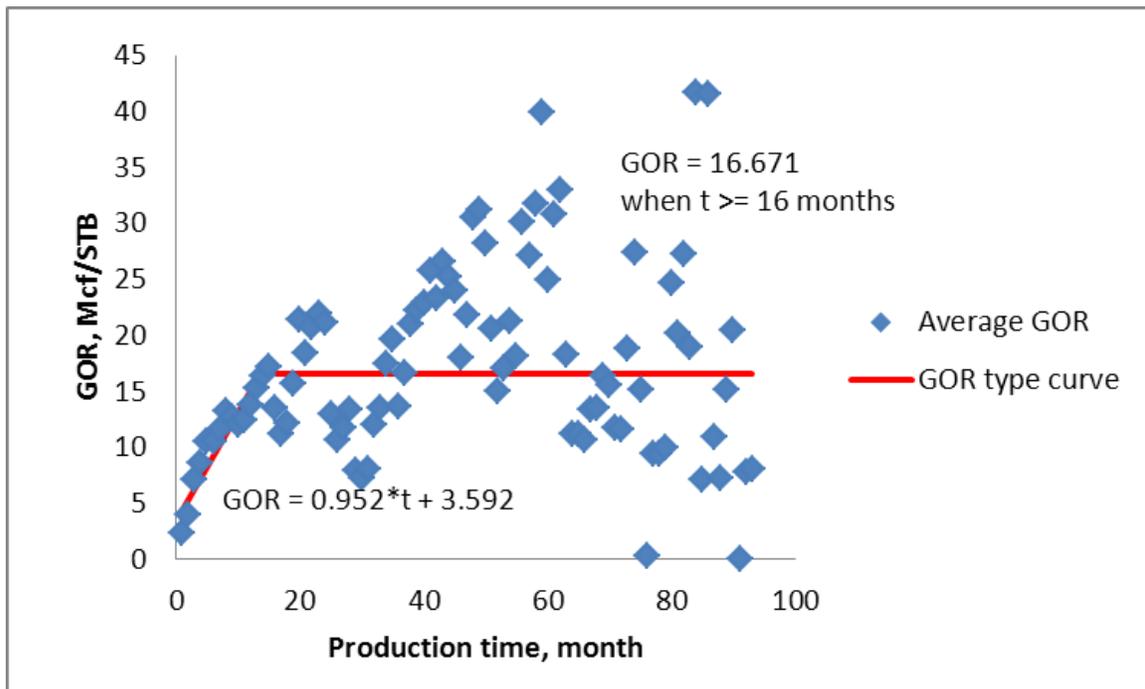
#### C.2.4 GOR Model

Since we elected to build the regression model in terms of oil production, we needed to be able to determine gas production from oil production. We first built a GOR type curve based on the average monthly GOR of the 64 horizontal wells (Fig. 38). The average monthly GOR of the 64 horizontal wells shown in Fig. 38 is the summation of gas production of the 64 wells divided by the summation of oil production of the 64 wells for each month. We assume the GOR curve of any well in the study area has the same shape as the type curve but with a different initial GOR. Second, we correlated the initial oil percentage (Eq. 10) with thermal maturity digitized from the thermal maturity map in Fig. 39 (correlation shown in Fig. 9). The error between the modeled initial oil percentage and actual initial oil percentage is assumed to be normally distributed and has a standard deviation = 0.2 (Eq. 11).

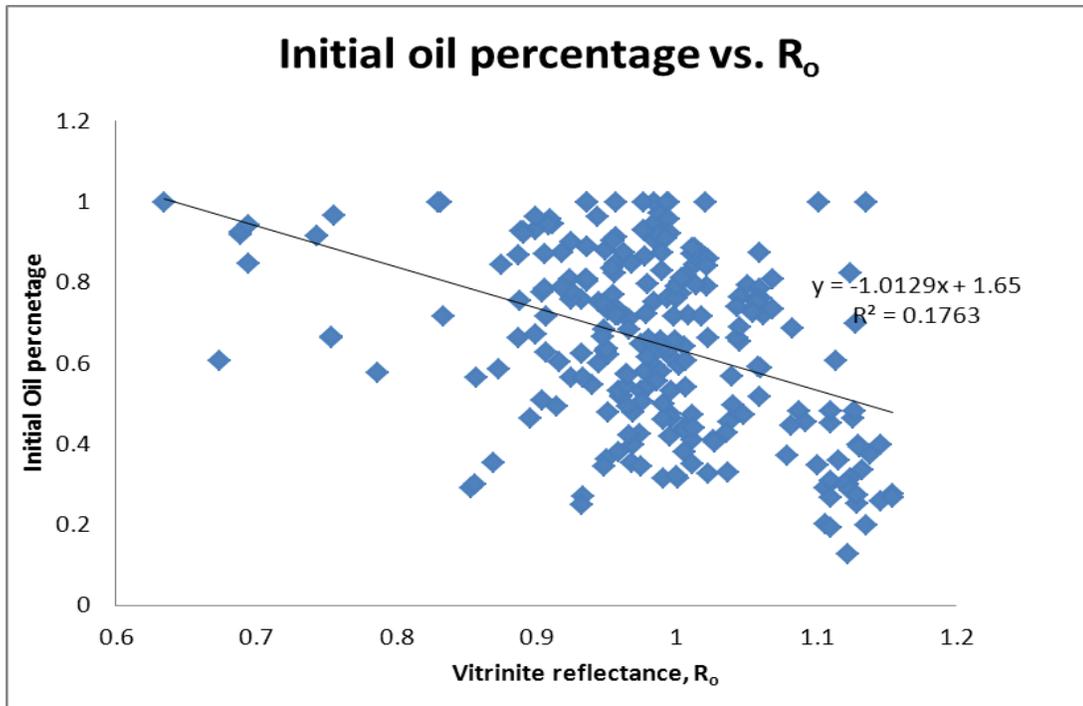
$$\text{Oil percentage} = (\text{oil production (STB)} + \text{gas production}/6(\text{Mscf})) \dots \dots \dots (10)$$

$$\text{Initial oil Percentage} = -1.0129 * R_0 + 1.65 + \varepsilon_o, \text{ where } \varepsilon_o \sim N(0, 0.201064) \dots \dots \dots (11)$$

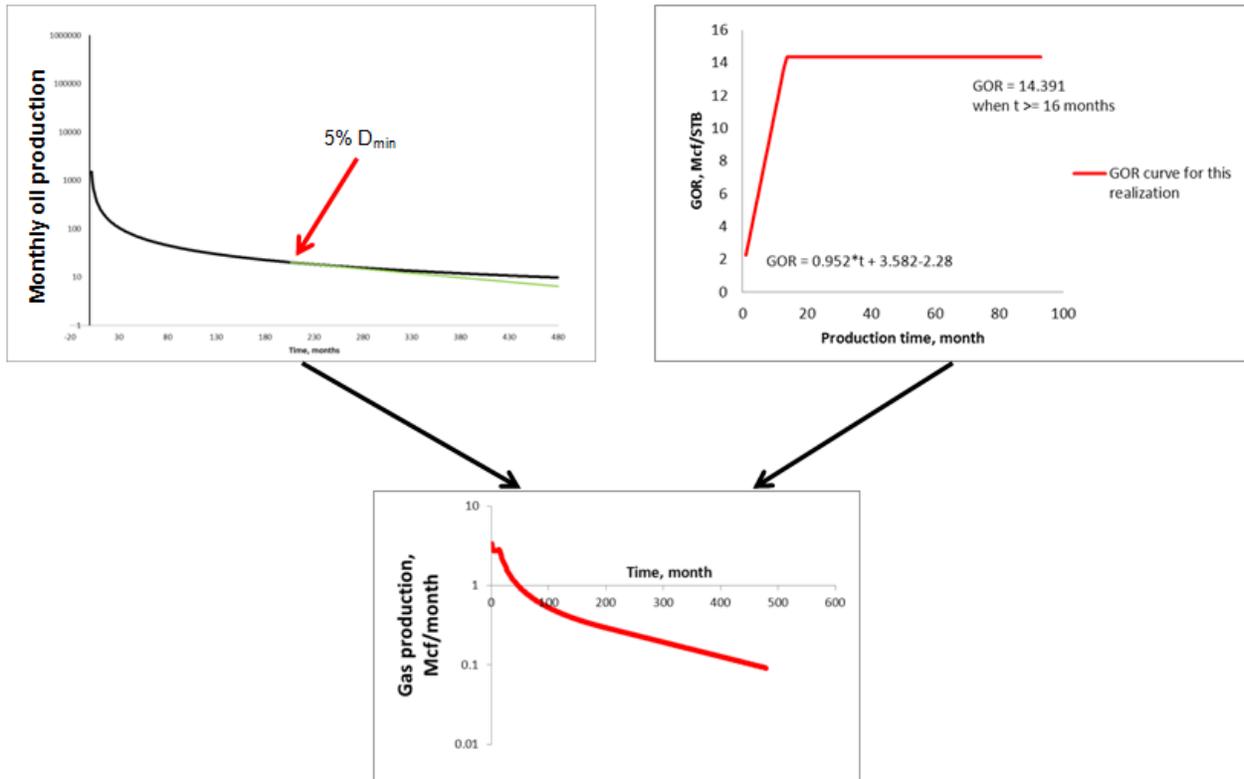
Monte Carlo simulation is used to generate oil and gas production forecasts in the decision model. For each realization, the decline curve parameters for oil production are based on Eqs. 2 through 5. The oil production forecasts are then generated using the decline-curve parameters and a 5% minimum decline rate. To obtain gas production forecasts, an error was first randomly generated from the normal distribution in Eq. 11 to obtain the initial oil percentage. The initial GOR was then calculated from the initial oil percentage (Eq. 10). The type curve in Fig. 38 is shifted vertically to match the initial GOR, and the shifted GOR is then combined with the oil production forecast to generate the gas production forecast (Fig. 40). The oil and gas production forecasts generated using Monte Carlo simulation become inputs to the decision model for identifying optimal development strategies.



**Fig. 38 – GOR type curve fitted to average GOR of 64 wells**



**Fig. 39 – Initial oil percentage is negatively related to vitrinite reflectance ( $R_o$ ).**



**Fig. 40 – Gas production forecast generated by multiplying oil production forecast and GOR curve.**

### C.2.5 Monte Carlo Simulation of Decline Curve Model

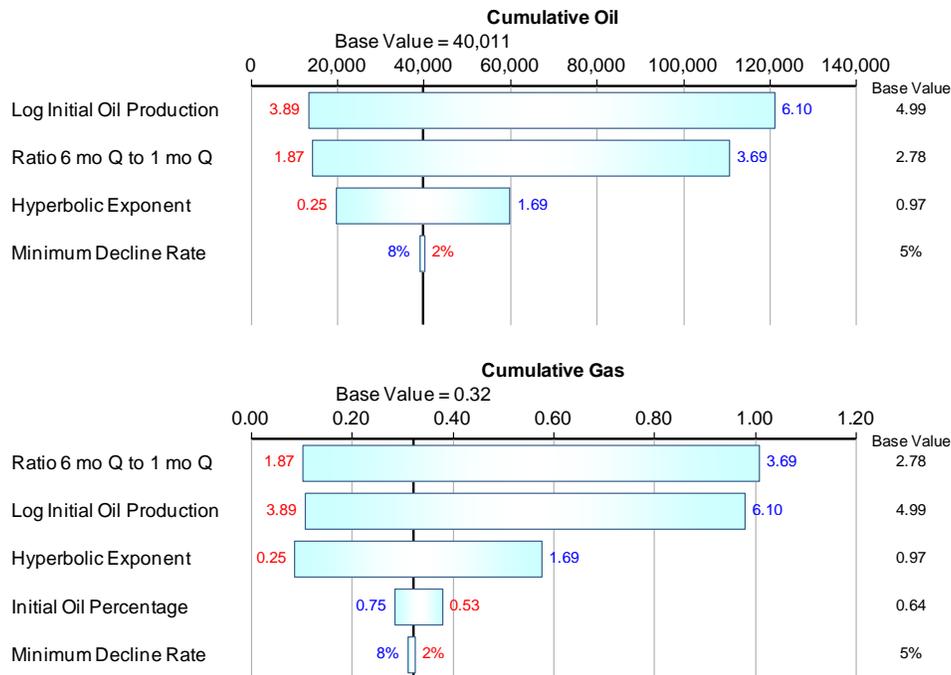
In this section, we use the decline curve model developed above, including the associated uncertainty, to perform a Monte Carlo simulation to determine the relative influence of decline-curve parameters on uncertainty in production. We consider a single well with average decision parameters determined from the study area: perforated interval of 3500 ft, 600-ft spacing, and 125,000 of stimulation fluid pumped. We do not vary any of the decision variables in this particular analysis, so we thus consider only uncertainty in the decline curve model.

Based on the regression analysis described above, we established P10, P50, P90 values for the decline curve parameters  $\ln(q_i)$ ,  $CP6to1$ ,  $b$ , and the initial oil percentage (Table 15). We also imposed a minimum decline rate for long-term production forecasts. Minimum decline rate was not derived from the regression analysis above because no wells had produced long enough to establish this value. P10, P50 and P90 values for minimum decline rate are also shown in Table 15.

**Table 15 – Reservoir Model Input Uncertainty**

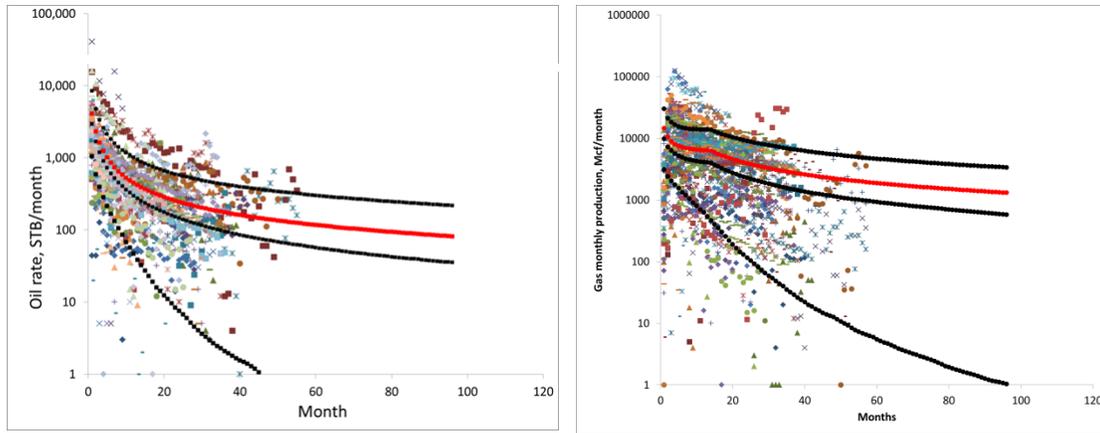
Variable	P10	P50	P90
Log Initial Oil Production	3.89	4.99	6.10
Ratio 6 mo Q to 1 mo Q	1.87	2.78	3.69
Hyperbolic Exponent	0.25	0.97	1.69
Minimum Decline Rate	2%	5%	8%
Initial Oil Percentage	53%	64%	75%

We can better understand the relative importance of each uncertainty by computing the cumulative oil and gas production at 25 years for the set of values for each parameter in Table 2. We plot these results in the “tornado” diagrams show in Fig. 41. In this case, we have run each input through its uncertainty range while holding all other variables at their P50, or base case. Taking each variable one at a time, we see that cumulative oil production could range from around 20,000 STB to 120,000 STB. Gas production is also highly uncertain, with a range of between 0.1 to 1.0 Bscf, or an order of magnitude. Uncertainty in 25-year cumulative production of both oil and gas is driven primarily by uncertainty in initial rate and initial decline rate, and to a lesser extent by the decline exponent.



**Fig. 41 – Tornado diagrams for cumulative oil and gas production at 25 years.**

The analysis summarized in Fig. 41 assumes that each input varies one at a time. We next allowed the inputs to vary simultaneously, and included the correlations described in Equation 5. The distribution of production profiles generated from the decline curve regression model are superimposed on actual oil and gas production profiles for all study area wells in Fig. 42. The good match indicates the decline curve regression model is reasonable and represents the uncertainty in production in the study area. The P10, P50 and P90 values, along with the mean, from the calculated results are presented in Table 16. The uncertainty ranges are wider than what was observed in Fig. 41, since we are varying all of the inputs simultaneously. Again, we see that production uncertainty is large; the range spans more than an order of magnitude.



**Fig. 42 – Decline curve regression model fits actual study-area oil and gas production profiles well.**

**Table 16 – Cumulative Oil and Gas Uncertainty Range**

Output	P10	P50	P90	Mean
Cumulative Oil Production (STB)	10,691	37,190	135,116	61,062
Cumulative Gas Production (BCF)	0.05	0.28	1.25	0.53

### C.3 Decision Analysis

In this section we show the results of Monte Carlo simulation that relate production performance to well development decisions, in particular, perforated interval, well spacing, and stimulation fluid volume (Table 17, Table 18, and Table 19 respectively). To allow proper comparison of performance between wells of different lengths and well spacings, and thus different drainage areas, we normalize production to a unit area 3500 ft by 1000 ft in size. This normalization essentially assumes that the horizontal laterals are laid end to end. The unit-area production reported in Table 17 – Table 19 for wells with average study-area properties (3500-ft perforated interval and 600-ft well spacing) are 1.67 times the production of a single well.

The impact of perforated interval on production performance is shown in Table 17. Well spacing and stimulation fluid volume are held constant at the study-area average values of 600 ft and 125,000 bbls, respectively. We show P10, P50, P90 and mean from the distributions of oil and gas production. Again, the ranges in production are wide, indicating significant uncertainty in well production response. Normalized 25-year production of both oil and gas increase as the perforated interval decreases. This may appear counterintuitive, since initial production rate increases with perforated interval (Fig. 35). However, in Fig. 37,  $CP_{6to1}$  increases with fluid

volume per foot of perforated interval. Since most of the sensitivity in Fig. 37 is in the perforated interval, this means that  $CP6to1$  decreases with perforated interval. Thus, wells decline faster as the perforated interval increases. This is consistent with LaFollette and Holcomb's (2011) findings that production per foot declines as the perforated interval increases.

The impact of well spacing on production performance is shown in Table 18. Perforated interval and stimulation fluid volume are held constant at the study-area average values of 3500 ft and 125,000 bbls, respectively. Normalized 25-year production of both oil and gas increase as the well spacing decreases. There is little difference in production between 1000-ft and 600-ft well spacing. However, normalized production increases significantly between 600-ft and 200-ft well spacing. Fig. 36 shows that  $CP6to1$  decreases (i.e., decline rates increase) as well spacing decreases. However, this is apparently outweighed by the impact of having 3 times as many wells at 200-ft spacing than 600-ft spacing.

The impact of stimulation fluid volume on production performance is shown in Table 19. Perforated interval and well spacing are held constant at the study-area average values of 3500 ft and 600 ft, respectively. Normalized production increases with fluid volume, as expected. Production increases 30-35% as fluid volume is increased 50% from 100,000 to 125,000 bbls.

These results demonstrate that combining reservoir modeling with decision modeling can be helpful in relating production performance to decision parameters in the field and should aid in optimization of spacing, completion and stimulation decisions in this area. While instructive, these results cannot be used directly to make decisions. Economics can and should be incorporated into the integrated reservoir/decision model. It is possible that the greater costs per unit area associated with having shorter laterals and shorter spacing could outweigh the benefits in production shown in Table 17 and Table 18.

**Table 17 – Impact of perforated interval on oil and gas production at 25 years**

Perf Int (ft)	Cumulative Oil (STB)				Cumulative Gas (BCF)			
	mean	P10	P50	P90	mean	P10	P50	P90
4500	59,618	9,911	36,699	132,140	0.50	0.05	0.27	1.15
3500	61,062	10,691	37,190	135,116	0.53	0.05	0.28	1.25
2500	78,663	14,465	50,922	171,483	0.70	0.07	0.40	1.60

**Table 18 – Impact of well spacing on oil and gas production at 25 years**

Well Space (ft)	Cumulative Oil (STB)				Cumulative Gas (BCF)			
	mean	P10	P50	P90	mean	P10	P50	P90
1000	61,071	11,058	39,697	134,118	0.55	0.06	0.33	1.30
600	60,489	10,892	37,727	132,803	0.52	0.05	0.29	1.19
200	118,391	18,453	70,116	269,229	1.00	0.08	0.50	2.38

**Table 19 – Impact of stimulation fluid volume on oil and gas production at 25 years**

Fluid (bbls)	Cumulative Oil (STB)				Cumulative Gas (BCF)			
	mean	P10	P50	P90	mean	P10	P50	P90
150,000	70,929	12,396	43,074	154,016	0.78	0.08	0.43	1.74
125,000	61,851	11,239	37,779	132,022	0.67	0.06	0.34	1.53
100,000	53,677	9,352	33,205	118,135	0.58	0.06	0.30	1.39

## C.4 Updated Model Using a Larger Data Set

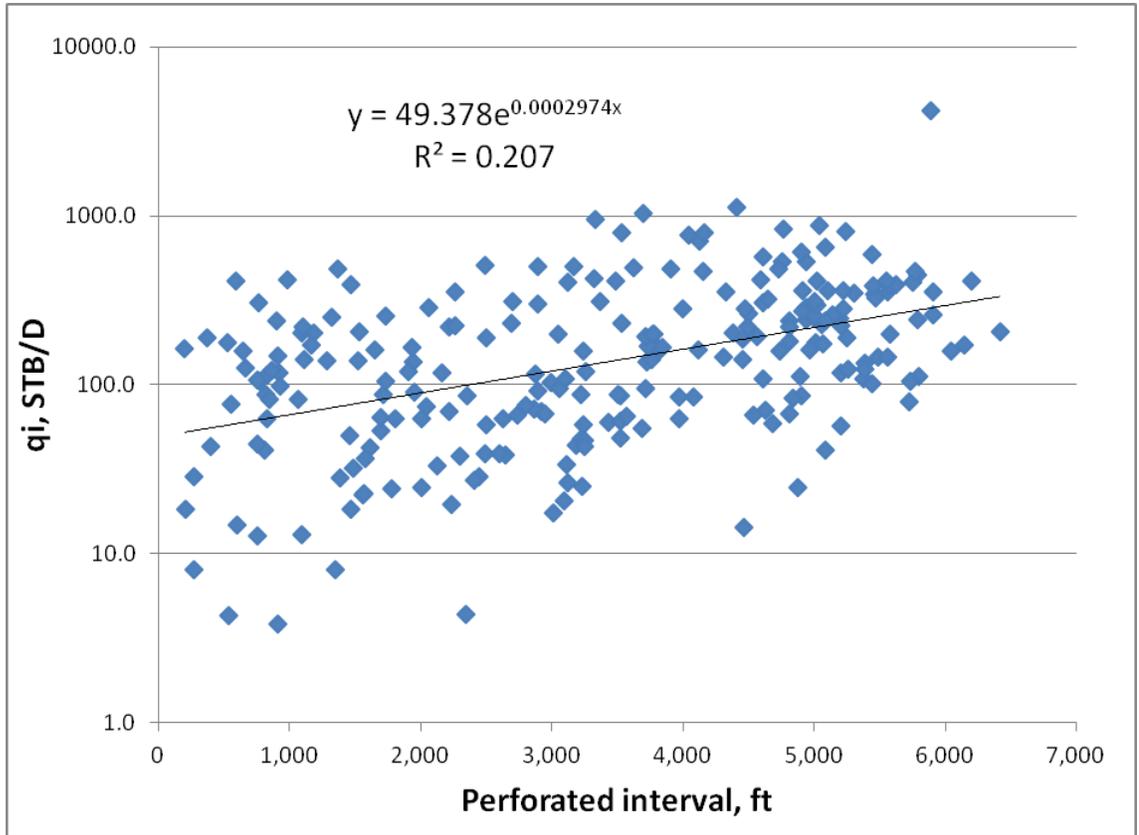
### C.4.1 Introduction

In this larger data set, we obtained 266 wells with 85 wells having well spacing information (compared with 64 total wells with 42 wells having well spacing data in the first data set) in the same area with the same selection criteria as the first data set (C.1). Unlike the first data set where average lease production was used, allocated production was available for individual wells in the second data set. Monthly well production data until January 2011 was analyzed using decline curve analysis.

### C.4.2 Decline Curve Based Reservoir Model

Deterministic decline curve fits were applied to the monthly oil production of the 266 wells. Decline-curve parameters  $\ln(q_i)$ ,  $CP6to1$ , and  $b$  were calculated based on the deterministic fits and then correlated with decision and completion parameters using linear regression.

According to Fig. 43,  $\ln(q_i)$  is positively related to perforated interval for the 266 wells. We observed the same relationship in the first data set, but with a different intercept and slope.



**Fig. 43 –  $\ln(q_i)$  is positively related to perforated interval (2<sup>nd</sup> data set)**

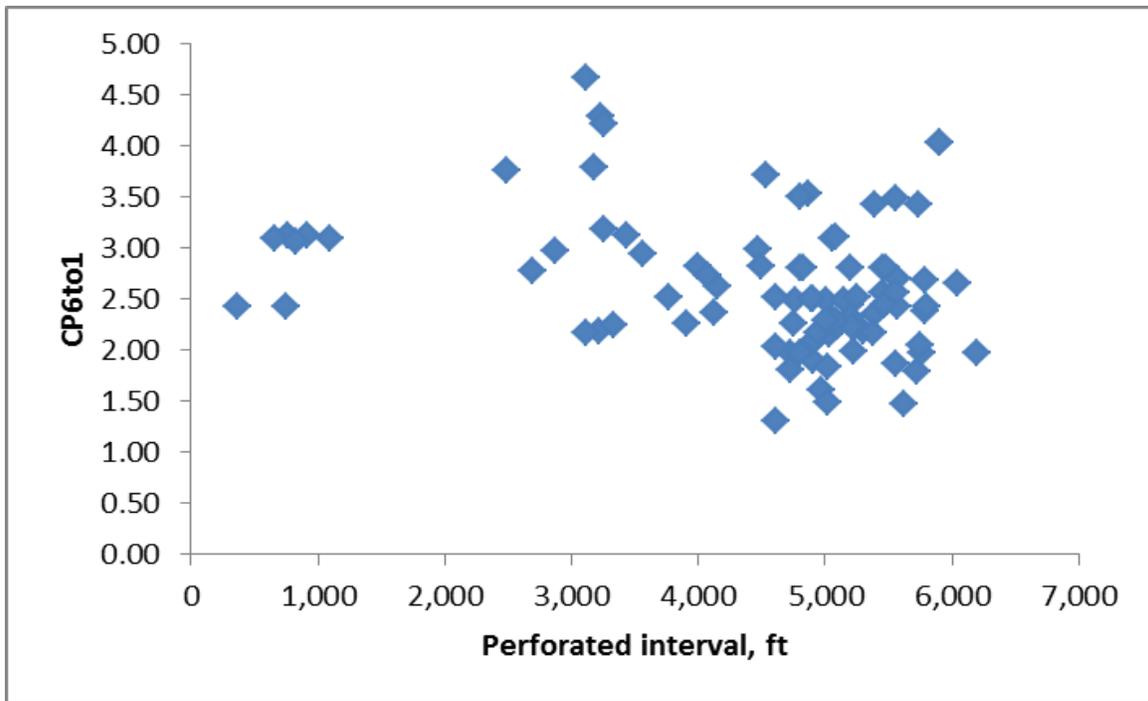
CP6to1 was plotted against perforated interval for the 85 wells with well spacing information (Fig. 44). Seven outliers with perforated interval less than 1500 ft were observed. They are not on the same trend in the plot as other data points, for reasons unknown. Since most operators are not targeting less than 1500-ft laterals in the northern Barnett shale reservoir, we decided to exclude the outliers in our analysis. After excluding the outliers, we observed a negative relationship between CP6to1 and perforated interval (Fig. 45). The first model exhibited a positive relationship between CP6to1 and fracturing fluid per perforated interval, where most of the correlation was in with perforated interval and less was with fracturing fluid volume. Thus, the second model is in agreement with the first model in that CP6to1 is negatively related to perforated interval.

The residual between CP6to1 and the linear regression model in Fig. 45 was calculated and named as CP6to1\_ex. CP6to1\_ex was then plotted against tighter well spacing (Fig. 46). There was no straight line relationship between CP6to1\_ex and tighter well spacing; thus, we

averaged CP6to1\_ex and CP6to1 by groups (Table 20). The average curve of CP6to1\_ex shows that CP6to1\_ex increases from 0 ft to 350 ft, decreases from 350 ft to 450 ft, and then levels up after well spacing exceeds 450 ft. A least-squares fit was then obtained using 340 ft as the peak and 450 ft as the leveling point (Fig. 46).

The decline exponent  $b$  was found to not be related to any decision/reservoir parameters, as with the first data set, but with different mean and standard deviation.

The regression models for  $\ln(q_i)$ , CP6to1,  $b$  and the covariance matrix between the residuals are shown in Eq. (12). The  $R^2$  between  $\ln(q_i)$  and perforated interval is 21%, the  $R^2$  between CP6to1 and both perforated interval and tighter well spacing is 20%



**Fig. 44 – CP6to1 vs. perforated interval. Seven outliers for wells with less than 1500-ft PI are observed.**

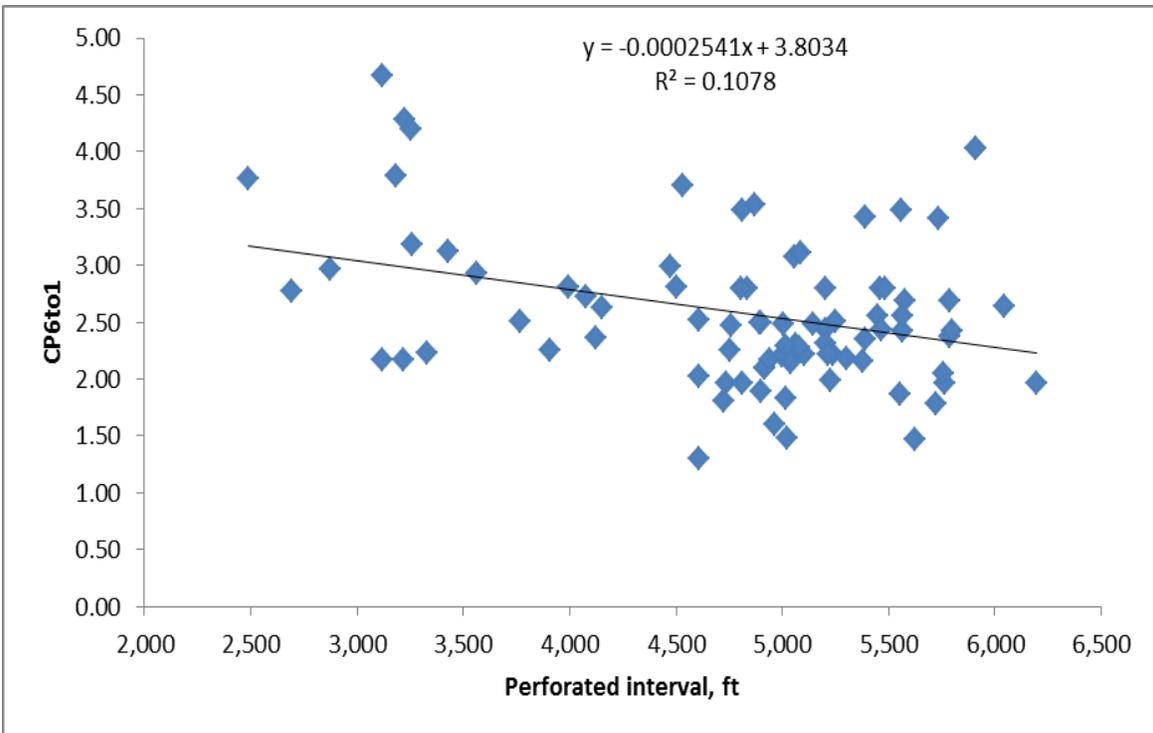


Fig. 45 – CP6to1 is negatively correlated to perforated interval, excluding 7 outliers.

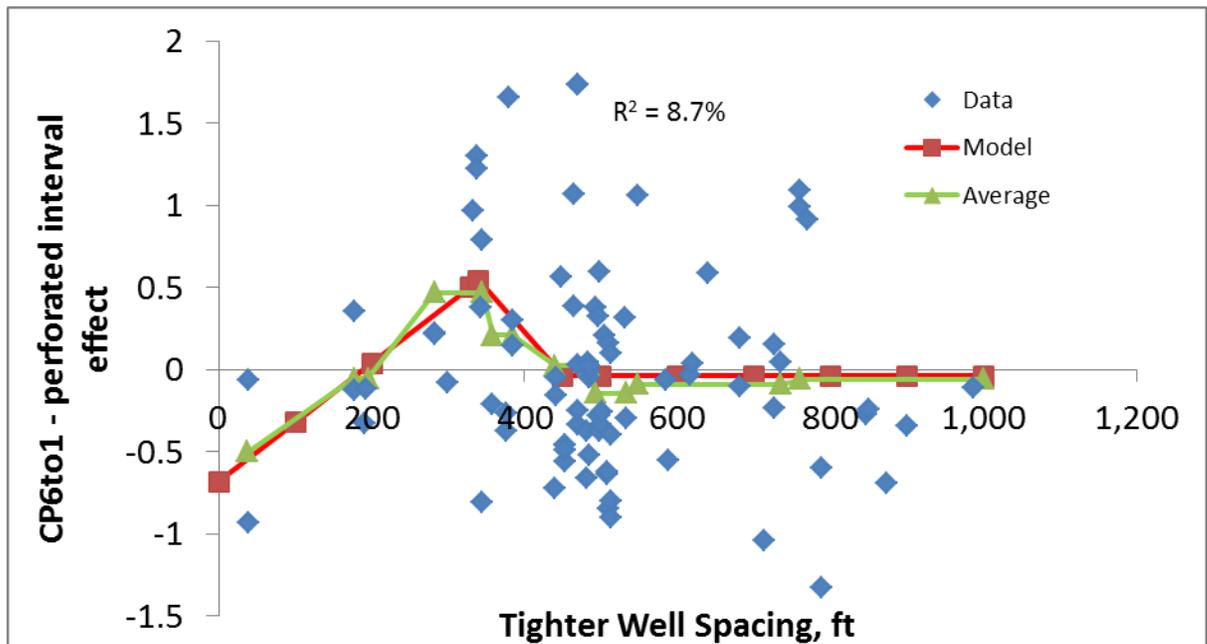


Fig. 46 – CP6to1\_ex peaks around 340 tighter well spacing

**Table 20 – CP6to1 and CP6to1\_ex average used for Fig. 47**

Well spacing	<40	175-195	282-345	358-384	440-484	493-540	547-735	>760
Count	2	4	9	6	23	20	11	10
CP6to1_Ex	-0.5	-0.053	0.467	0.21	0.021	-0.144	-0.09	-0.06
CP6to1	2.05	2.5	3.19	2.69	2.53	2.34	2.75	2.53

$$\ln(q_i) = 3.8995 + 0.0002974 * PI + \varepsilon_{\ln q_i}$$

$$CP6to1 = 4.137 - 0.0002215 * PI \begin{cases} +0.003421326 * (WS - 340) + \varepsilon_{CP6to1}, & \text{if } WS < 340 \\ -0.00507343 * (WS - 340) + \varepsilon_{CP6to1}, & \text{if } 340 < WS < 450 \\ -0.00507343 * (450 - 340) + \varepsilon_{CP6to1}, & \text{if } WS > 450 \end{cases}$$

$$b = 0.8625 + \varepsilon_b$$

$$\begin{pmatrix} \varepsilon_{\ln q_i} \\ \varepsilon_{CP6to1} \\ \varepsilon_b \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.953 & -0.270 & -0.127 \\ -0.270 & 0.375 & 0.027 \\ -0.127 & 0.027 & 0.339 \end{pmatrix} \right)$$

.....(12)

According to Fig. 47, there are significant differences between the two data sets in the relationship between CP6to1 and tighter well spacing. First, the wells are different. The second data set has 85 wells with well spacing information while the first data set has 42 with well spacing information, and not all the 42 wells in the first data set are included in the second data set. Second, the production data are from two different public data bases: the first data set is from HPDI and the second data set is from IHS. Third, IHS provided allocated individual-well production, while there is production from 8 multi-well leases in the first data from HPDI. The differences in the two data sets have resulted in the difference between the two regression models for CP6to1 vs. tighter well spacing. (Fig. 36 and Fig. 46, Eq.(7) and Eq.(12)).



Fig. 47 – Comparison between first and second data sets in CP6to1 vs. tighter well spacing

### C.4.3 GOR Model

The average monthly GOR of the 243 horizontal wells equals the summation of the gas production of all wells divided by the summation of oil production of all wells in the corresponding month. A GOR type curve was fit to the average monthly GOR in Fig. 48. We assume the GOR curve for each of the 243 horizontal wells will have the same shape as the type curve but with a different initial GOR. A correlation between oil percentage of peak three months’ production and thermal maturity was constructed (Fig. 49) for modeling the initial GOR of each well.

$$\text{Oil percentage} = (\text{oil production (STB)} + \text{gas production}/6(\text{Mscf})) \dots \dots \dots (13)$$

$$\text{Peak three months oil Percentage} = -0.6033 * R_0 + 1.0753 + \varepsilon_o \dots \dots \dots (14)$$

Where  $\varepsilon_o \sim N(0, SD = 0.188579)$

For each realization of Monte Carlo simulation, a random number from the normal distribution of  $\varepsilon_o$  was substituted in Eq. 14 to calculate peak three-months oil percentage, then the GOR of the first three months was calculated based on peak three-months oil percentage (Eq. 13). The type curve in Fig. 49 will then shift vertically to match the GOR of the first three months. The shifted GOR can then be combined with the oil production forecast to generate a

gas production forecast (Fig. 50). The oil production and gas production forecasts will be used to calculate net present value in Monte Carlo simulation in decision analysis.

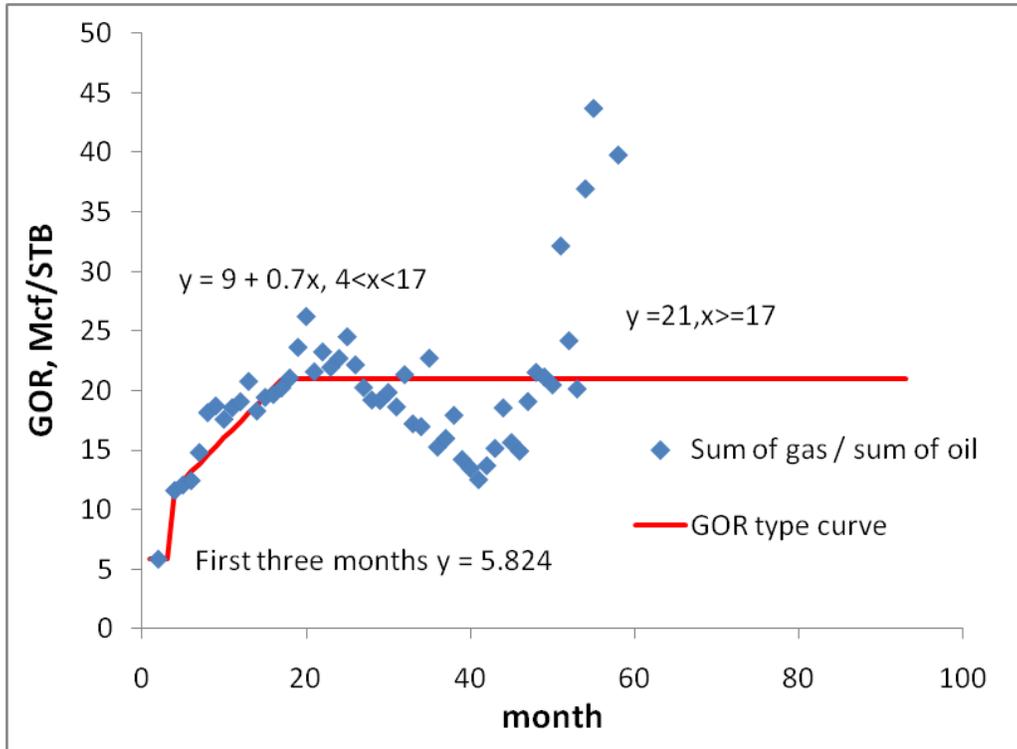


Fig. 48 – The average GOR was fit by a type curve.

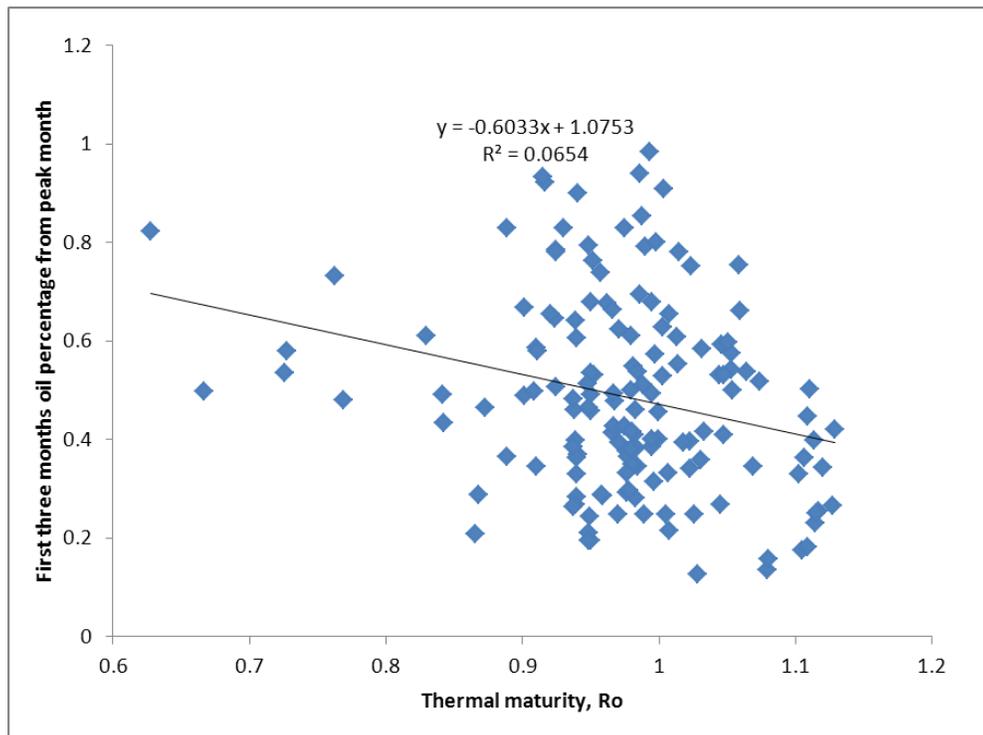


Fig. 49 – The peak three months' oil percentage were correlated with vitrinite reflectance ( $R_o$ ).

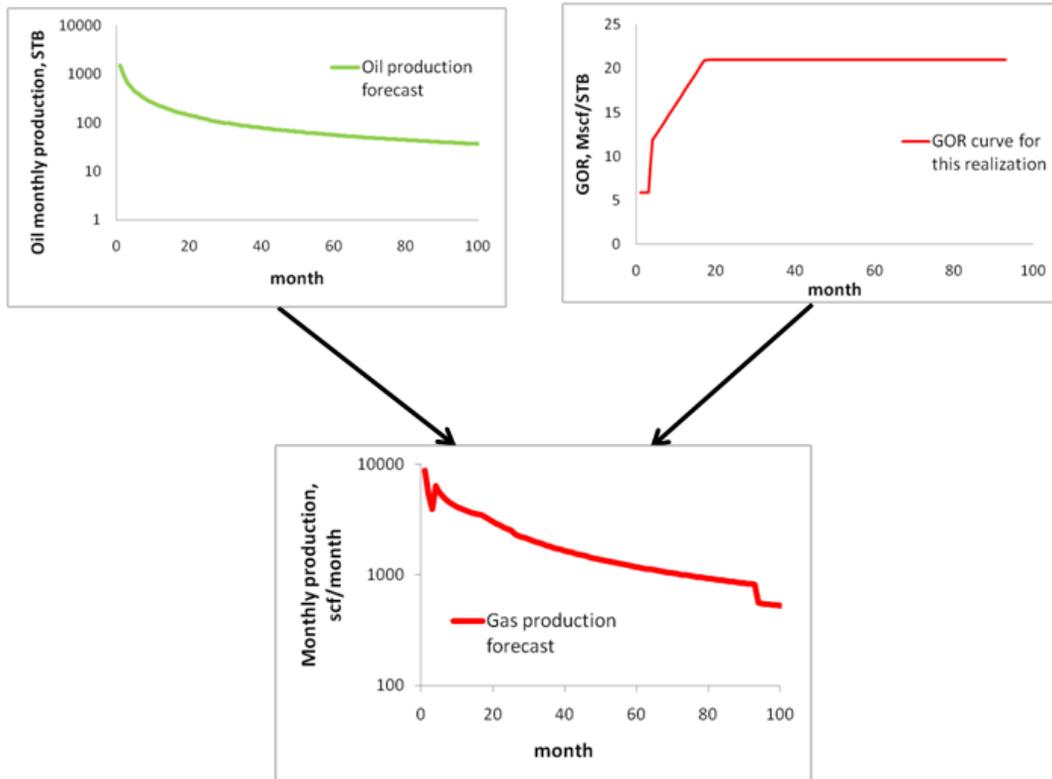


Fig. 50 Combine oil production forecast with GOR curve to generate gas production forecast

#### C.4.4 Illustrative Economic Example Using the Second Regression Model

##### C.4.4.1 Economics Inputs

Economic inputs used in this illustrative example are summarized in Table 21 to Table 23.

Table 21 – Overall economics and capital cost per well

Net Working Interest	100%
Yearly Discounted Rate	0.1
CAPEX Low	2.8 \$Million
Perforated Interval Low	2500 ft
CAPEX High	3.5 \$Million
Perforated Interval High	5500 ft

**Table 22 – Oil, gas and condensate prices**

<b><u>Oil Revenue</u></b>		
Price	80.00	\$/bbl
Price Differential	2.00	\$/bbl
<b><u>Gas Revenue</u></b>		
Price	4.00	\$/mmbtu
Price Differential	0.50	\$/mmbtu
Field and Processing Shrinkage	28.00%	
Residue BTU Adjustment	1.00	mmbtu/mcf
<b><u>NGL Revenue</u></b>		
NGL Yield	0.110	bbls/mmcf
Price	42.00	\$/bbl
Price Differential	4.00	\$/bbl

**Table 23 – Monthly OPEX per well**

Lease Operating Expense	5,000	
Cost of Operations	1,000	
Water Volume	0.0267	bbls/Mcf
Water Disposal	1.30	\$/bbl
Oil Severance Tax Rate	4.60%	
Gas & NGL Severance Tax Rate	7.50%	
Approximate Ad Valorem Tax Rate	2.00%	

#### C.4.4.2 Monte Carlo Simulation with DCA

In order to compare the monetary values among different perforated interval and well spacing combinations, we use a normalized area of 640 acres to calculate the normalized production and normalized net present value (NPV). We observed that when two groups of

wells are located next to each other, there is normally a non-zero distance between the wells parallel to the direction of the laterals; we assumed the distance between two wells end to end is 1000 ft. As a result, we assumed that each well will occupy an area equal to  $(PI+1000)*WS$ , where PI is perforated interval length and WS is well spacing. The number of wells using a certain perforated interval and well spacing in the normalized area equals the normalized area (640 acres) divided by area occupied by each well (Fig. 51).

Monte Carlo simulations with 5000 iterations were performed to generate economics output for different completion strategies. The decline curve parameters were generated from Eq. 12, where the expected values for  $\ln(q_i)$ ,  $CP_{6to1}$ , and  $b$  are calculated from linear regression with errors  $(\varepsilon_{\ln(q_i)}, \varepsilon_{CP_{6to1}}, \varepsilon_b)$  drawn from the multi-normal distribution shown in Eq. 12.  $Q_i$ ,  $D_i$ , and  $b$  for oil production are then calculated from  $\ln(q_i)$ ,  $CP_{6to1}$ , and  $b$ . Initial first three months oil percentage is calculated based on Eq. 14. Gas production forecasts were calculated based on the oil production and the GOR curve. Net present values were calculated based on the oil and gas production forecasts generated from the Monte Carlos simulation and the economics input shown in the previous section. A correlation of 50% was assumed among wells in the normalized area, which means there is a 50% chance that the new well will produce exactly the same as the neighbor well with another 50% chance that the new well is independent of the neighbor well.

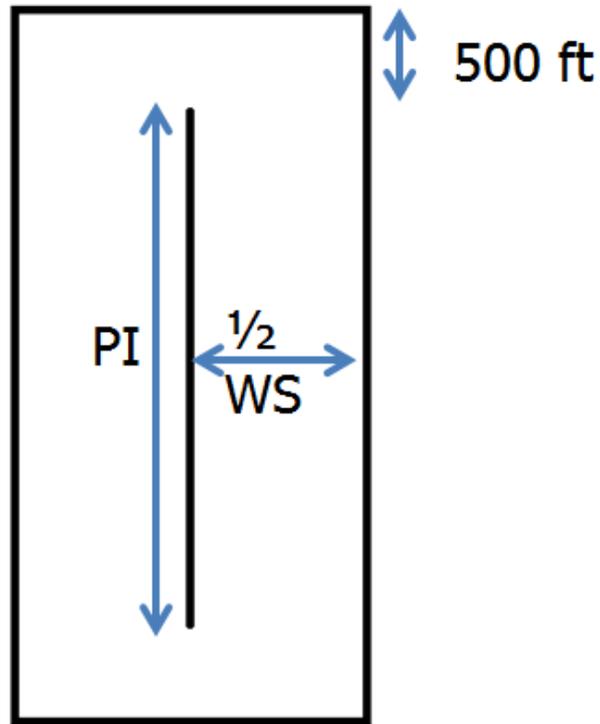


Fig. 51 – Occupied area by a single well in normalized area

### C.4.4.3 Sensitivity Analysis of Perforated Interval and Tighter Well Spacing

#### C.4.4.3.1 Sensitivity Analysis for Perforated Interval

In the sensitivity analysis, the tighter well spacing was kept at 340 ft (CP6to1 peaks at 340 well spacing) and  $R_o$  was kept at 0.93 (average of 266 wells). Monte Carlo simulations with 5000 iterations were performed to generate the results.

- Well count per 640 acres decreases when perforated interval increases because longer perforated interval results in greater occupied area by each well and thus less well count (Fig. 52).
- NPV @10% discount rate excluding capital decreases when perforated interval increases (Fig. 53).
- NPV/Capital increases as perforated interval increases (Fig. 54 and Fig. 55).
- NPV-Capital increases as perforated interval increases (Fig. 56 and Fig. 57).

- The uncertainty (P10 – P90) of NPV, NPV/Capital and NPV-Capital are all very large.

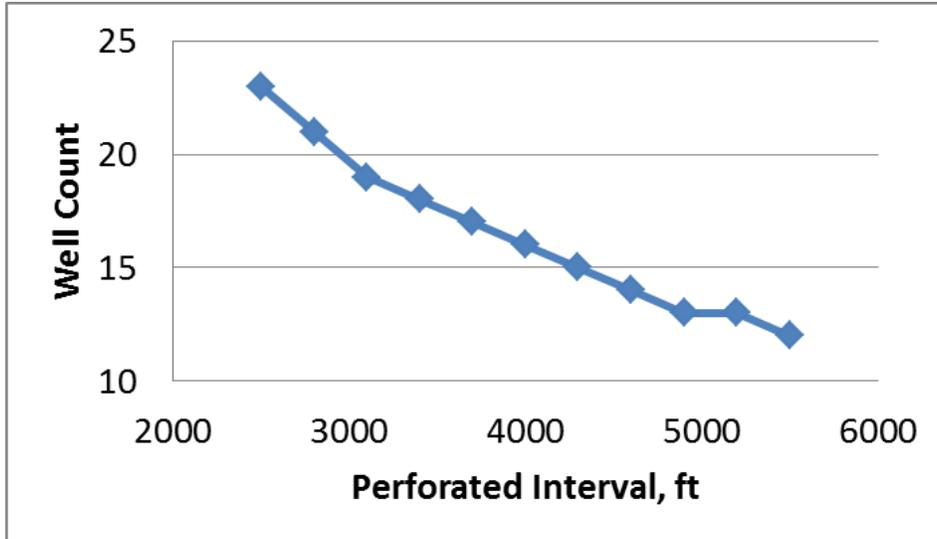


Fig. 52 – Well count per 640 acres vs. perforated interval

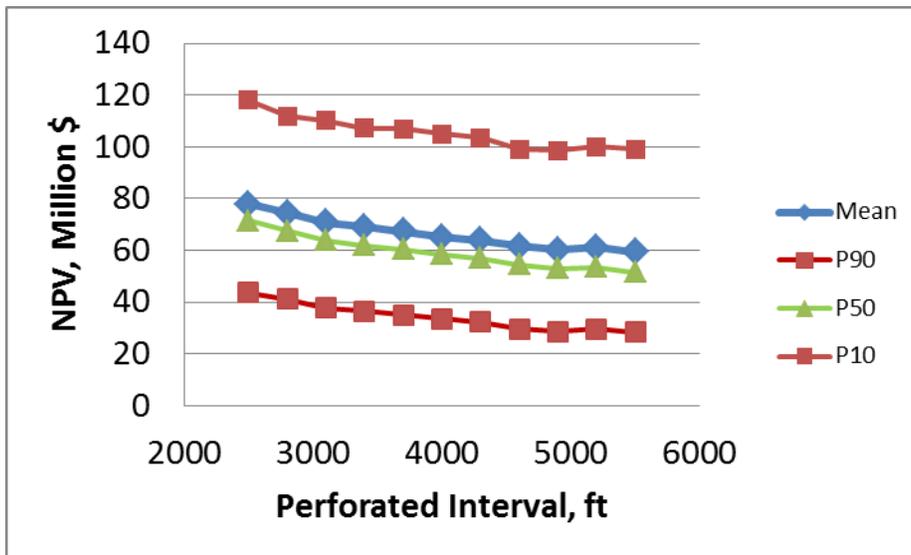


Fig. 53 – NPV(excluding capital) vs. perforated interval. Normalized to 640 acres.

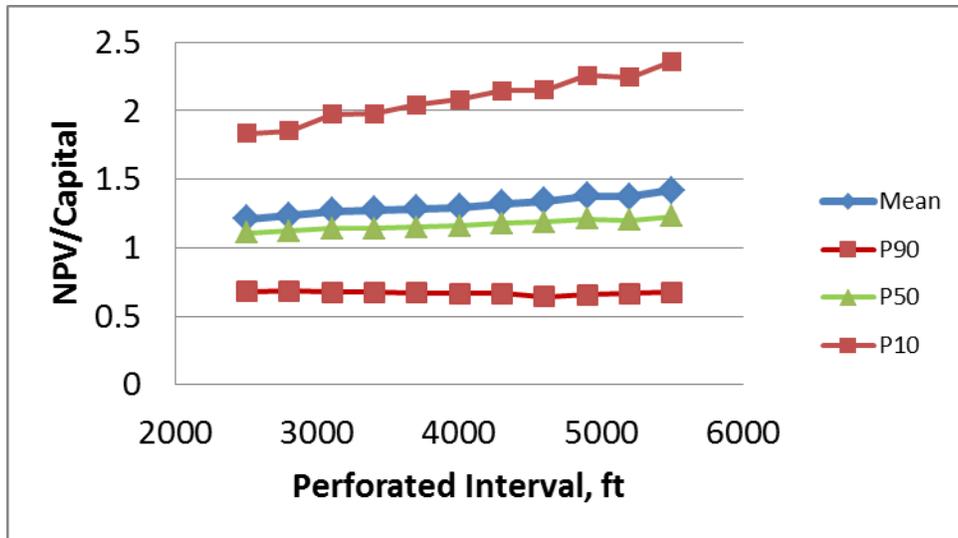


Fig. 54 – NPV/Capital vs. perforated interval. Normalized to 640 acres.

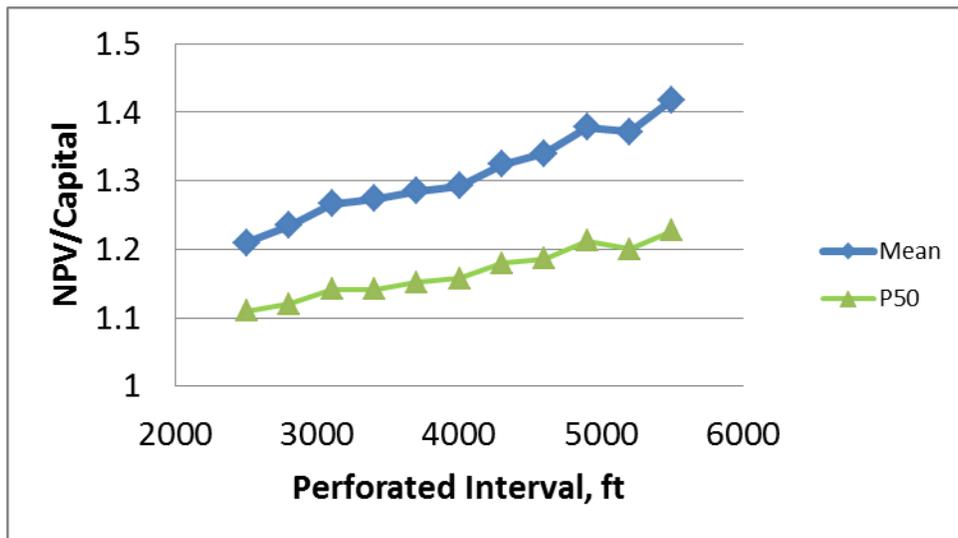


Fig. 55 – NPV/Capital vs. perforated interval (mean and P50). Normalized to 640 acres.

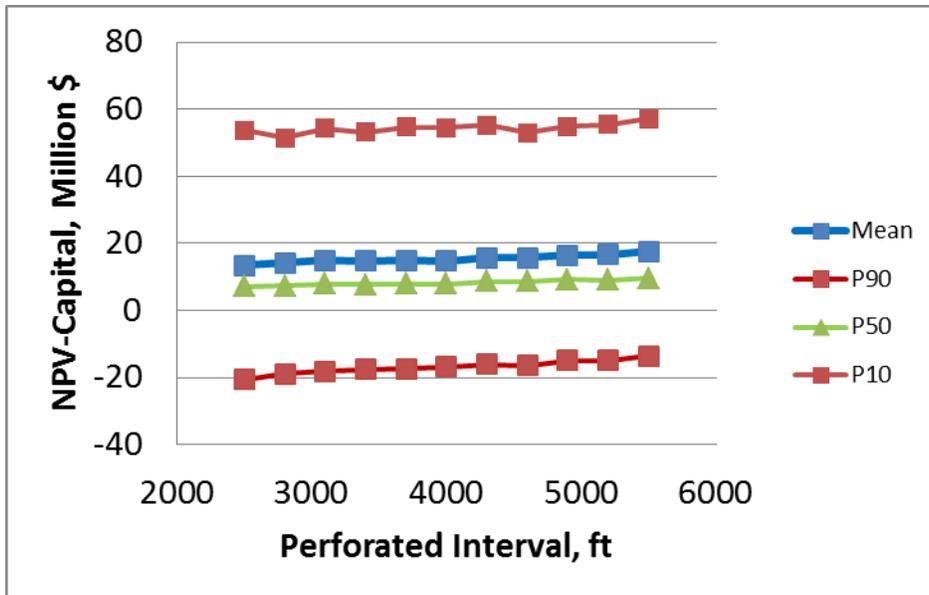


Fig. 56 – NPV-Capital vs. perforated interval. Normalized to 640 acres.

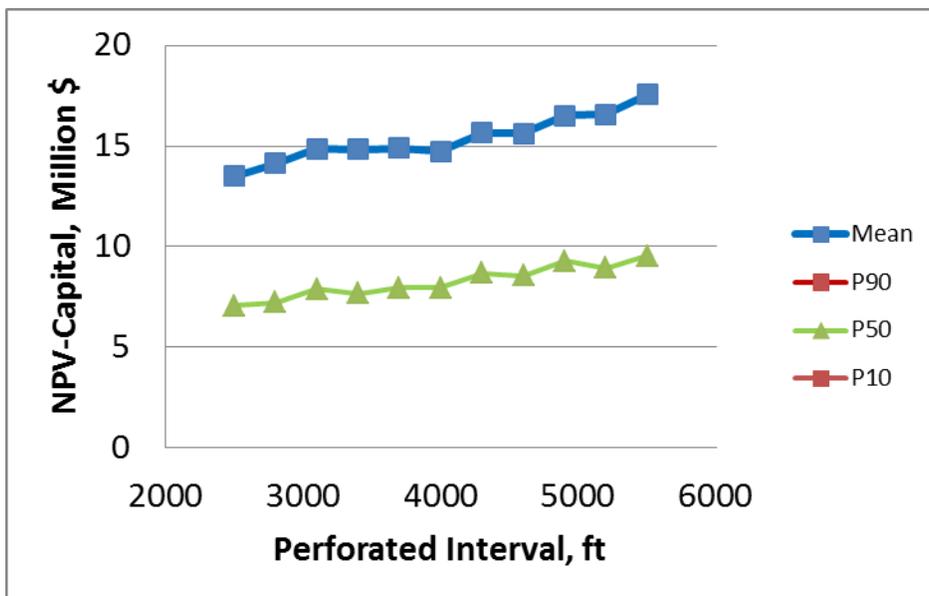


Fig. 57 – NPV-Capital vs. perforated interval (mean and P50). Normalized to 640 acres.

#### C.4.4.3.2 Sensitivity Analysis for Well Spacing

In the sensitivity analysis, we assume well spacing equals the distance to the nearest well, perpendicular to the direction of the laterals. The perforated interval was kept at 5500 ft (longest in this study area) and  $R_o$  was kept at 0.93 (average of 266 wells). Monte Carlo simulations with 5000 iterations were performed to generate the results.

- Well count per 640 acres decreases when well spacing increases because greater distance between wells results in greater occupied area by each well and thus less well count (Fig. 58).
- NPV @10% discount rate excluding capital decreases when well spacing increases (Fig. 59).
- NPV/Capital peaks at well spacing = 340 ft (Fig. 60 and Fig. 61).
- NPV-Capital peaks at well spacing = 340 ft (Fig. 62 and Fig. 63).
- The uncertainty (P10 – P90) of NPV, NPV/Capital and NPV-Capital are all very large.

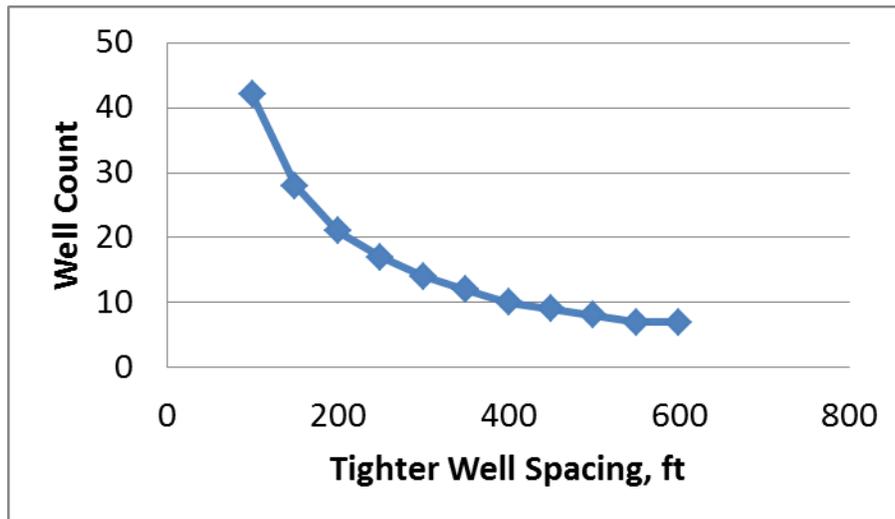


Fig. 58 – Well count per 640 acres vs. tighter well spacing.

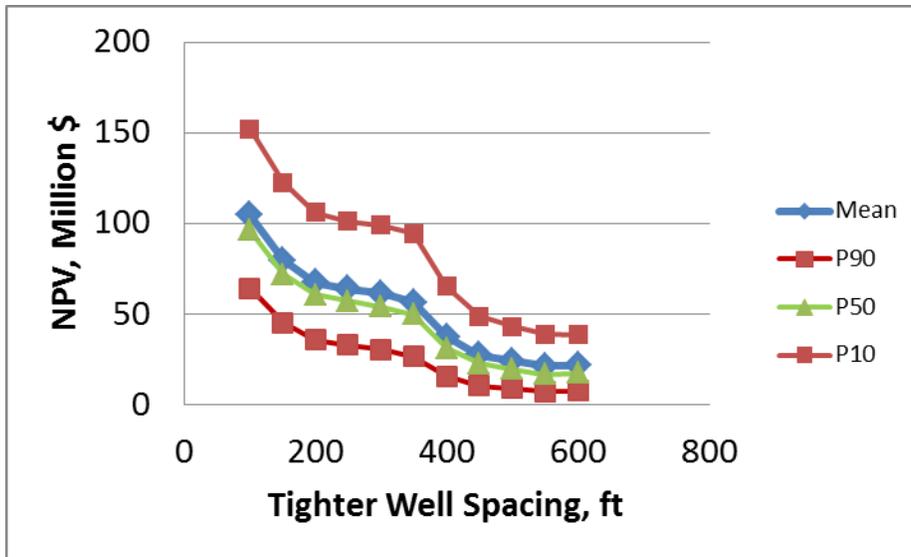


Fig. 59 –NPV(excluding capital) vs. tighter well spacing. Normalized to 640 acres.

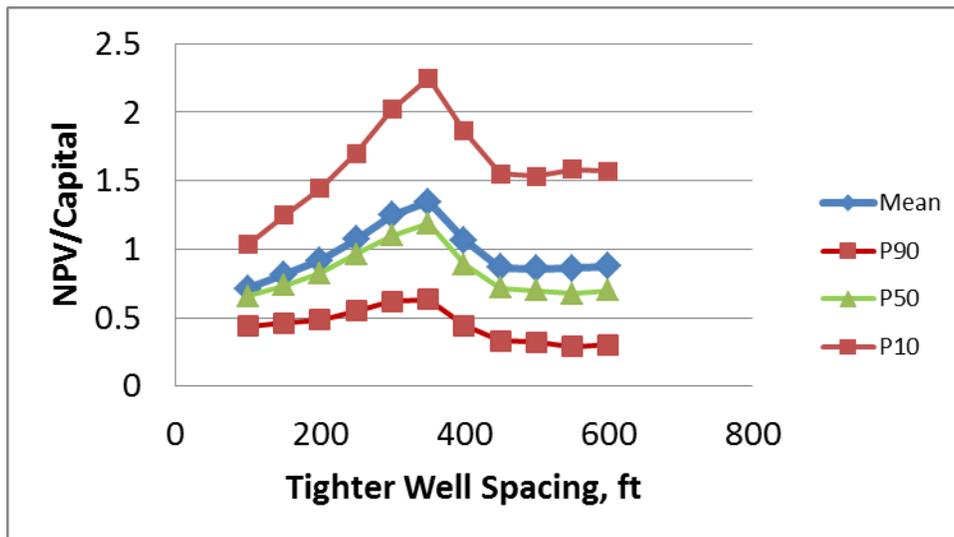


Fig. 60 – NPV/Capital vs. tighter well spacing. Normalized to 640 acres.

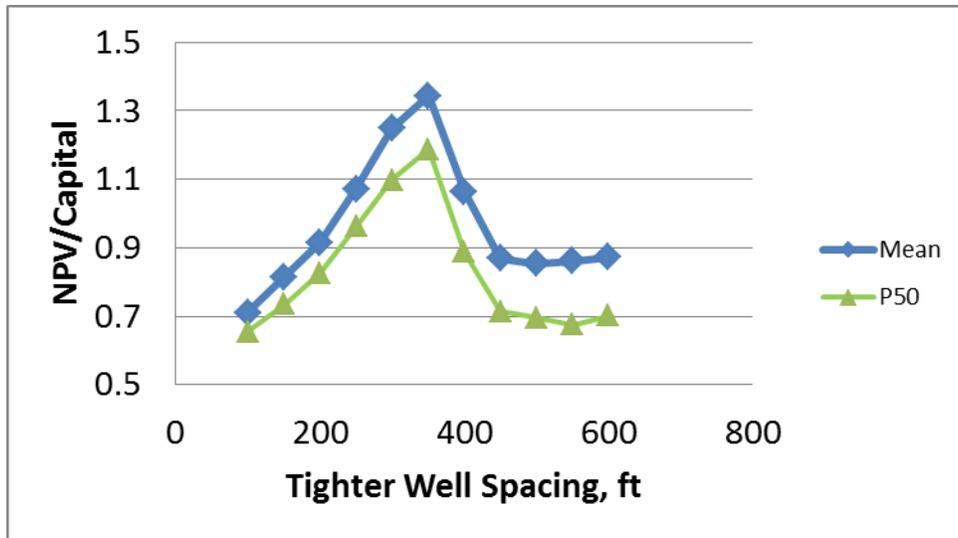


Fig. 61 – NPV/Capital vs. tighter well spacing (mean and P50). Normalized to 640 acres.

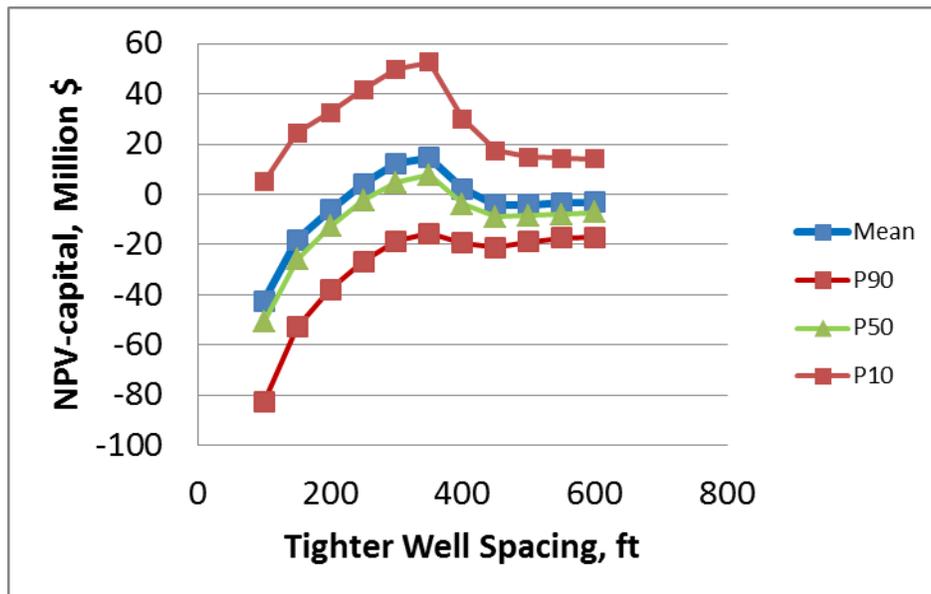


Fig. 62 – NPV-Capital vs. tighter well spacing. Normalized to 640 acres.

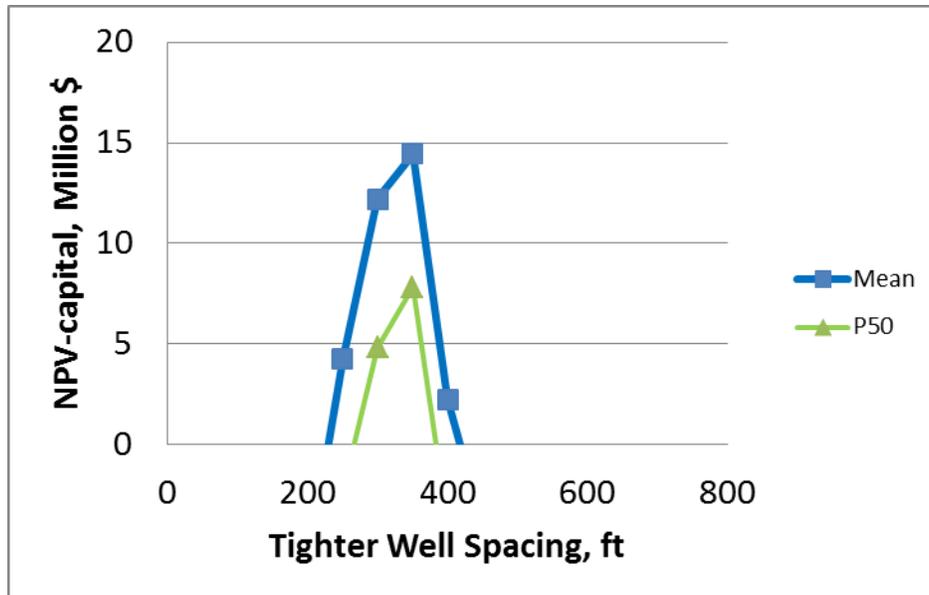
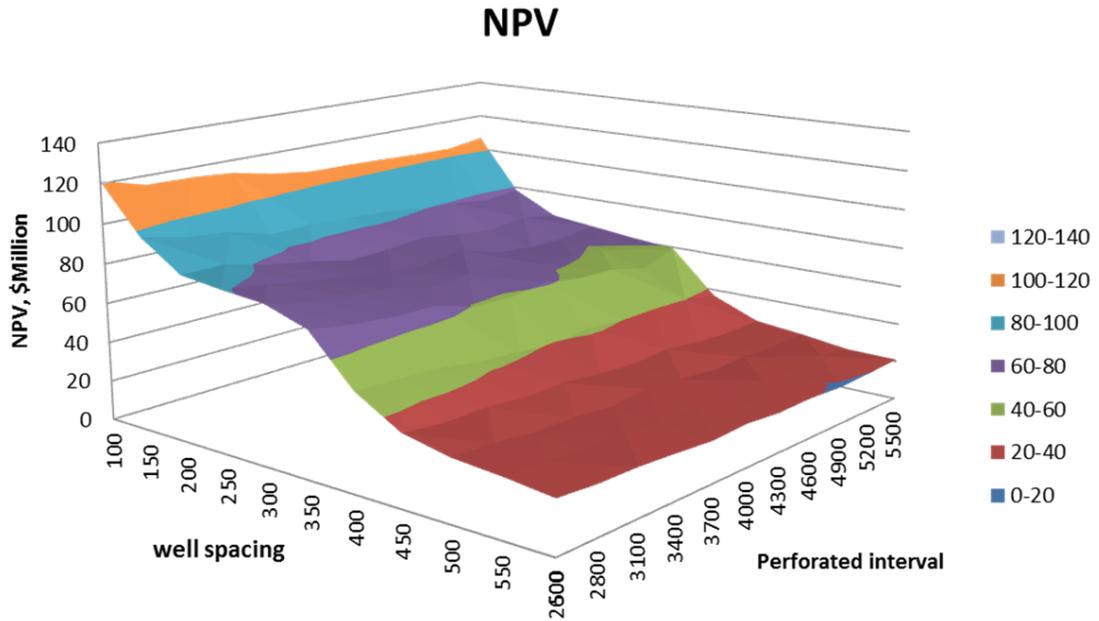


Fig. 63 – NPV-Capital vs. tighter well spacing (mean and P50). Normalized to 640 acres.

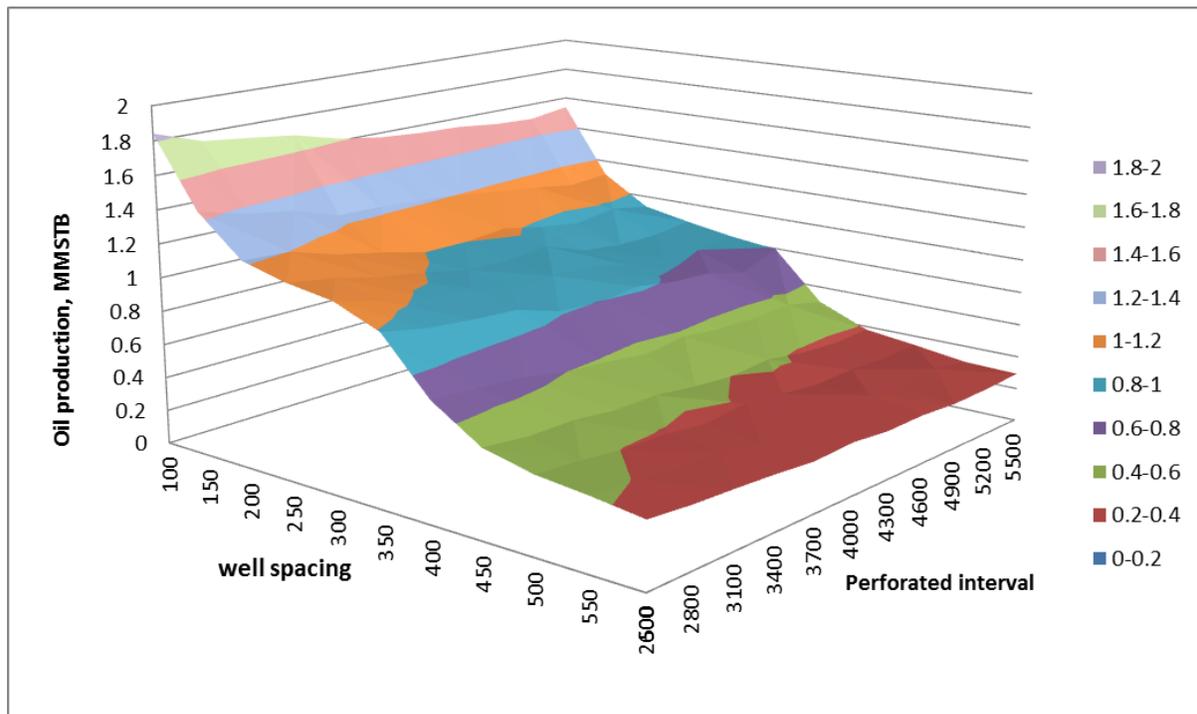
#### C.4.4.3.3 Sensitivity Analysis for Both Perforated Interval and Well Spacing

After we performed sensitivity analysis for both perforated interval and well spacing independently, we plotted the average of important objective functions such as NPV versus both perforated interval and well spacing on a 3D surface plot so the effects of perforated interval and well spacing can be easily visualized. We determined the following:

- NPV and oil production normalized to 640 acres decrease when the occupied area by a single well increases (perforated interval or well spacing increases) (Fig. 64 and Fig. 65).
- NPV/Capital and NPV-Capital peak at 340 ft well spacing and increase as perforated interval increases. Average NPV/Capital and NPV-Capital reach maximum values when the perforated interval equals 5500 ft (maximum observed in data set) and well spacing equals 340 ft (Fig. 66 and Fig. 67).



**Fig. 64 – NPV (excluding capital) per section vs. perforated interval and tighter well spacing.**



**Fig. 65 –Oil production per section vs. perforated interval and tighter well spacing.**

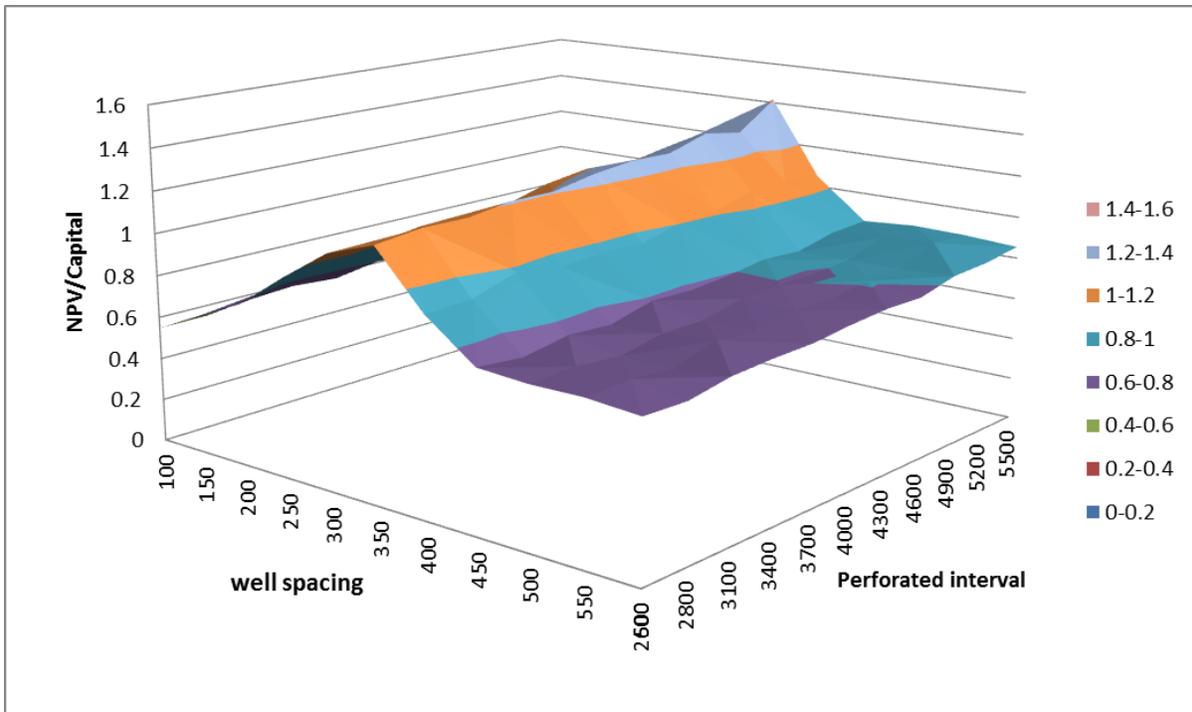


Fig. 66 –NPV/Capital per section vs. perforated interval and tighter well spacing.

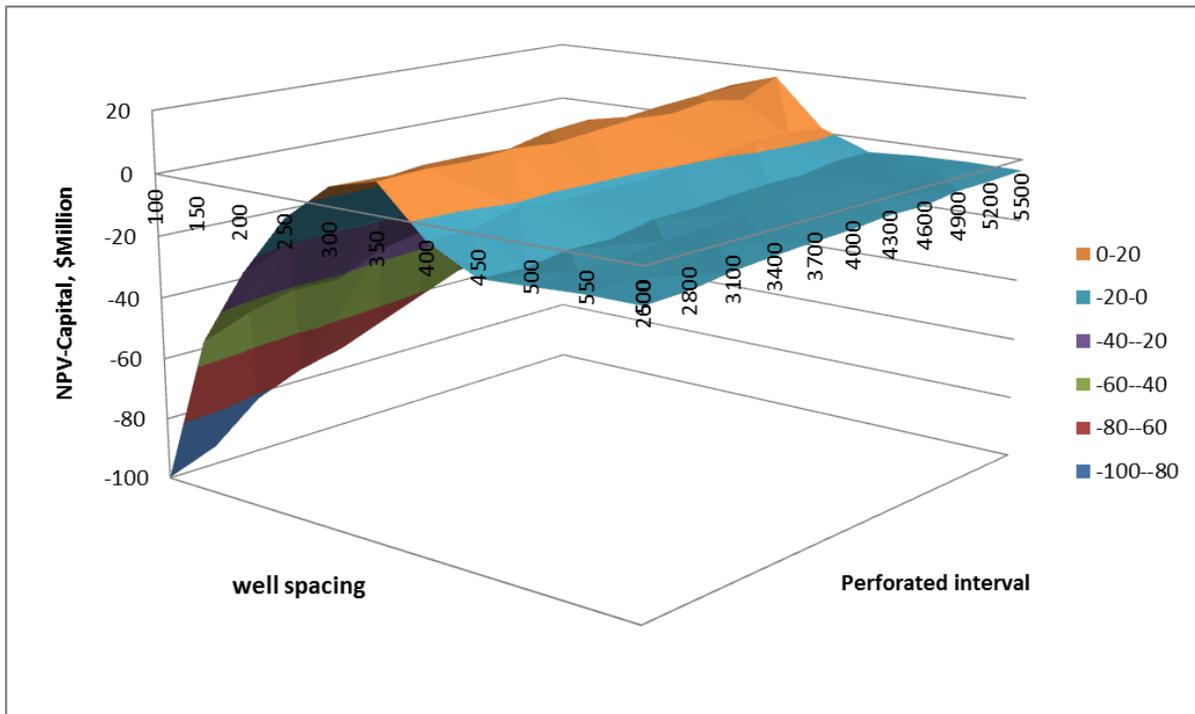


Fig. 67 –NPV-Capital per section vs. perforated interval and tighter well spacing.

#### C.4.4.3.4 Comparison of Models Based on First and Second Data Sets

A normalized area of 3500 ft \* 1000 ft was used to compare the results between the two models. Total NPV excluding capital cost of the normalized area generally decreases with perforated interval increase and single-well value generally increases with perforated interval increase for both models (Fig. 68).

The relationship between NPV excluding capital cost and well spacing is significantly different between the two models (Fig. 69). In the first model, total NPV and single-well NPV increase when well spacing increases. In the second model, single-well NPV peaks at 340 ft well spacing, while total NPV decreases with increasing well spacing with a bump at 340 ft.

The agreement between both models regarding the effect of perforated interval indicates that the relationship between NPV vs. perforated interval is likely to be present and individual-well NPV–Capital generally increases as perforated interval increases (within the range of perforated intervals investigated). The disagreement between the two models on the relationship between NPV and well spacing indicates there is a significant uncertainty regarding how well spacing will affect production performance and economics.

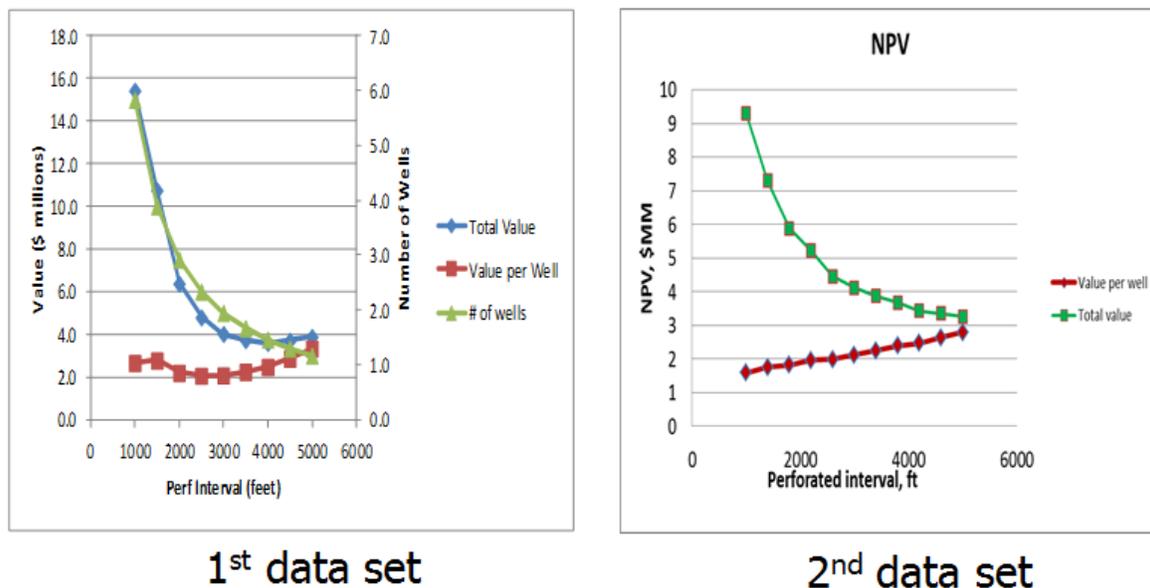
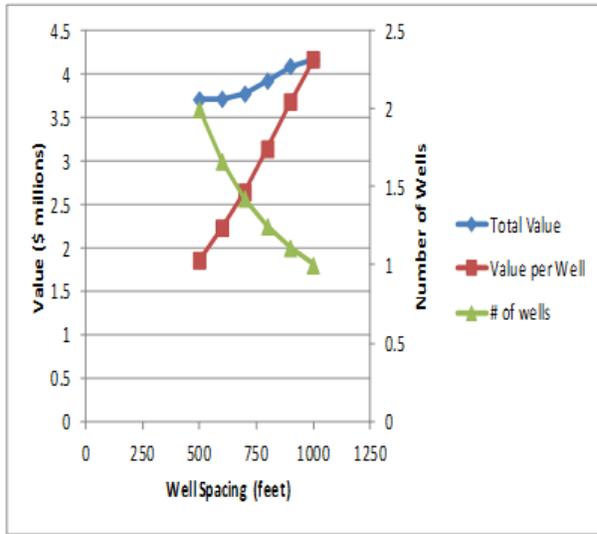
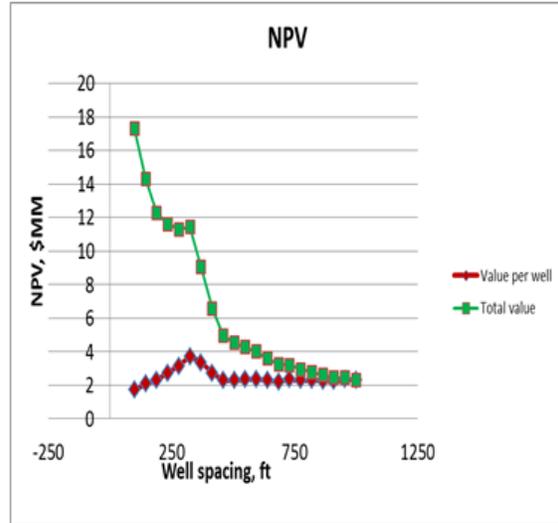


Fig. 68 – Comparison between 1<sup>st</sup> and 2<sup>nd</sup> data sets vs. perforated interval. Values exclude capital costs and are normalized to 3500-ft x 1000-ft area.



1<sup>st</sup> data set



2<sup>nd</sup> data set

Fig. 69 – Comparison between 1<sup>st</sup> and 2<sup>nd</sup> data sets vs. tighter well spacing. Values exclude capital costs and are normalized to 3500-ft x 1000-ft area.

#### C.4.5 Discussion

Significant uncertainty was observed when performing the sensitivity analysis to perforated interval and well spacing. While the Monte Carlo simulation and economic analysis indicate the optimal well spacing is around 340 ft and longer perforated intervals outperform shorter perforated intervals in terms of average NPV/Capital and NPV-Capital (Fig. 66 and Fig. 67), there is considerable uncertainty in these conclusions as well.

Single-well probabilistic oil production forecasts are shown in Fig. 70 for well spacing equal to 340 ft and perforated interval equal to 5500 ft. Considerable uncertainty is observed in the production forecast. Individual wells produce an average of 60 MSTB oil and generate an average NPV-Capital equal to \$1.37 million with P90, P50, and P10 values of -\$2.73 million, -\$0.44 million, and \$7.24 million, respectively, for oil and gas prices of \$80/STB and \$4/Mcf (Table 24). The average of single-well NPV/Capital = 1.42 with P90, P50, and P10 values of 0.22, 0.87 and 3.01, respectively (Table 24).

The average oil production normalized to 640 acres is 790,000 STB with P90, P50, and P10 values of 390,000 STB, 690,000 STB, and 1,260,000 STB, respectively. The average NPV-Capital equals \$17 million with P90, P50, and P10 values of -\$14 million, \$10 million, and \$54

million, respectively. Average NPV/Capital equals 1.41 with P90, P50, and P10 values of 0.67, 1.24, and 2.29, respectively (Table 25).

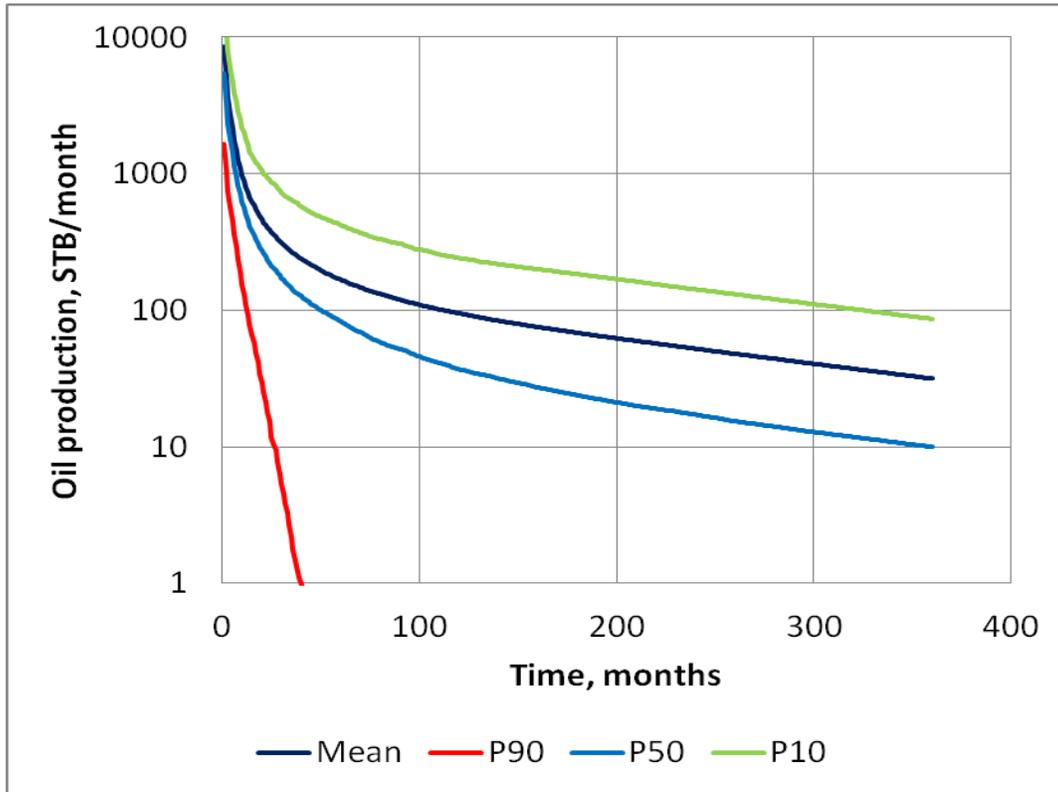


Fig. 70 – Probabilistic single-well production forecasts for 5500 ft perforated interval and 340 ft well spacing.

Table 24 – Single-well results using 5500-ft perforated interval and 340-ft well spacing.

Single Well	NPV10, \$Million	Capital, \$Million	NPV/Capital	Oil Production, MMSTB
Mean	4.87	3.5	1.42	0.06
90	0.77	3.5	0.22	0.01
50	3.06	3.5	0.87	0.04
10	10.74	3.5	3.07	0.14

Table 25 – Results normalized to 640 acres using 5500-ft perforated interval and 340-ft well spacing.

Normalized 640 acres (12 wells)	NPV10, \$Million	Capital, \$Million	NPV/Capital	Oil Production, MMSTB
Mean	59.09	42	1.41	0.79
90	28.11	42	0.67	0.39
50	51.94	42	1.24	0.69
10	96.13	42	2.29	1.26

## C.5 Conclusions

- Gas and oil production decline behavior of northern Barnett shale wells can be correlated to decision variables such as well spacing, perforated interval and stimulation fluid volume.
- The correlations are not high, indicating significant uncertainty in these decisions.
- The two models using two different data sets agree in that there is a positive correlation between initial production and perforated interval. Both models indicate that NPV-Capital and NPV/Capital increase as perforated interval increases, within the range of perforated intervals investigated.
- The two models resulted in two different relationships between decline curve parameters and well spacing, which indicates there is significant uncertainty on how well spacing affects well performance.
- Despite the considerable uncertainty in Barnett shale well performance, modeling these relations and their associated uncertainty can aid in optimization of spacing, completion and stimulation decisions in this area.

## D IMPACT TO PRODUCERS

We have transferred to the industry, through meetings with industry, workshops and conferences, and technical papers (see Technology Transfer Efforts below), technology that can help operators in unconventional gas reservoirs increase reserves and accelerate production, while protecting the environment, by determining the optimal well spacing and completion strategy as quickly as possible. It is impossible to develop an integrated reservoir and decision modeling tool that can work for all, or even many, different reservoirs. Thus, we have developed and transferred an array of integrated reservoir and decision modeling tools that can be adapted and tailored to specific reservoir and economic conditions (Table 20). These tools range from decline-curve based methods that can be applied more easily and quickly to single-well and multi-well simulation-based methods that require more data, time and expertise.

Operators can choose a preferred reservoir modeling approach based on their interests, expertise, time and data available, and the particulars of the reservoir being modeled.

This project was a “research” project, requiring 20% cost share, rather than a “development” project, requiring 50% cost share and field application. As such, we cannot directly quantify the impact of our work. However, we believe it to be quite positive, based on positive comments we have received from operators when our work has been presented at workshops, meetings and conferences. While we cannot yet report results of direct application of our technology, we are confident that adoption of the technology and tools we have developed and transferred to industry will enable operators to make better decisions with regard to spacing and completion/stimulation practices in unconventional reservoirs.

**Table 26 – Range of reservoir models developed**

<b>Production forecast methodology</b>	Decline-curve based reservoir model	Single-well reservoir simulation model	Multi-well reservoir simulation model
<b>Uncertainty quantification</b>	Monte Carlo simulation with decline curve forecasts	Monte Carlo simulation with reservoir simulation	Monte Carlo simulation with multiple geological models
<b>Time required to run 1000 realizations</b>	10 mins	8 hours	5 days

## **E TECHNOLOGY TRANSFER EFFORTS**

We have transferred technology developed in this project to industry specifically through Unconventional Gas Resources and Pioneer Natural Resources, whom we worked closely with during the two major phases of this project. We have transferred technology developed in this project to industry generally through workshops, presentations at conferences, and technical papers. A complete bibliography of technology transfer efforts is listed below in the section Bibliography of Technology Transfer Efforts.

## **F CONCLUSIONS**

- Three different integrated reservoir and decision modeling systems were developed in this project to help operators in unconventional gas reservoirs increase reserves and accelerate production, while protecting the environment, by determining the optimal well spacing and completion strategy as early in reservoir life as possible.
- These models provide decision makers with an understanding of the uncertainty inherent in unconventional gas reservoir development and allow them to evaluate tradeoffs between these risks and other factors such as increased capital spending.
- These integrated reservoir and decision modeling systems were applied in Deep Basin (Gething) tight sands and the Northern Barnett shale play to determine optimal well spacing and completion and stimulation strategies. While specific strategies were determined for these particular applications, these results are specific to the reservoir properties and economic assumptions used in these applications and cannot be generalized.
- The methodologies developed in this project can be applied to other unconventional reservoirs with minor changes.

## **G RECOMMENDATIONS**

- The integrated reservoir and decision models developed here should be applied to other unconventional reservoirs to validate and improve the methods developed in this project.
- Operators should strive to incorporate probabilistic and decision analytic tools and methods into their workflows. This is especially important in unconventional gas, where the uncertainties are quite large.

## H REFERENCES

- Bickel, J.E. and Bratvold, R.B. 2007. Decision Making in the Oil & Gas Industry: From Blissful Ignorance to Uncertainty-Induced Confusion Paper presented at the SPE Annual Technical Conference and Exhibition, Anaheim, California. SPE 109610.
- Bickel, J.E., Gibson, R.L., McVay, D.A., Pickering, S. and Waggoner, J. 2006. Quantifying 3D Land Seismic Reliability and Value. Paper presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas. SPE 102340..
- Bickel, J.E. and Smith, J.E. 2006. Optimal Sequential Exploration: A Binary Learning Model. *Decision Analysis* **3** (1): 16-32.
- Bickel, J.E., Smith, J.E. and Meyer, J.L. 2008. Modeling Dependence among Geologic Risks in Sequential Exploration Decisions. *SPE Reservoir Evaluation and Engineering* **11**(2) 352-361.
- Bratvold, R.B., Bickel, J.E. and Lohne, H.P. 2007. Value of Information in the Oil and Gas Industry: Past, Present, and Future. Paper presented at the SPE Annual Technical Conference and Exhibition, Anaheim, California.
- Cheng, Y., Lee, W.J. and McVay, D.A. 2007. Quantification of Uncertainty in Reserves Estimation from Decline Curve Analysis of Production Data for Unconventional Reservoirs. Paper presented at the International Conference on Offshore Mechanics and Arctic Engineering, San Diego, CA.
- Cheng, Y., McVay, D.A. and Lee, J.W. 2006a. Optimal Infill Drilling Design for Marginal Gas Reservoirs Using a Simulation-Based Inversion Approach. Paper presented at the SPE Eastern Regional Meeting, Canton, Ohio, USA. SPE 104574-MS
- Cheng, Y., McVay, D.A., Wang, J. and Ayers, W.B. 2006b. Simulation-Based Technology for Rapid Assessment of Redevelopment Potential in Stripper-Gas-Well Fields "Technology Advances and Validation in the Garden Plains Field, Western Canada Sedimentary Basin. Paper presented at the SPE Gas Technology Symposium, Calgary, Alberta, Canada. SPE 100583-MS.
- Cheng, Y., McVay, D.A., Wang, J. and Ayers, W.B. 2008. Simulation-Based Technology for Rapid Assessment of Redevelopment Potential in Marginal Gas Fields--Technology

- Advances and Validation in Garden Plains Field, Western Canada Sedimentary Basin. *SPE Reservoir Evaluation & Engineering* **11** (3): 521-534. DOI: 10.2118/100583-PA
- Cipolla, C.L. and Wood, M.C. 1996. A Statistical Approach to Infill-Drilling Studies: Case History of the Ozona Canyon Sands. *SPE Reservoir Engineering* **11** (3): 196-202. DOI: 10.2118/35628-PA
- Coulter, G.R., Benton, E.G. and Thomson, C.L. 2004. Water Fracs and Sand Quantity: A Barnett Shale Example. Paper presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas. SPE 90891-MS
- EIA (U.S. Energy Information Administration), *Annual Energy Outlook 2011*, April, 2011. Doc. DOE/EIA-0383(2011)
- Gao, H. 2005. Rapid Assessment of Infill Drilling Potential Using a Simulation-Based Inversion Approach. Ph.D. Dissertation, Texas A&M University.
- Gao, H. and McVay, D.A. 2004. Gas Infill Well Selection Using Rapid Inversion Methods. Paper presented at the SPE Annual Technical Conference and Exhibition, Houston, Texas. SPE 90545-MS.
- Gibson, R., Bickel, J.E., McVay, D.A. and Pickering, S. 2007. Model-Based Uncertainty Quantification and Seismic Information Value. In *SEG Annual Meeting*. San Antonio, Texas.
- Grayson, C.J., Jr. 1960. *Decisions under Uncertainty: Drilling Decisions by Oil and Gas Operators*. Boston, MA: Harvard Business School, Division of Research. Original edition. ISBN.
- Gong, X., McVay, D., Bickel, E. and Montiel, L. 2011. Integrated Reservoir and Decision Modeling to Optimize Northern Barnett Shale Development Strategies. Paper presented at the Canadian Unconventional Resources Conference, Calgary, Alberta, Canada. SPE 149459-MS.
- Guan, L. and Du, Y. 2004. Fast Method Finds Infill Drilling Potentials in Mature-Tight Reservoirs. Paper presented at the SPE International Petroleum Conference in Mexico, Puebla Pue., Mexico. Society of Petroleum Engineers 91755-MS.
- Guan, L., McVay, D.A. and Jensen, J.L. 2004. Parameter Sensitivity Study of a Statistical Technique for Fast Infill Evaluation of Mature Tight-Gas Reservoirs. Paper presented at

- the Canadian International Petroleum Conference, Calgary, Alberta. Petroleum Society of Canada 2004-164.
- Guan, L., McVay, D.A., Jensen, J.L. and Voneiff, G.W. 2002. Evaluation of a Statistical Infill Candidate Selection Technique. Paper presented at the SPE Gas Technology Symposium, Calgary, Alberta, Canada. Copyright 2002, SPE 75718-MS
- Haskett, W.J. and Brown, J. 2005. Evaluation of Unconventional Resource Plays. Paper presented at the SPE Annual Technical Conference and Exhibition, Dallas, Texas. SPE 96879-MS
- Hudson, J.W., Jochen, J.E. and Jochen, V.A. 2000. Practical Technique to Identify Infill Potential in Low-Permeability Gas Reservoirs Applied to the Milk River Formation in Canada. Paper presented at the SPE/CERI Gas Technology Symposium, Calgary, Alberta, Canada. SPE 59779-MS
- Hudson, J.W., Jochen, J.E. and Spivey, J.P. 2001. Practical Methods to High-Grade Infill Opportunities Applied to the Mesaverde, Morrow, and Cotton Valley Formations. Paper presented at the SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas. SPE 68598-MS
- Kyte, D.G. and Meehan, D.N. 1996. Horizontal Spacing, Depletion, and Infill Potential in the Austin Chalk. Paper presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado. SPE36721-MS.
- Lafollette, R.F. and Holcomb, W.D. 2011. Practical Data Mining: Lessons-Learned from the Barnett Shale of North Texas. Paper presented at the SPE Hydraulic Fracturing Technology Conference, The Woodlands, Texas, USA. SPE 140524-MS.
- Mayerhofer, M.J., Lolon, E.P., Youngblood, J.E. and Heinze, J.R. 2006. Integration of Microseismic Fracture Mapping Results with Numerical Fracture Network Production Modeling in the Barnett Shale. Paper presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas. SPE 102103-MS
- Mckinney, P.D., Rushing, J.A. and Sanders, L.A. 2002. Applied Reservoir Characterization for Maximizing Reserve Growth and Profitability in Tight Gas Sands: A Paradigm Shift in Development Strategies for Low-Permeability Gas Reservoirs. Paper presented at the SPE Gas Technology Symposium, Calgary, Alberta, Canada. SPE 75708-MS.

- Newsham, K.E. and Rushing, J.A. 2001. An Integrated Work-Flow Model to Characterize Unconventional Gas Resources: Part I - Geological Assessment and Petrophysical Evaluation. Paper presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana. SPE 71351-MS
- NPC. 2007. *Facing the Hard Truths About Energy*. Washington, D.C.: National Petroleum Council.
- Reeves, S.R., Bastian, P.A., Spivey, J.P., Flumerfelt, R.W., Mohaghegh, S. and Koperna, G.J. 2000. Benchmarking of Restimulation Candidate Selection Techniques in Layered, Tight Gas Sand Formations Using Reservoir Simulation. Paper presented at the SPE Annual Technical Conference and Exhibition, Dallas, Texas. SPE 63096-MS
- Rushing, J.A. and Newsham, K.E. 2001. An Integrated Work-Flow Model to Characterize Unconventional Gas Resources: Part II - Formation Evaluation and Reservoir Modeling. Paper presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana. SPE 71352-MS
- Salazar, J., McVay, D.A. and Lee, W.J. 2007. Development of an Improved Methodology to Assess Potential Unconventional Gas Resources. Paper presented at the SPE Rocky Mountain Oil & Gas Technology Symposium, Denver, Colorado. SPE 107974-MS
- Stabell, C. 2005. Integrated Risk, Resource and Value Assessment of Unconventional Petroleum Assets. Paper presented at the SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas. SPE 94667-MS
- Teufel, L.W., Chen, H.-Y., Engler, T.W. and Hart, B. 2004. *Optimization of Infill Drilling in Naturally-Fractured Tight-Gas Reservoirs: Phase Ii*. Final Report, Contract No. De-Fc26-98ft40486. US DOE: Washington, DC. Original edition. ISBN.
- Turkarslan, G., McVay, D.A., Bickel, J.E., Montiel, L.V. and Ortiz, R.R. 2010. Integrated Reservoir and Decision Modeling to Optimize Spacing in Unconventional Gas Reservoirs. Paper presented at the Canadian Unconventional Resources and International Petroleum Conference, Calgary, Alberta, Canada. SPE 137816-MS.
- Voneiff, G.W. and Cipolla, C. 1996. A New Approach to Large-Scale Infill Evaluations Applied to the Ozona (Canyon) Gas Field. Paper presented at the SPE Permian Oil and Gas Recovery Conference, Midland, Texas. SPE 35203-MS

Xiong, H. and Holditch, S.A. 2006. Will the Blossom of Unconventional Natural Gas Development in North America Be Repeated in China? Paper presented at the International Oil & Gas Conference and Exhibition in China, Beijing, China. SPE 103775-MS

Zahid, S., Bhatti, A.A., Khan, H.A. and Ahmad, T. 2007. Development of Unconventional Gas Resources: Stimulation Perspective. Paper presented at the Production and Operations Symposium, Oklahoma City, Oklahoma, U.S.A. SPE 107053-MS.

## **I BIBLIOGRAPHY OF TECHNOLOGY TRANSFER EFFORTS**

### Technical Papers and Journal Articles

Gong, X., McVay, D., Bickel, E. and Montiel, L. 2011. Integrated Reservoir and Decision Modeling to Optimize Northern Barnett Shale Development Strategies. Paper presented at the Canadian Unconventional Resources Conference, Calgary, Alberta, Canada. SPE 149459.

Turkarslan, G., McVay, D.A., Bickel, J.E., Montiel, L.V. and Ortiz, R.R. 2010. Integrated Reservoir and Decision Modeling to Optimize Spacing in Unconventional Gas Reservoirs. Paper presented at the Canadian Unconventional Resources and International Petroleum Conference, Calgary, Alberta, Canada. SPE 137816.

Bickel, J.E., L.W. Lake, and J. Lehman. 2011. "Discretization, Simulation, and Swanson's (Inaccurate) Mean." *SPE Economics & Management* 3(3) 128-140. SPE 148542-PA.

Bickel, J.E. 2011. "Discretization, Simulation, and the Value of Information." Paper presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, 30 October – 02 November. Under review at *SPE Economics & Management*.

### Workshops

McVay, D.A. and Bickel, J.E. April 2009, RPSEA Unconventional Gas Workshop. Denver, Colorado.

McVay, D.A. and Bickel, J.E. April 2010, RPSEA Unconventional Gas Workshop. Golden, Colorado.

McVay, D.A. and Bickel, J.E. April 2011, RPSEA Unconventional Gas Workshop. Golden, Colorado.

McVay, D.A. and Bickel, J.E. June 2011, RPSEA Unconventional Gas Software Solutions meeting. Houston, Texas.

#### Technology Transfer Meetings with Sponsors

McVay, D.A., Bickel, J.E and Turkarslan, G., October 2008, meeting with UGR

McVay, D.A., Bickel, J.E and Gong, X., October 2010, meeting with PNR

McVay, D.A., Bickel, J.E and Gong, X., March 2011, meeting with PNR

McVay, D.A., Bickel, J.E and Gong, X., September 2011, meeting with PNR

#### Other Technolgy Transfer Presentations

Crisman Institue For Petroleum Research, Texas A&M University, May 2010

Crisman Institue For Petroleum Research, Texas A&M University, May 2011

Center for Petroleum Asset Risk Management, The University of Texas, October 2008.

Center for Petroleum Asset Risk Management, The University of Texas, October 2009.

Center for Petroleum Asset Risk Management, The University of Texas, October 2010

## **J LIST OF ACRONYMS AND ABBREVIATIONS**

$a$  = the range beyond which variogram remains essentially constant, m

$a_C$  = main diagonal of coefficient Matrix A, scf . cp/psi<sup>2</sup> . D

$a_N, a_W$  = north and west flow coefficient, scf . cp/psi<sup>2</sup> . D

$a_E, a_S$  = east and south flow coefficient, scf . cp/psi<sup>2</sup> . D

$A$  = well spacing, acre

$b$  = hyperbolic exponent

$B$  = formation volume factor, reservoir ft<sup>3</sup>/scf

BY = best 12 consecutive months of production divided by 12, MSCM/M

$c_t$  = total system compressibility, psi<sup>-1</sup>

$C(h)$  = covariance function

CMG = Computer Modeling Group

$CP6to1$  = cumulative production ratio 6 months to 1 month

$d$  = right-side column vector of the 2D flow equation

$D$  = decline rate, 1/year  
 $DA$  = decision analysis  
 $DCA$  = Decline curve analysis  
 $Di$  = initial decline rate, 1/year  
 $ENPV$  = Estimated Net Present Value  
 $F_{cD}$  = dimensionless fracture conductivity  
 $GOR$  = gas oil ratio, scf/STB  
 $Gp$  = cumulative production, standard cubic feet  
 $h$  = lag distance between pairs, ft  
 $h$  = net pay thickness, ft  
 $IMEX$  = CMG's g Adaptive Implicit-Explicit Black-Oil Simulator  
 $J'$  = well index  
 $k$  = permeability, md  
 $k_{gas}$  = gas permeability, md  
 $k_f$  = fracture permeability, md  
 $L_f$  = fracture length, ft  
 $N$  = normal distribution  
 $Np$  = cumulative production, MMscf  
 $NPC$  = National Petroleum Council  
 $NPV$  = net present value  
 $p$  = the vector of well block pressure, psi  
 $p_i$  = initial reservoir pressure, psia  
 $PI$  = Principal Investigator  
 $Pioneer$  = Pioneer Natural Resources Company  
 $p_p$  = real-gas pseudopressure,  $m/Lt^3$ ,  $psi^2/cp$   
 $Pr > F$  and  $Pr > /t/$  = P-value, the probability that its corresponding relationship with decline curve parameters do not exist  
 $p_{wf}$  = flowing bottomhole pressure, psi  
 $p_{pi,j}^n$  = real-gas pseudopressure in (i,j) grid at n time step,  $m/Lt^3$ ,  $psi^2/cp$   
 $p_{pi,j-1}^{n+1}$  = real-gas pseudopressure in (i,j-1) grid at n+1 time step,  $m/Lt^3$ ,  $psi^2/cp$

$p_{pi-1,j}^{n+1}$  = real-gas pseudopressure in (i-1,j) grid at n+1 time step,  $m/Lt^3$ ,  $psi^2/cp$   
 $p_{pi,j}^{n+1}$  = real-gas pseudopressure in (i,j) grid at n+1 time step,  $m/Lt^3$ ,  $psi^2/cp$   
 $p_{pi+1,j}^{n+1}$  = real-gas pseudopressure in (i+1,j) grid at n+1 time step,  $m/Lt^3$ ,  $psi^2/cp$   
 $p_{pi,j+1}^{n+1}$  = real-gas pseudopressure in (i,j+1) grid at n+1 time step,  $m/Lt^3$ ,  $psi^2/cp$   
 $q$  = production rate, Mscf/D or STB/D  
 $q_i$  = initial production rate, Mscf/D or STB/D  
 $R_o$  = vitrinite reflectance  
 RPSEA = Research Partnership to Secure Energy for America  
 $r_o$  = equivalent radius of well gridblock, L, ft  
 $r_w$  = wellbore radius, L, ft  
 $s$  = skin factor, dimensionless  
 scf = standard cubic feet  
 SGS = Sequential Gaussian Simulation  
 STB = stock tank barrel  
 $t_o$  = transition time, years  
 $T$  = temperature, T, °R  
 TEES = Texas Engineering Experiment Station  
 $T_{sc}$  = temperature at standard condition, °R  
 TTW = Technology Transfer Workshop  
 $\Delta t$  = timestep, t, days  
 UGR = Unconventional Gas Resources Canada Operating, Inc.  
 UQ = Uncertainty Quantification  
 US DOE = U.S. Department of Energy  
 VBA = Visual Basic Application for MS Excel  
 $V_p$  = pore volume of gridblock,  $L^3$ ,  $ft^3$   
 VBY = virgin best year, MSCM/M  
 $w$  = fracture width, ft  
 WS1 = shorter well spacing, ft  
 WS2 = longer well spacing, ft  
 WSA = average well spacing, ft

- $z$  = z factors
- $\alpha$  = coefficient in the  $a_C$  equation
- $\varepsilon_b$  = error between linear regressed b and least square fit b
- $\varepsilon_{CP6to1}$  = error between linear regressed CP6to1 and least square fit CP6to1
- $\varepsilon_{\ln(q_i)}$  = error between linear regressed  $\ln(q_i)$  and least square fit  $\ln(q_i)$
- $\rho(h)$  = correlation function
- $\rho_{1,2}$  = correlation coefficient between stages
- $\gamma(h)$  = semi-variogram
- $\mu$  = viscosity, cp
- $\mu_1$  = expected value for discounted cumulative production for stage-1
- $\mu_2$  = expected value for discounted cumulative production for stage-2
- $\mu_{2/1}(a)$  = expected discounted cumulative production at stage-2 giving the discounted cumulative production of stage-1
- $\phi$  = porosity, fraction
- $\sigma$  = standard deviation

## K APPENDIX A: SAS OUTPUT

The tables in this appendix include the linear regression models of  $\ln(q_i)$  and  $CP6to1$ . Each linear regression model consists of three parts for the 1<sup>st</sup> data set. The first two parts include the statistics for the entire model, while the third part includes statistics for each independent parameter (decision/reservoir parameters). The “Pr > F” and “Pr>|t|” columns in the first and third parts are  $p$ -values, which represent the probabilities that the corresponding relationships do not exist. In these two models, all of the  $p$ -values are less than 0.01. This means that, for every relationship included in the linear regression models, there is only a 1% chance that the relationship does not exist, which gives us confidence that the regression models are valid.

The GLM Procedure

Dependent Variable: Inqi

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	18.72076276	18.72076276	25.09	<.0001
Error	62	46.26364387	0.74618780		
Corrected Total	63	64.98440662			

	R-Square	Coeff Var	Root MSE	logqi Mean
	0.288081	17.45479	0.863822	4.948909

Parameter	Estimate	Standard Error	t Value	Pr >  t
Intercept	3.782191508	0.25674152	14.73	<.0001
Interval	0.000346087	0.00006910	5.01	<.0001

Dependent Variable: CP6To1

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	14.47647626	4.82549209	9.78	<.0001
Error	60	29.60801375	0.49346690		
Corrected Total	63	44.08449001			

	R-Square	Coeff Var	Root MSE	MP6To1 Mean
	0.328380	23.55956	0.702472	2.981685

Parameter	Estimate	Standard Error	t Value	Pr >  t
Intercept	-0.756564310 B	0.95538163	-0.79	0.4315
WS1	0.001495744	0.00045032	3.32	0.0015
WS1_C N	1.843330080 B	0.64911530	2.84	0.0062
WS1_C WS2000	0.000000000 B	.	.	.
Fluid_Interval	0.022245455	0.00580230	3.83	0.0003

## L APPENDIX B: DISCRETIZATION AND SIMULATION

Most decision or risk analyses include continuous random variables (e.g., oil in place, oil price, or porosity). Analysts are frequently concerned with how to best structure, compute, and communicate decision models under these circumstances. While decision trees are well suited for discrete random variables with a few possibilities, they can become unmanageable when the number of outcomes is large. For example, trees may become too large to easily display or the large number of endpoints may require too many evaluations of a possibly costly (to evaluate) objective function (e.g., runs of a reservoir simulation model). The question we address in this paper is how one should include continuous random variables in a decision tree given these limitations.

Two approximation methods, which are closely related, have been developed to address this problem: discretization and Monte Carlo (MC) simulation. Specifically, suppose we are constructing a decision model that takes as an input the continuous random variable (rv)  $X$  (e.g., oil in place). Under either discretization or Monte Carlo simulation, we approximate the continuous probability density function (pdf)  $f(x)$ , or the cumulative distribution function (cdf)  $F(x)$ , with a set of values  $x_i \in X$ ,  $i = 1, 2, \dots, N$ , and associated probabilities  $p_i \equiv p(x_i)$ . In MC simulation,  $N$  is generally large (in the thousands) and the  $x_i$  are randomly drawn from the entire support of  $X$  using methods such as the inverse cumulative method (Clemen and Reilly 2001). In this case,  $X = F^{-1}(U)$ , where  $F^{-1}$  is the inverse cdf and  $U$  is uniformly distributed over  $[0, 1]$ . The inverse cumulative method assures that  $X$  is distributed according to  $F$ . Discretization methods, on the other hand, comprise only a few points (three is common), but seek to choose points that will preserve desired properties of  $X$ . The most natural and common properties of interest are the raw (e.g., the mean) and central moments of  $X$  (e.g., variance, skewness, kurtosis). As we discuss below, if an approximation fails to preserve the moments of  $X$  it is unlikely that it will accurately preserve the moments of the output distribution (Miller and Rice 1983; Smith 1993), including the mean, which are of interest to decision makers. Since the moments of a pdf do not uniquely determine the underlying pdf (an infinite number of distributions have a mean of zero and a variance of one, for example) the discretization

methods discussed in this paper, and used throughout the oil and gas industry, may fail to match particular percentiles of the pdf or extreme values even though they match many moments. This important and underappreciated distinction means that calculating the probability of exceeding particular values of the output variable (e.g., the probability of exceeding zero) based on a cumulative distribution that is comprised of a several discretized input uncertainties may not be justified. We do not address this issue here. Rather, our goal will be to match the moments of the input pdfs, with the caution that this process is designed only to estimate the moments of the output distribution.

Several discretization methods are in common use. For example, Swanson's Mean (SM), which weights the 10<sup>th</sup>, 50<sup>th</sup>, and 90<sup>th</sup> percentiles of  $F(x)$  by 0.30, 0.40, 0.30, is heavily used within the oil and gas industry (Megill 1984; Hurst et al. 2000; Rose 2001). However, SM is neither the only, nor, as we shall see the, the best choice.

Interest in discretization methods within the oil and gas industry has recently increased. For example, Arild et al. (2008) compared the use of SM to MC simulation in the context of a value of information (VOI) problem. They demonstrated that the two methods yield different results, but were unable to comment on which approach is closer to the true VOI, since their problem did not have an analytic (closed-form) solution. Prange et al. (2009) compared a numerical approximation method to a discretization method and found that the discretization was inaccurate. Thus, these two papers have suggested the use of MC simulation instead of discretization. Willigers (2009), on the other hand, recommends the use of discretization instead of MC in order to reduce the number of computations and speed the analysis of asset portfolios.

This appendix is organized as follows. In the next section we review the theory underlying discrete approximations and discuss the accuracy of common methods. In the third section, we discuss the accuracy of discretization methods and the demonstrably poor performance of SM. In the fourth section we contrast the use of discrete approximations to MC simulation and demonstrate that the best discretization methods are equivalent to hundreds of thousands of MC trials. We provide a simple example in section five, which demonstrates that different discretizations yield different value estimates and possibly different recommendations for action. Finally, in the sixth section we conclude and provide recommendations for practice.

## L.1 Discrete Approximations: Methods and Shortcuts

Before we discuss specific discretization methods, we begin with a brief review and introduce the notation we use throughout the paper. The  $k$ -th raw (or uncentered) moment of rv  $X$  with pdf  $f(x)$  is

$$\mu_k = E[X^k] = \int_X x^k f(x) dx \quad k = 0, 1, 2, \dots, \quad (1)$$

where  $E[-]$  is the expectation operator. The zeroth raw moment specifies that the probabilities must integrate (or sum) to one. The first raw moment  $\mu_1$  is the mean. Raw moments measure the distribution of a rv about the origin. Central moments, on the other hand, measure the distribution about the mean.<sup>1</sup> The  $k$ -th central moment of  $X$  is

$$m_k = E[(X - \mu_1)^k] = \int_X (x - \mu_1)^k f(x) dx \quad k = 0, 1, 2, \dots \quad (2)$$

The second central moment is the variance and is given the special symbol  $\sigma^2$ .

There is a one-to-one relationship between the raw and central moments. For example, if one knows the raw moments  $\mu_k$  the central moments can be found as follows (Papoulis 1984, p. 110):

$$m_k = \sum_{j=0}^k \binom{k}{j} (-1)^{k-j} \mu_j \mu_1^{k-j} \quad (3)$$

Hence, the second central moment, the variance, expressed in terms of the raw moments, is

$$m_2 = \mu_2 - \mu_1^2 \equiv E[X^2] - (E[X])^2$$

Two normalized central moments, skewness and kurtosis, measure other properties of the distribution. Skewness is a measure of asymmetry and is given by  $\gamma_3 = m_3 m_2^{-3/2}$ . Kurtosis measures the distribution's degree of "peakedness" or how thick its tails are. Kurtosis is defined as  $\gamma_4 = m_4 m_2^{-2} - 3$ . The minus three normalizes kurtosis relative to the normal distribution, which has a kurtosis of three.

---

<sup>1</sup> Central moments are simply raw moments of the transformed variable  $Y = X - \mu_1$ .

### L.1.1 The Motivation for Moments.

Suppose we are interested in a value function  $v(x)$ , which is a function of a rv  $X$ . This value function might be net present value (NPV) or ultimate hydrocarbon recovery. If this value function is sufficiently differentiable, such that it can be approximated by a polynomial  $P(x)$  of degree  $N$ , then we have

$$v(x) \approx P(x) = \sum_{k=0}^N w_k x^k, \quad (4)$$

for the weights  $w_k$ .<sup>2</sup> The mean, or expected value of  $v(x)$  is given by

$$E[v(X)] \approx E[P(X)] = E\left[\sum_{k=0}^N w_k X^k\right] = \sum_{k=0}^N w_k E[X^k]. \quad (5)$$

Thus, the expected value of  $v(x)$  is an expansion of the raw moments of  $X$ . Therefore, if one wants to accurately compute the expected value of the value function (e.g., expected NPV), *all* the moments of  $f(x)$  must be represented accurately. It is not sufficient, as is sometimes assumed, to accurately reflect only the mean value of input uncertainties. To make this concrete, suppose that  $v(x) = x^2$  and that  $X$  is a normal random variable with zero mean and unit variance. The expected value of  $v(x)$  is equal to 1, while  $v$  evaluated at the expectation is equal to zero. Our goal then will be to develop discretization methods which preserve the moments of input distributions. Before we discuss particular discretizations, however, we will formally state the approximation problem.

### L.1.2 Formal Statement of Approximation Problem.

Given a pdf  $f(x)$  and function  $\Omega(x)$ , we wish to approximate a definite integral with a finite sum,

$$\int_x f(x)\Omega(x)dx \approx \sum_{i=1}^N p_i \Omega(x_i), \quad (6)$$

by choosing values  $x_i$  and weights (or probabilities)  $p_i$ . In other words, we seek to approximate the pdf with a discrete probability mass function (pmf). This procedure, originally developed by Gauss in the early 1800s (Gauss 1866), is known as Gaussian quadrature. If  $\Omega$  is a polynomial then the approximation in (6) holds with equality (Davis and Rabinowitz 1984).

---

<sup>2</sup> This discussion closely follows Smith (1993).

In this appendix, we analyze the particular case  $\Omega(x_i) = x_i^k$ , where  $k = 0, 1, 2, \dots$  represents the  $k$ -th raw moment of  $X$ . Thus, we have

$$E[X^k] = \int_X f(x)\Omega(x)dx = \int_X f(x)x^k dx = \sum_{i=1}^N p_i x_i^k \quad \text{for } k = 0, 1, 2, \dots, N, \quad (7)$$

for a set of probabilities  $p_i$  and values  $x_i$ ; our challenge is to find these parameters. As an example, consider  $k = 0$ , in which case (7) requires that the discrete probabilities sum to 1. When  $k = 1$ , Equation (7) requires that we match  $E[X]$ , or the mean, and so on. This application of Gaussian quadrature is referred to as *moment matching* (Miller and Rice 1983; Smith 1993). An approximation with  $N$  points can match  $2N$  moments of  $X$ , including the zeroth moment (Stroud and Secrest 1966; Miller and Rice 1983). For example, a three-point approximation can match the zeroth through the fifth moment of  $X$ . This remarkable fact enables one to preserve many characteristics (mean, variance, skewness, kurtosis, etc.) of commonly used distributions with a limited number of points, assuming these moments are finite.

Gaussian quadrature provides both an organizing framework for understanding the objective of discretization and a computational method for determining the best approximation. However, many commonly used approximations fail to match some or all of the moments of  $X$ . We discuss the differing approaches below, which we divide into general methods and shortcuts. The general methods are intended for use on a case-by-case basis and are tailored to the underlying pdf. The shortcuts are “non-parametric” in that they are discretizations that are applied to any pdf, irrespective of its shape. Typically, these are discretizations that were found to work reasonably well over some set of commonly encountered pdfs, but are applied more generally than might be warranted.

### L.1.3 General Discretization Methods.

In this section we describe three general discretization methods: Moment Matching, Bracket Median, and Bracket Mean.

*Moment Matching.* The most accurate general method for matching moments is, of course, Gaussian quadrature or moment matching itself.<sup>3</sup> This approach can be applied to any pdf for which the moments are known. Quadrature rules have been calculated and published for several commonly encountered pdf families. For example, Miller and Rice (1983) presented the two-, three-, and four-point Gaussian quadratures for the standard uniform, normal, and exponential distributions. We repeat these results in **Table B1**.<sup>4</sup> However, instead of providing the values  $x_i$ , we provide the percentiles of the excess distribution function (edf), also called the complementary cdf, which is used more frequently than the cdf in oil and gas settings. These percentiles are  $\alpha_i \equiv G(x_i) \times 100 = (1 - F(x_i)) \times 100$ .

We summarize an approximation as the set of probabilities and percentiles ( $p_i; \alpha_i$ ). As shorthand, we will refer to the  $Z$ -th percentile of the edf as the PZ. For example, the 50<sup>th</sup> percentile is the P50. The two-point approximation of the normal distribution may be summarized as (0.500, 0.500; P84.1, P15.9). The three-point approximation of the normal is (0.167, 0.667, 0.167; P95.8, P50.0, P4.2). This simple approximation will match the first six moments of the normal distribution, including the requirement that the probabilities sum to one. It may help the reader to think of the discretizations in Table B1 as probability trees. For example, the three-point approximation for the normal is shown in **Figure B1**.

In the case of the uniform distribution, Gaussian quadrature does not result in equal weights on uniformly dispersed values. For example, one might think that an equal weighting of the P75, P50, P25 would perfectly match the uniform distribution since both the values and the probabilities have been divided uniformly. In fact, this approximation underestimates the variance by 50%.

*Bracket Median and Bracket Mean (or Equal Areas).* In two intuitively appealing methods, Bracket Median (Clemen and Reilly 2001) and Bracket Mean (McNamee and Celona 1990), the excess

---

<sup>3</sup> While we are using a special case of Gaussian quadrature, where  $\Omega(x_i) = x_i^k$ , we will use the terms moment matching and Gaussian quadrature interchangeably.

<sup>4</sup> Miller and Rice provided their results to six significant figures, but we round these to three significant figures to facilitate communication. Readers requiring more accuracy should consult Miller and Rice (1983).

distribution  $G(x)$ , is divided horizontally into  $N$  intervals as shown in Figure B2. It is common for the intervals to be of unequal size in the Bracket Mean approach and of equal size in the Bracket Median approach, although in neither case is it required. The probability that  $X$  will be in interval  $i$  is  $G_i - G_{i-1}$ . Given that  $X$  is in interval  $i$ , we summarize the conditional distribution of  $X$  (conditional on  $X$  being within the interval) by a single number. Bracket Median takes this number to be the median, while Bracket Mean uses the mean. Since these conditional distributions are generally skewed, the median and the mean differ and the two approaches may result in (very) different discretizations. For example, applying the three-point Bracket Median and Bracket Mean methods with intervals of 0.25, 0.50, 0.25 to the normal distribution yields discretizations of (0.25, 0.50, 0.25; P87.5, P50.0, P12.5) and (0.25, 0.50, 0.25; P89.8, P50.0, P10.2), respectively. The difference between the approaches is greater for skewed distributions such as the lognormal.

The advantage of the Bracket Median approach is that the median of each interval is simply that interval's midpoint,  $G_{i-1} + (G_i - G_{i-1})/2$ , which can be read directly off of the excess distribution, whereas, the mean of each interval must be calculated. For certain distributions, such as the normal, this calculation is straightforward. For other distributions, including subjectively assessed distributions that may not belong to any family, this calculation may be complex. However, there is a graphical method to estimate the interval means that gives Bracket Mean its more common name of "Equal Areas." As shown in Figure B2, the mean of each interval  $i$  is the point at which the area  $A_i$  to the left and above the excess distribution is equal to the area below and to the right. The interested reader should see McNamee and Celona (1991, pp. 18-21) for a proof of this fact. In our experience, individuals are skilled at finding these points by eye.

#### **L.1.4 Discretization Shortcuts.**

The general methods discussed above can be tailored to the shape of the distribution. For example, the intervals in Equal Areas need not be symmetric and extra attention can be placed on particular portions of the distribution, such as the tails. Yet, this flexibility comes at a cost—one must compute the appropriate discretization. For this reason, several shortcut methods

have been developed. In what follows, we discuss the approximations in the order in which they were originally proposed.

Pearson and Tukey (1965) developed a three-point discretization meant to closely approximate the mean of common pdfs, including the normal, beta, gamma, inverse gamma, and Student's  $t$  distributions (the uniform was not included). Their three-point approximation of the mean is (0.185, 0.630, 0.185; P95, P50, P5), which was found to work well except for distributions that were highly skewed. Pearson and Tukey did not recommend this approximation for higher moments and instead provided a more complex approximation for the standard deviation. However, Keefer and Bodily (1983), based on work by Keefer and Pollock (1980) and their own supporting analysis, suggested treating the Pearson-Tukey approximation as a complete pmf and referred to this as "Extended Pearson-Tukey" (EPT).

The Equal Areas method discussed above was developed by Jim Matheson and his colleagues at the Stanford Research Institute (SRI) between the late 1960s and the early 1970s.<sup>5</sup> Application of this method to the normal distribution produces an approximation of (0.25, 0.50, 0.25; P89.8, P50, P10.2). Based on this, SRI began using a shortcut of weighting the P90, P50, P10 by 0.25, 0.50, 0.25, which is sometimes referred to as the 25-50-25 approximation. This method was then heavily used and popularized by Strategic Decisions Group (SDG), which was founded by individuals from SRI's decision analysis group. SDG trained hundreds of oil and gas professionals in decision analysis methods and helped to establish existing decision analysis programs at several major corporations, including Chevron; this explains the use of 25-50-25 in oil and gas settings. The shortcut is described in McNamee and Celona (1990, pp. 32-33) and has come to be known as the "McNamee and Celona" shortcut or MCS. McNamee and Celona, SDG consultants at the time, cautioned that this shortcut should be used in the "early stages" of analyzing a decision and that one needs to carefully assess the distribution and develop a full discretization (using Equal Areas) "more carefully later on!" [emphasis in original]. Over time, this guidance has been widely forgotten and today MCS is commonly applied without regard

---

<sup>5</sup> Personal communication with Jim Matheson and Peter McNamee.

for the shape of the underlying distribution and is not followed with a secondary and more careful assessment and discretization.

While working for Exxon, Roy Swanson, in a 1972 internal memo, proposed approximating mean reserves, by weighting the P90, P50, P10 of the reserves edf by 0.30, 0.40, 0.30 (Megill 1984, Appendix B ; Hurst et al. 2000). Following the discussion above, we will refer to this as a 30-40-30 weighting. According to Megill (1984, p. 187), Swanson arrived at this rule empirically and found that it reasonably approximated the mean of “modestly skewed” distributions. Like Pearson and Tukey before him, Swanson (apparently) did not propose using his approximation to estimate higher moments or as a complete pmf. However, Keefer and Bodily (1983) proposed treating Swanson’s 30-40-30 rule as a complete pmf and referred to it as “Extended Swanson-Megill” (ESM). The general use of ESM, especially in the case of the lognormal, has been advocated by Pete Rose and his colleagues at Rose & Associates (Hurst et al. 2000; Rose 2001). As we will see below, ESM is close to a Gaussian quadrature for the normal distribution that matches the mean and variance. However, when directly applied to a lognormal distribution, ESM fails to match the mean and *significantly* underestimates the variance and the skewness.

Before discussing the other discretization methods, we pause briefly so that we can emphasize the difference between developing an approximation of the mean and an approximation of the pdf. A mean does not uniquely determine a pdf; there are an infinite number of pdfs with the mean of 5, for example. Therefore, one could very easily find an approximation that matches the mean, but fails to faithfully represent the underlying pdf. This is the case with Swanson’s Mean. Swanson did not suggest using his method to approximate the *pdf* of reserves; he only suggested that it be used to approximate the mean. However, approximations that only work well for the mean are not particularly useful in most decision analyses. Using a discretization in a probability or decision tree as shown in Figure B1, implicitly assumes that it is an accurate representation of the pdf, not just the mean of that pdf. Thus, strictly speaking, Swanson’s method cannot be used in a decision tree because it was intended merely to be a method for estimating the mean of a distribution. When Swanson’s

values (0.30, 0.40, 0.30; P90, P50, P10) are used in a decision tree it is really an application of Extended Swanson-Megill which describes a probability mass function (pmf).

Returning to our discussion of the various shortcuts, Miller and Rice (1983) introduced the use of Gaussian quadrature to the decision analysis literature. As detailed above, this method can exactly match as many moments of any pdf as desired, as long as the moments are finite. However, in practice it would be helpful to have these discretizations pre-calculated. This is possible in the case of known pdf families, as shown in Table B1. When one is dealing with a distribution that is not from a known family (as might happen if the distribution is directly assessed by an expert, for example) Miller and Rice proposed several generic discretizations based on Gaussian quadrature. For a three-point approximation, Miller and Rice proposed weighting the P91.5, P50.0, and P8.5 by 0.248, 0.504, 0.248. This approximation has become known as the “Miller-Rice One Step” (MRO). Notice that MRO is very close to the Bracket Mean approximation applied to a normal distribution and MCS. This correspondence further supported the MCS shortcut and SDG’s use of it.

D’Errico and Zaino (1988) and Zaino and D’Errico (1989) used Taguchi’s method (Taguchi 1978) to develop two approximations. First, equally weighting the P89, P50, P11, which we refer to as the ZDT approximation. Second, applying the three-point Gaussian quadrature formula for the normal distribution (0.167, 0.667, 0.167; P95.8, P50.0, P4.2), displayed in Table B1, more generally. We will refer to this approximation as GQN. We summarize each of the discretization shortcuts in **Table B2**.

#### **L.1.5 Moment Matching with Fixed Values or Fixed Probabilities.**

After many years of use, approximations with values fixed at the P90, P50, P10 (MCS and ESM) or weightings of 25-50-25 (MCS) or 30-40-30 (ESM) have become common. Yet, the ESM and MCS shortcuts are not distribution specific and therefore may induce unnecessary errors. In this section, in an effort to improve practice, we apply moment matching to develop rules for (a) weighting the P90, P50, and P10 and (b) values with fixed weights of 25-50-25 or 30-40-30. We consider the uniform, normal, exponential, and triangular distributions. These results appear in Table B3, Table B4, and Table B5.

As seen in Table B3, the weights for the normal are nearly identical to ESM. As such, ESM almost matches the mean and variance of a normal distribution. The listed triangular results are for a mode,  $c$ , of 0.5. However, the weights are a very weak function of  $c$ . For example, when  $c = 0.1$  the weights are (0.277, 0.451, 0.272). Thus, the values listed in Table B3 should provide satisfactory results in most situations.

It is not always possible to fix the values and find a feasible solution (Smith 1993). For example, a weighting of a lognormal's P90, P50, P10 that will simultaneously match the mean and variance is very often impossible to find if the underlying distribution is even modestly skewed. This underscores the futility of the common practice of applying ESM or MCS without regard for the underlying distribution or, indeed, in lieu of considering the nature of that distribution. To remedy this, in **Table B6** we present several approximations for the lognormal with any mean but the listed standard deviation (in terms of  $\ln X$ ), using four points. These approximations will match the mean and variance when applied directly to lognormal distributions with the stated standard deviation; as we discuss more fully below, matching the skewness of a lognormal is very difficult. The values that we have found to work well are the P90, P50, P5, and the mode (most likely value).<sup>6</sup>

## L.2 Accuracy of Discretization Methods

Given the variety of discretization methods, the question naturally arises as to which approximation is best. Or, perhaps more correctly, in which situations do the various approximations perform well or poorly? When the pdf is from a family given in Table B1 (or Tables K3-K6), the Gaussian quadrature formulas exactly match the first  $2N$  underlying moments of the original pdf, by definition. Thus, they can be taken as a “gold-standard” of accuracy.

Miller and Rice (1983) proved that Equal Areas will always underestimate the even moments of the original distribution. This occurs because  $x^k$  is convex when  $k$  is even and, by

---

<sup>6</sup> If  $X$  is lognormally distributed then the mode of  $X$  is  $\text{Exp}[\mu - \sigma]$ , where  $\mu = E[\ln X]$  and  $\sigma^2 = E[(\ln X - \mu)^2]$ .

Jensen's Inequality<sup>7</sup>, the expectation of a convex function will exceed the value of the function evaluated at the expectation. In the case of odd  $k$ , the direction of the error is more difficult to sign. If  $x$  is positive (negative) then the odd moments will be underestimated (overestimated), since  $x^k$  is convex (concave). Thus, direct application of Equal Areas to a lognormal distribution (which cannot take negative values) will underestimate all moments.

Keefer and Bodily (1983) tested EPT and ESM (among others) across a range of beta distributions, which can assume a wide variety of shapes and thus higher moments. They concluded that EPT is the "clear winner." Both EPT and ESM approximated the mean reasonably well; the average (maximum) errors were 0.02% (0.07%) and 0.05% (0.33%) for EPT and ESM, respectively. However, the two approaches differed in their ability to approximate the variance. The average (maximum) errors were 0.46% (-1.6%) and 2.7% (11.1%). Keefer (1994) extended the analysis of Keefer and Bodily (1983) by analyzing the accuracy of MCS, MRO, GQN, and ZDT across a range of beta distributions. In terms of estimating the mean and variance, he found that EPT slightly outperformed GQN, and that they both dominated ESM, MRO, MCS, and ZDT.

### L.2.1 Swanson's (Inaccurate) Mean.

Given the widespread use of ESM in the oil and gas industry, and the fact that it is being used as intended, which is not the case for MCS<sup>8</sup>, it seems appropriate to scrutinize its accuracy. The studies discussed above demonstrated the rule's inability to accurately approximate the variance of many distributions. In this section we more carefully analyze the use of ESM to approximate lognormal distributions; since its inception, ESM has been used to summarize reserve distributions, and reserves are widely held to be lognormally distributed (Rose 2001). Likewise, Rose has advocated the use of ESM specifically for use with lognormal distributions (Rose 2001, Appendix A).

Megill (1984, Appendix B) directly applies ESM to the lognormal and finds that ESM underestimates the mean by around 10% for modestly skewed distributions, which he

---

<sup>7</sup>  $E[v(X)] \geq v(E[X])$  if  $v$  is a convex function.

<sup>8</sup> The reader should bear in mind that MCS was never recommended as a final approximation, as ESM has been.

associates with “typical prospect ranges.” However, he also finds that ESM underestimates the mean by 45% for more skewed distributions, which he associates with “typical basin-play ranges for field size distributions.” Megill concludes that “Swanson’s rule should not be applied to obtain the mean of play or basin assessments.” We extend Megill’s analysis in **Figure B3**, which plots the error in ESM’s estimate of the mean, variance, and skewness for a lognormal distribution, against the ratio of the P10 to the P50, which Megill intended to be a measure of skewness. As Megill stated, ESM underestimates the mean by up to 45% in this example. What Megill did not mention is that it also underestimates the variance by 80% for the typical prospect range and by 100% for more skewed distributions. ESM’s estimation of the skewness is even worse.

Rose (2001) supports his use of ESM by arguing that reserves above the P1 of a lognormal occur with much less than a 1% frequency, and therefore the lognormal should be truncated above this point. Doing so reduces skewness and does improve the accuracy of ESM. However, Rose supports his argument by analyzing only the mean and examining a single lognormal distribution with a mean of 15.1 and a standard deviation of 28.2—implying a skewness of 3.9. In this case, ESM underestimates the true mean by 1.5%. However, it also underestimates the variance by 59% and the skewness by 78%. Furthermore, under a different truncated lognormal distribution with a skewness of 4.9, ESM underestimates the mean by 10%, the variance by 77% and the skewness by 83%.

Swanson’s Mean can be directly applied to lognormal distributions by applying it to the *logarithm* of  $X$ , instead of directly to  $X$ . If  $X$  is lognormally distributed then  $\ln X$  is normally distributed and the moments of  $X$  are functions only of the mean  $\mu$  and variance  $\sigma^2$  of  $\ln X$ . The equations for the first four moments of  $X$  appear below<sup>9</sup>:

$$\mu_1(X) = \text{Exp}\left[\mu + \frac{1}{2}\sigma^2\right] \quad (8)$$

$$m_2(X) = \text{Exp}\left[2\mu + \sigma^2\right]\left(\text{Exp}\left[\sigma^2\right] - 1\right) \quad (9)$$

$$\gamma_3(X) = \left(\text{Exp}\left[\sigma^2\right] + 2\right)\left(\text{Exp}\left[\sigma^2\right] - 1\right)^{1/2} \quad (10)$$

---

<sup>9</sup> These formula can be found in most probability textbooks or online at <http://mathworld.wolfram.com/>.

$$\gamma_4(X) = \text{Exp}[4\sigma^2] + 2\text{Exp}[3\sigma^2] + 3\text{Exp}[2\sigma^2] - 6. \quad (11)$$

To use SM we would simply determine  $\mu$  and  $\sigma^2$  by applying the 30-40-30 approximation to  $\ln X$ , which is normal. Substituting these values into Equations (8) through (11) would almost perfectly match the first four moments of  $X$ , which is lognormal. This approach would reduce all the errors in Figure B3 to zero. This “log-SM” approximation is a significant improvement over SM and is only slightly more complicated—requiring the use a logarithm and an exponential.

If we want to truncate the lognormal distribution at, say, the P1, we simply need to truncate the underlying normal distribution at the P1, which is elementary. We can then find a Gaussian quadrature for this truncated normal. For example, if we want to truncate the lognormal at the P1, a Gaussian quadrature of (0.32, 0.37, 0.31; P89.1, P50, P9.9) of the original untruncated normal will exactly match the first three moments of the normal and, therefore, the lognormal distribution. This approximation would reduce all of the errors in Rose’s (2001) example to zero.

### L.3 Why Not Just Simulate?

Thus far, we have compared discretization methods among themselves and found that some, such as Extended Swanson-Megill and naïve uses of McNamee-Celona, produce significant errors. One may wonder then, given the widespread availability and use of Monte Carlo methods, why use discretization at all?<sup>10</sup> We must remember that MC is also an approximation. Discretization methods induce *approximation error*, while MC methods include *sampling error*. The relevant question is whether or not MC is more accurate than discretization and which situations lend themselves to each method. To address this, this section offers some reasons as to why one may prefer to use discretization rather than simulation. More importantly, we determine how many MC samples are required to achieve the same accuracy as each discretization method.

---

<sup>10</sup> As one analyst argued, “After all, my Monte Carlo software package includes the lognormal distribution.”

### L.3.1 Modeling Difficulties.

From a modeling perspective, MC simulation is not easy to implement in situations that include downstream decisions or options. A generic example appears in Figure B4. In this case, a decision  $D1$  is made at the beginning of Stage 1. Then, a possibly continuous uncertainty  $X$  is revealed yielding a particular value  $x$ . At the start of Stage 2, decision  $D2$  is made, with knowledge of  $D1$  and  $x$ . Then, the possibly continuous uncertainty  $Y$  is realized. To make this situation concrete, suppose that  $D1$  is a decision about the acquisition of seismic data,  $X$  is the result of the seismic survey,  $D2$  is the decision to drill, and  $Y$  is the economic value of the reservoir (Stibolt and Lehman 1993; Bickel et al. 2008).

To evaluate this decision tree, we start at the end and roll-back (McNamee and Celona 1990; Clemen and Reilly 2001). Suppose we only seek to maximize expected values (i.e., we are risk neutral) and that  $X$  and  $Y$  are dependent. In this case, we would need to compute  $E[Y | X = x]$  to choose the optimal alternative at Stage 2. If this conditional mean cannot be expressed as a function of  $x$  we must simulate  $Y$  for each realization of  $X$ . If we are performing 1000 MC trials, for each of the 1000 trials for  $X$  we would have to perform an additional 1000 trials for  $Y | X = x$ , requiring  $1000^2$  or 1 million trials. This nested MC is not straightforward to implement and may require a long time to evaluate.

### L.3.2 Costly Evaluation.

The second reason one may prefer discretization methods is that they require fewer points (generally *many* fewer points) to match the underlying moments. Since each point or MC trial requires the evaluation of an output function (e.g., NPV) this process may be computationally costly. To understand this more fully, we can determine the *number of MC samples that would be required to achieve the same accuracy as a discretization method* (Pfeifer et al. 1991). We refer to this as  $S$ -equivalence.

The  $N$ -point Gaussian quadratures in Table B1 exactly match the first  $2N$  moments of the uniform, normal, and exponential distributions. Even in the case when  $N = 2$ , these approximations exactly match the mean, variance, and skewness of the original distributions. As such, MC could not do a better job approximating these moments. The shortcuts listed in Table B2 may, on the other hand, fail to match some moments of certain distributions. We will

determine  $S$ -equivalents for these shortcut-distribution combinations. Based on Equation (5) and to ease of exposition, we focus only on the raw moments.

*Uncertainty in Moments.* Imagine creating a new distribution  $Y_k = X^k$  from which we will sample in order to estimate the raw moments of  $X$ . The mean of  $Y_k$ ,  $\mu_k$ , is the  $k$ -th raw moment of  $X$ . For example, if  $k = 1$  then we would simply sample from  $X$ . Each set of  $S$  samples would produce an estimate of the mean of  $X$ . However, each time we rerun our MC simulation using  $S$  samples we would compute a different mean. In other words, the sample raw moments are random variables. We quantify their uncertainty by computing their central moments (e.g., the variance). The  $j$ -th central moment of  $Y_k$  is

$$m_j^k \equiv E[(Y_k - E[Y_k])^j] = E[(X^k - \mu_k)^j]. \quad (12)$$

If  $k$  equals 1 and  $j$  equals 2 then we have

$$m_2^1 \equiv E[(X - \mu_1)^2] = \mu_2 - \mu_1^2, \quad (13)$$

which is the familiar formula for the variance written in terms of raw moments. When referring specifically to the variance of the  $k$ -th raw moment, we will use the symbol  $\sigma_k^2 \equiv \sigma_k \sigma_k$ .

*Discretization Accuracy.* Let  $c_k$  be the difference between the true moment  $\mu_k$  and the approximate moment  $\hat{\mu}_k$ , obtained via a discretization. It will be useful to normalize this difference by the standard deviation of the moment, which may be obtained via Equation (12).

We write the accuracy of the approximation as

$$\delta_k = \frac{c_k}{\sqrt{m_k^2}} = \frac{\mu_k - \hat{\mu}_k}{\sigma_k}. \quad (14)$$

*Computing  $S$ -Equivalence.* Now, suppose that instead of using a discretization, we estimate  $\mu_k$  via MC simulation. Let the estimated value, which is an average, given  $S$  samples be  $\bar{\mu}_{k:S}$ . According to the Central Limit Theorem (CLT),  $\bar{\mu}_{k:S}$  is normally distributed with mean  $\mu_k$  and variance  $\sigma_k^2 / S$  for large  $S$ . Thus, the probability that the simulation will more accurately estimate  $\mu_k$  than the discretization method is (see section K6 for derivation)

$$P\left(\left|\bar{\mu}_{k:S} - \mu_k\right| \leq c_k\right) = 2\Phi(x) - 1. \quad (15)$$

where  $x = \delta_k \sqrt{S}$  and  $\Phi$  is the standard normal cdf. If we want this probability to be  $\tau$  then we must take

$$S = \delta_k^{-2} \Phi^{-1} \left( \frac{\tau+1}{2} \right)^2 \quad (16)$$

samples (see section K6 for derivation). For example, if we want the simulation to have a 95% chance of estimating the  $k$ -th raw moment more accurately than the discretization we must take  $S = \delta_k^{-2} \Phi^{-1} (1.95/2)^2 \approx \delta_k^{-2} \cdot 1.96^2$  samples. Thus, Equation (16) establishes an equivalence between discretization and simulation (Pfeifer et al. 1991).<sup>11</sup>

Equation (16) will not work well when the underlying random variable  $X^k$  is highly skewed or kurtotic. In these cases, the CLT approximation may not be accurate for the number of trials that Equation (16) says we need to perform.<sup>12</sup> In this case, we must rely on the Edgeworth expansion (Cramér 1946; Hall 1992), which corrects the CLT for higher underlying moments. For the distributions we consider here, this issue primarily affects the lognormal distribution and, to a lesser degree, the exponential. In this case, the probability that the MC simulation will more accurately estimate  $\mu_k$  than the discretization method is (see section K6 for derivation)

$$P(|\bar{\mu}_{k,S} - \mu_k| \leq c_k) = 2\Phi(x) - 1 - \frac{1}{S} \left[ \frac{1}{12} \gamma_3^k (x^3 - 3x) \varphi(x) + \frac{1}{36} (\gamma_4^k)^2 (x^5 - 10x^3 + 15x) \varphi(x) \right] + O(S^{-2}), \quad (17)$$

where  $x = \delta_k \sqrt{S}$ , as before, and  $\varphi$  is the standard normal pdf. The term  $O(S^{-2})$  signifies that the remaining terms are at most of order  $S^{-2}$ . The terms  $\gamma_3^k = m_k^3 m_k^{-3/2} = E[(X^k - \mu_k)^3] E[(X^k - \mu_k)^2]^{-3/2}$  and  $\gamma_4^k = m_k^4 m_k^{-2} = E[(X^k - \mu_k)^4] E[(X^k - \mu_k)^2]^{-2} - 3$  are the skewness and kurtosis, respectively. If the skewness and kurtosis are zero (or small), or if  $S$  is very large, then Equation (17) reduces to Equation (15) and we can use Equation (16) to find the equivalent number of samples. On the other hand, if the skewness and kurtosis are non-negligible then we must numerically solve Equation (17) for  $S$ .

Lerche and Mudford (2005a, b) investigate the number of samples required to estimate the mean of several distributions including lognormal and exponential and base their estimates

---

<sup>11</sup> Pfeifer et. al (1991) determined  $S$ -equivalence for matching the mean. They did not consider higher moments or use the Edgeworth expansion, as we do here.

<sup>12</sup> The CLT holds under quite general conditions and states that the sum of an infinite number of random variables converges in distribution to the normal. However, it does not specify how quickly this convergence will take place.

on the CLT. This approximation, while not fully appropriate, should not have induced substantial errors in their case because of their focus on estimating only the mean. However, we are interested in estimating higher moments and these higher moments (e.g.,  $E[X^4]$ ) will induce additional skewness and kurtosis and the CLT will fail to work well in these cases. For example, for the lognormal distribution that we discuss here, we find that Equation (16) overestimates the required number of samples by almost 25%.<sup>13</sup>

Table B7 provides the 95%  $S$ -equivalences based on Equation (17) for the EPT, GQN, ESM, MCS, MRO, and ZDT shortcuts for the uniform  $U(0, b)$ , normal  $N(0, \sigma)$ , triangular  $T(0, b, b/2)$ , exponential  $E(\lambda)$ , and lognormal  $L(\mu, 1)$  distributions. These results hold for any (see section K6):

- Exponential distribution.
- Uniform distribution bounded by zero on one side,  $U(0, b)$ .
- Normal distribution centered at zero,  $N(0, \sigma)$ .
- Symmetric triangular distribution bounded by zero on one side,  $T(0, b, b/2)$ .
- Lognormal distribution with unit variance (of  $\ln X$ ),  $L(\mu, 1)$ .

While the results in Table B7 do not hold for all possible distributions within a given family, they should give the reader a sense for the magnitude of MC samples that are required to match a particular discretization.

These results are striking. First, setting aside the fact that all the approximations we consider perfectly match the mean of symmetric distributions ( $U, N, T(a, b, b/2)$ ), most of the shortcuts are equivalent to thousands, if not, *tens of thousands* of MC samples. For example, it would take 57,469 MC samples to have a 95% chance of estimating the second raw moment of an exponential with more accuracy than EPT, which Pearson and Tukey did not even suggest using to estimate any moment beyond the mean. Second, EPT and GQN are clearly dominant,

---

<sup>13</sup> In the case of the lognormal, negative values are not possible. This causes Equation (16), which is based on the CLT and assumes the distribution of the mean is normally distributed, to require more samples so as to reduce the variance of the mean enough such that negative values are very unlikely—at least less than  $(1 - 0.95) / 2 = 2.5\%$ . The Edgeworth expansion performs better by taking the kurtosis and the skewness into account. These normalized central moments are *very* large in the case of the lognormal—especially we when we are considering the distribution of  $X^3$  and  $X^4$ .

having larger  $S$ -equivalences than the other approximations. These approximations do very well on everything except the third moment of uniform distributions. Consider estimating the mean of lognormal distribution: EPT and GQN are equivalent to almost 12 and 65 times more samples, respectively, than ESM (29,499 or 161,943 compared to 2,495). Third, these results serve to emphasize our earlier conclusion that the performance of ESM and MCS is quite poor in some cases. For example, ESM and MCS are only equivalent to 941 and 676 MC samples, respectively, in terms of estimating the second moment of the  $L(\mu,1)$  distribution.<sup>14</sup> While the direction of this result is not surprising, given that we know these methods underestimate variance, its magnitude is shocking. Use of ESM to model the uncertainty of a lognormal distribution, which occurs whenever it is applied to estimations of oil or gas reserves, is equivalent to running less than 1000 MC trials!

A large  $S$ -equivalence is a result of two possible factors: (1) high accuracy of the discretization method and (2) the difficulty of simulating the underlying random variable. As a case in point, consider the EPT and GQN approximations of the lognormal distribution. The  $S$ -equivalences for estimating the mean are large because the discretizations estimate the mean closely. The  $S$ -equivalences for the third moment are large because approximating the third moment of a lognormal distribution via MC simulation is very difficult, requiring many tens of thousands of MC samples. Given the large number of samples that are required in most cases, we see that discretization is a viable, and in some cases preferable, alternative to MC simulation.<sup>15</sup>

#### L.4 Does Discretization Matter?

“New”<sup>16</sup> methodologies are often resisted on the grounds that they will not obviously make a material difference. While we do not have space to fully address this issue here, we note

---

<sup>14</sup> The discretization shortcuts do not provide unbiased estimates of the underlying moments. We do not address the issue of bias here. Rather, we assume that over- and under-estimates of the moments are equally costly.

<sup>15</sup> Of course, we are interested in preserving the moments of the output (e.g., NPV), which might be a function of many input uncertainties and several downstream decisions. We do not address this more complicated issue here.

<sup>16</sup> We place “new” in quotes because Person-Tukey predates Swanson-Megill by almost a decade and Gaussian quadrature predates all other methods by at least a century.

that most of the methods presented in this paper are no more complicated than existing approximations, but are more accurate. For example:

- EPT and GQN are a three-point approximations, like ESM and MCS, but are more accurate across a range of distributions.
- The approximations given in Tables K3 through K5 are three-point approximations that match the first three moments of the listed distributions—again, exceeding the performance of ESM and MCS.
- The four-point approximations given in Table B6 match the mean and variance of lognormal distributions, resulting in better performance than the three-point ESM and MCS shortcuts.
- The log-SM method that we outline will perfectly match the all the moments of the lognormal and is only slightly more complex than SM.
- Other methods that we discuss, such as moment matching for distributions not listed in Tables K4 through K7, are more complex than current practice. However, this increase in complexity will increase the accuracy of the results (much in the same way that 3-dimensional reservoir simulation is more complex, but more accurate, than 2-dimensional simulation).

However, we realize that some will require convincing before changing their practice. To address this, we offer the following, necessarily simple, example.

#### **L.4.1 Illustrative Example.**

Suppose an oil company is considering the purchase of a prospect that contains an uncertain volume of oil. For the sake of argument, suppose the reserves are believed to be lognormally distributed with a mean of 90 MMBOE and a standard deviation of 118 MMBOE. These parameters correspond to a standard deviation for log-reserves of 1.0 and, thus, we can make use of the 4-point approximation given in Table B6. To illustrate the accuracy of each approximation method, we will consider three different valuation scenarios. In the first, the company believes, based on market transactions (Howard and Harp, Jr., 2009), that proved reserves in this geographic location and depositional environment are worth \$5 per BOE. This case is labeled “Linear” in Figure B5. We also consider a case where the value function is convex

and another that is concave. The concave case could correspond to a situation where the host government takes a small share of small fields to increase the probability of development, but will take a proportionally larger share as field size increases. The convex case might represent large initial fixed costs, perhaps for infrastructure, that do not scale in proportion to field size. Whatever the case may be, the important point, as far as the example is concerned, is that discretization accuracy depends upon the shape of the value function.

We next apply the discretization methods discussed in this paper to the reserves distribution and compute the expected NPV of the prospect. For the bracket median and bracket mean discretizations, we consider 3- and 4-point approximations with weightings of 25-50-25 and 10-40-40-10, respectively. We compare the approximate value from each discretization to the “exact” value, which we obtain via 100,000 Monte Carlo samples. The error for each approximation is presented in Table B8.

The errors in the linear case are simply the errors of each discretization in estimating mean reserves. We see that the 4-point lognormal approximation, EPT, GQN, and the bracket mean approaches estimate the expected NPV to within 1%. The other methods underestimate the mean by at least 5%; the bracket median approach is especially poor. Errors in the concave and convex cases differ from the linear case because in these situations the variance (and other moments) of the input distribution is important. Since most of discretization methods misestimate the variance, they will misestimate the expected value of the output (NPV). The errors are less in the concave case because the concavity of the value function serves to reduce impact of misestimating the variance of the input distribution. In the convex case, we see that the 4-point lognormal and the 4-point bracket mean perform very well. Table B9 presents the error in the variance of NPV. In this case, we see that all the approximations underestimate the output variance. The performance of ESM, MCS, and the bracket median approaches are especially poor. While the 4-point lognormal and the 4-point bracket mean discretizations are significantly better. Of course, the 4-point lognormal and the bracket mean approximations have been specifically tailored to the underlying distribution, which demonstrates the importance of this practice.

Is underestimating the mean by 10% and the variance by 50% a problem? Clearly, if the company is setting a purchase (or sale) price, underestimating the value of the prospect by 10% could undermine the opportunity. This would be especially true in a competitive situation, such as bidding. In addition, the optimal bid amount depends critically upon the company's estimate of uncertainty. Underestimating the variance by 50-80% may increase the probability of overpaying for the property. Thus, we see that discretization matters and could have a material impact on decision making. In what other part of the business would misestimating key performance metrics by 10-80% be acceptable?

## **L.5 Conclusion: Recommendations and Discussion**

We conclude with a set of recommendations and observations regarding the oil and gas industry's approach to probabilistic modeling.

### **L.5.1 Recommendations.**

If one is interested in estimating the moments of an output distribution, then closely matching the moments of the input distributions is a necessary requirement. In this case, the best discretization methods presented in this paper are very accurate, matching  $2N$  moments (including the zeroth) with only  $N$  points. They could therefore be used instead of Monte Carlo simulation, at least early in an analysis, to quantify uncertainty with only a few evaluations of the output function. The primary recommendations that follow from our work are:

- For maximum accuracy, we should use moment matching and apply it directly to each input distribution. For several distribution families, the appropriate weights and values have already been calculated (see Table B1 and Tables K3 through K7). In other cases, the quadrature could be calculated on a case-by-case basis using methods detailed in Stroud and Secrest (1966), Miller and Rice (1983), Smith (1993), or Davis and Rabinowitz (1984), for example.
- If moment matching is too difficult to implement or communicate then the Equal Areas approach is not an unreasonable alternative. However, one must bear in mind that this approach will tend to underestimate the variance (and other higher moments).

- If one wishes to fix the values at the P90, P50, P10 or weights at 25-50-25 or 30-40-30 for communication, assessment, or computational reasons, then using the moment matching discretizations in Tables K3 through K5 will result in maximum accuracy.
- When dealing with lognormal distributions, work with the log of the random variable, which translates it to a normal random variable. Then apply the discretization methods to the transformed variable.
- Swanson’s Mean (or ESM), the McNamee-Celona shortcut, and ZDT should not be used as part of a final analysis (recall McNamee and Celona explicitly warned against using their shortcut in this way). Direct application of ESM to lognormal variables, as is common, to estimate their moments (mean, variance, etc.), should be considered an unacceptable professional practice.
- Well chosen discrete approximations are equivalent to tens of thousands of MC samples. This argues for greater use of discretization methods.

### L.5.2 Discussion and Conclusion.

Despite attempts to justify it (Hurst et al. 2000; Rose 2001, Appendix A), Swanson’s Mean has no theoretic justification for use with any distribution other than normal (direct use with the lognormal is especially error prone). Megill noted its problems nearly 40 years ago. Why would we expect MCS or ESM/SM (or the other shortcuts), which are *symmetric*, to preserve the mean of skewed distributions?<sup>17</sup> Why have we institutionalized a method known to be biased? Megill’s answer was that Swanson’s Mean offers “protection” from uncertainty since its estimates are known to be biased low (Megill 1984, p. 188). This is in much the same spirit as using a high discount rate or a low oil price to account for uncertainty. We do not support such adjustments; decision makers should be provided with unbiased estimates of the risks facing the company.

Another explanation for the acceptance of Swanson’s Mean is the industry’s focus on the mean and “risking” a prospect (Bickel and Bratvold 2008), to the exclusion of other moments.

---

<sup>17</sup> Some analysts have told us that MCS (25-50-25) is good for symmetric distributions, but that ESM (30-40-30) should be used when the distribution is skewed!

Paraphrasing Steven Jay Gould, “the mean is not the message” (Gould 1985). Risking a prospect does nothing to either understand the risk or to help manage it. It simply implies that the mean, the probability weighted average, was calculated. Decision makers are not indifferent to all projects with the same mean. Rather, they want to understand the surrounding uncertainty and the risk. Furthermore, as demonstrated by Equation (5) and illustrated in the previous section, one may not accurately determine the mean of an output distribution without accurately representing the mean and higher moments of the input distributions.

Why do we continue to use ESM when more accurate alternatives exist? Why do we spend millions of dollars on reservoir simulation models and then represent the output of those models in a decision tree with simplistic discretizations? Perhaps it is simply a matter of tradition and path dependence: “We’ve always done it this way”, “That’s the way everybody else does it”, or “That’s the way I was taught.” In fact, some companies mandate that either McNamee-Celona or Extended Swanson-Megill must be used in project valuations. We hope that our paper will encourage improved practice.

## L.6 Derivations

### L.6.1 Probability the MC Simulation will be more Accurate than the Approximation

$$\begin{aligned}
 P\left(\left|\bar{\mu}_{k:S} - \mu_k\right| \leq c\right) &= P\left(\bar{\mu}_{k:S} - \mu_k \leq +c\right) - P\left(\bar{\mu}_{k:S} - \mu_k \leq -c\right) \\
 &= P\left(\frac{\bar{\mu}_{k:S} - \mu_k}{\sigma_k / \sqrt{S}} < \frac{+c}{\sigma_k / \sqrt{S}}\right) - P\left(\frac{\bar{\mu}_{k:S} - \mu_k}{\sigma_k / \sqrt{S}} < \frac{-c}{\sigma_k / \sqrt{S}}\right) \\
 &= P\left(\frac{\bar{\mu}_{k:S} - \mu_k}{\sigma_k / \sqrt{S}} < \delta\sqrt{S}\right) - P\left(\frac{\bar{\mu}_{k:S} - \mu_k}{\sigma_k / \sqrt{S}} < -\delta\sqrt{S}\right) \\
 &= \Phi\left(\delta\sqrt{S}\right) - \Phi\left(-\delta\sqrt{S}\right) \\
 &= \Phi\left(\delta\sqrt{S}\right) - \left[1 - \Phi\left(\delta\sqrt{S}\right)\right] \\
 &= 2\Phi\left(\delta\sqrt{S}\right) - 1.
 \end{aligned}$$

### L.6.2 S-Equivalence for CLT

$$\begin{aligned}
 2\Phi\left(\delta\sqrt{S}\right) - 1 &= \tau \\
 \delta\sqrt{S} &= \Phi^{-1}\left(\frac{\tau+1}{2}\right) \\
 S &= \delta^{-2}\Phi^{-1}\left(\frac{\tau+1}{2}\right)^2.
 \end{aligned}$$

### L.6.3 Edgeworth Expansion

The first three terms of the Edgeworth expansion are

$$F_S(x) = \Phi(x) - \frac{1}{S^{1/2}} \left( \frac{1}{6} \frac{\kappa_3}{\kappa_2^{3/2}} \Phi^{(3)}(x) \right) + \frac{1}{S} \left( \frac{1}{24} \frac{\kappa_4}{\kappa_2^{4/2}} \Phi^{(4)}(x) + \frac{1}{72} \left( \frac{\kappa_3}{\kappa_2^{3/2}} \right)^2 \Phi^{(6)}(x) \right) \\ - \frac{1}{S^{3/2}} \left( \frac{1}{120} \frac{\kappa_5}{\kappa_2^{5/2}} \Phi^{(5)}(x) + \frac{1}{144} \frac{\kappa_3}{\kappa_2^{3/2}} \frac{\kappa_4}{\kappa_2^{4/2}} \Phi^{(7)}(x) + \frac{1}{1296} \left( \frac{\kappa_3}{\kappa_2^{3/2}} \right)^3 \Phi^{(9)}(x) \right) \\ + O(S^{-5/2})$$

where,  $\kappa_r$  is the  $r$ -th cumulant.  $\kappa_2$  is the variance.  $\kappa_1$  is the mean.

$\Phi^{(j)}$  is the  $j$ -th derivative of  $\Phi^{(j)}$ . Thus,  $\Phi^{(j)} = \varphi^{(j-1)}$ , where  $\varphi$  is the standard normal pdf.

The  $j$ -th derivative of the standard normal pdf is  $\varphi^{(j)} = (-1)^j H_j(x)\varphi(x)$ , where  $H_j(x)$  is the Hermite polynomial of order  $j$ . Thus,  $\Phi^{(j)} = \varphi^{(j-1)} = (-1)^{j-1} H_{j-1}(x)\varphi(x)$ . The Hermite polynomials are an even (odd) function when  $j - 1$  is even (odd). Thus, we have

$$F_S(x) = \Phi(x) - \frac{1}{S^{1/2}} \left( \frac{1}{6} \frac{\kappa_3}{\kappa_2^{3/2}} H_2(x)\varphi(x) \right) - \frac{1}{S} \left( \frac{1}{24} \frac{\kappa_4}{\kappa_2^{4/2}} H_3(x)\varphi(x) + \frac{1}{72} \left( \frac{\kappa_3}{\kappa_2^{3/2}} \right)^2 H_5(x)\varphi(x) \right) \\ - \frac{1}{S^{3/2}} \left( \frac{1}{120} \frac{\kappa_5}{\kappa_2^{5/2}} H_4(x)\varphi(x) + \frac{1}{144} \frac{\kappa_3}{\kappa_2^{3/2}} \frac{\kappa_4}{\kappa_2^{4/2}} H_6(x)\varphi(x) + \frac{1}{1296} \left( \frac{\kappa_3}{\kappa_2^{3/2}} \right)^3 H_8(x)\varphi(x) \right) + O(S^{-2}).$$

In the discretization work, we are looking at the probability of being within the interval  $[-x, x]$ . Which is  $F_S(x) - F_S(-x)$ . Since  $H_j$  is even when  $j$  is even and  $\varphi(x)$  is an odd function, the terms of order  $S^{-j/2}$  will cancel in the subtraction when  $j$  is odd and we will be left with

$$F_S(x) - F_S(-x) = 2\Phi(x) - 1 - \frac{1}{S} \left[ \frac{1}{12} \frac{\kappa_4}{\kappa_2^{4/2}} H_3(x)\varphi(x) + \frac{1}{36} \left( \frac{\kappa_3}{\kappa_2^{3/2}} \right)^2 H_5(x)\varphi(x) \right] + O(S^{-2}) \\ = 2\Phi(x) - 1 - \frac{1}{S} \left[ \frac{1}{12} \gamma_4 H_3(x)\varphi(x) + \frac{1}{36} \gamma_3^2 H_5(x)\varphi(x) \right] + O(S^{-2}),$$

where  $\gamma_3$  is the skewness and  $\gamma_4$  is the kurtosis. The third and fifth Hermite polynomials are  $H_3(x) = x^3 - 3x$  and  $H_5(x) = x^5 - 10x^3 + 15x$ . Substituting these values into the above yields Equation (17).

### L.6.4 Exponential S-Equivalence

The inverse cdf for the exponential is  $F^{-1}(\alpha_i) = -\lambda^{-1} \ln(1 - \alpha_i)$ . The approximate moment for approximation  $A$  is then

$$\hat{\mu}_k = \sum_{i=1}^N p_i^A \left( -\lambda^{-1} \ln(1 - \alpha_i^A) \right)^k = -\lambda^{-k} \sum_{i=1}^N p_i^A \left( \ln(1 - \alpha_i^A) \right)^k.$$

The  $k$ -th raw moment of the exponential is

$$\mu_k = -\lambda^{-k} \Gamma(k+1).$$

The variance of the  $k$ -th raw moment is

$$\sigma_k^2 = -\lambda^{-2k} (\Gamma(2k+1) - \Gamma(k+1)^2).$$

Thus, we then have

$$\delta_k = \frac{\mu_k - \hat{\mu}_k}{\sigma_k} = \frac{\lambda^{-k} \Gamma(k+1) - \lambda^{-k} \sum_{i=1}^N p_i^A (\ln(1-u_i^A))^k}{\lambda^{-k} (\Gamma(2k+1) - \Gamma(k+1)^2)^{1/2}} = \frac{\Gamma(k+1) - \sum_{i=1}^N p_i^A (\ln(1-u_i^A))^k}{(\Gamma(2k+1) - \Gamma(k+1)^2)^{1/2}}.$$

We see that  $\delta$  is independent of  $\lambda$  and the  $S$ -equivalences that we list in Table B7 hold for any exponential pdf.

### L.6.5 Uniform S-Equivalence

The inverse cdf for the uniform is  $F^{-1}(\alpha_i) = a + \alpha_i(b-a)$ . The approximate  $k$ -th raw moment for approximation  $A$  is then

$$\hat{\mu}_k = \sum_{i=1}^N p_i^A (a + \alpha_i^A(b-a))^k.$$

The true  $k$ -th raw moment of the uniform is

$$\mu_k = \frac{a^{k+1} - b^{k+1}}{(a-b)(k+1)}.$$

The variance of the  $k$ -th raw moment is

$$\sigma_k^2 = \frac{(a^{2k+2} + b^{2k+2})k^2 - a^{2k+1}b(k+1)^2 + ab(2(ab)^k(2k+1) - b^{2k}(k+1)^2)}{(a-b)^2(k+1)^2(2k+1)}.$$

Thus, we have

$$\delta_k = \frac{\mu_k - \hat{\mu}_k}{\sigma_k} = \frac{\frac{a^{k+1} - b^{k+1}}{(a-b)(k+1)} - \sum_{i=1}^N p_i^A (a + u_i^A(b-a))^k}{\left( \frac{(a^{2k+2} + b^{2k+2})k^2 - a^{2k+1}b(k+1)^2 + ab(2(ab)^k(2k+1) - b^{2k}(k+1)^2)}{(a-b)^2(k+1)^2(2k+1)} \right)^{1/2}}.$$

If  $a = 0$  then we have

$$\delta_k = \frac{\mu_k - \hat{\mu}_k}{\sigma_k} = \frac{\left( \frac{1}{k+1} - \sum_{i=1}^N p_i^A (u_i^A)^k \right)}{\frac{k}{k+1} \frac{1}{(2k+1)^{1/2}}}$$

and we see that  $\delta$  is independent of  $b$  and the  $S$ -equivalences that we list in Table B7 hold for any  $U(0,b)$ . If  $a$  is nonzero then  $S$ -equivalences will differ from those shown in Table B7.

### L.6.6 Normal S-Equivalence

The inverse cdf for the normal is  $F^{-1}(\alpha_i) = \mu + \sigma\Phi^{-1}(\alpha_i)$ . The approximate  $k$ -th raw moment for approximation  $A$  is then

$$\hat{\mu}_k = \sum_{i=1}^N p_i^A \left( \mu + \sigma\Phi^{-1}(\alpha_i^A) \right)^k.$$

In the case of the normal, the raw moments are quite complex and we restrict our attention to the case where  $\mu = 0$ . In this case, the true  $k$ -th raw moment of the normal is

$$\mu_k = \pi^{-1/2} \sigma^k 2^{k/2-1} (1 + e^{ik\pi}) \Gamma\left(\frac{k+1}{2}\right).$$

Similarly, the variance of the  $k$ -th raw moment is

$$\sigma_k^2 = \pi^{-1} 2^{k-2} \sigma^{2k} \left( 2(1 + e^{2ik\pi}) \sqrt{\pi} \Gamma\left(k + \frac{1}{2}\right) - (1 + e^{ik\pi})^2 \Gamma\left(\frac{k+1}{2}\right)^2 \right).$$

Thus, we have

$$\delta_k = \frac{\mu_k - \hat{\mu}_k}{\sigma_k} = \frac{2^{k/2-1} (1 + e^{ik\pi}) \Gamma\left(\frac{k+1}{2}\right) - \sqrt{\pi} \sum_{i=1}^N p_i^A \left( \Phi^{-1}(u_i^A) \right)^k}{\left( 2^{k-2} \left( 2(1 + e^{2ik\pi}) \sqrt{\pi} \Gamma\left(k + \frac{1}{2}\right) - (1 + e^{ik\pi})^2 \Gamma\left(\frac{k+1}{2}\right)^2 \right) \right)^{1/2}},$$

and see that  $\delta$  is independent of  $\sigma$  and the  $S$ -equivalences that we list in Table B7 hold for any  $N(0, \sigma)$ .

### L.6.7 Triangular S-Equivalence

The inverse cdf for the triangular is

$$F^{-1}(\alpha_i) = \begin{cases} a + \sqrt{(b-a)(c-a)}\alpha_i & 0 \leq \alpha_i < F(c) \\ b - \sqrt{(b-a)(b-c)}(1-\alpha_i) & F(c) \leq \alpha_i \leq 1 \end{cases}.$$

The approximate  $k$ -th raw moment for approximation  $A$  is then

$$\hat{\mu}_k = \sum_{i=1}^{N_c} p_i^A \left( a + \sqrt{(b-a)(c-a)}\alpha_i \right)^k + \sum_{i=N_c+1}^N p_i^A \left( b - \sqrt{(b-a)(b-c)}(1-\alpha_i) \right)^k,$$

where  $N_c$  is the approximation point  $i$  at which  $F(c) = \alpha_i$ .

The raw moments are complex in the case of the triangular and we restrict our attention to the case where  $a = 0$ . Writing  $c$  as a fraction  $z$  of  $b$ , the approximate  $k$ -th raw moment is then

$$\hat{\mu}_k = b^k \left( \sum_{i=1}^{N_c} p_i^A \left( \sqrt{z^{-1}}\alpha_i \right)^k + \sum_{i=N_c+1}^N p_i^A \left( 1 - \sqrt{(1-z^{-1})(1-\alpha_i)} \right)^k \right).$$

The true  $k$ -th raw moment is

$$\mu_k = \frac{2(bc^{k+2} - b^{k+2}c)}{b(b-c)c(2+3k+k^2)} = b^k \frac{2(z^{-k-2} - z^{-1})}{(z^{-1} - z^{-2})(k^2 + 3k + 2)}.$$

The variance of the  $k$ -th raw moment is

$$\sigma_k^2 = b^{2k} \frac{[k^2(5+k) + z^{-2k-1}(k^2(5+k) + 8(1+2k)) - (z^{-1} + -z^{-2k-1})(1+k)(2+k)^2]}{z^{-2}(z-1)^2(1+2k)(2+3k+k^2)^2}.$$

Thus, we have

$$\delta_k = \frac{\mu_k - \hat{\mu}_k}{\sigma_k} = \frac{\frac{2(z^{-k-2} - z^{-1})}{(z^{-1} - z^{-2})(k^2 + 3k + 2)} - \left( \sum_{i=1}^{N_c} p_i^A \left( \sqrt{z^{-1}\alpha_i} \right)^k + \sum_{i=N_c}^N p_i^A \left( 1 - \sqrt{(1-z^{-1})(1-\alpha_i)} \right)^k \right)}{\sqrt{\frac{[k^2(5+k) + z^{-2k-1}(k^2(5+k) + 8(1+2k)) - (z^{-1} + -z^{-2k-1})(1+k)(2+k)^2]}{z^{-2}(z-1)^2(1+2k)(2+3k+k^2)^2}}},$$

and we see that  $\delta$  is independent of  $b$  and the S-equivalences that we list in Table B7 hold for any  $T(0, b, b/2)$ .

### L.6.8 Lognormal S-Equivalence

The inverse cdf for the lognormal is  $F^{-1}(\alpha_i) = \text{Exp}[\mu + \sigma\Phi^{-1}(\alpha_i)]$ . The approximate  $k$ -th raw moment for approximation A is then

$$\hat{\mu}_k = \sum_{i=1}^N p_i^A \left( \text{Exp}[\mu + \sigma\Phi^{-1}(\alpha_i^A)] \right)^k = \sum_{i=1}^N p_i^A \text{Exp}[k\mu + k\sigma\Phi^{-1}(\alpha_i^A)].$$

The true  $k$ -th raw moment of the lognormal is

$$\mu_k = \text{Exp}\left[ k\mu + \frac{1}{2}k^2\sigma^2 \right].$$

The variance of the  $k$ -th raw moment is

$$\sigma_k^2 = \text{Exp}\left[ 2k\mu + k^2\sigma^2 \right] \left( \text{Exp}\left[ k^2\sigma^2 \right] - 1 \right).$$

Thus, we have

$$\delta_k = \frac{\mu_k - \hat{\mu}_k}{\sigma_k} = \frac{\text{Exp}\left[ \frac{1}{2}k^2\sigma^2 \right] - \sum_{i=1}^N p_i^A \text{Exp}[k\sigma\Phi^{-1}(\alpha_i^A)]}{\sqrt{\text{Exp}\left[ k^2\sigma^2 \right] \left( \text{Exp}\left[ k^2\sigma^2 \right] - 1 \right)}},$$

and we see that  $\delta$  is independent of  $\mu$  and the S-equivalences that we list in Table B7 hold for any  $L(\mu, 1)$ .

## L.7 References

Arild, Ø., Lohne, H.P., and Bratvold, R. 2008. A Monte Carlo Approach to Value of Information. Paper SPE IPTC 11969 presented at the International Petroleum Technology Conference, Kuala Lumpur, Malaysia, 3-5 December.

- Bickel, J.E. and Bratvold, R.B. 2008. From Uncertainty Quantification to Decision Making in the Oil and Gas Industry. *Journal of Energy Exploration and Exploitation* **26** (5) 311-325.
- Bickel, J.E., Gibson, R.L., McVay, D.A., Pickering, S., and Waggoner, J. 2008. Quantifying the Reliability and Value 3D Land Seismic. *Reservoir Evaluation and Engineering* **11** (5) 832-841.
- Clemen, R.T. and Reilly, T. 2001. *Making Hard Decisions*. Pacific Grove, CA: Duxbury.
- Cramér, H. 1946. *Mathematical Methods of Statistics*. Princeton, NR: Princeton University Press.
- Davis, Philip and Philip Rabinowitz. 1984. *Methods of Numerical Integration, Second Edition*. Orlando, FL. Academic Press, Inc.
- D'Errico, J.R. and Zaino, N.A. 1988. Statistical Tolerancing Using a Modification of Taguchi's Method. *Technometrics* **30** (4) 397-405.
- Gauss, C.F. 1866. Methodus nova integralium valores per approximationem inveniendi. In C. F. *Gauss Werke* 163-196. Koniglichen Gesellschaft der Wissenschaften.
- Gould, S.J., *The Median Isn't the Message*, in *Discover Magazine*. 1985. p. 40-42.
- Hall, P. 1992. *The Bootstrap and Edgeworth Expansion* Springer Series in Statistics. New York, NY: Springer-Verlag.
- Howard, Alex W. and Harp, Jr., Alan B. 2009. Oil and Gas Company Valuations. *Business Valuation Review* **28** (1) 30-35.
- Hurst, A., Brown, G.C., and Swanson, R.I. 2000. Swanson's 30-40-30 rule. *AAPG Bulletin* **84** (12) 1883-1891.
- Keefer, D.L. 1994. Certainty Equivalents for Three-Point Discrete-Distribution Approximations. *Management Science* **40** (6) 760-773.
- Keefer, D.L. and Bodily, S.E. 1983. Three-Point Approximations for Continuous Random Variables. *Management Science* **29** (5) 595-609.
- Keefer, D.L. and Pollock, S.M. 1980. Approximations and Sensitivity in Multiobjective Resource Allocation. *Operations Research* **28** (1) 114-128.
- Lerche, I. and Mudford, B.S. 2005a. How Many Monte Carlo Simulations Does one Need to do? *Energy Exploration & Exploitation* **23** (6) 405-427.

- Lerche, I. and Mudford, B.S. 2005b. How Many Monte Carlo Simulations Does one Need to do? II. Lognormal, Binomial, Cauchy, and Exponential Distributions. *Energy Exploration & Exploitation* **23** (6) 429-461.
- McNamee, P. and Celona, J. 1990. *Decision Analysis with Supertree*, Second edition. South San Francisco, California: The Scientific Press.
- McNamee, P. and Celona, J. 1991. *Decision Analysis with Supertree: Instructor's Manual*, Second edition. South San Francisco, CA: The Scientific Press.
- Megill, R.E. 1984. *An Introduction to Risk Analysis*, Second edition. Tulsa, Oklahoma: PennWell.
- Miller, A.C., III and Rice, T.R. 1983. Discrete Approximations of Probability Distributions. *Management Science* **29** (3) 352-362.
- Nicholas A. Zaino, J. and D'Errico, J. 1989. Optimal Discrete Approximations for Continuous Outcomes with Applications in Decision and Risk Analysis. *Journal of the Operational Research Society* **40** (4) 379-388.
- Papoulis, A. 1984. *Probability, random variables, and stochastic processes*. New York, NY: McGraw-Hill.
- Pearson, E.S. and Tukey, J.W. 1965. Approximate Means and Standard Deviations Based on Distances between Percentage Points of Frequency Curves. *Biometrika* **52** (3/4) 553-546.
- Pfeifer, P.E., Bodily, S.E., and Jr., F., Sherwood C. 1991. Pearson-Tukey Three-Point Approximations Versus Monte-Carlo Simulation. *Decision Sciences* **22** 74-90.
- Prange, M., Bailey, W., and Kuchuk, F. 2009. Quantifying the Value of a Future Well Test with Measurement Uncertainty. Paper SPE IPTC 13595-PP presented at the International Petroleum Technology Conference, Doha, Qatar, 7-9 December.
- Rose, P. 2001. *Risk Analysis and Management of Petroleum Exploration Ventures* AAPG Methods in Exploration Series. Tulsa, OK: American Association of Petroleum Geologists.
- Smith, J.E. 1993. Moment Methods for Decision Analysis. *Management Science* **39** (3) 340-358.
- Stibolt, R.D. and Lehman, J. 1993. The Value of a Seismic Option. Paper SPE 25821 presented at the SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas, 29-30 March.
- Stroud, A.H. and Secrest, D. 1966. *Gaussian Quadrature Formulas*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Taguchi, G. 1978. Performance Analysis Design. *International Journal of Production Research* (16) 521-530.

Willigers, B.J.A. 2009. Practical Portfolio Simulation: Determining the Precision of a Probability Distribution of a Large Asset Portfolio when the Underlying Project Economics are Captured by a Small Number of Discrete Points. Paper SPE 124180 presented at the Annual Technical Conference and Exhibition, New Orleans, Louisiana, 4-7 October.

## L.8 Tables

**Table B1: Two-, Three-, and Four-Point Gaussian Quadrature Formulae for Common Distributions**

Distribution	Two Points		Three Points		Four Points	
	$p_i$	$\alpha_i$	$p_i$	$\alpha_i$	$p_i$	$\alpha_i$
Uniform $f(x) = 1$ $0 \leq x \leq 1$	0.500	78.9	0.278	88.7	0.174	93.1
	0.500	21.1	0.444	50.0	0.326	67.0
			0.278	11.3	0.326	33.0
				0.174	6.9	
Normal $f(x) = (2\pi)^{1/2} \text{Exp}[-\frac{1}{2}x^2]$ $-\infty \leq x \leq \infty$	0.500	84.1	0.167	95.8	0.046	99.0
	0.500	15.9	0.667	50.0	0.454	77.1
			0.167	4.2	0.454	22.9
				0.046	1.0	
Exponential $f(x) = \text{Exp}[-x]$ $x \geq 0$	0.854	55.7	0.711	66.0	0.603	96.8
	0.146	3.3	0.279	10.1	0.357	17.5
			0.010	0.2	0.039	1.1
				0.001	0.01	

**Table B2: Three-Point Discretization Shortcuts**

EPT		MCS		ESM		MRO		GQN		ZDT	
$p_i$	$\alpha_i$										
.185	5.0	.250	0.0	.300	0.0	.248	1.5	.167	5.8	.333	9.0
.630	0.0	.500	0.0	.400	0.0	.504	0.0	.667	0.0	.333	0.0
.185	.0	.250	0.0	.300	0.0	.248	.5	.167	.2	.333	1.0

**Table B3: Three-Point Moment Matching Weights for P90, P50, P10 Values**

Uniform	Normal	Exp.	Triangular*
0.260	0.304	0.465	0.273

0.480	0.392	0.175	0.454
0.260	0.304	0.360	0.273

\* Note: Triangular discretization is a function of  $c$ , but this relationship is weak enough to ignore.

**Table B4: Three-Point Moment Matching Values for 25-50-25 Weights (same pdfs as Table 3)**

Uniform	Normal	Exp.	Triangular
P75.0	P92.1	P93.3	P78.9
P62.1	P50.0	P44.6	P50.0
P0.9	P7.9	P11.2	P21.1

**Table B5: Three-Point Moment Matching Values for 30-40-30 Weights (same pdfs as Table 3)**

Uniform	Normal	Exp.	Triangular
P74.9	P90.2	P91.8	P76.4
P64.0	P50.0	P42.3	P50.0
P6.5	P9.8	P9.5	P23.6

**Table B6: Representative Four-Point Discretizations for Lognormal**

$\sigma = .25$		$\sigma = .50$		$\sigma = .75$		$\sigma = 1.0$	
$\gamma_3 = 0.78$		$\gamma_3 = 1.75$		$\gamma_3 = 3.26$		$\gamma_3 = 6.18$	
$p_i$	$\alpha_i$	$p_i$	$\alpha_i$	$p_i$	$\alpha_i$	$p_i$	$\alpha_i$
.203	0.0	.151	0.0	.186	0.0	.436	0.0
.288	9.9	.265	9.1	.235	7.3	.274	4.1
.303	0.0	.378	0.0	.357	0.0	.017	0.0
.206	.0	.206	.0	.222	.0	.272	.0

**Table B7: 95% S-Equivalences for Discretization Shortcuts**

Raw	EPT					GQN				
Moment	$U(0,b)$	$N(0,\sigma)$	$T(0,b,b/2)$	$E(\lambda)$	$L(\mu,1)$	$U(0,b)$	$N(0,\sigma)$	$T(0,b,b/2)$	$E(\lambda)$	$L(\mu,1)$
First (mean)	$\infty$	$\infty$	$\infty$	>1MM*		$\infty$	$\infty$	$\infty$	>1MM	161,943
Second	4,830	>1MM	67,654	57,470	2,781	1,930	$\infty$	644,360	>1MM	4,583
Third	1,940	$\infty$	24,388	3,918	48,591	775	$\infty$	231,941	14,592	56,102
Raw	ESM					MCS				
Moment	$U(0,b)$	$N(0,\sigma)$	$T(0,b,b/2)$	$E(\lambda)$	$L(\mu,1)$	$U(0,b)$	$N(0,\sigma)$	$T(0,b,b/2)$	$E(\lambda)$	$L(\mu,1)$
First (mean)	$\infty$	$\infty$	$\infty$	>1MM	2,495	$\infty$	$\infty$	$\infty$	1,451	560
Second	2,128	36,165	9,745	1,674	941	30,732	240	14,068	407	676
Third	855	$\infty$	3,508	498	34,475	12,347	$\infty$	5,064	207	35,520
Raw	MRO					ZDT				
Moment	$U(0,b)$	$N(0,\sigma)$	$T(0,b,b/2)$	$E(\lambda)$	$L(\mu,1)$	$U(0,b)$	$N(0,\sigma)$	$T(0,b,b/2)$	$E(\lambda)$	$L(\mu,1)$
First (mean)	$\infty$	$\infty$	$\infty$	13,529	1,702	$\infty$	$\infty$	$\infty$	120,295	2,722
Second	77,873	1,840	115,500	1,264	988	1,046	904,694	5,986	1,716	895
Third	31,293	$\infty$	41,584	554	35,758	420	$\infty$	2,155	448	33,688

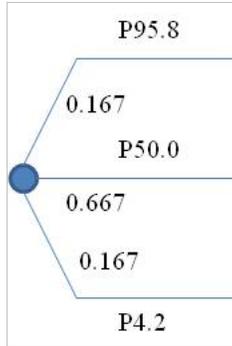
\*Note: >1MM = more than 1 million.

**Table B8: Approximation Error in Mean for Illustrative Example**

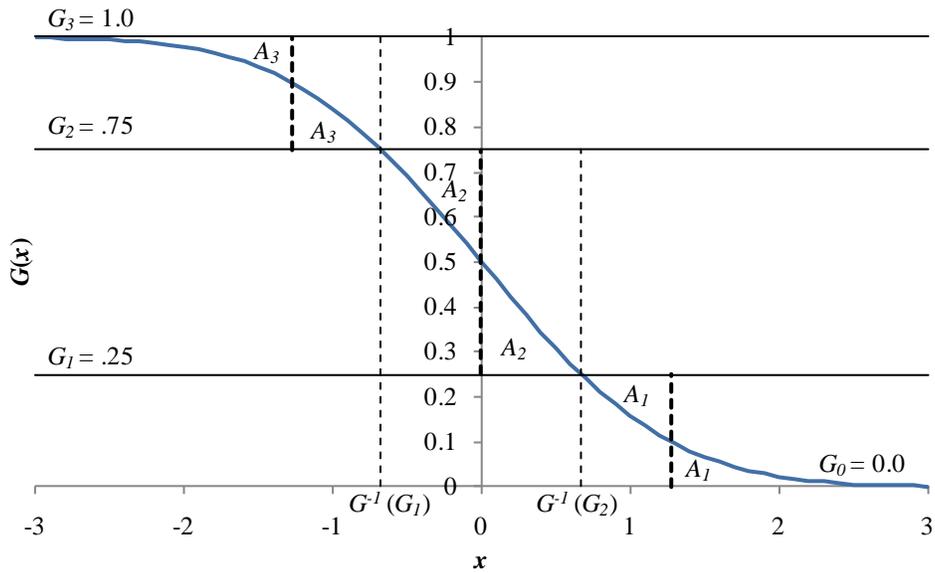
Approximation	Concave	Linear	Convex
4-pt Lognormal	-1%	0%	0%
EPT	0%	-1%	-4%
GQN	0%	-1%	-3%
ESM	-3%	-5%	-9%
MCS	-8%	-11%	-15%
MRO	-4%	-6%	-10%
ZDT	-2%	-5%	-9%
3-pt Bracket Median	-14%	-17%	-21%
4-pt Bracket Median	-10%	-12%	-16%
3-pt Bracket Mean	2%	0%	-4%
4-pt Bracket Mean	1%	0%	-2%

**Table B9: Approximation Error in Variance for Illustrative Example**

Approximation	Concave	Linear	Convex
4-pt Lognormal	30%	0%	-57%
EPT	-18%	-36%	-72%
GQN	-11%	-30%	-69%
ESM	-45%	-60%	-83%
MCS	-53%	-66%	-86%
MRO	-42%	-57%	-82%
ZDT	-47%	-62%	-84%
3-pt Bracket Median	-66%	-75%	-90%
4-pt Bracket Median	-48%	-60%	-83%
3-pt Bracket Mean	-38%	-54%	-80%
4-pt Bracket Mean	-21%	-37%	-71%



**Figure B1**– Gaussian quadrature for the normal distribution.



**Figure B2** – Illustration of Equal Areas discretization method.

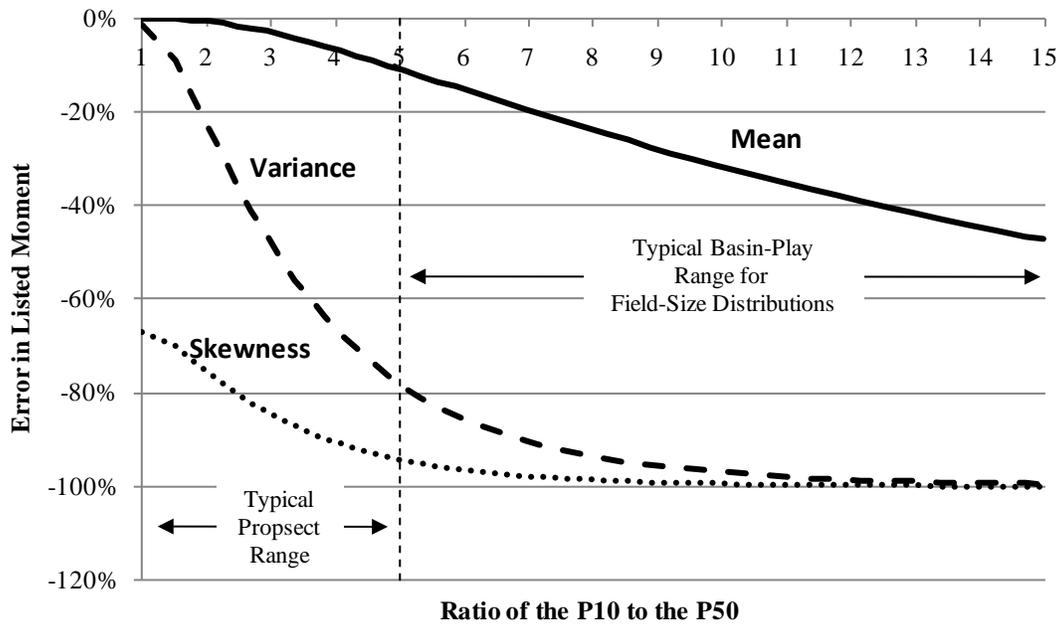


Figure B3: Error in Swanson's Mean with direct application to lognormal distribution. Compare to Megill (1984, Fig. B.1)

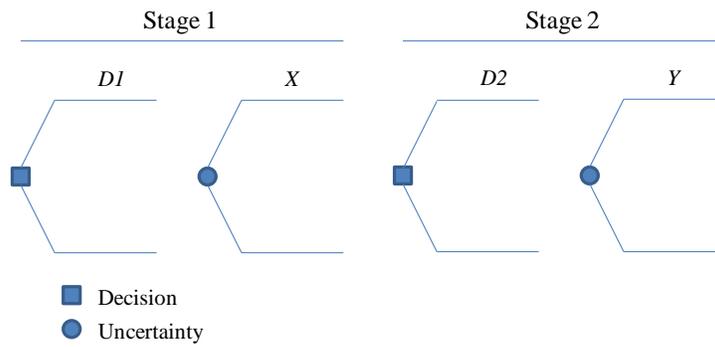


Figure B4: Example of decision situation that may not lend itself well to Monte Carlo methods.

## M APPENDIX C: DISCRETIZATION AND THE VALUE OF INFORMATION

Most decision and risk analyses include a mixture of continuous (e.g., oil and place, porosity, oil price) and discrete uncertainties (e.g., geologic success). How to best represent these uncertainties in decision models is a continuing question along multiple dimensions. From a communication perspective, continuous uncertainties complicate the discussion of particular scenarios that facilitate understanding. For example, it is unclear how to represent continuous uncertainties in a decision tree. From a computational perspective, continuous uncertainties may require many more calculations or model runs (they still must ultimately be discretized into many points to allow for computation). Yet, this complexity is believed to result in greater accuracy.

Bickel et al. (2011) recently summarized the most common discretization methodologies and demonstrated that some methods commonly used within the oil and gas industry are inaccurate. Bickel et al. focused on matching the moments (e.g., mean, variance, skewness) of input uncertainties, with the goal of more accurately estimating moments of the output distribution (e.g., expected net present value). In this paper, we extend Bickel et al. by investigating the accuracy of discretization techniques within the context of an information gathering example.

Our interest in this problem was motivated by Arild et al. (2008), who compared the use of one particular discretization (Extended Swanson-Megill) to Monte Carlo (MC) simulation within the context of a value-of-information (VOI) problem. They demonstrated that these two methods produced different results and implicitly assumed MC was more accurate. Prange et al. (2009) compared a numerical approximation method to a discretization method and also found that the discretization was inaccurate. Thus, Arild et al. and Prange et al. suggested the use of MC simulation instead of discretization. Willigers (2009), on the other hand, recommends the use of discretization instead of MC in an effort to reduce the number of computations and speed the analysis of asset portfolios.

This appendix is organized as follows. In the next section, we review three commonly used discretization methods. In the third section, we introduce an illustrative example that we use throughout the paper. In the fourth section, we derive exact and approximate VOI formulas for our illustrative example. In the fifth section, quantify and discuss the accuracy of the discretization methods. We next compare discretization to MC simulation. Finally, in the seventh section we conclude.

## **M.1 Discretization Methods**

Several discretization methods are in common use. These include general methods such as “Equal Areas” and shortcuts such as Extended Swanson-Megill. In this paper, we focus only upon commonly used shortcuts and summarize them below. The interested reader should see Bickel et al. (2011) for a more detailed discussion of these methods and others.

Pearson and Tukey (1965) developed a three-point discretization to closely approximate the mean of common probability density functions (pdfs). Their three-point approximation of the mean weights the 95<sup>th</sup> (P95), 50<sup>th</sup> (P50), and the 5<sup>th</sup> (P5) percentiles of the excess distribution function (edf) with probabilities of 0.185, 0.630, and 0.185, respectively. Pearson and Tukey did not recommend this approximation for higher moments and instead provided a more complex approximation for the standard deviation. However, Keefer and Bodily (1983), based on work by Keefer and Pollock (1980) and their own analysis, suggested treating the Pearson-Tukey approximation as a complete probability mass function (pmf) and referred to this as “Extended Pearson-Tukey” (EPT).

In the late 1960s, the Stanford Research Institute and later Strategic Decisions Group (SDG) popularized the practice of weighting the P90, P50, and P10 by 0.25, 0.50, and 0.25, respectively. This shortcut is described in McNamee and Celona (1990, pp. 32-33) and has come to be known as the “McNamee and Celona” shortcut or MCS.

Roy Swanson, an Exxon geologist, proposed, in a 1972 internal memo, approximating mean reserves, by weighting the P90, P50, P10 of the reserves edf by 0.30, 0.40, 0.30 (Megill 1984, Appendix B; Hurst et al. 2000). Following the discussion above, we will refer to this as a 30-40-30

weighting. Keefer and Bodily (1983) proposed treating Swanson’s 30-40-30 rule as a complete pmf and referred to it as “Extended Swanson-Megill” (ESM).

We summarize each of the discretization shortcuts in **Table C1**.

**Table C1: Three-Point Discretization Shortcuts**

EPT		MCS		ESM	
Probability	Fractile	Probability	Fractile	Probability	Fractile
0.185	P95	0.250	P90	0.300	P90
0.630	P50	0.500	P50	0.400	P50
0.185	P5	0.250	P10	0.300	P10

### M.1.1 Illustrative Example

Throughout this appendix, we test the accuracy of the differing discretization methods using a simple example.

Suppose a risk-neutral oil company is considering drilling a well in an undeveloped area. Based on core samples and well logs, the company estimates that the net present value (NPV) of the well is normally distributed with a mean  $\mu$  of \$10 million and a standard deviation  $\sigma$  of \$20 million. Evaluating the standard normal cumulative at a value of 0.5 ( $\$10/\$20$ ), the company finds that the probability of the well being a good investment ( $NPV > \$0$ ) is 69%. If the company does not drill, it will earn a sure  $v = \$0$ .

In an effort to improve its decision, the company is considering the acquisition of a seismic survey. Modeling the value of seismic surveys is complicated and time consuming (Pickering and Bickel 2006; Bickel et al. 2008). However, based on previous results in similar areas, the company's geologist believes the seismic results are correlated with the true value of the well with a correlation coefficient  $\rho$  of 0.6.

The structure of the example we consider here is known as a “two-action problem with normal priors” (TALL-N). As we will see below, the advantage of this simple example is that the VOI can be written in closed form and calculated exactly. This will allow us to compare the discrete approximations to a true value.

### M.1.2 Value of Information

*Exact Solution.* Define  $c = (\mu - v) / \sigma$ , which is a measure of the distance or divergence between the two alternatives in the TALL-N problem, measured in units of standard deviation. Bickel (2008) has shown that the expected value of information (EVOI) with a correlation coefficient of  $\rho$  in the TALL-N problem is

$$EVOI_{\rho} = \begin{cases} \rho\sigma \left[ \phi(\rho^{-1}c) - \rho^{-1}c\Phi(-\rho^{-1}c) \right] & \mu \geq v \\ \rho\sigma \left[ \phi(\rho^{-1}c) + \rho^{-1}c\Phi(\rho^{-1}c) \right] & \mu < v, \end{cases} \quad (18)$$

where  $\phi$  and  $\Phi$  are the standard normal pdf and cumulative distribution function (cdf), respectively.

Applying this formula to our example, with  $c = (\$10 - \$0) / \$20 = 0.5$ , we find that the value of perfect information ( $EVOI_1$ ) is \$3.96 million and the value of imperfect information with a correlation coefficient of 0.6 is \$1.36 million. It might be surprising that the value of test with a correlation coefficient of 0.6 is only worth about 34% of perfect information, rather than 60%. This result is typical. The value of an imperfect test is generally much less than  $\rho$  times the value of perfect information. In fact, as can be seen here,  $\rho^2$  times  $EVOI_1$  is a better estimate. See Bickel (2008) for a full discussion of these issues.

The EVOI of obtained via Equation (1) is the mean or the expectation of a distribution that describes the uncertainty in the VOI. The variance of the VOI with correlation coefficient  $\rho$  when  $\mu \geq v$  is (see section L4 for the derivation)

$$VVOI_{\rho} = (v^2 - 2v\mu + \mu^2 - \rho^2\sigma^2)\Phi(-\rho^{-1}c) - \rho\sigma(\mu - v)\phi(-\rho^{-1}c) + \rho^2\sigma^2 - \left( \rho\sigma \left[ \phi(\rho^{-1}c) - \rho^{-1}c\Phi(-\rho^{-1}c) \right] \right)^2. \quad (19)$$

This formula will be useful later in the paper. The case where  $\mu < v$  can be derived similarly and is omitted in the interest of space.

*Discrete Approximation.* We now derive a formula to calculate the EVOI under each of the approximations described above. We then compare these values to the exact value provided by Equation (1). We use the following additional notation:

- $\Theta$ . A random variable that represents the signal provided by our test (e.g., the seismic amplitude).
- $\theta$ . A particular realization of  $\Theta$ .
- $X$ . Random variable for the reservoir property of interest (e.g., porosity).
- $x$ . A particular realization of  $X$ .
- $\mu$ . Mean of  $X$ .
- $\sigma$ . Standard deviation of  $X$ .
- $\rho$ . Correlation between  $X$  and  $\Theta$ .
- $\alpha_i$ . A fractile of a random variable.
- $\theta_i^A \equiv \theta(\alpha_i^A)$ . The  $\alpha_i$  fractile of the random variable  $\Theta$ , under approximation  $A$ .
- $p_i^A$ . The weight placed on fractile  $\alpha_i$  under approximation  $A$ .
- $N$ . The number of points in our discrete approximation.

Given that  $X$  and  $\Theta$  are jointly normally distributed, the exact conditional mean of  $X$  given signal  $\theta_i^A$  is

$$\mu_{\theta_i^A} \equiv \mu + \rho\sigma\Phi^{-1}(\theta_i^A). \quad (20)$$

Since the approximations we consider are symmetric about the P50, the approximate mean, conditional on observing  $\theta_i^A$  under approximation  $A$ , is

$$\bar{x}_i^A = \mu_{\theta_i^A} \equiv \mu + \rho\sigma\Phi^{-1}(\theta_i^A). \quad (21)$$

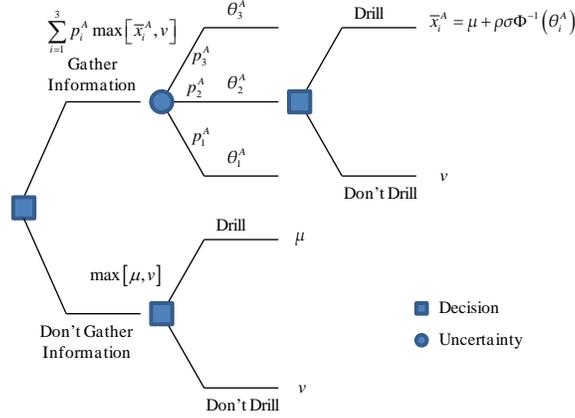
The approximate EVOI under approximation  $A$  is the expected value with information less the expected value without information (Bratvold et al. 2009):<sup>18</sup>

$$EVOI_{\rho}^A = \sum_{i=1}^N p_i^A \max[\bar{x}_i^A, v] - \max[\mu, v]. \quad (22)$$

The derivation of this formula is given graphically in **Figure C1**.

---

<sup>18</sup> This definition of VOI is correct only if the decision maker is risk neutral or has constant risk attitude.



**Figure C1:** Decision tree used to value information under three-point discrete approximation.

In this paper, we consider only three-point approximations ( $N = 3$ ). We define  $\theta_1^A$  as the smallest fractile,  $\theta_2^A$  as the P50, and  $\theta_3^A = 1 - \theta_1^A$ . Equation (5) then becomes

$$EVOI_{\rho}^A = \begin{cases} p_1^A \sigma (-\rho \Phi^{-1}(\theta_1^A) - c) & \frac{c}{\rho} < -\Phi^{-1}(\theta_1^A) \\ 0 & \frac{c}{\rho} > -\Phi^{-1}(\theta_1^A). \end{cases} \quad (23)$$

The ratio of the approximate to the exact VOI is then

$$\frac{EVOI_{\rho}^A}{EVOI_{\rho}} = \begin{cases} \frac{p_1^A (-\Phi^{-1}(\theta_1^A) - \rho^{-1}c)}{\varphi(\rho^{-1}c) - \rho^{-1}c \Phi(-\rho^{-1}c)} & \frac{c}{\rho} < -\Phi^{-1}(\theta_1^A) \\ 0 & \frac{c}{\rho} > -\Phi^{-1}(\theta_1^A). \end{cases} \quad (24)$$

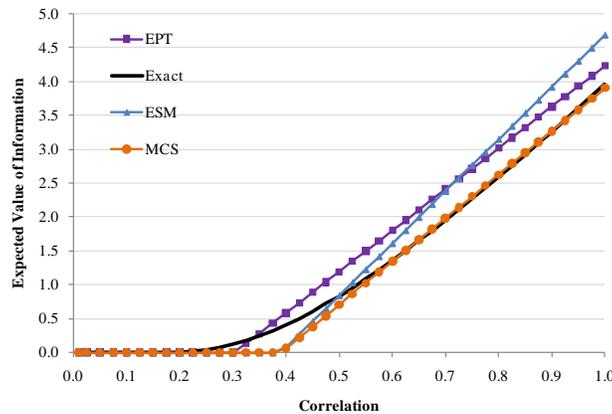
Thus, the ratio of the approximate EVOI to the exact EVOI is independent of the scale of the decision problem, as specified by  $\mu$  or  $\sigma$ . That is, this ratio depends only upon the accuracy of the test  $\rho$  and the divergence between the two alternatives  $c$ . As we will see, this result is quite helpful since it will allow us to quantify the accuracy of various discretization methods across a wide range of model parameters.

## M.2 Accuracy of Discretization Methods

We now use Equations (6) and (7) to compare the accuracy of the discrete approximations to the exact EVOI given in Equation (1). We begin by applying them to our illustrative example. Recall, that in this case  $\mu = \$10$  million and  $\sigma = \$20$  million and, thus,  $c = 0.5$ .

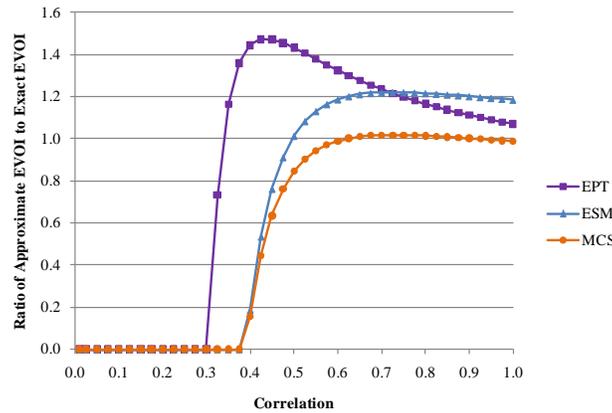
When  $\rho$  equals 0.6 the EVOI for the EPT, ESM, and MCS approximations are \$1.80, \$1.61, and \$1.34 million, respectively. Clearly, the approximations can produce very different VOI estimates. For example, the EPT and MCS approximations differ by over 34%. Whether or not these differences will lead to differing recommendations regarding information gathering (i.e., the differences are material) will depend upon the cost of the test. For example, if the cost of the test was \$1.5 million, both EPT and ESM would recommend gathering the information, while MCS would not. As discussed above, the exact EVOI (based on Equation (1)) is \$1.36 million. Thus, the ratio of the approximate to the exact EVOI for the EPT, ESM, and MCS approximations are 1.33, 1.19, and 0.99, respectively.

The EVOI as a function of the correlation coefficient is plotted in **Figure C2**. The solid black line is the exact EVOI. While the EVOI is small for low correlations, it is exactly zero only when the correlation is zero. When  $\rho = 1$ , perfect information, the EVOI equals \$3.96 million, as we found above. The remaining curves are the approximate EVOI for the EPT (squares), ESM (triangles), and MCS (circles) approximations. In this case, MCS produces EVOI's very close to the true EVOI for correlations above about 0.55. EPT and ESM tend to overestimate EVOI to varying degrees. On the low end, the approximations have differing thresholds below which EVOI is estimated to be zero. This threshold correlation  $\rho^*$  is given by Equation (7) and equals  $\rho^* = -c / \Phi^{-1}(\theta_1^A)$ . In the case of ESM and MCS,  $\rho^* = -0.5 / \Phi^{-1}(0.1) \approx 0.39$ . The EPT threshold is  $\rho^* = -0.5 / \Phi^{-1}(0.05) \approx 0.30$ .



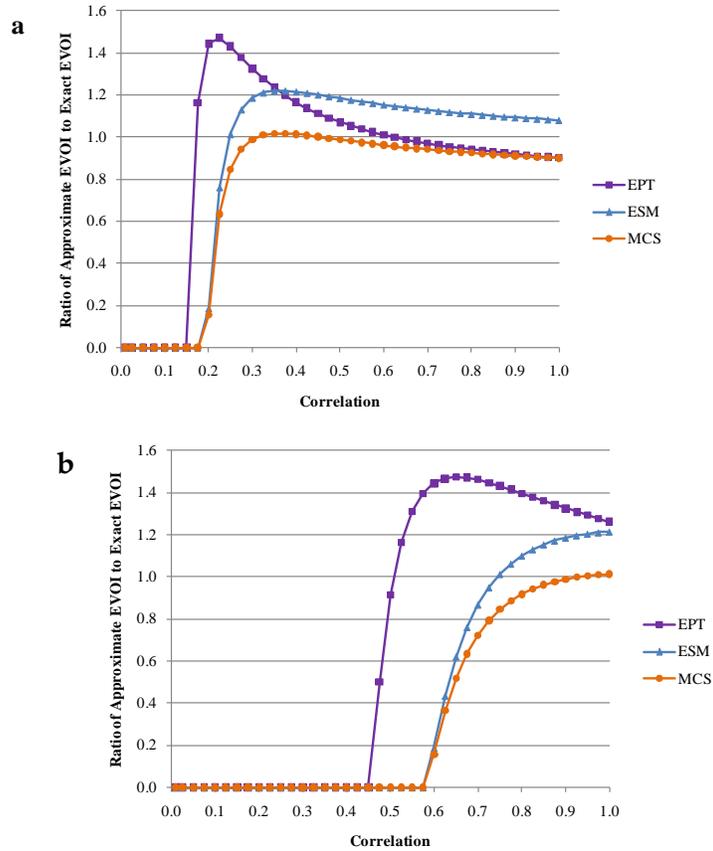
**Figure C2:** Comparison of approximate EVOIs to exact EVOI ( $\mu = \$10$  million and  $\sigma = \$20$  million).

**Figure C3** plots the ratio of the approximate EVOI to the exact EVOI (Equation (7)) for  $c = 0.5$ . As Equation (7) makes clear, these results hold for any jointly normal setting for which the  $c = 0.5$ . The best approximation depends upon the correlation. We conjecture that most tests considered in actual situations have correlations above 0.6. Tests with accuracies less than this are likely to be screened out before making it to market. Thus, MCS performs very well, in this case, over most of the range one might expect to find in practice.



**Figure C3:** Ratio of approximate to exact EVOI ( $c = 0.5$ ).

**Figure C4** tests the sensitivity of these results by considering two different values for  $c$ : 0.25 (panel **a**) and 0.75 (panel **b**). As  $c$  increases, EVOI decreases, and the region in which the approximations estimate a positive EVOI also decreases. Overall, the discretization errors are on the order of +/- 20%. They are, of course, larger than this in some cases and smaller in others. Which approximation is best depends upon the setting, but MCS appears to perform well over a wide range.



**Figure C4:** Sensitivity of the ratio of the approximate to exact EVOI to  $c$ .  $c = 0.25$  (panel **a**) and  $c = 0.75$  (panel **b**).

### M.2.1 Comparison of Discretization to Simulation

Often the “competition” is not among differing discretization methods, but rather between discretization and Monte Carlo (MC) simulation. This was the comparison made by Arild et al. (2008), who implicitly assumed that MC was more accurate. We address this issue next.

One must keep in mind that MC is also an approximation. Discretization methods induce *approximation error*, while MC methods include *sampling error*. The relevant question is whether or not MC is more accurate than discretization. To address this, we determine how many MC samples are required to achieve the same accuracy as each discretization method and refer to this as  $S$ -equivalence, where  $S$  is the equivalent number of MC samples.

### M.2.2 Discretization Accuracy

Let  $k$  be the difference between the approximate EVOI obtained via a discretization and the true EVOI. It will be useful to normalize this difference by the standard deviation of the VOI, which may be obtained via Equation (2). We write the accuracy of the approximation as

$$\delta = \frac{k}{\sqrt{VVOI_\rho}} = \frac{EVOI_\rho^A - EVOI_\rho}{\sqrt{VVOI_\rho}}. \quad (25)$$

*Simulation Procedure.* Now, suppose that instead of using a discretization, we estimate EVOI via MC simulation. We accomplish this by simulating the signal  $\Theta$ , which is assumed to normal; we assume it is standard normal without loss of generality. We then follow this procedure (Bickel et al. 2008):

1. Sample a signal  $\theta$  from  $\Theta \sim N(0,1)$ .
2. Compute the posterior mean  $\mu_\theta$  via Equation (3).
3. Choose the maximum of  $\mu_\theta$  and  $v$ . Return to Step 1.

We repeat this procedure  $S$  times. The average of the  $S$  trials is an estimate of the true EVOI, which we denote  $EVOI_\rho^S$ . Each time we perform a MC trial we will estimate a different  $EVOI_\rho^S$ . In other words, our estimate of the EVOI via MC simulation is a random variable.

*Computing S-Equivalence.* According to the Central Limit Theorem (CLT),  $EVOI_\rho^S$  is normally distributed with mean  $EVOI_\rho$  and variance  $VVOI_\rho / S$  for large  $S$ . Thus, the probability that the simulation will more accurately estimate EVOI than the discretization method is (see Appendix for derivation)

$$P(|EVOI_\rho^S - EVOI_\rho| \leq k) = 2\Phi(\delta\sqrt{S}) - 1, \quad (26)$$

where  $\Phi$  is the standard normal cdf. If we want this probability to be  $\tau$  then we must take

$$S = \delta^{-2} \Phi^{-1}\left(\frac{\tau+1}{2}\right)^2 \quad (27)$$

samples (see section L4 for derivation). For example, if we want the simulation to have a 95% chance of estimating the EVOI more accurately than discretization we must take  $S = \delta^{-2} \Phi^{-1}(1.95/2)^2 \approx (1.96/\delta)^2 \approx 4/\delta^2$  samples. Thus, Equation (10) establishes an equivalence between discretization and simulation (Pfeifer et al. 1991).

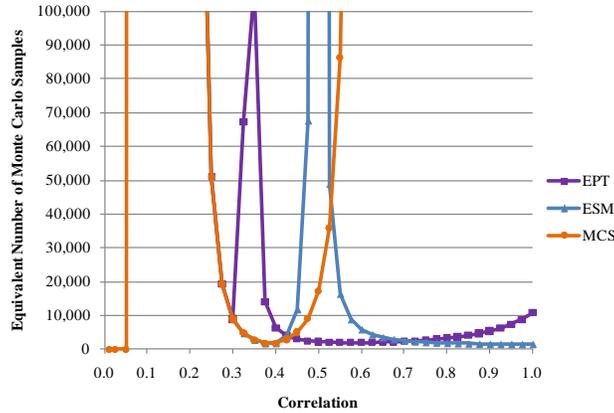
*Illustrative Example.* Returning to our illustrative example, we compute  $\delta$  and the equivalent number of MC samples, assuming that  $c = 0.5$  and  $\rho = 0.6$ . These results are presented in Table C2. In this case, EPT, ESM, and MCS are equivalent to 1,956, 5,923, and 1,696,499 samples, respectively. Thus, we see that even simple three-point approximations can be equivalent to thousands and possibly millions of MC samples. This makes it clear that one should not assume that MC is necessarily more accurate than three-point approximation methods. Even the poorly-performing EPT is equivalent to almost two thousand MC samples. Whether or not one would be better off using an approximation or MC depends upon how long it takes to compute the model under each MC sample.

**Table C2: 95% S-Equivalences for the Illustrative VOI Example ( $c = 0.5$  and  $\rho = 0.6$ )**

Approximation	$\delta$	Equivalent Number of MC Samples, $S$
EPT	0.044	1,956
ESM	0.025	5,923
MCS	-0.002	1,696,499

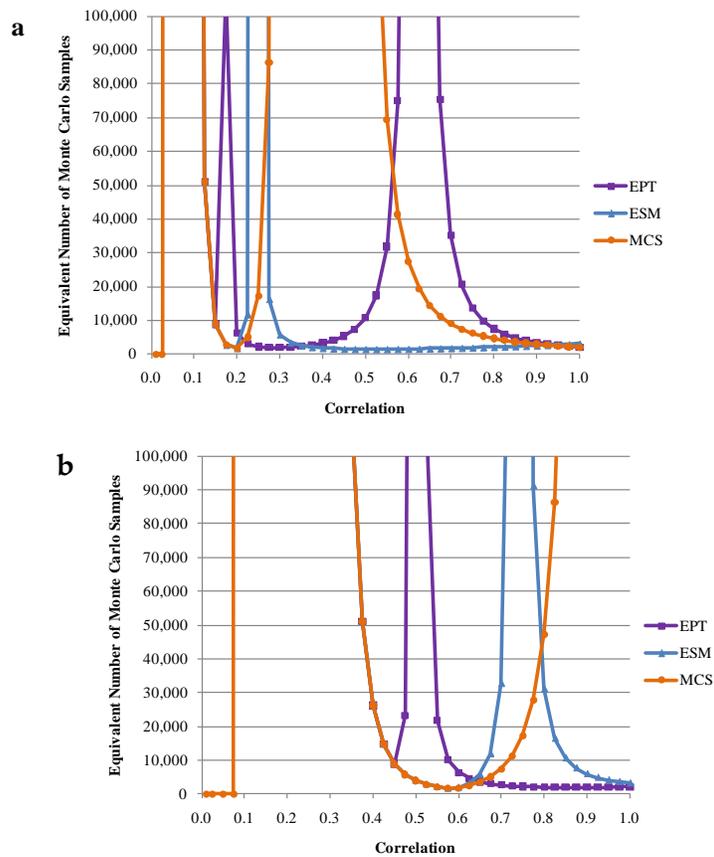
**Figure C5** plots the equivalent number of samples as a function of the correlation coefficient for  $c = 0.5$ . Values above 100,000 are omitted to save space. The equivalent number of samples is not a simple function of the correlation. This is driven by the fact that neither the difference in EVOIs in Equation (8) nor the standard deviation of the VOI (Equation (2)) are simple functions of  $\rho$ . For correlations above about 0.55 MCS is equivalent to over 100,000 MC samples. Large  $S$ -equivalences can stem from the fact that the approximate EVOI is close to the true EVOI or from a low value for  $VVOI_\rho$ . For example, when  $\rho = 0$  the standard deviation of the VOI is zero. EPT and ESM are equivalent to far fewer samples than MCS for correlations above 0.55. For example, for correlations above 0.7, ESM is equivalent to between 1,500 and 2,500 MC

samples. This result is surprising and important, given the oil and gas industry’s heavy use of this particular approximation.



**Figure C5:** Equivalent number of Monte Carlo samples ( $c = 0.5, \tau = 0.95$ ). Values above 100,000 are not shown.

**Figure C6** tests the sensitivity of these results by considering two different values for  $c$ : 0.25 (panel **a**) and 0.75 (panel **b**). If  $c$  is low (i.e., the value alternatives are not very different) then ESM is equivalent to about 2,000 MC samples over a very wide range. EPT and MCS are equivalent to tens of thousands of samples for correlations between about 0.5 and 0.8. Above about 0.8, EPT and MCS, like ESM, are equivalent to considerably fewer MC samples. When  $c$  equals 0.75, EPT is equivalent to around 2,000 MC samples and MCS and ESM perform better.



**Figure C6:** Sensitivity of the equivalent number of Monte Carlo samples ( $\tau = 0.95$ ) to  $c$ .  $c = 0.25$  (panel **a**) and  $c = 0.75$  (panel **b**). Values above 100,000 are not shown.

### M.3 Discussion and Conclusion

Our results make it clear that it is difficult to draw general conclusions regarding the performance of discrete approximations in VOI settings. Their performance is a strong function of the accuracy of the test and the divergence between the two alternatives. We can, however, draw a few conclusions:

1. Simple, three-point, discrete approximations are likely equivalent to thousands and possibly tens of thousands of MC samples. Therefore, one should not assume that MC simulation is necessarily more accurate than discretization. In fact, given the significant computational time required by many oil and gas

models (e.g., reservoir simulators) the burden of proof is on those using MC methods to show that they are more accurate than simple discretization methods. Unfortunately, Arild et al. (2008) did not specify the number of MC samples used in their analysis.

2. MCS performs well, and in many regions of interest (e.g. correlations above 0.5), better than the EPT and ESM discretizations. This has important practical implications as the industry relies heavily on ESM. See Bickel et al. (2011) for a further discussion of these issues.

The discretizations discussed in this paper are not a function of the underlying decision problem. For example, the fractiles used are not a function of the project payoffs. As such, the approximations will always induce some error. It is possible to tailor the discretization to the underlying decision and this will result in greater accuracy. Arild et al. (2008) provide an example of this and Merkhofer (1975;1977) discusses it in detail.

Other approaches to discretization include graphical methods such as Equal Areas and techniques that seek to match the moments of the input uncertainties. How well these methods perform within the VOI setting we present here, or more generally, is an open question. We hope that this research provides the foundation needed to explore this important area of

## M.4 Derivations

### M.4.1 Variance of VOI

Suppose we can conduct a test  $\Theta$  that yields a signal  $\theta$  regarding the value  $X$ . Assume further that  $\Theta$  and  $X$  are jointly normally distributed. The marginal pdf of  $\Theta$  (i.e., the preposterior in a Bayesian setting) is

$$f(\theta) = \frac{1}{\sqrt{2\pi}\sigma_{\Theta}} e^{-\frac{1}{2}\left(\frac{\theta - \mu_{\Theta}}{\sigma_{\Theta}}\right)^2}, \quad (28)$$

where  $\mu_{\Theta}$  and  $\sigma_{\Theta}$  are the prior mean and standard deviation of  $\Theta$ , respectively. The mean of  $X$  given a signal  $\theta$  is

$$\mu_{X|\theta} = \mu_X + \rho\sigma_X \frac{\theta - \mu_\Theta}{\sigma_\Theta}, \quad (29)$$

where  $\mu_X$  and  $\sigma_X$  are the prior mean and standard deviation of  $X$ , respectively. Notice that the conditional mean of  $X$  depends upon  $\theta$ 's normalized difference from its mean. Thus, we may assume that  $\Theta$ 's pdf is the standard normal ( $\mu_\Theta = 0$  and  $\sigma_\Theta = 1$ ). In which case, we have

$$\mu_\theta = \mu + \rho\sigma\theta, \quad (30)$$

where we have dropped the  $X$  subscripts since there is no possibility of ambiguity. Assume  $\rho > 0$ . If  $\rho = 0$  then  $\mu_\theta = \mu$  and  $\text{VOI} = 0$ . If  $\rho < 0$  the test just needs to be relabeled. The conditional mean will be greater than  $v$  as long as  $\mu + \rho\sigma\theta > v$  or  $\theta > (\rho\sigma)^{-1}(v - \mu)$ . Let  $a = (\rho\sigma)^{-1}(v - \mu) = -\rho^{-1}c$  and  $b = \infty$ . The second raw moment of the VOI is

$$\begin{aligned} E[\text{VOI}_\rho^2] &= \int_{-\infty}^a (v - \mu)^2 f(\theta) d\theta + \int_a^\infty (\mu_\theta - \mu)^2 f(\theta) d\theta = \int_{-\infty}^a (v - \mu)^2 f(\theta) d\theta + \int_a^\infty (\mu + \rho\sigma\theta - \mu)^2 f(\theta) d\theta \\ &= \int_{-\infty}^a (v - \mu)^2 f(\theta) d\theta + \int_a^\infty \rho^2 \sigma^2 \theta^2 f(\theta) d\theta \\ &= \int_{-\infty}^a (v^2 - 2v\mu + \mu^2) f(\theta) d\theta + \int_a^\infty \rho^2 \sigma^2 \theta^2 f(\theta) d\theta \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a (v^2 - 2v\mu + \mu^2) e^{-\frac{1}{2}\theta^2} d\theta + \frac{1}{\sqrt{2\pi}} \int_a^\infty \rho^2 \sigma^2 \theta^2 e^{-\frac{1}{2}\theta^2} d\theta \\ &= (v^2 - 2v\mu + \mu^2) \Phi(a) + \rho^2 \sigma^2 \frac{1}{\sqrt{2\pi}} \int_a^\infty \theta^2 e^{-\frac{1}{2}\theta^2} d\theta \\ &= (v^2 - 2v\mu + \mu^2) \Phi(a) + \rho^2 \sigma^2 \frac{1}{\sqrt{2\pi}} \left( a e^{-\frac{1}{2}a^2} + \sqrt{\frac{\pi}{2}} \text{Erfc}\left(\frac{a}{\sqrt{2}}\right) \right) \\ &= (v^2 - 2v\mu + \mu^2) \Phi(a) + \rho^2 \sigma^2 (a\phi(a) + 1 - \Phi(a)) \\ &= (v^2 - 2v\mu + \mu^2 - \rho^2 \sigma^2) \Phi(a) + \rho^2 \sigma^2 a\phi(a) + \rho^2 \sigma^2. \end{aligned}$$

Since  $a = -\rho^{-1}c$  we have

$$E[\text{VOI}_\rho^2] = (v^2 - 2v\mu + \mu^2 - \rho^2 \sigma^2) \Phi(-\rho^{-1}c) - \rho\sigma(\mu - v)\phi(-\rho^{-1}c) + \rho^2 \sigma^2. \quad (31)$$

The variance of the VOI is then

$$\begin{aligned} V\text{VOI}_\rho &= E[\text{VOI}_\rho^2] - (E\text{VOI}_\rho)^2 \\ &= (v^2 - 2v\mu + \mu^2 - \rho^2 \sigma^2) \Phi(-\rho^{-1}c) - \rho\sigma(\mu - v)\phi(-\rho^{-1}c) + \rho^2 \sigma^2 - \left( \rho\sigma \left[ \phi(\rho^{-1}c) - \rho^{-1}c \Phi(-\rho^{-1}c) \right] \right)^2. \end{aligned} \quad (32)$$

#### M.4.2 Probability the MC Simulation will be more Accurate than Discrete Approximation

$$\begin{aligned}
 P(|EVOI_{\rho}^s - EVOI_{\rho}| \leq k) &= P(EVOI_{\rho}^s - EVOI_{\rho} \leq +k) - P(EVOI_{\rho}^s - EVOI_{\rho} \leq -k) \\
 &= P\left(\frac{EVOI_{\rho}^s - EVOI_{\rho}}{\sqrt{VVOI_{\rho}/S}} < \frac{+k}{\sqrt{VVOI_{\rho}/S}}\right) - P\left(\frac{EVOI_{\rho}^s - EVOI_{\rho}}{\sqrt{VVOI_{\rho}/S}} < \frac{-k}{\sqrt{VVOI_{\rho}/S}}\right) \\
 &= P\left(\frac{EVOI_{\rho}^s - EVOI_{\rho}}{\sqrt{VVOI_{\rho}/S}} < \delta\sqrt{S}\right) - P\left(\frac{EVOI_{\rho}^s - EVOI_{\rho}}{\sqrt{VVOI_{\rho}/S}} < -\delta\sqrt{S}\right) \\
 &= \Phi(\delta\sqrt{S}) - \Phi(-\delta\sqrt{S}) \\
 &= \Phi(\delta\sqrt{S}) - [1 - \Phi(\delta\sqrt{S})] \\
 &= 2\Phi(\delta\sqrt{S}) - 1.
 \end{aligned}$$

#### M.4.3 S-Equivalence for CLT

$$2\Phi(\delta\sqrt{S}) - 1 = \tau$$

$$\delta\sqrt{S} = \Phi^{-1}\left(\frac{\tau+1}{2}\right)$$

$$S = \delta^{-2}\Phi^{-1}\left(\frac{\tau+1}{2}\right)^2.$$

#### M.5 References

- Arild, Ø., Lohne, H.P., and Bratvold, R. 2008. A Monte Carlo Approach to Value of Information. Paper SPE IPTC 11969 presented at the International Petroleum Technology Conference, Kuala Lumpur, Malaysia, 3-5 December.
- Bickel, J.E. 2008. The Relationship between Perfect and Imperfect Information in a Two-Action Risk-Sensitive Problem. *Decision Analysis* 5 (3) 116-128.
- Bickel, J.E., Gibson, R.L., McVay, D.A., Pickering, S., and Waggoner, J. 2008. Quantifying the Reliability and Value 3D Land Seismic. *Reservoir Evaluation and Engineering* 11 (5) 832-841.
- Bickel, J.E., Lake, L.W., and Lehman, J. 2011. Discretization, Simulation, and Swanson's (Inaccurate) Mean. Forthcoming in *SPE Economics and Management*.
- Bratvold, R.B., Bickel, J.E., and Lohne, H.P. 2009. Value of Information in the Oil and Gas Industry: Past, Present, and Future. *SPE Reservoir Evaluation & Engineering* 12 (4) 630-638.
- Hurst, A., Brown, G.C., and Swanson, R.I. 2000. Swanson's 30-40-30 Rule. *AAPG Bulletin* 84 (12) 1883-1891.
- Keefer, D.L. and Bodily, S.E. 1983. Three-Point Approximations for Continuous Random Variables. *Management Science* 29 (5) 595-609.
- Keefer, D.L. and Pollock, S.M. 1980. Approximations and Sensitivity in Multiobjective Resource Allocation. *Operations Research* 28 (1) 114-128.
- McNamee, P. and Celona, J. 1990. *Decision Analysis with Supertree*, Second Edition. South San Francisco, CA: The Scientific Press.

- Megill, R.E. 1984. *An Introduction to Risk Analysis*, Second edition. Tulsa, Oklahoma: PennWell.
- Merkhofer, M.W., *Flexibility and Decision Analysis*, PhD Dissertation, Department of Engineering-Economic Systems. 1975, Stanford University: Stanford, CA.
- Merkhofer, M.W. 1977. The value of information given decision flexibility. *Management Science* **23** (7) 716-727.
- Pearson, E.S. and Tukey, J.W. 1965. Approximate Means and Standard Deviations Based on Distances between Percentage Points of Frequency Curves. *Biometrika* **52** (3/4) 553-546.
- Pfeifer, P.E., Bodily, S.E., and Jr., F., Sherwood C. 1991. Pearson-Tukey Three-Point Approximations Versus Monte-Carlo Simulation. *Decision Sciences* **22** 74-90.
- Pickering, S. and Bickel, J.E. 2006. The Value of Seismic Information. *Oil & Gas Financial Journal* **3** (5) 26-33.
- Prange, M., Bailey, W., and Kuchuk, F. 2009. Quantifying the Value of a Future Well Test with Measurement Uncertainty. Paper SPE IPTC 13595-PP presented at the International Petroleum Technology Conference, Doha, Qatar, 7-9 December.
- Willigers, B.J.A. 2009. Practical Portfolio Simulation: Determining the Precision of a Probability Distribution of a Large Asset Portfolio when the Underlying Project Economics are Captured by a Small Number of Discrete Points. Paper SPE 124180 presented at the Annual Technical Conference and Exhibition, New Orleans, Louisiana, 4-7 October.