Characterizing Shale Gas and Tight Oil Drilling and Production Performance Variability

by

Justin B. Montgomery

B.S. Mechanical Engineering, Texas A&M University (2013)

Submitted to the Engineering Systems Division
in partial fulfillment of the requirements for the degree of

Master of Science in Technology and Policy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

Author .... Signature redacted ....
Engineering Systems Division
May 14, 2015

Certified by ... Signature redacted ...
Francis M. O'Sullivan
Director of Research, MIT Energy Initiative
Thesis Supervisor

Accepted by .... Signature redacted ....
Dava J. Newman
Professor of Aeronautics and Astronautics and Engineering Systems
Director, Technology and Policy Program

Characterizing Shale Gas and Tight Oil Drilling and Production Performance Variability

by

Justin B. Montgomery

Submitted to the Engineering Systems Division on May 14, 2015,
in partial fulfillment of the requirements for the degree of
Master of Science in Technology and Policy

Abstract

Shale gas and tight oil are energy resources of growing importance to the U.S. and the world. The combination of horizontal drilling and hydraulic fracturing has enabled economically feasible production from these resources, leading to a surge in domestic oil and gas production. This is providing an economic boon and reducing reliance on foreign sources of energy in the U.S., but there are still a number of environmental, economic, and technical challenges that must be overcome to unlock the resource's full potential.

One key challenge is understanding variability in individual well performance—in terms of both drilling time (a key driver of well cost) and well productivity—which has led to greater than anticipated economic risk associated with shale gas and tight oil development. Thus far, more reliable forecasting has remained elusive due to its prohibitive cost and the poorly understood nature of the resource. There is an opportunity to make use of available drilling and production data to improve the characterization of variability. For my analysis, I use publicly-available well production data and drilling reports from a development campaign.

In order to characterize variability, I use a combination of graphical, statistical, and data analytics methods. For well productivity, I use probability plots to demonstrate a universality to the distribution shape, which can accurately be described as lognormal. Building on this distributional assumption, I demonstrate the utility of Bayesian statistical inference for improving estimates of the distribution parameters, which will allow companies to better anticipate resource variability and make better decisions under this uncertainty. For drilling, I characterize variability in operations by using approximate string matching to compare drilling activity sequences, leading to a metric for operational variability. Activity sequences become more similar over time, consistent with the notion of standardization. Finally, I investigate variability of drilling times as they progress along the learning curve, using probability plots again. I find some indication of lognormality, with implications for how learning in drilling should be measured and predicted.

This thesis emphasizes the relevance of data analytics to characterizing performance variability across the spectrum in shale gas and tight oil. The findings also demonstrate the value of such an approach for identifying patterns of behavior, estimating future variability, and guiding development strategies.

Thesis Supervisor: Francis M. O'Sullivan
Title: Director of Research, MIT Energy Initiative

Acknowledgments

My advisor, Frank O'Sullivan, told me when I first arrived in Cambridge that I was not going to have a "typical" research experience during my master's if I worked for him. This has proved more true than I could have imagined! Working with Frank has been a pleasure and I am grateful to him for always supporting my crazy ideas, challenging me, and making it easy to be excited about the work that I do every day. It is important to mention here that some parts of this thesis, especially Section 3.2.2, reflect not only my own analysis and interpretation but Frank's as well. I am grateful for the confidence he places in me as a researcher to collaborate closely with him on this work, and to build my own research efforts on shared insights. I have benefited tremendously from this relationship both as a researcher and as a person and I am excited to continue working together in the future!

I want to thank Accenture for financially supporting my analysis of drilling performance variability. Being a part of the MIT-Accenture Business Analytics Consortium has been a unique and rewarding research experience. Additionally, the interactions I have had throughout this research with Melissa Stark, Brian Richards, John Tolle, Patrick Johnston, and Bob Palermo have been wonderful and I am grateful to them for their time, suggestions, and encouragement. While at MIT, I have been helped along the way by too many people to list here, but I especially want to thank Bob Kleinberg, Jessika Trancik, Gordon Kaufman, Youssef Marzouk, and Maite Pena-Alcaraz for their interest and help with my research. To all the people of the MIT Energy Initiative, thanks for making it a great environment to work in!

I also want to thank the Society of Petroleum Engineers for permission to base parts of my thesis (specifically Sections 4.1 and 4.2) on material from a paper titled *Measuring Drilling Standardization Using Approximate String Matching* that I presented at the SPE ATCE 2014 conference in Amsterdam[1].

I want to express my gratitude to some mentors, prior to my time at MIT, that have had a profound impact on who I am professionally and personally: Mike Mayekar, Calvin Yao, Gregg Isenmann, Richard Malak, Ed Harris, Scott Austin (your memory and influence lives

on!), Kyle O'Leary, Charlie Hensel, Shane Aldrich, Craig Brown, and Michael Helms. This is not a list I make carelessly—I would not be where I am today if it was not for the time and attention invested in me by each of these individuals.

I have been surrounded by some truly incredible people in the last two years at MIT. I have learned so much from my fellow TPP-ers[2] and have thoroughly enjoyed being a part of this unique social experiment that places eccentric, social science-loving engineering students together. I can't imagine a more interesting and fun group of people to do a master's degree with!

Speaking of social experiments, my home over the last two years has been a strange living arrangement called 'Murica. The transition from Texas to Massachusetts is a hard one in many ways and I would not have gotten through it without my roommates: Alex, Jacob, Chris, and John. They have been fantastic people to grow close with over this time and I know regardless of where we all end up, there will be a lifelong bond between us from this shared experience. Some of my fondest memories from these two years will be of sitting around our kitchen together in lawn chairs, drinking PBR, and watching the snow pile up outside. This is the graduate experience! But seriously, my mind has been stretched and my ideas challenged by all of our crazy conversations, and my understanding and appreciation of life is greater having known you guys.

I am incredibly fortunate that Cecilie came into my life a year ago. She has made me a much happier person here and she continues to amaze me with how supportive and understanding she can be of me. Thank you so much for being by my side throughout this thesis and the last year.

Finally, I want to thank my family for their love and support. You all have taught me to enjoy hard work, be passionate about everything I do, and above all, to always have the courage to do things differently than everyone around me. This thesis is dedicated to you!

Now, enough with the sentimentalism. On with the show!

[2]Technology and Policy Program cohort

Contents

List of Figures

List of Tables

Chapter 1

Setting the stage

1.1 Preview

In recent years, development of shale gas and tight oil resources has proceeded at a rapid pace. However, they have thus far been associated with widely varying development costs and initial production rates. These factors, combined with their rapid decline rates in production, make development of these resources an economically risky business, often only profitable at relatively high oil and gas prices. The nature of economic risk in these unconventional fields is still poorly understood, and in recent years this has led to a number of operating companies writing-down or selling shale assets due to highly variable, disappointing development results. There is a need to improve the process for characterizing uncertainty in these resources and ultimately forecasting the variability of their economic performance.

Alongside the development of these new and challenging resources, there is a growing interest in the answers that may be found in prolific field-generated data. Specialized techniques and expertise for extracting useful patterns from field data are still in their nascency though. This thesis aims to address some of the immediate questions about variability in shale gas and tight oil development, while also demonstrating the relevance of data analytics to this endeavor. I consider two distinct areas—well productivity and operations involved in drilling wells—which helps to illustrate the breadth of opportunity across many different aspects of shale gas and tight oil development.

In this chapter, I provide a brief background on the growth of shale gas and tight oil production in North America and discuss some of the important implications and challenges associated with this. In this context, I discuss the general motivation behind my thesis question—how to characterize performance variability in these resources.

In Chapter 2, I provide additional background on shale gas and tight oil development and on characterizing and managing performance variability. I first describe the basic aspects of the development process, which are important to be cognizant of throughout this thesis. Then I discuss past philosophies and techniques for characterizing variability in conventionals and how these have been extended to unconventionals (a somewhat more general term that I will use interchangeably with "shale gas and tight oil"). In many ways, unconventionals are of a different nature than the conventional oil and gas resources dominant in the past. As I discuss, they require a different set of development strategies due to the unique economic and technical challenges associated with them, and these strategies are still evolving in the industry. Finally, I elaborate on the role of empirical models in both resource types and delve deeper into some of the motivations for improving our understanding of variability.

Chapter 3 introduces and defines a probability distribution of initial production rates in shale gas and tight oil—both between and within areas of a field. This new paradigm for production variability has implications for how operating companies approach the early stages of development, in which they acquire costly information to inform development campaign decision-making. To demonstrate the power of this variability characterization, I develop an application of Bayesian statistical inference that improves the ability to estimate acreage quality.

In Chapter 4, I investigate different aspects of performance variability, specifically the operational consistency and time-based performance of drilling unconventional wells. Organizational learning leads to variation of drilling procedures and improvement of efficiency over time and the learning curve has thus far proved an important framework for understanding performance improvement in the development process. It is important to develop additional performance indicators for abstract dimensions of the learning process. As an example, I turn to the phenomenon of "standardization," by which experimentation with drilling procedures over time yields standard operating procedures that are more uniform from one well to the next. I present a new technique for measuring standardization, which also illustrates the importance of bringing analytics techniques from disparate disciplines to oilfield data, a hitherto under-utilized resource in the petroleum engineering discipline. I also discuss the challenges associated with measuring progress along the learning curve from an empirical perspective since literature in this area appears to be limited in its application to real world data and the natural variability that comes with it. I analyze the variability, or "noise," within the learning curve and discuss how consideration of this source of uncertainty can improve current attempts to measure progress.

Chapter 5 serves as a concluding section to the thesis. I discuss the important elements of a more general framework for unconventional performance variability and I re-emphasize the importance of such considerations in light of the findings in this thesis. I include some general recommendations for the industry, public policymakers, and researchers based on the salient role of variability in managing the development of these resources.

This thesis answers one fairly abstract question but raises a number of other practical challenges that will require insight from a range of often silo-ed disciplines. Characterizing variability is really about charting a path toward promising ways to improve the management of variability. In this way, my thesis is not a conclusion of investigation but a jumping off point for the topic.

1.2 The growing importance of shale gas and tight oil

In the last ten years, the combination of horizontal drilling and hydraulic fracturing has pushed production from North American shale gas and tight oil plays to new levels and greatly expanded known domestic hydrocarbon resources [50, 126]. As a result, over the past decade US tight-oil production has increased six-fold while shale gas production has increased by more than a factor of four [50]. All of this has led many to predict a more hydrocarbon abundant future and to forecasts of the US becoming a net exporter of natural gas by 2018 and even a net oil exporter by 2030 [50, 36, 154, 87, 70].

The technical innovations that brought about this energy revolution about are in fact quite old. Horizontal drilling is a technique that allows drillers to gradually turn the direction of a well ninety degrees after drilling down thousands of feet vertically, and continue drilling a well by a mile or more in the horizontal direction [152, 102]. Horizontal drilling was first accomplished in 1929 and became more widely used in the 1980s; its importance in shale gas development comes from drastically improving the economics of a well by accessing more of the resource, which is trapped within thin horizontal layers of rock formation [102].

Hydraulic fracturing is the other key enabler. It involves pumping large volumes of water mixed with chemicals and solids, at high pressures into a well to fracture rock within the formation, providing conduits for gas and oil to flow [152]. This reservoir stimulation technology has been used extensively since its first commercial application on a Kansas well in 1947 [102]. It is critical to shale gas production because of the extraordinarily low rock permeability (a measure of a rock's ability to allow fluid flow) found in these shale and "tight" (impermeable) sandstone formations [102]. It has been recognized for decades that prolific resources were present in these rocks but because of the extremely low rock permeability it was generally considered inaccessible (at least commercially) until recently [102]. Strong policy support for shale gas development combined with years of high domestic gas prices set the stage, but it was George Mitchell, a Texas entrepreneur who ignored industry naysayers and invested millions of dollars into years of experimentation in the Barnett shale of North Texas [139, 102, 42]. Eventually, his efforts paid off and he found a combination of horizontal drilling and hydraulic fracturing that made economic production of shale gas possible [102, 42]. Having already made a fortune, he abstained from patenting the process for developing these resources because he saw the potential good this knowledge could accomplish in the world as part of the public domain [42]. Thus was born one of the largest resource booms in recent history.

The economic benefits of these newly accessible resources are substantial, including lower energy prices for consumers, increased state revenues, reduced national energy imports, and even a shift toward some additional exporting of liquefied natural gas (LNG) and oil condensates [81, 115]. Furthermore, historically low natural gas prices in the United States are leading to a resurgence of manufacturing and associated jobs. Cheaper natural gas-derived feedstocks and low energy prices will increase domestic chemical and energy-intensive manufacturing by an estimated 20% over the next decade [37]. Additionally, oil prices have fallen by around half in the past year, driven in large part by surging U.S. oil production from tight oil [43]. The windfall to consumers created by this leads to an estimated 0.2% boost in the world GDP for every 10% drop in the oil price [46]. Of course, the economic benefit is primarily for oil importing countries, while oil exporting nations like Venezuela and Russia are being fiscally pummeled by this price change, and find themselves with a diminished influence abroad [46, 9]. Low oil prices also present an excellent opportunity for fuel subsidies to be phased out, as they recently have been in Indonesia [9, 46, 45].

Another trend driven by the surge in shale gas production is a shift in electricity generation in the U.S. from coal to natural gas [50]. Although natural gas emits half as much CO_2 per unit of energy as coal, some of this climate benefit may be offset by fugitive methane emissions [86, 16]. There is currently a great deal of uncertainty about the amount of fugitive methane emissions emitted at the well site or at various stages within the natural gas supply chain [16]. Because of the large global warming potential of CH_4 relative to CO_2, this may erode or even undermine the climate benefit for certain end uses [16, 47]. Still, many argue that natural gas has an important role to play as a "bridge" fuel, easing the transition away from coal and complementing the intermittent nature of solar and wind energy due to its flexibility as a generation source [86, 136].

The shift from coal to natural gas will also lead to a reduction in mercury, NO_x, SO_2, black carbon, CO, and particulates emitted from the power sector, thereby reducing associated illnesses and deaths and generally improving public health [86]. However, there are substantial concerns regarding local environmental risks associated with the extraction process including potential groundwater and surface-water contamination, local air quality issues, and induced seismicity from wastewater disposal in injection wells [83, 149, 72, 156]. It is essential that industry, researchers, and policymakers continue to work toward a better understanding and management of these risks through improved technologies and regulation.

In contrast to the growing role of shale gas and tight oil production in the U.S., efforts to build a similar unconventional boom elsewhere have so far foundered [24, 129]. Nevertheless, this resource has been touted as a world game-changer economically, environmentally, and geopolitically, given the widespread unconventional basins of the world (such as in China, Poland, and Argentina) and the world's insatiable appetite for inexpensive and secure sources of energy [44, 106, 49, 84].

1.3 Variability and economic risk

1.3.1 Under-appreciated variability and economic risk

In spite of this hubris about the resource potential, concerns abound regarding the long-term economic viability of shale gas and tight oil development [144, 100, 74, 80, 27]. There is now growing concern about the variability of well productivity and rapid production decline rates, which necessitate huge capital-expenditure on drilling campaigns to maintain rates of production [100, 80, 27]. It has recently been estimated that as many as 40% of unconventional wells in the U.S. are uneconomical due to variability of drilling, completion practices, and reservoir characteristics [54]. There is a wide range of production variability in fields, persisting even in recent years with supposedly "mature" development practices [54]. These production challenges have contributed to some major operating companies writing down the value of unconventional assets and even exiting from major plays [75, 145, 123]. Internationally, there have also been signs of tapering expectations, including an exodus of energy companies from Poland, and China halving its shale gas production targets for 2020 [24, 129]. The recent plummet in oil prices is casting additional doubts on the extent to which industry will continue to invest in developing these resources [1].

Despite the mounting evidence that economic risk in shale gas and tight oil development has thus far been under-appreciated, the nature of performance variability is still poorly understood. Increasing economic pressure from low oil and gas prices means it will be necessary to improve forecasting and decision-making in fields early on, taking into account this performance variability and making greater use of available data [55, 71]. Deterministic forecasting of expected field performance (even if this is scaled up and down to provide a spread of economic outcomes) is misleading because of important statistical features like distribution skewness [67]. More transparent models of field performance variability are also needed to aid decision-making in light of nuances like physical production constraints (i.e. shared pipeline capacity) and the non-linearity of value with shifts in variables [67]. Decision-making in unconventional development campaigns should be guided not only by traditional engineering, geological, and economic assessments but also by simple, short-term predictive performance metrics that allow visualization and description of variability patterns within large datasets [147].

1.3.2 Key drivers of resource economics

A major driver in the economics of unconventional wells is development cost, including the cost of drilling and completions (hydraulic fracturing) [35]. Well construction is the highest cost in onshore oil and gas development (70-90% of overall costs) and is one of the main areas for onshore operating companies to compete with each other on extracting petroleum resources at the lowest per-unit cost [89]. The horizontal wells necessary in unconventional fields are much more expensive to drill than vertical wells, in part because the specialized drilling equipment and techniques required present numerous technical challenges and significant outcome variability [89]. In fact, well cost is lognormally distributed and unpredictable due in large part to the costly unplanned activity needed to address equipment failures, geological uncertainties, and inconsistency of implementation in drilling operations [89]. Between 2007 and 2012, the average cost to drill and complete a well in the Haynesville shale play was $9.8 million with a standard deviation of $2.0 million [89]. Average drilling costs per well were around 45% of these costs but the unitized risk, or coefficient of variation, was higher for drilling than completions (0.3, compared with 0.24) [89]. The variation in completions cost is also more likely than the variation in drilling costs to have been planned and expected by operators.

Learning is critical for improving the efficiency of well construction in shale gas and tight oil development campaigns. An analysis of over 200 horizontal wells drilled in the Eagle Ford shale play by 31 different operators between 2008 and early 2011 found a trend toward decreased drilling days per footage over time, but there was large variation in total drilling days for similar well depths and and trajectories [62]. This suggests that greater consistency of implementation may lead to even larger improvements [62]. Indeed, a focus on greater consistency of learning efforts led Shell to work toward improving the organizational structure for capturing and sharing drilling best practices, allegedly reducing well delivery times by a factor of three in a relatively short time [148]. The specific levers behind the "learning curve"—which describes increased efficiency with increased experience—are not clear, but the process generally involves experimentation to adapt to local conditions, and standardization of the drilling approach [148]. The ability to spread learning from one rig or field to another and to quantify high-level metrics of performance improvement and consistency are critical steps toward effectively managing this temporal reduction in drilling costs [17, 148].

Another major source of uncertainty about economic success is the production profile of a well. The initial production rate and the rate at which the production declines from the peak rate achieved during the first few months of the well's life are highly variable and challenging to predict [35]. As much as two-thirds of the uncertainty about a well's net present value may be related to the production profile [68]. A well that has begun producing still has uncertainty about the future rate of decline in production, but is also a significant source of information about the area in which it was drilled. In addition to any well log data and core samples collected during drilling, there is immediately available information in the form of production data, which can be used to assess the acreage quality and evaluate completions techniques. Production from an early well can be used to support the critical decision of whether to continue drilling additional wells or abandon an area (the "go/no-go decision") [68]. However, even wells drilled from the same well pad may have widely different results, and the "data point" the well provides is useless for decision making if it is not considered more broadly within the potential variability of the area [116].

1.3.3 The need for new policies

The current shortcomings at understanding and managing variability may partly stem from an inability to adapt a "conventional mindset" for oil and gas development to the new challenges presented by unconventionals [55]. Economic success in shale plays requires a combination of efficiency (low development costs) and effectiveness (high well productivity) [54]. Many in industry have focused primarily on the need for greater efficiency by advocating a "factory approach" to development [53]. This philosophy may have some merits, but taken alone, it gives inadequate consideration to the consistency of the product. A fantastically efficient factory that is unpredictable and delivers widely varying results should hardly be an aspiration for the oil and gas industry. Indeed, in this sustained downturn in oil prices, questions have been raised as to whether a factory-inspired approach to drilling a large number of wells at the lowest cost possible is really a better strategy than focusing on achieving fewer, more-productive wells [55]. Part of this shift will require a replacement of unsystematic field experimentation with thorough analysis and planning [55].

The continued success of unconventional resource development within the U.S. and the spread of development to other parts of the world depends in large part on the ability to characterize, forecast, and manage the variability associated with these resources. Further development of this capability will support reliable assessments of a field's economics early on, more optimal planning, and faster learning during a field development campaign.

1.4 Thesis question

Concern about this important and poorly understood dimension of shale gas and tight oil development leads me to my thesis question: *How should an operating company characterize the variability of drilling and production performance of wells using existing field data?* There is a substantial opportunity to develop approaches that makes use of "free" unused field data to supplement and guide other more costly approaches, improve forecasting early in a development campaign, and provide a foundation for more advanced understandings of uncertainty and techniques for mitigation. This thesis makes an important contribution to this area by identifying some key patterns of variability and providing tools for assessing them.

1.4.1 Metrics

There are many ways in which this topic could be taken on. At this point though, I should define clearly the performance metrics I use, with some justification for each. These metrics attempt to address some important aspects discussed in the previous section, while also making use of the data that I had access to.

For production performance, I focus on the peak production rate of a well, which typically occurs within the first few months of production. Peak production is correlated with, and is commonly used as a proxy, for early-life production, and to a lesser degree, lifetime production [147, 125, 100]. Due to the rapid decline rates in production and the discounting of revenue from future production, early life production is critical to shale gas and tight oil resource economics[1] [147]. Initial production rates may be more indicative of the quality of the stimulated fracture network created than the underlying reservoir quality, but it's economic importance is still key. I have also considered an alternative metric for production performance—production in the first 12 months, which is likely more reflective of the underlying reservoir quality—and Appendix B includes a handful of figures extending my analysis to this.

For drilling performance I rely first upon drilling time (or, in some cases, time per footage of well drilled). More broadly, I am interested in variability of cost, but drilling time is a particularly important and substantial source of variability for this. Rig rental is the largest component of drilling costs (followed by labor, which could also be thought of as having a "daily" cost) [89]. Haynesville rigs between 2008 and 2012 had a base day-rate of $12,500-$22,000, excluding fuel and add-on costs[2] [89]. The only larger component of an overall well's costs is stimulation and sand control (associated with completions, not drilling) but these are static material inputs less susceptible to change over time or unpredictability. Reduction of these inputs might also negatively impact production whereas a well can be drilled just as effectively in different amounts of time depending on the efficiency of operations. Analysis of drilling time also allows me to investigate learning effects. Time is considered a better proxy of learning than directly measuring cost since it is not confounded by changing rig or equipment day-rates [17, 89].

Finally, I develop an additional metric for drilling performance that addresses the consistency of drilling procedures from one well to the next. I call this "operational variability" and use it to detect procedural "standardization" which takes place during the learning process. Additional work will be needed to realize the full potential for field deployment of this metric, but nevertheless it provides an interesting example of how analytics can be used to visualize hidden behaviors in a sociotechnical system and allows for immediate insights into how the drilling process changes over time in a field development campaign.

[1]Production flow rates in wells fall by around 50% in the first year, and substantially more after three years—by some estimates, to as low as 5-20% of the peak production rate [142, 74].

[2]The "effective" day rate for drilling (including all drilling costs divided by days of drilling) ranges from $71,000 to $92,000/day [89].

Chapter 2

The background

Unconventionals present the oil and gas sector with a fundamentally different type of economic risk than it is used to dealing with [74, 80, 27]. The high degree, and distribution of well-to-well production variability means companies are exposing themselves to appreciable economic risk each time they roll the die by drilling a well, even after many nearby wells have been drilled. This contrasts with economic risk in conventional oil and gas development, which is driven more by uncertainty about the size and presence of hydrocarbon accumulations and can be mitigated by geological data collection and reservoir simulation carried out during exploration [131, 143]. Characterizing the distribution of well productivity in shale gas and tight oil fields is critical to better understanding the production variability—a major factor in the economic risk associated with these resources [116].

Furthermore, observed variability of drilling operations and time-based performance suggests companies are often investing significantly more capital over the course of a field's development than would be necessary with greater operational consistency and faster organizational learning [148, 89]. Better understanding the nature of the variability present in the drilling process will improve the ability to forecast future drilling costs in a field, promote best practices and measure learning within a development campaign, leading to greater efficiency and effectiveness at earlier stages of development.

This chapter reviews some relevant literature to the issues of performance variability facing developers of unconventional assets in North America today. Because of the broad, systems perspective of my thesis, a breadth of interrelated topics are involved, and in reviewing them here I can only skim the surface of these prolific areas of existing knowledge and ongoing research. As in the rest of this thesis, precise causal relationships and full appreciation of the many involved nuances is of less concern than framing the bigger picture question, of how generally to think about the issue of variability in production and drilling performance in unconventionals and develop new techniques that can be incorporated into the overall development process to improve its management.

2.1 Background on development process

In this section I provide a basic overview of the operations for drilling and completing a massively hydraulic-fractured horizontal well typical in North American shale oil and gas fields. I then discuss some of the differences between the geology and development strategy of unconventionals and conventionals and the challenge that this is presenting to the oil and gas industry.

2.1.1 Drilling the well

An excellent reference on the process of drilling a well is *A Primer of Oilwell Drilling* by Bommer [13]. The following description is drawn from this source, unless otherwise noted, and should be referred to for greater clarity and additional illustrations.

The actors — An operating company owns, manages, and engineers the wells being developed. The fieldwork is carried out by drilling and service contractors.

Site preparation — The well location is determined based on a variety of legal and economic factors including the need to obtain a lease to drill there. If other wells have been drilled in the area, the placement of a well may be informed by knowledge obtained from them. In any case, geologists and geophysicists will generally suggest areas they think likely to be productive based on promising expected geology from available data. Prior to bringing the drilling rig on location, the drilling site is cleared and flattened if necessary. A reserve pit is dug to store waste drilling fluids. Other earthen pits may also be dug alongside to store water and rock cuttings. The primary drilling fluid, "mud," is generally stored in metal tanks to maintain quality. A small, secondary rig might at this point drill any nearby holes necessary for temporarily storing drill pipes. Additionally, in some cases this smaller rig might "spud," or drill the initial hole which is wide but shallow. This conductor hole is lined with casing and cemented in place. Alternatively, this last step of spudding the well might be done by the main rig once it is on site.

Rig move (from previous well) and rig up — The disassembled rig is moved to the well site on trucks, or if it is a "skid" rig, it is towed to the site. During rig up, the process varies based on rig type. For instance a "slingshot" rig can be elevated using hydraulic pistons. The "drawworks," which hoists or lowers a cable called a "drilling line," and other substructure equipment are set up and the mast or derrick is raised into position. This system will be used to raise and lower everything in and out of the wellbore [152]. Additional equipment is "rigged up," including electric generators, steel mud tanks, pumps, stairways and walkways. Living quarters are brought on site along with drill pipe, pipe racks and other drilling equipment. A metal building called a "doghouse" is set up next to the rig floor for the driller and crew to use as an office. A representative but highly simplified schematic of an onshore drilling rig is shown in Figure 2-1.

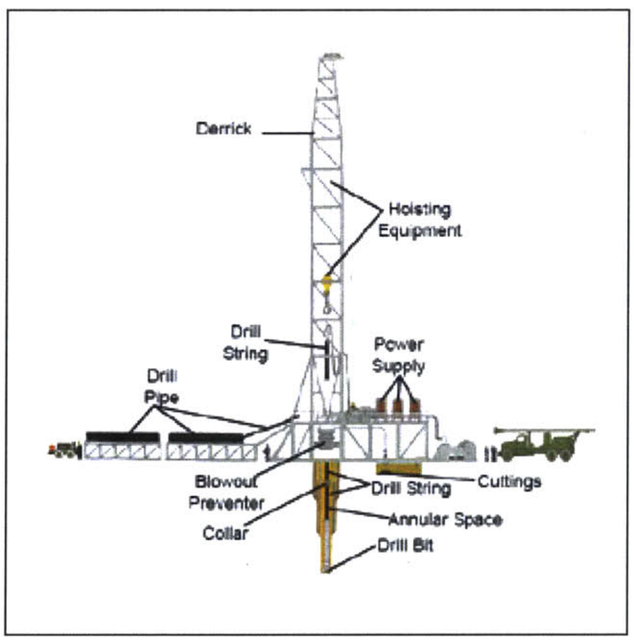


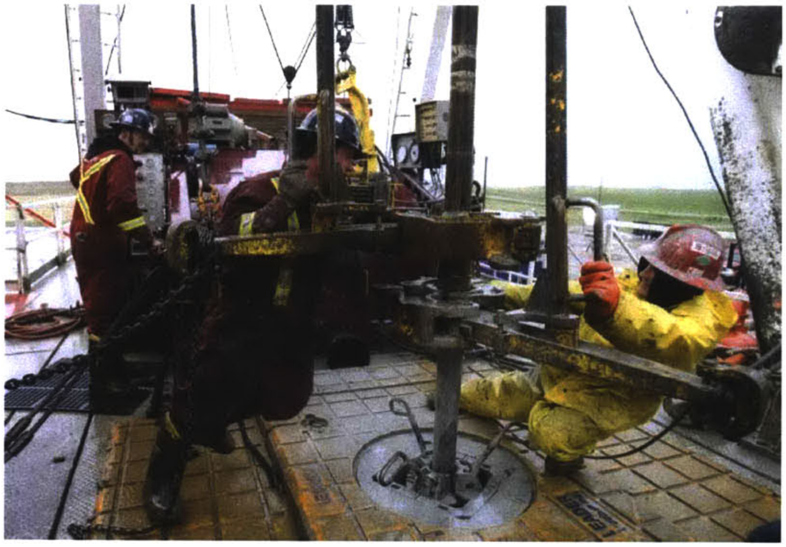
Figure 2-1: Simplified schematic of onshore drilling rig. Source: Tidal Petroleum

Drilling the surface hole — Once rigging up is complete, the well is spudded and the conductor casing cemented in place (if this was not carried out previously by another rig).

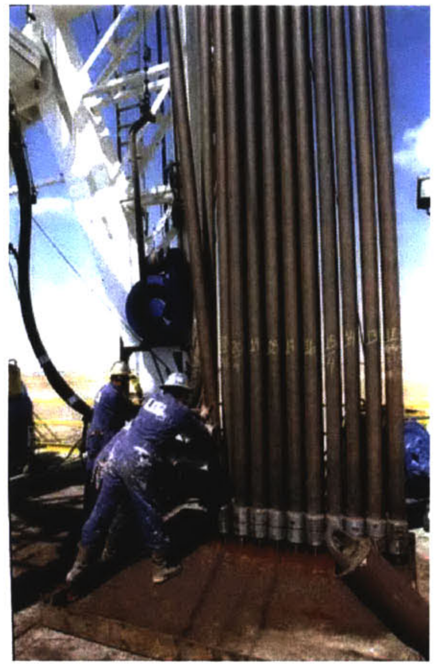
The surface hole is then drilled to a depth of a few thousand feet typically, using a large diameter surface drill bit "made up" to a "bit sub." As the hole is drilled, "drill collar" is added above the bit sub. This creates the "bottom hole assembly" (BHA) used to drill the well, which is the bottom portion of the "drill string," or column of connected drill pipe and equipment at any given time. Once the BHA is made up, drill pipe is added to lengthen the drill string and allow the bit to reach the bottom of the deepening wellbore. To make up (i.e. connect) the segments of drill pipe, "roughnecks," or fieldworkers, guide suspended drill pipe into place at the top of the drill string once it is flush with the rig floor. They then use tongs or specialized equipment to apply torque to the drill pipe and tighten the screw threads between pipes. Drive systems vary but they all allow rotation of the entire drill string [152]. It is this rotation, referred to as "rotate drilling," that is generally used to turn the bit and crush rock at the bottom of the wellbore. Mud is circulated through the drill string and back up the annulus of the hole in order to return drill cuttings and to maintain bottomhole pressure [152].

Tripping out — Once the planned depth for the surface hole is reached, the drill pipe is "tripped out" and stored. In order to trip out pipe, the drill string is pulled out gradually to expose a stand of pipe. Pipe is then "broken out" using equipment to rotate it and spin out the threads (Figure 2-2a). The pipe is then racked back (Figure 2-2b).

Running and cementing surface casing — Next, metal casing will be put in place and cemented to isolate the surface hole from the surrounding rock and surface water. The casing joints are added in a similar manner to that of drill pipe until the casing reaches the bottom of the surface hole. A cementing company brings cement-pumping equipment to the site and mixes the cement slurry. The cement is pumped into a cementing head that is attached to the top joint of casing. A "bottom plug" is dropped from the head just before the cement in order to keep mud in the wellbore from contaminating the cement slurry and to push the mud on through and back up the annulus. When the plug reaches the bottom of the casing, it sits in a "float collar" that is in place there. This causes a membrane to be broken and cement flows out of the bottom of the casing and fills the annulus space around the casing from the bottom up. When the last of the cement is pumped, the "top plug" is dropped to separate the cement from the displacement fluid, which is pumped after it to fill the inside



(a) Source: Troy Fleece Photography



(b) Source: OSHA

Figure 2-2: A rig crew uses tools to break out drill pipe and then rack it back.

of the casing and displace the cement. A pressure jump indicates that the top plug has "bumped" the bottom plug and the cement has filled the annulus of the casing, ending the cement job (Figure 2-3).

A period of a few hours follows for waiting on cement. After this, the wellhead can be installed and the blow-out preventer (BOP) placed, or more formally, "nippled up." Pressure testing of the BOP stack, wellhead, and well is carried out before further drilling can occur.

Tripping in and drilling out the shoe — The BHA for the next section of the well is picked up first and the bit is returned to the bottom of the hole by adding stands of pipe back to the drill string as it is lowered. Once the bit reaches the bottom of the surface casing, the bit drills out the equipment and cement at the bottom then starts making new hole. Before proceeding to drilling the next section of hole, the casing and cement job is pressure tested.

Constructing other sections of the well and downhole maintenance — The well may be designed to have two, three, or even four well sections (the surface hole is the first as the conductor hole is not generally counted). This is generally determined by drilling engineering calculations taking into account the target depth and pressure regime. The other sections are drilled, cased, and cemented in essentially the same manner as the surface hole. Casing for

Figure 2-3: Diagram of top and bottom plugs for cementing casing. Source: Schlumberger

these additional well sections may be run all the way to the surface or it may be suspended from a hanger and act as a liner beyond the first string of casing. During drilling of any of the sections, additional round trips may (and likely will) be required because of equipment failures or a need to change out the bit or BHA to achieve different drilling parameters. If a tool or other item is lost downhole it may be retrieved using "fishing equipment." Small items may also be drilled using "junk mills" if they are not retrieved. In addition to these downhole issues, problems with procedures or equipment at the surface may lead to nonproductive time (NPT), which may be interspersed throughout the overall process and is a frequent source of wasted time in the drilling process. There are many other activities that may be carried out throughout the drilling process, either to recover from unintended outcomes, or to attempt to preempt them. An example of this is pumping through the well different mixtures of drilling fluids, in precise volumes and densities, and sometimes containing solids[1] in order to control the wells behavior. The variability of these procedures from one well to the next will be relevant in the later discussion of operational variability in Chapter 4.

Drilling the curve section — Directional drilling in shale wells can be accomplished using either "slide drilling" with a "mud motor" or "rotary steerable assemblies." With slide drilling the BHA contains a section of bent housing that orients the bit at a small angle (Figure 2-4). The bit can then be pointed by holding the drill string steady and turning the bit

[1]Peanut shells may even be used as "lost-circulation material" to prevent loss of drilling fluids through the wellbore sides [134]!

by pumping mud through the hydraulically powered downhole mud motor. A "measurement while drilling" (MWD) tool placed higher up in the BHA detects the direction of the hole and sends pressure pulses back to the surface to relay the information. Once the desired direction is achieved the directional BHA can be rotated conventionally with the drill string and the current direction will be maintained because the bend in the mud motor has no preferred direction (Figure 2-4). This is much faster than slide drilling. If a rotary steerable assembly is used instead of slide drilling then the wellbore path is programmed into a computer in the assembly. The computer steers the drill bit while the drill string is being rotated allowing for more rapidly drilled curves.

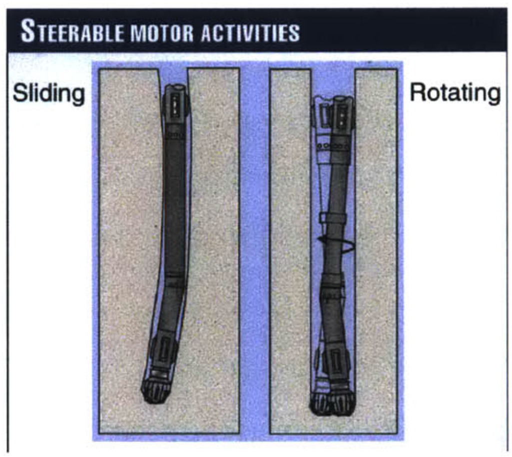


Figure 2-4: Slide and rotate drilling with a bent mud motor. Source: Oil & Gas Journal

Drilling the lateral section — In the lateral section a different BHA with a smaller bend angle may be used or the curve assembly may remain in place if the bend angle is not too large. The lateral section is drilled primarily by rotating the drill string. If the BHA "walks" and veers from the planned path or if the target zone is changed then some slide drilling may be required to change the wellbore direction.

Rig down/skid/slide — After the well total depth (TD) has been reached and the production casing installed the well is fully drilled. The rig is either disassembled and trucked to another site or skidded to a location nearby. For multi-well pads, the rig may be on slides that allow it to be quickly shifted in order to drill other wells alongside before it is removed from the site.

The drilling process is complicated and highly variable, even for wells in the same field that are intended to be homogeneous and standardized to promote efficiency. Almost every step described here has either some caveat, slightly different alternative approach, or more detailed set of optional and required procedures that I have neglected to mention here in the interest of brevity. The process is also not as linear as I have described it here. Many of these procedures are in reality mixed and matched in various orders, particularly when the operating company and crew engaged are less familiar with the field being developed (i.e. early in a drilling campaign). Variations may enter the process due to unforeseen conditions (and the need to mitigate them), differences in opinions and 'styles' of drilling engineers, company men and superintendents (field managers), planned experimentation with procedures to determine bast practices for the field, and inconsistency of the field crew at implementing the procedures. Chapter 4 will consider how to measure this operational variability and the variability in time-based performance improvements as an operating company and crew "learns" the best way to implement consistent wells in a field. Some further background on learning in oil and gas drilling will be provided in Section 2.3.

2.1.2 Completing the well

I will not go into as great of detail describing the process of completing, and hydraulically fracturing the well, as I have for the drilling operations of a well because the step-by-step procedure is of less interest to this thesis.

Completing the well includes casing and cementing the well[2], perforating the horizontal section of well with shape charges from a perforation gun, installing production equipment (production tubing and production wellhead), and stimulating the reservoir with hydraulic fracturing, or "fracking" (Figure 2-5) [13]. Hydraulic fracturing is essential to production of oil and gas from extremely low permeability unconventional reservoirs. Hydraulic fracturing involves pumping between 2 and 20 million gallons of water, mixed with specialized chemicals, and proppants such as sand or other solids into the well, and out of the perforations [83]. The volume and pressure (10,000-20,000 psi typically) is sufficient to fracture the rock around the well and drive the propagation of these cracks out into the rock formation to provide a flow path for fluids trapped in the shale reservoir [83]. The solid proppant is wedged into the

[2] Because this occurs alongside drilling the well I have included this activity in the previous section.

fractures and helps to prevent them from closing up once the pressure has been removed [13]. The chemical additives in the mixture adjust certain fluid properties, such as viscosity, to improve delivery of proppants into the formation; some additives address other down-hole concerns such as bacteria growth [41]. Hydraulic fracturing of a well is carried out in "stages" (sometimes as many as 40) in which a portion of the horizontal section of well is isolated and fracked. After the well has been stimulated and the waste-fluids flowed-back for disposal, the well enters its productive life.

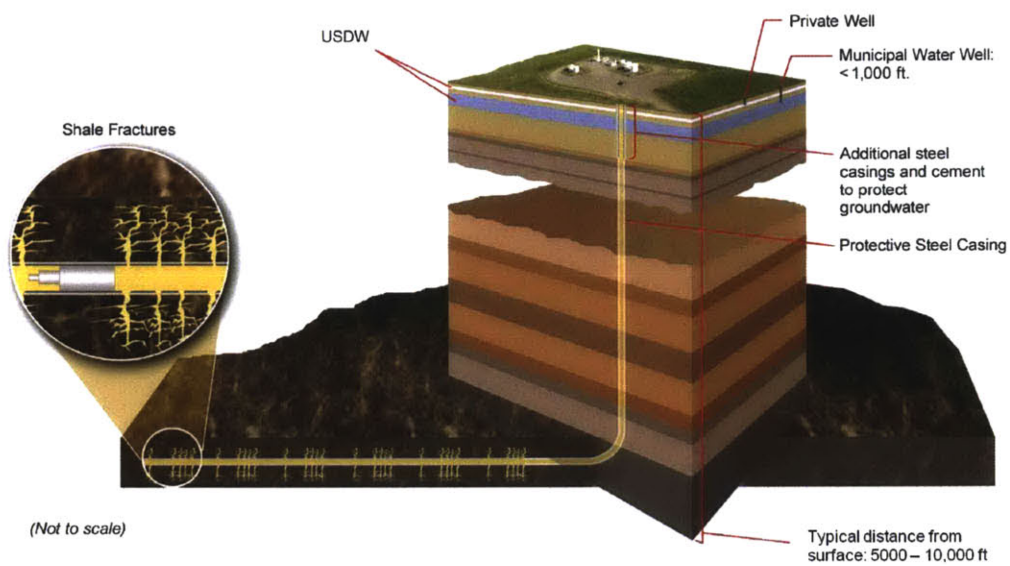


Figure 2-5: Schematic of hydraulically-fractured shale gas system. Source: US DOE-NETL

2.1.3 Unconventional oil and gas—a new set of development challenges

The differences between shale gas and tight oil and conventional gas and oil reservoirs begin with the geological processes forming the hydrocarbon accumulations. Source rock is a rock formation, often shale or limestone, which is rich in organic matter that becomes oil or gas under the right combination of heat and pressure [152]. Because the formed hydrocarbons are less dense than the water typically filling underground pore space, over time they will travel upward until they either reach the surface through seeps or accumulate in geological "traps." [152] These trap accumulations are the conventional reservoirs that have long been the exclusive target of petroleum exploration; with shale gas and tight oil we are going to the source rock itself, and "raiding the kitchen," in a sense (Figure 2-6) [152]. Conventional reservoirs can be thought of as discrete deposits of petroleum, defined spatially by the geological structures trapping them in place, while unconventional reservoirs are really source rock formations through which trapped petroleum is somewhat continuously dispersed [8]. See Figure 2-6 for a diagram of this. More generally, continuous petroleum systems spread throughout a regional area with a common geological history are often referred to as resource "plays." [65, 91]

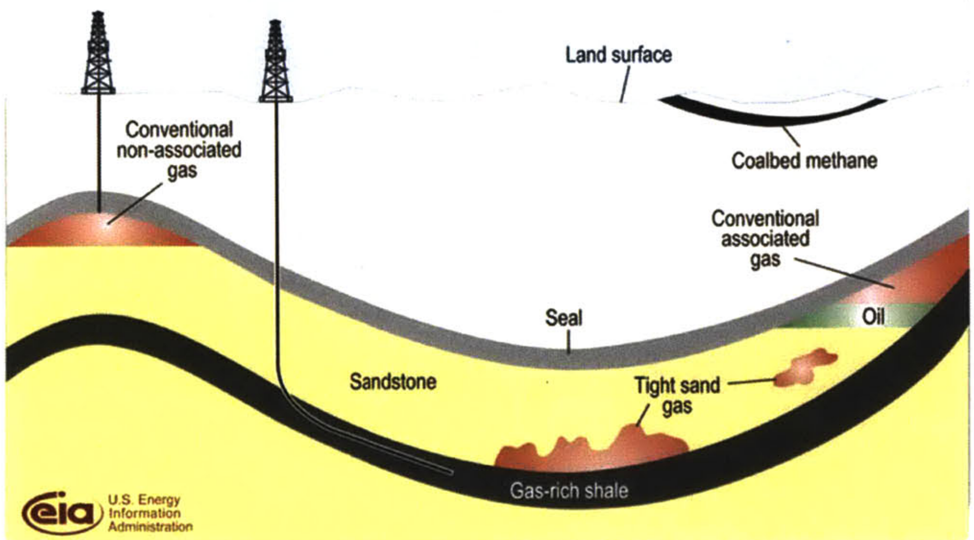


Figure 2-6: Schematic geology of natural gas resources (Oil is similar). Note the conventional petroleum accumulations formed by seal "traps." Coalbed methane is another type of unconventional gas that does not migrate from shale but is generated in separate processes related to coal. Source: US EIA

This geological story aside, the reservoir characteristics are very different between conventional and unconventional resources (which leads to differences in the development challenges associated with each). Porosity (a measure of the pore space, which can hold fluid in the rock) and permeability are both necessary for production in conventionals [152]. In unconventionals, the rock has voids (porosity) filled with petroleum, but the voids are not very connected (and therefore, the rock has low permeability), and the ability of fluid to flow is very limited [85, 8]. In conventional development, rocks are naturally more favorable to flow, so much of the economic success of a project is associated with the size of the reservoir, especially thickness, which will impact how long wells will be productive [152, 131]. Important geologic factors are greater in number and have more complex interrelationships in unconventionals, making rates of production and life of wells both highly unpredictable. [142] Furthermore, heterogeneity of rock properties is greater in unconventionals, and properties like permeability, or oil and gas saturation are widely distributed throughout even relatively small areas [116, 8].

Although risk and uncertainty is always present in oil and gas exploration and production, in conventionals it is substantially reduced over the course of developing a field, and can be described as sequential investment decisions under decreasing uncertainty [131, 143]. A major goal of exploration and appraisal efforts in conventionals is to reduce the scope of uncertainty as much as possible early on, which leads to extensive data-collection, simulation, and planning for conventionals (as described in the next section), particularly for large scale offshore projects [152, 131]. In contrast to this, a manufacturing, or factory, approach to development has been adopted in unconventional fields with less concern directed toward addressing key uncertainties [53]. Shale gas and tight oil plays are often referred to as "engineering plays," due to a greater emphasis in industry on drilling and completions strategies rather than thorough geological assessment, simulation, and reservoir optimization [25]. This frenetic development approach would benefit from more accurate predictions early on [142], but thus far there have been limitations to forecasting in shale gas and tight oil, as the next section will explore.

2.2 Resource evaluation methods

In this section, I discuss some of the key sources of data on conventional petroleum reservoirs, and some limitations seen in data collection for unconventionals. I delve further into the limits to different approaches for reserve estimation in unconventionals stemming from insufficient data and other challenges. As this thesis takes a more statistical perspective of well productivity, I present some past efforts toward this in conventionals and background on the lognormal distribution in natural systems generally.

2.2.1 Reservoir data collection

There are a wide range of surveying methods used to detect the potential presence of a conventional petroleum reservoir and estimate properties of rock and fluids in the reservoir. These include surface observations of field geology and hydrocarbon seeps, but in recent years technological innovation has vastly improved exploration efforts through geophysical surveying of subsurface structures [152, 131]. Geophysical surveying, such as gravity, electromagnetic, and radiometric surveys, are used in exploration to identify anomalies that vary from the background geology of an area [152]. A technology that has become extremely important for reservoir characterization in recent years is seismic imaging, which maps the subsurface by monitoring the reflections of sound waves propagated into the ground [152]. These sound waves are produced by airguns in offshore exploration, or vibrating trucks onshore [152]. This data is costly to obtain because it requires specialized equipment and personnel and the process takes months [152]. After being collected, the data must be processed extensively and requires subjective interpretation by specialists [152].

Ultimately, the best source of information on a conventional reservoir, and the only way to know that hydrocarbons are present, is drilling. Surveying data is used by reservoir engineers to strategically place appraisal wells in order to determine presence of a petroleum reservoir, and estimate the size and commercial viability of a field [152]. In addition to determining the existence and extant of hydrocarbon accumulations, a number of other properties can be sampled or estimated from data collected in the wellbore while drilling is taking place, through "logging while drilling." Drilling mud returns are analyzed to determine pressure, depths of formation change, and hydrocarbon presence [152]. "Cores," or rock samples collected in the reservoir zone of the wellbore using a cylindrical tube, are subjected to numerous laboratory tests to measure rock porosity, permeability, hydrocarbon and water saturation, and lithology (rock type); Similarly, reservoir fluids can be collected and tested to determine their properties and reservoir conditions [152]. Electrical resistivity, acoustic logging (similar to seismic), and nuclear logging in the wellbore can provide guidance on a number of rock properties correlated with these measures [152, 128]. The aim of many of these studies is to determine fluid and rock properties that can be used for numerical reservoir simulations or to estimate a "recovery factor," the percentage of petroleum volumes in place that can be recovered, helping to reduce the uncertainty about well productivity [152].

There are some limitations to the deployment of these approaches in unconventional resources. Some of this is associated with deployment cost. The value of information must be balanced with cost of data collection, and often there is less value in gathering more data for an individual unconventional well through surveying technology [25]. The development cost and potential production for an individual unconventional well is substantially less than for many conventional wells, such as those in large offshore projects; as a result, less data is gathered and less analysis goes into optimization of production. Seismic for instance, can be used to optimize a well location according to subsurface structure and geomechanics [138], but may not be feasible or cost effective in an onshore environment [131]. To make matters worse, rock heterogeneity places limits on the usefulness of samples and tests carried out to ascertain properties of the reservoir. Many parameters, such as porosity, are more complex and widely distributed throughout shale and tight formations than they are in conventionals, with large variations often occurring within relatively small areas [25]. This makes it nearly impossible to sample a field extensively enough to understand properties and how they vary throughout the formation. Finally, there are issues with obtaining accurate measurements from reservoir samples taken, including cores, and cuttings. The rocks found in these formations are susceptible to irreversible changes at atmospheric conditions, and it is difficult to retrieve, handle, and prepare samples for tests in a manner that can replicate *in situ* conditions reasonably well [25].

2.2.2 Reserve estimation in unconventionals

One category of approaches used for forecasting reserves and production in shale gas and tight oil fields is bottom up analysis of geological parameters and modeling reservoir production behavior based on this [116]. As the previous section reviewed, there are challenges associated with acquiring sufficient and reliable data in unconventionals, making it difficult to reduce uncertainty about reservoir properties [116]. In addition to data limitations, the "extremely low aggregate permeability" of these petroleum systems makes conventional reservoir engineering models and analysis inadequate and inappropriate for unconventionals [142]. For instance, a straightforward material balance approach to reservoir modeling is not reliable because the low permeability precludes the use of an "average" reservoir pressure [142, 31]. It is also difficult to establish a reasonable recovery factor from reservoir parameters, which could be used to obtain volumetric estimates of reserves [142, 31].

Numerical simulations of flow, which are used prodigiously in conventionals, may prove their worth in unconventionals, but only if large advances can be made in our scientific understanding of fluid storage and transport mechanisms in these types of rock formations [122, 25]. Furthermore, the natural and induced fracture networks are a major factor governing production behavior, but there is significant uncertainty about the fracture properties for a given well and how to model them[3] [25, 142]. There have also been efforts to adapt rate-transient analysis, by which flow regimes are inferred from production history, to unconventional reservoirs and completion geometries but this is an area of ongoing work and suffers from many of the same limitations as reservoir modeling, because of our limited understanding of flow behavior in shale and tight formations[142, 25]. It is also difficult to reliably estimate properties at early stages of well production [142, 25].

An alternative approach to forecasting is extrapolation of production, or decline curve analysis, in which parameters of an existing decline curve model are fitted to the early months of production decline in a well. [116, 100]. The fitted decline curve model is then projected into the future to forecast future production and technically or economically achievable ultimate recovery. As a result of the challenges in physical and geologically-informed models, decline curve analysis is one of the most widely used approaches for forecasting future production from shale gas and tight oil wells [116, 100]. This approach is simple and accessible, even to those without any understanding of reservoir engineering, but there is also huge room for error especially early on in a well's life when limited production data is available [142]. Decline curve analysis is perceived as lacking transparency and rigor (due to a lack of physical explanation for the curves), and the use of this method by operating companies to book reserves is somewhat controversial because of the potential to overstate resources [116].

Different decline curves exist in literature, including the traditional Arps decline curve [6],

[3]Some uncertainty can be reduced through monitoring of fracture propagation using micro-seismic but the use of this technology in unconventional fields is still limited and many uncertainties remain about the observations it provides [111, 112].

and newer models intended to capture unique decline behavior observed in unconventionals, like those by Valkó [146] and Ilk et al. [79]. There are concerns about the applicability of the Arps decline curve with shale gas and tight oil because it has a tendency to overestimate future production (especially when it is applied early in a well's production life) [100, 116]. Arps decline curves are really not appropriate for unconventional wells, because the rate of decline changes over time and the fitting parameter often exceeds physically realistic values [142, 116]. The methods proposed by Valkó and Ilk et al. may be better suited for use with shale gas and tight oil [100].

Even with these newer models, it remains unclear how valid the assumption is that future decline will follow early decline behavior and a probabilistic approach to decline curves would be a great improvement [116]. Thus far, there is no widely accepted approach for this [116, 100]. Some attempts have been made to adapt existing decline curve models for these purposes but they lack reliability and produce extraordinarily wide uncertainty bands making them less useful [116, 100, 60, 61, 21]. A better understanding of the statistical variability of well productivity in shale gas and tight oil may help to improve probabilistic models of production.

The importance of empirical production models is not only in prediction, but also potentially improving our understanding of the physics of shale gas and tight oil production. After years as a widely accepted empirical model, a physical explanation for the Arps decline curve was eventually developed for conventionals, which remains useful in rate transient analysis [52]. These models can also be used for computational analysis of the role of different production parameters, allowing for greater incorporation of available data [132]. A notable example of this from Patzek et al. incorporated a large database of unconventional production data and developed not only an empirical model of production but also a physical theory to explain it [125].

Production forecasting and reserve estimation in unconventionals remains a challenging area of work today. To date though, extensive drilling of a field is the only way to really assess the resources in a field [116]. Defining a probability distribution for productivity may enable attainment of knowledge about a field earlier in a drilling campaign, before many inefficiencies of development are built in.

2.2.3 Statistical approach to resource evaluation

The petroleum industry has historically preferred to avoid a "casino" analogy to development, in favor of a more deterministic perspective[4] [131]. However, such a view does not eliminate the large uncertainty associated with petroleum development projects, and may lead to inadequate consideration of risk and poor decision making as compared to more probabilistic approaches [131]. Fortunately, in recent years, there has been growing recognition that quantification of uncertainty and statistical models can improve estimates and decision making, especially in resource plays, which are often considered to have more statistical risk than geological risk [91, 131, 65, 116].

One frequently useful probability distribution in petroleum resource evaluation is the lognormal distribution[5]—a distribution which is Gaussian when the natural logarithm is taken of it. This concept was introduced by Arps and Roberts in 1958 and gained popularity in the 1960s due to further data analysis by others, most notably Kaufman[6] [5, 90, 38]. Initially, this distribution was used to describe the distribution of field sizes[7] in a region. Over time, the theory was further refined to better take into account aspects such as economic truncation (small fields are less likely to be developed) and size bias (larger fields are more likely to be discovered first) and a number of statistical methods for exploration were developed based on lognormal and related distributions [92, 137, 7, 4]. More recently, the lognormal distribution continues to be advocated as useful for constraining or informing estimates of resource properties and reserves, although it has fallen out of favor with unconventional resource plays because of the perceived low risk associated with their development[8] [131, 65, 64, 68]. In unconventionals, uncertainty about field size may have simply been replaced by uncertainty about the productive quality of rock and completion effectiveness. A rigorous test or application of lognormality with shale gas and tight oil has yet to be reported.

Lognormality is commonly found in natural systems due to the salient role of multiplica-

[4]The industry has a tendency to view deterministic models as more "scientific" for some reason [131].

[5]This has also been used to describe the exploration economic risk, a related but more complex assessment [15].

[6]The lognormal distribution was also applied in the mining industry for assessment of mineral deposits [95, 96, 97].

[7]Note the importance of field size with exploration for conventional petroleum fields, as discussed in Section 2.1.3.

[8]This assumption is only recently being questioned, as discussed in Chapter 1.

tive effects (as opposed to additive effects which generate more Gaussian distributions), and the lower bounding at zero of values that must be nonnegative [11, 105, 141]. An example of multiplicative effects is the amount of petroleum in a reservoir, which is the product of bulk rock volume, porosity, and pore petroleum saturation. A normal distribution in each of these parameters leads to a lognormal distribution for the product [141]. This is, in effect a variant of the central limit theorem, and the lognormal distribution should be thought of as just as fundamental as the normal distribution, despite its derivative name [2]. Although some have gone as far as advocating that this theory is explained by geochemical processes leading to the formation of mineral deposits, a more supported explanation is that the process itself of classifying elements of a population based on shared characteristics leads to an emergent, approximately lognormal distribution [91, 2, 18]. For the case of resource plays (particularly the consideration of lognormality at different scales, which is investigated in Chapter 3), it is relevant that Crovelli demonstrated that the convolution of lognormal random variables within petroleum plays produces a nearly lognormal distribution, with noticeable differences only in the extreme tails, implying a scalability to the application of the theory within fields [28, 29, 91].

2.3 Learning in unconventional drilling

The upstream petroleum business is a unique engineering venture because of the confluence of vast amounts of advanced technology, complex natural systems and phenomena, direct involvement of humans at various levels (from analysis and design, to manual deployment of equipment into miles-deep wells), and most of all, immense uncertainty. The natural-artificial hybrid systems being designed and engineered are never directly observed, and conditions and behavior must be inferred from a range of data sources. This makes the process of learning and inference particularly important for understanding drilling performance variability. In this section, I provide background on learning curves, discuss some important organizational aspects of learning, and highlight some aspirations used to direct learning in drilling, such as standardization and continuous improvement. I also outline some of the limitations in all of these areas, especially in understanding and managing system variability.

2.3.1 Empirical models of learning

A "learning curve" is frequently used to empirically describe the effect of learning by doing, in which cost of production is reduced with increasing units of production [104, 155, 99]. It has also been identified as describing progress (both in terms of time and cost) in the drilling of oil and gas wells in a field [77, 17]. The form of the drilling learning curve suggested by Ikoku[9] [77] is shown in Equation 2.1.

$$y_n = an^b \tag{2.1}$$

In this expression, n is the well number in the sequence, y_n is the time of drilling the n^{th} well, and a and b are parameters fitted with nonlinear regression. An additional model for learning in drilling, which is still widely used [82], was presented by Brett and Millheim[10] [17] as shown in Equation 2.2.

$$y_n = C_1 e^{(1-n)C_2} + C_3 \tag{2.2}$$

In this expression, n is the well number in the sequence, y_n is the time of drilling the n^{th} well, and C_1, C_2, and C_3 are estimated parameters. More specifically, C_3 is the technical limit once no more efficiency gains can be made (in a plot of the learning curve, this appears as an asymptote as $n \to \infty$), C_2 is the rate of learning, and C_1 is the initial time it takes to drill the first well.

Although the learning curve is often applied to historical data to understand trends, there are challenges with using it as a basis for planning and decision-making. To begin with, past progress is no guarantee of future performance (as is an issue with any curve fitting exercise). Particularly early on in production, there is significant uncertainty about future rates of learning, although probabilistic methods can help to forecast a range of outcomes [113]. Beyond prediction though, one of the primary motivations for studying the learning curve is the development of tools and organizational structures for better managing

[9]In the broader learning curve literature, this was the originally proposed power function learning curve used by Wright to describe airplane manufacturing cost [153].

[10]In the broader learning curve literature, this is similar to the exponential function learning curve proposed by Levy [104].

and speeding up the learning process. In order to remain competitive in a dynamic and uncertain environment, companies need to constantly monitor their learning process, properly evaluate and incentivize progress of teams, measure the outcome of trials, and establish best practices. There is great interest in using the learning curve as a foundation for addressing many of these concerns, but the inputs to the learning curve need to be better understood both conceptually and quantitatively [99, 155, 121, 94]. One important step toward this is better linking the learning curve with understandings of how organizations learn, codify, and implement knowledge in practice, and the challenges of doing this, which will be discussed in the next section [103, 99, 94].

2.3.2 Learning in organizations

Organizational learning is rooted in the learning of individuals but is more complex due to the multiple interacting levels of learners, including individuals, teams, and the overall organization [99]. Not only must individuals make inferences from observations, but knowledge must be shared through formal and informal means within the organization, and the capacity to take effective action must consequently be enabled [94]. As a result, the learning process is substantially different for organizations than for individuals, and tends to be "routine-based, history dependent, and target-oriented" because of the rigidity of these exchanges [103, 94].

There are generally two types of learning an organization can engage in. The first of these, *operational learning*, refers to a "know-how" for implementing procedures and making minor adjustments to routines, which tends to lead to short-term incremental improvements in efficiency [94, 121]. The second is *conceptual learning*, which instead addresses the "know-why" of a process and challenges prevailing ways of doing things in order to develop new frameworks [94]. Conceptual learning seeks to understand causes and effects using testable hypotheses and experimentation [121]. This type of learning can lead to design changes or breakthrough innovations, but in order to have a lasting effect it must be codified in some way, such as in a standard operating procedure (SOP) [103, 121].

Memory is tightly linked with learning, and takes the form of individual and shared mental models. In addition to SOPs and other documents, organizations may rely on informal shared understandings, particularly in settings (like oil and gas drilling) where complex uncertainties make it harder to formally document concepts [124, 103]. There is a risk with poorly defined SOPs of enshrining procedures in a way that inhibits future changes [94]. Further work is needed to understand the way that mental models, such as SOPs, influence and are influenced by the learning curve [94, 99].

Another key component of learning is experimentation, or "deliberate learning activities," but there is a need for further study to understand optimal strategies for this in different settings [99]. Experiments can be a source of great conceptual learning, but if they are not managed properly, they can actually be detrimental to learning [99]. Bohn and Lapre identified four types of experiments an organization can engage in [12]. *Controlled experiments* introduce deliberate changes on different groups, while *natural experiments* attempt to infer causation from variations in data under normal operations. *Ad hoc experiments*[11] use deliberate changes but without a control group or experimental design, leading to limitations in the ability to reach strong conclusions about findings, especially when there are confounding variables present. Finally, *evolutionary operation* is a hybrid approach between controlled and natural experiments, where very small changes are introduced that do not create necessarily different groups. This is common with continuous variables that should be optimized, such as the density of drilling mud used in a well. Instead of comparing effects between groups, statistical measures are used to infer effects and further changes are made to investigate in the more favorable direction of the parameter. In choosing an experimental approach, some important considerations are the speed and cost of each "learning cycle" (the time from deployment of a trial to measurable outcome, analysis, and procedural revision), the signal-to-noise ratio (the ratio of the unknown effect to the standard deviation of outcomes), the fidelity of experiments if they are simulations of real world conditions, and the value of information being investigated [99].

Frequently, because of the many points of exchange in an organization, learning may break down or be inhibited by aspects of the organization or environment. March and Olsen identified four incomplete learning cycles, shown in Figure 2-7, where limitations arise in an organizations ability to learn [107, 94]. In *role-constrained learning*, individuals are limited

[11] Based on numerous conversations with engineers and managers at operating companies, this type of experimentation is very common in unconventional drilling.

by constraints on their roles and as a result are unable to act upon learning. With *audience learning*, individuals are able to take action but are limited in their ability to impact the broader organization's actions. *Superstitious learning* occurs when the action taken as a result of learning does not have a clear effect on the environment. This occurs because of a "misattribution" of cause and effect [121]. Finally, *learning under ambiguity* means that actions have an effect on the environment but the causal connection is not clear. This may lead to the formulation of "myths," which are inaccurate beliefs based on interpretations of data [121]. Multiple myths may also engender conflict as individuals compete to assign accountability or attribution to outcomes, be they negative or positive [103]

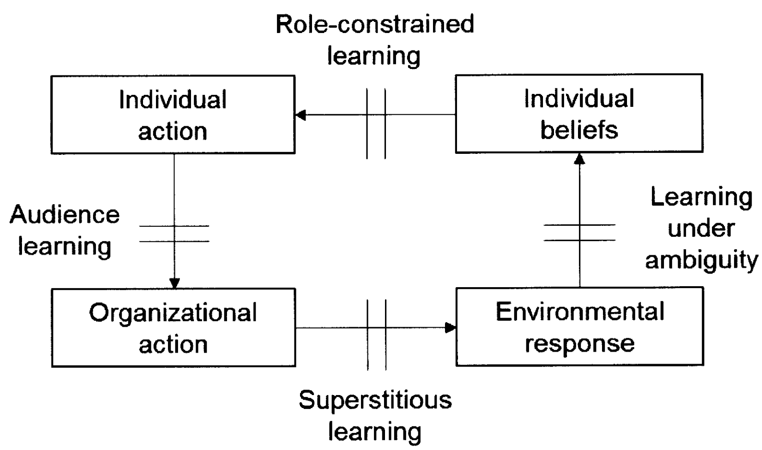


Figure 2-7: Model of organizational learning and incomplete learning cycles. Adapted from March and Olsen [107]

In addition to March and Olsen's identified incomplete learning cycles, Kim discusses three others that deal with insufficient utilization of mental models. Briefly, these are *situational learning*, where an individual improvises to solve a problem but does not codify or generalize the learning, *fragmented learning*, in which the link between individual and shared mental models is broken, and *opportunistic learning*, where shared mental models are intentionally bypassed to allow an individual or small group of individuals to implement their own mental model instead [94].

There are additional non-organizational considerations that make learning particularly difficult in the setting of unconventional oil and gas drilling. Learning is difficult in unstable environments because it is difficult to effectively implement changes or learn from ambiguous outcomes [121] This "incomplete technological knowledge" makes it impossible to have a full understanding of the effects of input variables on output [99]. With drilling, *detail complexity* means there are too many variables to comprehend the problem entirely, and *dynamic complexity* presents a particularly nasty hurdle to learning because the uncertainty of subsurface drilling conditions and the lengthy learning cycle for drilling a well makes it difficult to establish cause-and-effect [99].

2.3.3 Procedural and performance variability and learning

Despite these challenges, learning happens in drilling. In unconventionals, the rapid and repetitive approach to development requires an incessant drive toward *standardization* of well designs and drilling procedures in order to improve efficiency [34]. Additionally, operators engage in *continuous improvement* of techniques based on systematically testing procedures to determine those that lead to more optimal outcomes in a field [34]. The combination of standardization and continuous improvement drives progress along the learning curve. This development strategy, which has been referred to as a "flexible factory," relies on data "triggers" to inform cyclical well redesigns, striving toward greater efficiency of implementation through increasingly repetitive practices (from standardization) with better frameworks for understanding best practices (through continuous improvement) [53]. However, the success of this effort depends on the ability to understand and manage sources of operational and performance variability throughout this process, a dimension rarely appreciated within discussions of learning from experience in drilling [34].

Standardization of drilling procedures has been found to improve the efficiency of task execution and reduce variation in time spent on identical tasks [3]. It requires process definition, ensuring that shared mental models are established, and avoiding some learning pitfalls described in the previous section. It also engages learning at multiple levels, from planning to implementation [103]. For example, the directional drilling plan in a horizontal well consists of an equipment selection and a strategy for combining sliding and rotating drilling to achieve the desired curvature. During execution, the directional driller is guided by this plan but may slightly alter the approach as needed in the field. If the strategy of the plan is sound, over time the directional driller will become more proficient at executing it in the given field, resulting in more standardized combinations of sliding and rotating. If variations introduced by the directional driller improve the plan, then they may be incorporated formally, or informally into future wells, also leading to more standardized drilling. Thus, there are both top-down and bottom-up sources of standardization and it is important to consider not only standardization of designs and plans, but the standardization of actual field practices. It is important to understand the role that standardization plays in learning in order to inform planning of drilling campaigns and the design of proper incentives for those driving improvement [63]. Detection and description are the first steps related to this effort.

Continuous improvement on the other hand is a form of experimentation that seeks to introduce and evaluate new drilling approaches to a field in a structured way [34]. These changes may be evaluated in absolute terms, or relative to the learning curve, as Zangwill and Kantor propose [155]. It is important to consider both of these measures, because of the obvious economic implications associated with absolute performance improvements, but also the improved ability to manage and disseminate the learning process by considering the impact of variables on the slope of the learning curve. However, due to natural variability introduced between cause and effect, and noise associated with measuring performance, it is important to statistically assess the impact of a continuous improvement [82]. Learning curves are useful and important in drilling campaigns but the practical issue of scatter, or noise, within the learning curve drives the need for probabilistic methods and better characterization of performance variability [12, 99, 69, 103, 82].

2.4 Benefits to characterizing the nature of performance variability

A critical development in the understanding and management of shale gas and tight oil resources is the characterization of performance variability, both in well productivity and well drilling operations. Here, I will summarize some of the benefits to doing this.

There is a growing awareness of a high level of productivity variation both between and within areas of unconventional resource plays, due to immense heterogeneity of reservoir properties. The extent and nature of this variation is not well understood but analysis of production data may help to define this better, as this thesis aims to address [116]. At this time, the wide range of well productivity, even within areas of seemingly similar geology, makes it difficult to reliably predict individual well performance in advance [65]. Characterization of this variability as a distribution of outcomes will help to improve economic forecasting and development planning in fields and can also underlie any reserve analysis [142]. At the same time, it is important (especially within an operating company) to use a consistent approach, such as a probability model, to estimate the chance of success, in order to accurately compare opportunities and investments [131]. According to McGlade et al.:

> "[A] key weakness of existing resource estimates is the absence of a rigorous approach to handling uncertainty... few studies provide thorough analysis of the sources and consequences of uncertainty or present their results *in the form of a probability distribution*[12]" [116]

The appraisal and development phases overlap more in unconventionals than in conventionals, and it is important to incorporate quantification of risk into the early stages of field development. This will give a more complete economic picture throughout development, as more is learned about a field. It is also important to move away from static forecasts toward forecasts that incorporate uncertainty so that *value of information* strategies can be employed [20].

Furthermore, characterization of production variability will support future efforts to identify drilling, completion, and geological sources of variability in order to advance knowledge of unconventional reservoir behavior and operational proficiency. Variability may be a salient feature and should be incorporated into theories if this is the case.

With regards to learning, improving knowledge about drilling performance variability will contribute both in unconventional drilling applications and also more broadly to the fields of organizational learning and learning curve theory. There have been attempts to address variability in learning rates [82, 99], but little work has been done to address variability in

[12]Italics added by me for emphasis.

the system itself, which makes it difficult to assess the progress of an organization and make decisions accordingly [155]. The method for evaluating continuous improvement, presented by Zangwill and Kantor, is very limited in application because it fails to take into account error present in the observations relative to the theoretical learning curve [155]. Characterizing variability will provide a statistical basis for more effective evaluation of experimentation.

Many of the development strategies currently employed in unconventional well development may be inappropriate or may differ in practice from theory [20]. It is not clear at this point that a low-cost manufacturing approach is the best strategy for development campaigns, although lessons may be drawn from this strategy, such as the benefits of standardization as opposed to one-off "craft" wells typical of conventionals [55, 34, 33]. To identify new development strategies and measure the efficacy of existing ones, additional measures must be developed to understand facets of learning beyond the simple metrics of time, cost, or number of defects [3, 99]. As with the learning curve, these additional performance metrics can be used to compare different teams, establish best practices, and design better incentives [3].

Finally, there is great interest in advancing data-driven models and learning in oil and gas development, and characterizing variability is an important part of this [71]. Dramatic expansion in data acquisition, transmission, and storage capabilities over the past few years has enabled oil and gas operating companies to collect unprecedented amounts of data from drilling and other operations. These large and diverse structured and unstructured datasets are leading to an increased "digitization" of the oilfield [51]. This data-rich paradigm is creating both opportunities and challenges in terms of the effective utilization of this data in enhancing field operations. Unfortunately, for many operators, this new data abundance has not led to operational gains owing to "data paralysis," a situation where the available data is not utilized to the fullest extent possible owing to insufficient analytics expertise and an inability to engage in sophisticated data-driven learning [133].

There is a great opportunity to improve forecasting abilities, decision-making and development planning, and promote more effective learning by making use of abundant existing field data from sensors and reports. Automated data analysis techniques are essential for effective learning from plentiful feedback because they allow the incorporation of greater quantities of useful data than can be achieved manually, they can identify latent trends, and they can automate parts of the learning process; all of which leads to more useful real-time guidance for the planning of future wells [19]. Obtaining data is not sufficient—it is necessary to tightly manage the process of experimentation and close the learning feedback loop by correctly interpreting effects and encoding inferences in shared memory models [130]. Understanding the patterns of variability in both drilling and production will provide a rigorous testing framework for experimentation with drilling and completion parameters and is a critical step toward the optimization, automation, and application of machine learning in field practices.

Chapter 3

Characterizing production performance variability

The variability of early-life well productivity is a major driver of economic uncertainty in a shale gas or tight oil field development campaign. The peak rate of production from wells provides a suitable proxy for early life productivity and is a useful metric for statistical assessment and development of forecasting tools.

In this chapter, I characterize the nature of production performance variability in major shale gas and tight oil fields. This analysis includes the Barnett, Marcellus, Bakken, Eagle Ford, and Haynesville, but I place the Barnett at the forefront of the study because of the abundant drilling of wells in that play and availability of data across a longer timeframe. In Section 3.1, I describe the production history dataset used for the study, including information about how it was processed and prepared for the analysis. I introduce the graphical method of probability plots and use them to evaluate the shape of the distribution of well peak production rates in Section 3.2. Using the distributional assumption I develop here, distribution parameters for a region, such as μ, can be estimated using Bayesian statistical inference with a known distribution family. I outline the procedure for this and then demonstrate the utility of this tool in Section 3.3.

3.1 Description of data and processing

3.1.1 Production data and initial processing

The production data I have used in this analysis was accessed on July 3, 2014 from the drillinginfo HPDI online database of US oil and gas production data [73]. This is a service that aggregates production data for oil and gas wells in 33 US states, drawing on publicly available repositories managed by the respective states. The information in these databases has been reported by oil and gas operating companies according to state requirements. There are challenges associated with utilizing well production databases like HPDI because the requirements for the categories and quality of data reported vary from state to state. Monthly production rates and perforation locations were reporting requirements in all the states analyzed though, and formed the basis of this study.

It was necessary to clean the dataset prior to analysis. The raw HPDI dataset contained some erroneous "junk" wells with obvious misreporting (i.e. values outside the range of reasonable values for some or all fields), which were thus excluded from the analysis. I used a set of data filters to obtain a cleaned dataset of active wells producing from the plays of interest. The criteria used for filtering, and the number of remaining valid wells in each play, is summarized in Appendix A in Table A.1.

The plays included in this analysis—Barnett, Marcellus, Bakken, Eagle Ford, and Haynesville—are major US shale plays with high levels of drilling activity in recent years, providing abundant production datasets [14, 26, 30, 39, 48, 56, 66, 120, 127, 140, 151]. They also have been developed similarly enough to warrant comparison, yet have known geological differences. They include a range of reservoir fluid types, from the dry gas of Haynesville to the liquids-rich gas condensate of the Eagle Ford[1] and the black oil of the Bakken [73, 114]. Some descriptive characteristics of these plays are summarized in Appendix A in Table A.2.

[1] I analyzed the gas production rates from Eagle Ford to avoid complicating the analysis with consideration of varying gas-oil-ratios.

3.1.2 Transformation and normalization of production data

As discussed in Chapter 2, there are theoretical and historical reasons to expect that well productivity may be lognormally distributed. The probability density function of a lognormal distribution is shown in Equation 3.1.

$$f_X(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, x > 0 \tag{3.1}$$

If the lognormal distribution in Equation 3.1 is used to describe the distribution of peak production rates in a well ensemble, x becomes the well's peak production rate, μ becomes the arithmetic mean of the log-transformed well ensemble peak production rates, and σ is the standard deviation of the log-transformed well ensemble peak production rates in the well ensemble. Put another way, if each value in a lognormal distribution has the natural logarithm taken the resulting distribution is normal (Gaussian) with mean, μ, and standard deviation, σ. The parameters μ and σ are also sometimes referred to as location and scale parameters respectively. It is worth noting here that in lognormal distributions, μ also corresponds to the natural logarithm of the median value. For a thorough examination of the lognormal distribution and all of its peculiarities, see the classic monograph on the subject from Aitchison and Brown [2].

It is important to use an appropriate system of normalization when searching for common patterns within large oil and gas datasets [71]. In order to compare the productivity distribution shape for different well ensembles, the production rates must be normalized within well ensembles. This allows for comparison of the shape of the distribution separately from central tendency and spread of data, which may vary between well ensembles. To accomplish this, I took the log-transformation of the peak production rate for each well and then calculated the "standard score" for each well relative to the distribution of wells in the ensemble under consideration[2] [93]. The standard score, z, is calculated using Equation 3.2.

$$z = \frac{\ln x - \mu}{\sigma} \tag{3.2}$$

[2]This means I am effectively assuming a log-normal distribution and the analysis in the following section evaluates the validity of this model.

In addition to the absolute peak production rate, I also consider the specific peak production rate, which normalizes production to the length of the productive section of the well. This is an important metric because many companies have been moving toward longer lateral sections and more stages of hydraulic fractures to increase absolute production, potentially influencing patterns found in the data [59]. In order to calculate the specific peak production rate I use Equation 3.3.

$$Q_{spec.,peak} = \frac{Q_{peak}}{D_{perf.,lower} - D_{perf.,upper}} \tag{3.3}$$

$Q_{spec.,peak}$ is the monthly specific production rate in the peak month, Q_{peak} is the (absolute) monthly production rate, $D_{perf.,lower}$ is the measured depth of the lowest perforation in the well, and $D_{perf.,upper}$ is the measured depth of the highest perforation in the well. The denominator in this expression represents the productive length of the well that has been perforated.

3.2 Evaluating the shape of the productivity distribution with probability plots

In this section, I assess the validity of the distributional assumption that well productivity in shale gas and tight oil is lognormal. There are four reasons for making distributional assumptions, as identified by Chambers et al. [22]:

1. Allow for compact description of data—mean and standard deviation are adequate parameters to describe normal or lognormal distributions
2. Enable useful statistical procedures (e.g. analysis-of-variance)
3. Characterize sampling distributions for parameters and make statistical inferences from data
4. Improve understanding of physical mechanisms generating data

All of these reasons apply to the case of a lognormal distributional assumption for well productivity. Simple, yet accurate, descriptions of well productivity (enabled by reason 1) will lead to better resource assessments. Statistical procedures like analysis-of-variance for lognormal distributions (as in reason 2) can be used in the future to evaluate different completion (hydraulic fracturing) techniques for effectiveness. Making statistical inferences (from reason 3) is the focus of Section 3.3, and enables operators to assess the productivity of resource areas early on. It is my hope that the findings of this study will help guide and inform fundamental research into understanding the complex physical processes governing shale gas and tight oil production (leading to future realizations of reason 4).

In the remainder of this section, I introduce probability plots, the technique used to evaluate the distributional assumption of lognormality, and present results to demonstrate the general validity of this distributional assumption.

3.2.1 Probability plots

In order to evaluate distribution shape I use the graphical method of probability plots, also called theoretical quantile-quantile plots. This technique allows for a graphical comparison of a univariate empirical distribution with an ideal distribution. This test is invariant to location and spread, allowing it to be used to test for an entire family of distributions [22]. This property also means that the transformation in 3.2 has no impact on the shape of the resulting plot, except that in order to test for lognormality of the original data, we need to test the z values for normality. The transformation does enable us to use a common scale for different well ensembles and plot them on the same graph for convenient comparison.

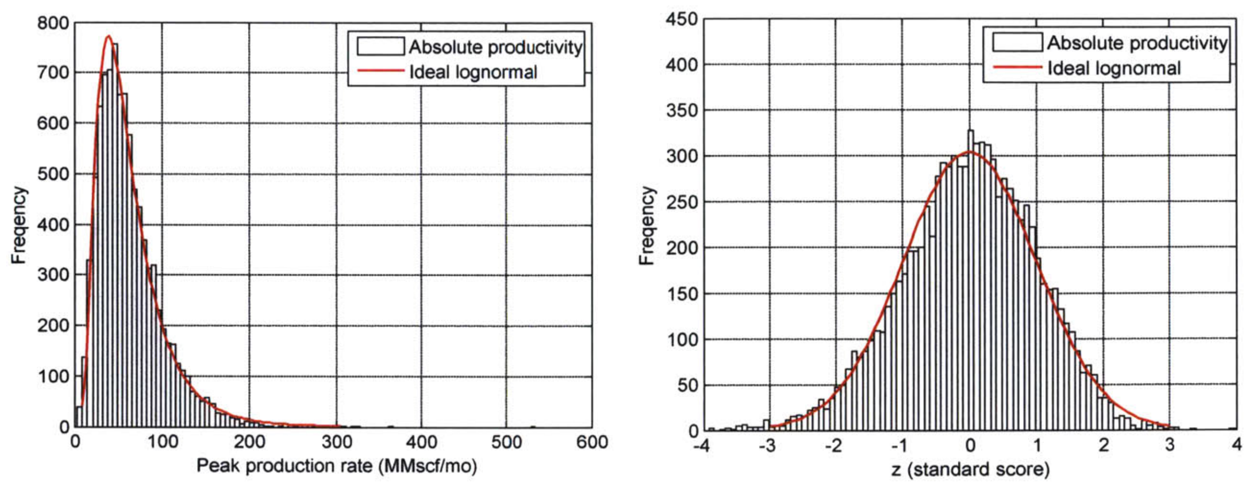
A probability plot is constructed by first ordering data by increasing value and determining the empirical quantile of each data point. For example, if there are $n = 100$ data points, the lowest value has a quantile of $p = 0.01$. The empirical quantile of each data point is then matched to the value that would occur at the same quantile in an ideal distribution of the same size as the data. The empirical values are then plotted against the theoretical values. [22]

In this thesis, the y-axis corresponds to the theoretical distribution and the x-axis corresponds to the data[3]. For ease of comparison, the plots have been set to a standard horizontal scale of $z = -3$ to $z = 3$. This pushed some left-hand extreme values off the graph but allows for easier visual inspection over a suitably wide-range of the data. I did all of the probability plotting in MATLAB, using the "probplot" function with the default midpoint probability plotting positions [110].

The interpretation of probability plots is extensively explored in literature, and I will provide only a brief discussion of the basics here [22, 58, 118]. A good fit of data with a distribution is indicated by the straightness of the data points in the plot. Often, the line is shifted or multiplied by a constant to allow it to be compared to a diagonal line. This is done automatically in MATLAB. In addition to checking for goodness of fit, much can be learned from any systematic departures from the line. Often, outliers exist at either end of the data. Additionally, there tends to be greater variation in tails for distributions that have density that gradually tapers to zero in the tails (such as normal or lognormal distributions) [22]. Defined and systematic curvature at the ends may indicate longer or shorter tails in the data than the ideal distribution. With the axis selection I have chosen, curvature upward at the left tail or downward at the right tail indicate longer tails at those ends of the distribution [22]. The opposite orientations indicate shorter tails at either end. Asymmetry can also be identified. Convexity of the plotted data indicates that the empirical distribution is more left-skewed than the ideal distribution (and contrarily, concavity indicates right-skewness) [22]. Additionally, gaps and plateaus may tell something about data rounding or acquisition issues [22].

It is important to remember that, in the real world, no truly normal or lognormal distribution of data exists. Measurement error, physical and practical limitations on data generation and collection, along with the discrete nature of sampling preclude such a possibility [22]. We are really interested in evaluating whether a distributional assumption is adequate for the purposes of describing data and using statistical tools with the data. Probability plots are thus more appropriate for testing distributional assumptions with real data than frequentist

[3]It is important to be clear about which axis is the data and which is the theoretical distribution because interpretations of any deviations have opposite meaning depending on the orientation of the plot. There is not a standard axis selection and the different scales and units that might be used for each axis can lead to ambiguity.



(a) Before normalization

(b) After normalization

Figure 3-1: Histograms showing distribution of well productivity in Barnett before and after normalization to standard score. MMscf/mo is million standard cubic feet per month.
[Summary statistics for specific productivity of counties in Barnett (fig 3-4b).]Summary statistics for specific productivity of counties in Barnett (fig 3-4b). Units of mscf/mo/ft are 1000 standard cubic feet per month per foot

hypothesis tests, which estimate the probability that differences between the data and ideal distribution could have been caused by sampling error alone [58, 71, 93, 118]. For large, real datasets—like the ones used in this thesis— frequentist hypothesis testing is also unreliable because the large sample size makes the measurable threshold for statistical significance too small [118].

3.2.2 Lognormality of productivity

Before applying probability plots to the production history data, it is worth plotting the distribution of productivity as a histogram for illustrative purposes. Figure 3-1 shows histograms for well productivity in the Barnett Shale play, the most extensively developed (and thus sampled) unconventional resource play, both before and after the transformation and normalization described in Section 3.1.2.

It can be seen from the histograms in Figure 3-1 that the distribution of well productivity is right-skewed and resembles a lognormal distribution. To more reliably test this distributional assumption though, we turn to probability plots.

The probability plot in Figure 3-2 demonstrates the lognormality of well productivity across the Barnett play for the entire population of over 9,000 wells. The adherence to this distribution is remarkable for both absolute productivity and specific productivity, the length normalized metric defined in Section 3.1.2.

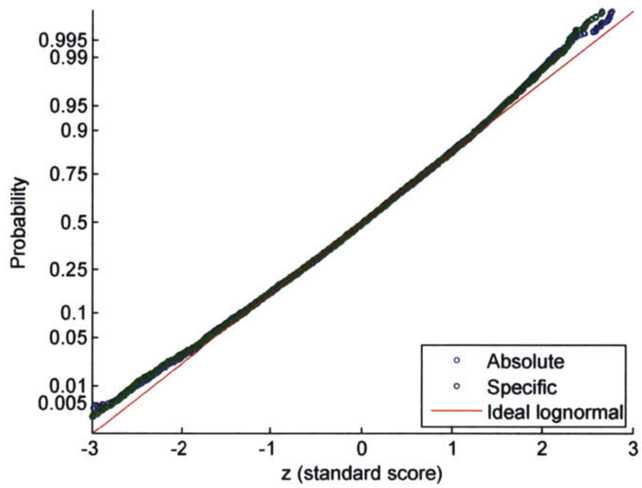


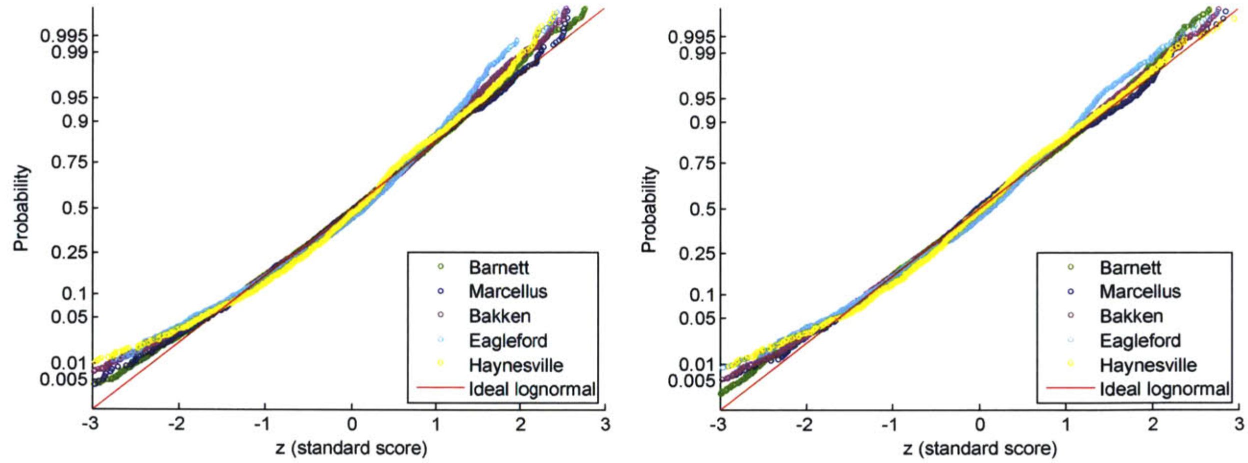
Figure 3-2: Probability plot comparing absolute and specific productivity in the Barnett to an ideal lognormal distribution.

This characterization also applies to the other resource plays in the dataset. Figure 3-3 shows probability plots of major U.S. unconventional plays, including Barnett, Marcellus, Bakken, Eagle Ford, and Haynesville. Across the well ensembles of all of these plays, the empirical productivity distribution shape can consistently be characterized as lognormal, especially between the fifth and ninety-fifth quantile. The curvature at the ends will be discussed in Section 3.2.3.

Interestingly, because this characterization of the distribution shape is invariant to location and scale parameters, it also applies at multiple spatial resolutions within plays, despite known differences between areas. First, an examination of counties within the Barnett play reveals that the lognormal distribution assumption holds when the play is broken into its constituent counties (Figure 3-4). The same distribution shape can be seen both in core areas, such as Johnson county, and lower performing non-core areas, like Parker county.

Although core counties have higher median well productivity, the *absolute variability*—measured as the P90-P10 ratio[4]—is actually extraordinarily consistent within all of the Barnett counties, as can be seen in Table 3.1. The same is generally true for absolute variability of

[4]This is a ratio of the 90^{th} percentile value over the 10^{th} percentile value.



(a) Absolute productivity

(b) Specific productivity

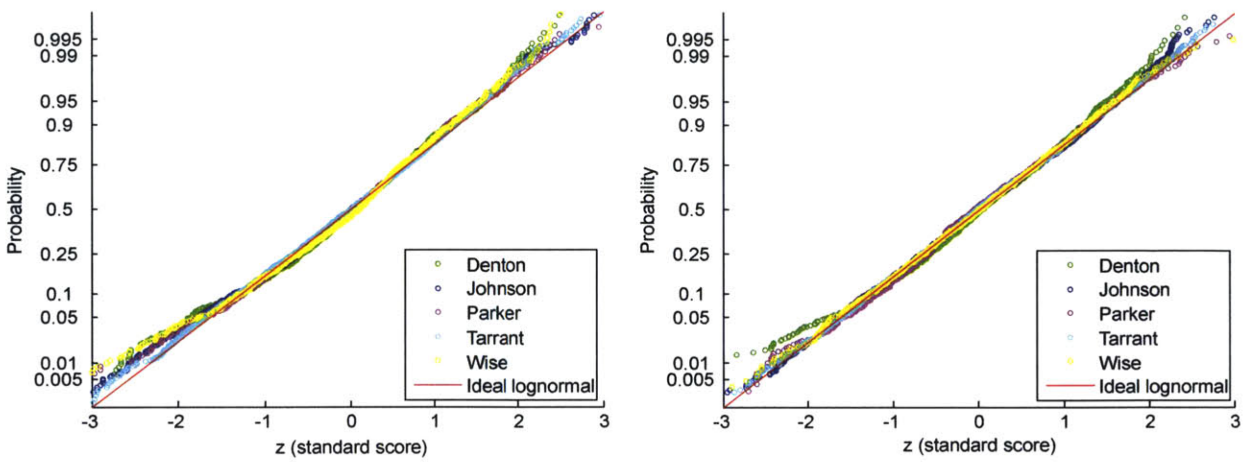
Figure 3-3: Probability plots comparing absolute and specific productivity of all plays to an ideal lognormal distribution.

counties in other unconventional plays, as shown in Appendix A in Table A.3.

Table 3.1: Summary statistics for specific productivity of counties in Barnett (fig 3-4b). mscf/mo/ft is thousand standard cubic feet per month per foot.

County	P90-P10	P50 (mscf/mo/ft)	Mean (mscf/mo/ft)
Denton	3.38	19.63	21.51
Johnson	3.95	23.56	27.25
Parker	3.22	13.93	16.07
Tarrant	3.37	25.90	29.29
Wise	3.92	13.91	16.00

This characterization can be carried to still finer spatial resolutions than the county-level. In order to do this, I develop artificial spatial divisions within Johnson county in the Barnett to spatially group wells together. I divide the county into uniformly-dimensionsed "blocks" of ten square miles and then use the latitudinal and longitudinal coordinates for each well in that county to determine the block that it sits in. I only include in my analysis the blocks that have enough wells in them to reasonably be examined for their productivity distribution (I use an arbitrary cutoff of 16 or more wells for this). The wells in each of these blocks is treated as a separate well ensemble that can be analyzed with a probability plot, as in Figure 3-5. The limited sampling of these smaller well ensembles makes it unsurprising that the consistency of the distribution shape from one well ensemble to another is less



(a) Absolute productivity (b) Specific productivity

Figure 3-4: Probability plots comparing absolute and specific productivity of well ensembles in different Barnett counties to an ideal lognormal distribution.

than those at larger spatial resolutions. However, viewed collectively, the symmetry of the noise about the 45° line suggests that this characterization still captures the core behavior of productivity variability in shale gas and tight oil at these smaller geographical scales. It is certainly an adequate characterization for practical purposes of forecasting, as I will demonstrate in Section 3.3. Further investigation of this phenomenon will be necessary to understand the relationship between this noise and the spatial scale.

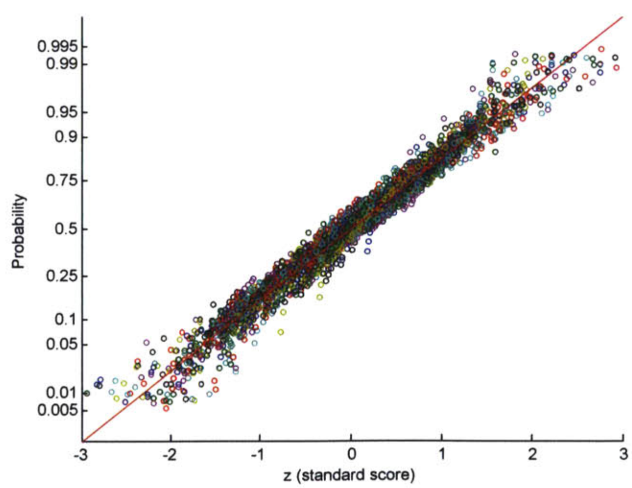
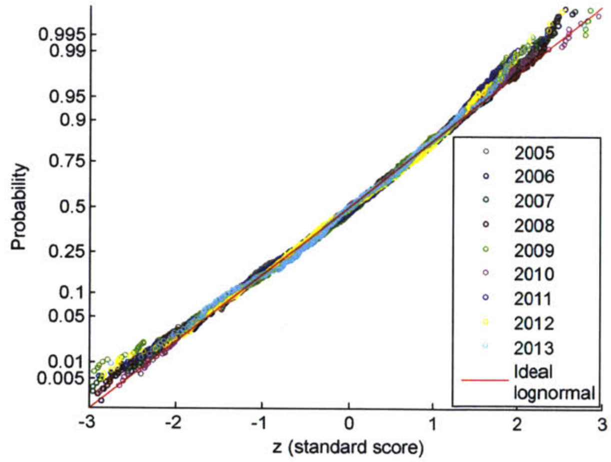
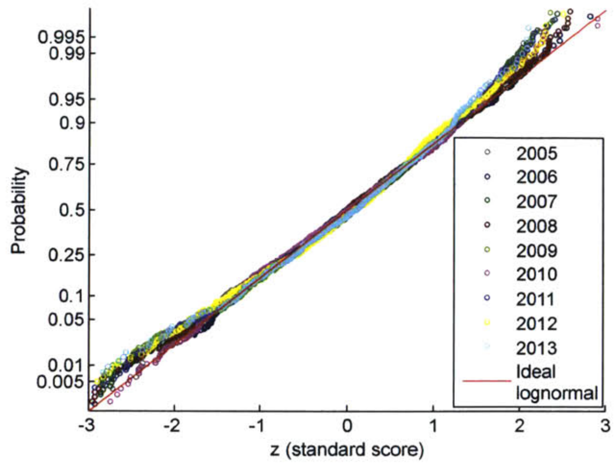


Figure 3-5: Probability plot comparing specific productivity of well ensembles in ten square mile blocks with 16 or more wells within Johnson county of the Barnett to an ideal lognormal distribution.

How reliable is this distribution assumption over time with potentially evolving technol-



(a) Absolute productivity



(b) Specific productivity

Figure 3-6: Probability plots comparing absolute and specific productivity of different vintages in the Barnett to an ideal lognormal distribution.

ogy? To answer this, we look at some additional sub-groupings of well ensembles in plays. First, we consider each vintage, or development year, as a well ensemble. The probability plots examining different vintages in the Barnett are shown in Figure 3-6.

In addition to the consistency of the distribution shape, it is worth noting that the absolute variability has remained fairly consistent over this period of development for both absolute and specific productivity as well, as shown in Table 3.2. The same trend of persisting, wide variability of well productivity for other shale gas and tight oil plays in this analysis can be seen in Appendix A in Table A.4.

Table 3.2: Spread of productivity by vintage in Barnett (Figure 3-6).

Vintage	Absolute productivity P90-P10	Specific productivity P90-P10
2005	4.48	5.08
2006	4.59	4.49
2007	4.33	4.58
2008	3.90	4.43
2009	4.06	3.97
2010	4.00	4.66
2011	4.81	4.46
2012	4.14	3.88
2013	5.22	4.26

Furthermore, categorizing wells by the length of the perforated section of well yields the same distribution shape. In general, wells have been increasing in length (and number of frack stages) over time, making this a relevant consideration for understanding how technology evolution might impact the shape of the productivity distribution. Figure 3-7 shows the results of this categorization for absolute productivity (specific productivity is excluded due to the redundancy of both categorizing and normalizing based on length).

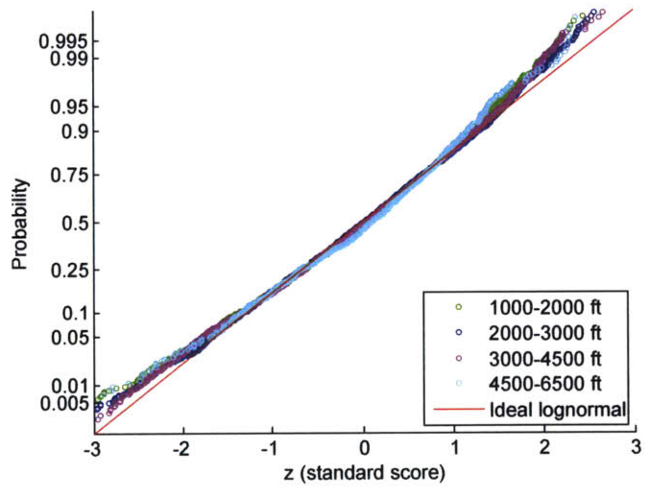
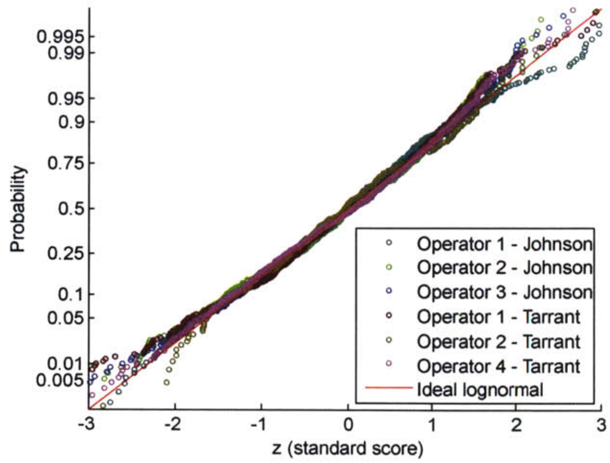


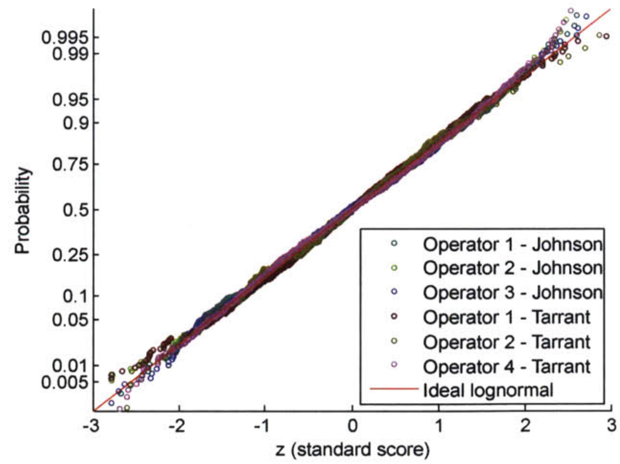
Figure 3-7: Probability plot comparing absolute productivity for Barnett well ensembles categorized by perforation length to an ideal lognormal distribution.

Finally, if this distributional assumption is to be used by individual operating companies, it is worth considering the distribution for a specific company's portfolio of wells in a given area. Figure 3-8 shows the distribution of well productivity for specific county-level individual company well portfolios in the Barnett. These samples represent the largest individual company well portfolios in the Barnett.

The figures presented in this section provide a narrative that demonstrates the lognormality of unconventional well productivity. The Barnett shale gas play has been the focus here because the broadness of the dataset facilitates division into many sub-categories of wells. Many additional figures are included in Appendix B (Figures B-1 to B-19) to reinforce this characterization with similar analysis for other plays, as well as for an alternate metric to the peak production rate—the first twelve months of production.



(a) Absolute productivity



(b) Specific productivity

Figure 3-8: Probability plots comparing absolute and specific productivity of different operating company well portfolios in Barnett counties to an ideal lognormal distribution.

3.2.3 Discussion of deviations at the ends

It is worth considering some potential explanations for the deviations from lognormality at the extremes of the data. The adherence to a lognormal distribution was found to be generally strong between the fifth and ninety-fifth percentile. Outside this range, the right tail is thinner than lognormal and the left tail is heavier than lognormal. There are three categories of explanations for these deviations that I will address:

1. The pattern of productivity is better described by a different distribution.
2. Lognormality is only applicable to non-extreme samples; geological differences lead to different behavior.
3. The extremes of the data are influenced by differences in sampling.

Given the consistency of the deviations, we admit that a distribution with a similar shape to the lognormal distribution but with slightly different tail behavior might describe the data better. One such possibility is the gamma distribution, which has a smaller right tail than the lognormal distribution [23, 135]. However, a probability plot comparing the normalized absolute productivity in the Barnett to the ideal gamma distribution in Figure 3-9 shows that the gamma distribution actually overcompensates for tail behavior and does not provide an obviously better model. Whether other distributions provide suitable models for productivity in unconventionals is a topic for further consideration elsewhere. My purpose here is to provide a generally descriptive and useful model for the productivity distribution shape, for which the lognormal seems a good candidate. Additionally, the analytical tractability of the lognormal probability density function, and the numerous statistical methods useful with normal (or lognormal) distributions make a strong case for sticking with this characterization[5].

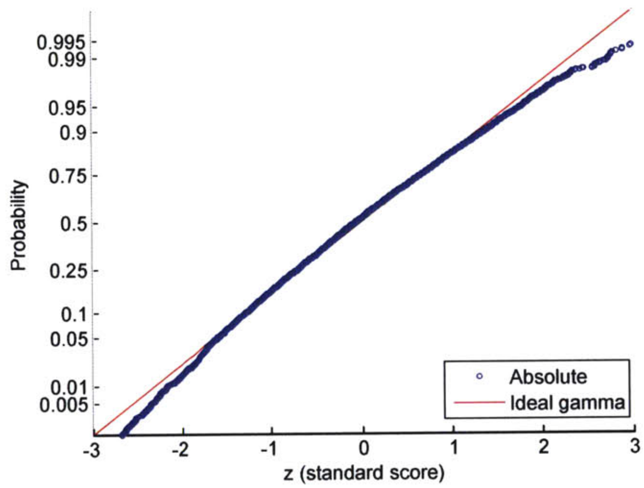


Figure 3-9: Probability plot comparing absolute productivity for Barnett wells to an ideal gamma distribution.

In the second explanation, the lognormal distribution is applicable to the bulk of the data but at the extremes, different natural processes lead to behavior deviating from the general pattern. A lognormal distribution is indicative of numerous "errors" being multiplied together in a process [2]. In the extreme cases, certain outlier geological properties may dominate other factors in the system, overpowering the generally multiplicative nature of outcomes. Although shale gas and tight oil are often thought of as continuous resources with heterogeneously distributed properties, there may be "dry" spots with exceptionally low organic matter. If there is little to no organic matter, than this quality will dominate productivity and other heterogeneously varying qualities will have no discernible effect. If this was the case for the entire lateral section of the well, it would yield a "dry hole" and

[5]Eberhardt and Gilbert found that the use of a lognormal distribution for similar distributions, such as the gamma distribution, could be justified by convenience, as it provides equal levels of power and robustness regardless of the actual distributional origin of data [40].

would not be in our dataset. It could apply only to a portion of a well's lateral length though, leading to much lower than expected output. Unfortunately, given the lack of geological data for this production dataset, I am limited to this type of conjecture.

The third and most intriguing explanation for deviations is that "human factors" are behind it, with sampling differences influencing results in some way [150]. This has been an important consideration in past statistical evaluations of conventional field size distributions. To begin with, a process of economic truncation was identified, in which the smallest fields might not be reported, leading to a thinning of the left tail [137]. Additionally, the sampling of conventional oil drilling was not random and there was a tendency for larger fields to be targeted initially, also known as proportional sampling, size bias, or "creaming" [92, 117].

However, when it comes to deviations it is difficult to reconcile what is observed in our production data with the analogue of conventional field size distributions. These past explanations were not only dealing with a petroleum resource of a very different nature, the deviations they sought to explain were in fact different. The deviations observed in conventional field size were generally the opposite of ours, with a thin tail on the left and a heavy tail on the right (relative to a lognormal distribution) [92, 4]. In spite of the censoring out of inactive wells in our dataset, there remains a greater abundance of low-performing wells than an ideal lognormal distribution would imply. Economic truncation may be a more complex issue in unconventionals, leading to different behavior than with conventionals but it is difficult to say at this point.

With regard to size bias, this takes a slightly different form in unconventionals. There is a definite targeting of the best acreage first, as can be seen by the general decline in specific productivity over the different vintages of wells in the Barnett (fig 3-10). However, as Figure 3-6 shows, the shape of the distribution in the different vintages does not change. This means that, relative to conventionals, there is a relative inability for operators to target the most productive wells first without also sampling the lower end of the distribution in an area. This implies a very different nature of risk with each well drilled than in conventionals, which is further discussed in Chapter 5. Additionally, past concerns about the influence of size bias on statistical inference methods are less relevant with shale plays than with conventional field sizes [92].

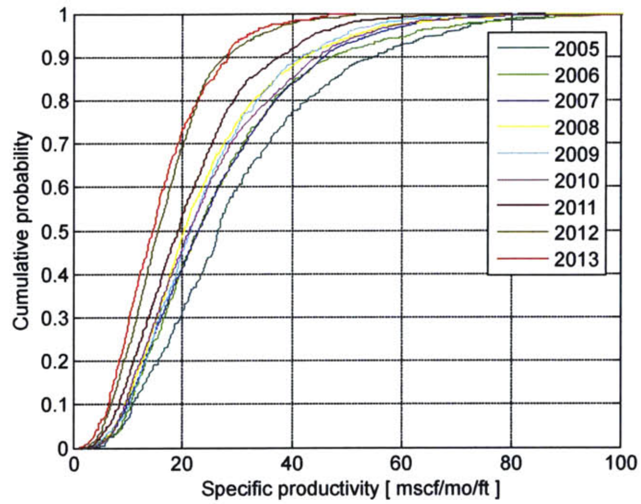


Figure 3-10: Cumulative distribution function plot showing a general decline in specific productivity by vintage in the Barnett.

There are additional potential paths leading to differences in sampling at the extremes, such as the choice of drilling locations. Individual well production may be reduced when wells are drilled in close proximity to each other and share the same pressure drawdown area, leading to interference. Ikonnikova et al. estimated that in the Barnett, around 15% of wells had overlapping drainage areas and might suffer from interference effects [78]. At the upper end of our empirical distribution, higher performing wells tend to slightly underperform what would be in line with an ideal lognormal distribution and, due to the tendency for high density development in productive "sweet-spot" areas, interference may be a reasonable explanation. Figure 3-11 shows wells drilled in the Barnett, with the top fifth and bottom fifth percentile identified in red and blue, respectively. The top fifth percentile of wells tend to occur in more productive areas that are heavily drilled.

This tendency can be expressed more precisely by calculating the number of wells drilled within a radius of 0.564 miles[6] of each well (which creates a circle of one square mile around the well). On average, this is notably higher for the most productive wells, leading to greater potential for interference with these wells[7]. Table 3.3 shows some summary statistics that

[6] I calculate this by assuming 1° latitude=69 mi. and 1° longitude=58.5 mi. (for 32.5° N and 97.5° W).

[7] I am ignoring an important temporal aspect of this here. Interference can only weaken the peak production rate of a well if the pressure in the area has already been drawn down by nearby wells drilled earlier. Analysis of the timing of the start of production for nearby wells could provide further insight but this is beyond the scope of my discussion here.

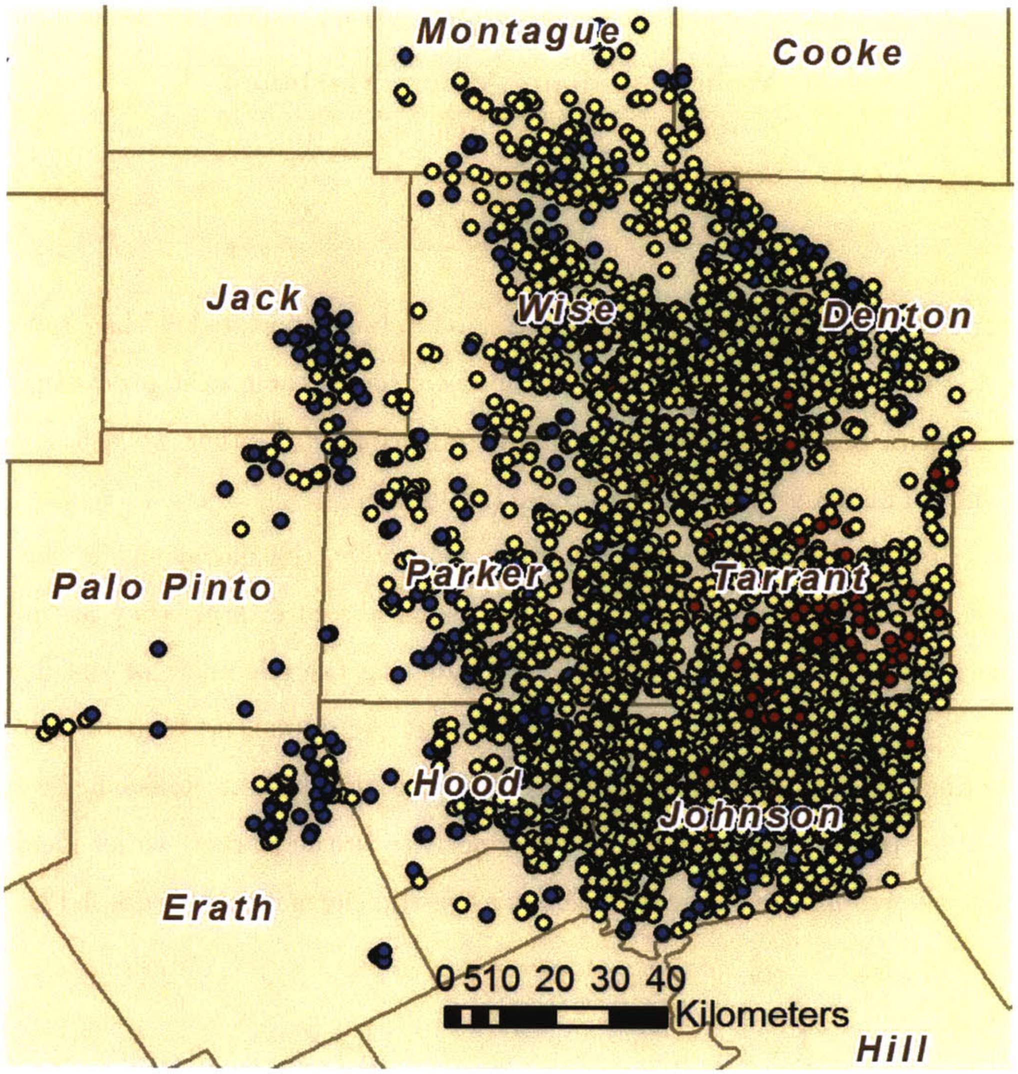


Figure 3-11: Barnett shale play wells. Red indicates peak production in the top fifth percentile. Blue indicates peak production in the bottom fifth percentile.

capture this dynamic.

To put these numbers in context, Iknonnikova et al. estimated that 10–15 wells within a square mile area was the maximum density of wells that could be drilled in the Barnett without interference effects. Although this description of well proximity is not proof of interference , it suggests that the circumstances for a number of more productive wells to have their peak production lowered by interference may be present, at least in the Barnett. This effect may partially explain a thinning of the right tail of the distribution.

Another possibility that could lead to sampling differences at both extremes, is choking back of wells either intentionally or unintentionally. Wells produce into shared pipelines and

Table 3.3: Average number of nearby wells (within 0.564 mi. radius) for wells in the Barnett.

Wells	Median	Mean	Maximum
<P5	1	2.4	16
Overall	5	6.2	31
>P95	8	9.4	31

it may be the case that the strongest wells have to be choked back below their full production potential due to capacity constraints. On the other hand, the lowest producing wells have inadequate initial pressure to compete with production from other wells flowing into the same pipeline. This may inhibit their production as well.

It is worth noting that from an economic perspective, the deviations at the upper end of the distribution are of greater concern because in absolute terms they are much larger. Taking the log-transformation is helpful for evaluating the adequacy of the distributional assumption but it masks the fact that deviations at the upper end tend to be of a greater magnitude due to the positively-skewed nature of the distribution. This can be clearly seen in a probability plot comparing absolute productivity in the Barnett to an ideal lognormal distribution, shown on a linear scale without a log-transformation (Figure 3-12).

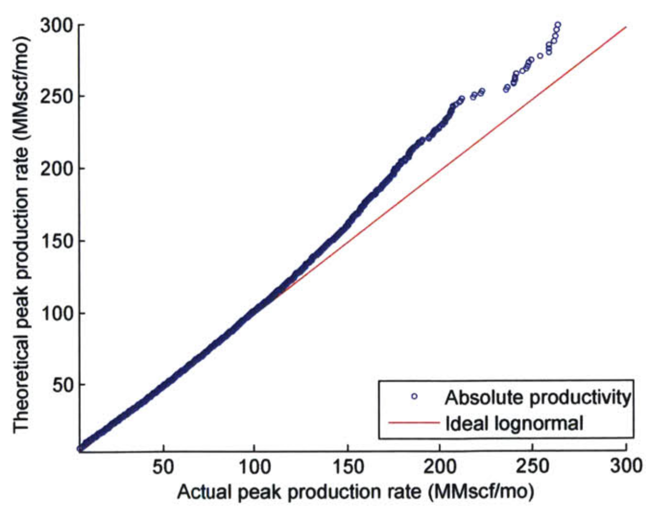


Figure 3-12: Probability plot showing absolute productivity of Barnett wells compared to an ideal lognormal distribution. Axes are linear scale of raw data without a log-transformation.

Despite the observed differences in tail behavior discussed in this section, I conclude that the lognormal distribution is a suitable distribution for describing productivity in shale gas and tight oil except at the most extreme values. This allows us to proceed to Bayesian statistical methods of inference.

3.3 Bayesian inference of productivity distribution parameters

3.3.1 Background

In general, statistical inference is a process that uses observations, or data, to provide information about unknown variables or an unknown model [10]. Bayesian statistical inference seeks to accomplish this using probability models fitted to the data, with results summarized as probability distributions for the model parameters and for predicted future observations [57]. It is based on Bayes' rule (Equation 3.4), with Θ representing the paramater(s) of interest and X as the observations [10].

$$f_{\Theta|X}(\theta|x) = \frac{f_{\Theta}(\theta) f_{X|\Theta}(x|\theta)}{\int f_{\Theta}(\theta') f_{X|\Theta}(x|\theta') d\theta'} \tag{3.4}$$

In Equation 3.4, $f_{\Theta}(\theta)$ is referred to as the *prior*, $f_{X|\Theta}(x|\theta)$ is the *likelihood*, the denominator is referred to either as the *evidence*, or the *marginal likelihood*, and the left term, $f_{\Theta|X}(\theta|x)$, is called the *posterior* [98].

The procedure for Bayesian analysis involves first fully defining the probability model for the quantities of the system. To update the knowledge about the system and incorporate new observations, this probability model is conditioned on the observed data to form the posterior distribution, which effectively combines the beliefs of the prior with the observed data. The marginal likelihood serves to normalize the posterior distribution. The fit of the model and implications of the posterior should always be evaluated and, if necessary, the process can be repeated with a different model or set of assumptions. [57]

Put simply, the posterior distribution is proportional to the prior distribution and the likelihood of the data being observed given this prior. After it is calculated, the posterior distribution can be reused as a prior distribution and the process repeated when new observations becomes available. This makes Bayesian analysis ideal for situations in which data is gathered sequentially, with some time lag (and potentially decisions) in between observations. [98]

Bayesian statistical inference is a powerful approach to statistical inference with some advantages over classical (frequentist) statistical inference, which views unknown quantities as deterministic rather than as random variables. The ability to clearly state and incorporate distributional assumptions and beliefs about parameters into a prior distribution makes Bayesian analysis both transparent and well-suited to situations where subjective beliefs can improve the accuracy of inferences. Alternatively, when little is known about the system in advance, or the data should "speak for itself," an uninformative prior can be selected that will be dominated by the observations in the posterior calculation. [10, 98]

Bayesian inference may yield similar or identical results to traditional approaches in certain, simple cases. However, it's generality and flexibility provide a framework that can be extended to more advanced approaches. For instance, Bayesian hierarchical modeling allows inferences to be made when multiple parameters have some dependence on each other. This is useful in situations where information about observations is available at different levels, such as reservoir or well parameters influencing production from a well. [57]

Conclusions drawn from Bayesian statistical inference are also more intuitive, and sometimes more robust, than those from classical approaches. For instance, a probability distribution for an unknown value provides more information and is less ambiguous than a confidence interval defined by rigid hypothesis testing [57]. These factors make it a good approach for estimating unknown parameters of well productivity distributions, understanding the uncertainty associated with the distributions, gauging the influence of data collected, and allowing for additional gathered information to be incorporated through prior distributions and more advanced methods like hierarchical modeling.

3.3.2 Estimating the mean of a normal distribution

As it turns out, Bayesian statistical inference builds quite nicely on our distributional assumption that well productivity in a field is lognormally distributed. By taking the natural logarithm of productivity, we can transform our distribution into a normal distribution. An observation x thus has the likelihood defined by the Gaussian probability density function in Equation 3.5 [98].

$$p(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{3.5}$$

I focus here on the problem of estimating μ when σ is known. Estimating both μ and σ together requires multiple parameter Bayesian inference, an exercise worthy of development but not necessary to go into for this thesis. One convenient way to approach this, is to choose the conjugate prior for the Gaussian likelihood function in Equation 3.5. A conjugate prior will combine with the likelihood function to yield a posterior distribution of the same family as the prior. Because the product of two normal distributions is also normal, we can thus choose a normal distribution for the prior distribution on μ. [98]

We define the prior distribution on μ to be normal with mean M_μ and standard deviation S_μ. Following the derivation presented by Kruschke [98], the likelihood times the prior is:

$$\begin{aligned}
p(x|\mu,\sigma)p(\mu) &\propto \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right)\exp\left(-\frac{1}{2}\frac{(\mu-M_\mu)^2}{S_\mu^2}\right) \\
&= \exp\left(-\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2}+\frac{(\mu-M_\mu)^2}{S_\mu^2}\right]\right) \\
&= \exp\left(-\frac{1}{2}\left[\frac{S_\mu^2(x-\mu)^2+\sigma^2(\mu-M_\mu)^2}{\sigma^2 S_\mu^2}\right]\right) \\
&= \exp\left(-\frac{1}{2}\left[\frac{\sigma^2+S_\mu^2}{\sigma^2 S_\mu^2}\left(\mu^2-2\frac{\sigma^2 M_\mu+S_\mu^2 x}{\sigma^2+S_\mu^2}\mu+\frac{\sigma^2 M_\mu^2+S_\mu^2 x^2}{\sigma^2+S_\mu^2}\right)\right]\right) \\
&\propto \exp\left(-\frac{1}{2}\left[\frac{\sigma^2+S_\mu^2}{\sigma^2 S_\mu^2}\left(\mu^2-2\frac{\sigma^2 M_\mu+S_\mu^2 x}{\sigma^2+S_\mu^2}\mu\right)\right]\right)
\end{aligned} \tag{3.6}$$

The transition to the final line above was possible because the last term in the innermost parentheses was a constant and could be dropped. Following this same rationale in reverse, we insert a different constant in order to "complete the square" and achieve an expression that resembles our original normal probability density function (Equation 3.5) in its structure. [98]

$$
\begin{aligned}
p(x|\mu,\sigma)p(\mu) &\propto \exp\left(-\frac{1}{2}\left[\frac{\sigma^2+S_\mu^2}{\sigma^2 S_\mu^2}\left(\mu^2 - 2\frac{\sigma^2 M_\mu + S_\mu^2 x}{\sigma^2+S_\mu^2}\mu + \left(\frac{\sigma^2 M_\mu + S_\mu^2 x}{\sigma^2+S_\mu^2}\right)^2\right)\right]\right) \\
&= \exp\left(-\frac{1}{2}\left[\frac{\sigma^2+S_\mu^2}{\sigma^2 S_\mu^2}\left(\mu - \frac{\sigma^2 M_\mu + S_\mu^2 x}{\sigma^2+S_\mu^2}\right)^2\right]\right)
\end{aligned} \tag{3.7}
$$

Equation 3.7 is the numerator for Bayes' rule, which can be normalized in order to get a probability density function. The form of Equation 3.7 is a normal distribution on μ with mean M'_μ and standard deviation S'_μ as shown in Equation 3.8 and 3.9. [98]

$$M'_\mu = \frac{\sigma^2 M_\mu + S_\mu^2 x}{\sigma^2 + S_\mu^2} \tag{3.8}$$

$$S'_\mu = \sqrt{\frac{\sigma^2 S_\mu^2}{\sigma^2 + S_\mu^2}} \tag{3.9}$$

This can be repeated, with the posterior distribution on μ becoming the new prior distribution for the next observation x. Alternatively, if N observations are made at once, we can use the fact that the mean of the observations $\bar{x}$ should be normally distributed with a standard deviation of $\sigma/\sqrt{N}$ [98]. Thus for N samples the posterior distribution on μ becomes normal with mean $M'_{\mu,N}$ and standard deviation $S'_{\mu,N}$ as shown in Equation 3.10 and 3.11.

$$M'_{\mu,N} = \frac{(\sigma^2/N)M_\mu + S_\mu^2 \bar{x}}{\sigma^2/N + S_\mu^2} \tag{3.10}$$

$$S'_{\mu,N} = \sqrt{\frac{(\sigma^2/N)S_\mu^2}{(\sigma^2/N) + S_\mu^2}} \tag{3.11}$$

Equation 3.10 and 3.11 form the basis for a study in the following subsection.

3.3.3 Estimating the mean for a ten square mile block of wells

It was shown in Figure 3-5 that even at the scale of ten square miles, the distribution of well productivity was approximately lognormally distributed. In this subsection, I demonstrate how this distributional assumption, combined with Bayesian statistical inference tools, provides a powerful method for estimating the mean of the distribution of wells in an area.

Consider a situation in which an operating company would like to estimate the value of μ for the productivity distribution of wells that will be drilled in a ten square mile area. This company may ultimately decide to drill fifty wells in this area, which each cost \$5 million, making this a \$250 million investment decision. If the company expects the value of μ for this area to be too low, it may choose to drill wells in a different area instead, or not at all. Traditionally, the company would drill a small number of appraisal wells in this area in order to gather geological data, for instance through core plug samples. It would then run reservoir simulations to generate production estimates aiding the decision of whether to continue drilling. As I discussed in Chapter 2, this can be a costly and imperfect process. The company could additionally use initial production rates from early wells drilled in the area to form a statistical inference of μ and use this knowledge to inform their decision to continue drilling the area.

We will assume that the operating company has some knowledge about well productivity in the broader area, such as other wells in the county in which this ten square mile block lies. This is not unreasonable given the availability of production data, such as through HPDI [73]. This knowledge allows the company to use an informative prior distribution on μ. The company does not know the distribution for the current block that they are in though, and must infer this from observations (wells drilled).

To evaluate the efficacy of Bayesian statistical inference in this situation I return to the same geographical divisions for Johnson county in the Barnett that led to Figure 3-5. I choose the prior distribution on μ to be normal with a mean equal to the mean of the log-transformed production rates for a random[8] selection of 250 wells in the county (this represents less than 10% of the total wells in our dataset for this county). Additionally, I choose the prior standard deviation to be equal to the standard deviation for mean values of log-transformed productivity in a random selection of 7 ten square mile blocks in the county (this represents less than 10% of the total ten square mile blocks in the county).

In order to use Equation 3.10 and 3.11, developed in Section 3.3.2, we will treat σ as a

[8]I use the Mersenne twister algorithm in MATLAB to generate random integers.

constant. The value assumed for σ is again drawn from a random sample of 7 ten square mile blocks in the county. The standard deviation for wells in each of these blocks is taken and the median value is chosen for σ.

This is a somewhat arbitrary procedure for determining the prior parameters and σ but it allows us to illustrate the use of imperfect knowledge about the area as a basis for developing an informative prior distribution. An operating company would of course make use of any available production data, both internally and externally, and tailor these numbers based on expert opinions in the company (this incorporation of subjectivity is a strength of Bayesian statistical inference).

There are a total of 72 ten square mile blocks in Johnson county that have 8 or more wells in our dataset. For each of these blocks, I carry out Bayesian statistical inference with initial wells drilled in the block and estimate the value of μ for the productivity distribution of all wells in that block. As a measure of accuracy, I calculate the squared error between the estimate of μ and the actual μ in each block. For each possible sample size N between 1 and 8, I sum the squared error across all 72 blocks. For comparison, I also use a maximum likelihood estimator (MLE; the sample mean, since we are working with normally distributed samples after log-transformation) and calculate the sum of squared errors. I also find the sum of squared errors if a static estimate of the county average is used (since this knowledge was used for the prior distribution). The result of this analysis is shown in Figure 3-13.

The Bayesian inference systematically balances observed data with prior knowledge about the area, and as a result of this, does better than both MLE and the static estimate based on prior knowledge. There is a particular advantage over MLE when only a small number of wells have been drilled. This makes this technique particularly valuable as it can provide insight about the value of drilling in an area without having to invest as much in drilling appraisal wells.

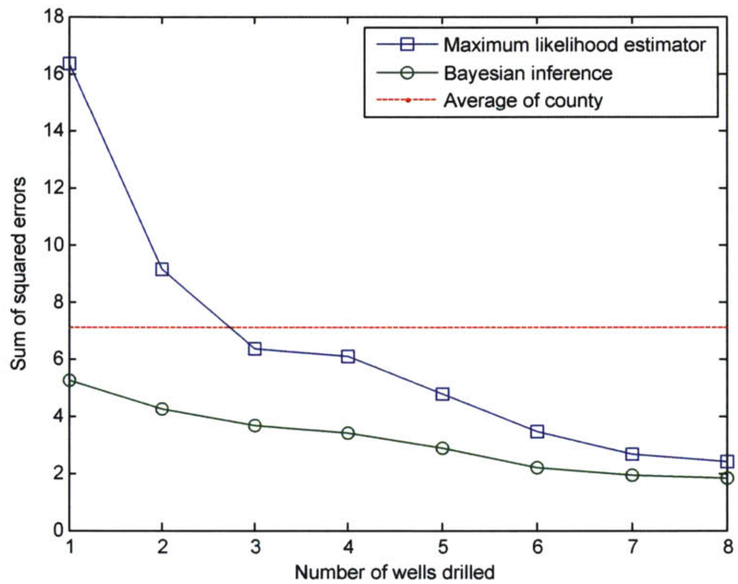


Figure 3-13: Sum of squared error with different numbers of wells drilled for Bayesian inference, maximum likelihood estimator, and county average (the prior knowledge) as a static estimate.

Chapter 4

Characterizing drilling performance variability

Drilling wells is essential to development of shale gas and tight oil resources and is often a focal point for well cost reduction. Because of its procedural complexity, as discussed in Chapter 2, there is a large amount of uncertainty about development costs, both for present and future wells. The learning curve, discussed in Section 2.3.1, provides a model for projecting changes in efficiency of drilling operations over time driven by organizational learning, particularly through the use of "experiments," or field trials of new or altered procedures.

The variability in operations and performance presents challenges for understanding the inputs to the learning curve. There is variability in performance around the learning curve, which appears as "noise" in a chronological plot of well construction times with a fitted learning curve. This noise makes it challenging to reliably measure progress on the learning curve and forecast future improvements. Variability in the sequence of operations used to drill a well represents an additional indicator of learning, capturing elements of both "standardization"—an unconventional drilling aspiration I introduced in Section 2.3.3—and experimentation with new procedures, especially early on in the development of a field. These forces of learning need to be carefully balanced. Thus, it is worthwhile to consider a way to quantify the level of standardization and procedural consistency between wells drilled sequentially in a field.

Performance and operational variability in well drilling operations is a challenging and ambiguous topic. In addition to geotechnical physical uncertainties, there is a critical social dimension to drilling. The data I have used, rather than being direct physical measurements, is based on human-reported conditions of the drilling system, which adds to the challenge of identifying underlying trends. Furthermore, in contrast to the production dataset used in Chapter 3, the scope of the drilling dataset used in this chapter also makes it difficult to draw broadly applicable conclusions about variability beyond the dataset examined. The production data spanned all wells drilled in a number of plays over the course of several years, allowing for ubiquitous trends to be identified. The drilling report data used in this chapter is for a single drilling campaign in a field, including around 200 wells drilled by a single operating company. The drilling report data is neither granular enough for specific conclusions about individual operations, or broad enough to develop an understanding of how operations evolve more generally across multiple campaigns. Nevertheless, the data provides a useful analysis test-bed, providing insight into how performance variability can be characterized, interesting trends worth further investigation in other drilling campaigns, and how future efforts toward understanding drilling variability and managing the transformative learning process might be improved.

In the first section, I describe the available data in greater detail, as well as how it was processed. I then discuss a method for measuring *operational variability* between wells drilled sequentially and present the results of applying this to wells from a single drilling campaign. The relationship between this measure and learning will be explored along with related implications for managing the drilling learning process. In the last section, I discuss variability of performance around the learning curve and the challenge this presents for evaluating and managing the learning process in drilling. I characterize variability in the residuals of both a power and exponential learning curve fit to the overall wells drilled in this campaign and discuss the implications of both the specific findings and more generally the benefits of considering this dimension of variability in drilling performance.

4.1 Description of data and processing

4.1.1 Description of drilling report data

The drilling report data that I have analyzed consists of a database of detailed drilling reports from nearly two hundred horizontally-drilled development wells drilled in a major North American unconventional field by a single operator over the course of about three years. Each well's collection of drilling reports is comprised of sequential entries spaced at an hourly resolution, including structured data like process time and bit depth along with unstructured text entries describing activities and measurements. Each entry in the drilling report is identified with an operation code, which indicates the category of activity associated with that drilling report. In total, approximately 60 different operation codes are used in the reports, corresponding to activities that include testing the blow-out preventer stack, rotary drilling, slide drilling, casing placement, pumping cement, etc. Combined, these describe with a reasonable level of granularity the sequence of activities taking place during the drilling of each well.

In an unconventional field, the delivered well is essentially a commodity, and unlike with hydraulic fracturing, the drilling procedure is unlikely to have a direct impact on well production. The exception to this is well placement and length. Although the well is constructed essentially the same, a longer well may allow for additional stages of fracturing and access to more of the reservoir. Longer wells can also be more time-consuming to drill. The placement of the well also matters, both in terms of the area that it accesses (although the quality of rock is difficult to predict as Chapter 3 identified) and whether the well is effectively drilled to stay within the target zone. In this analysis, I occasionally treat wells as being of identical length, a valid assumption since the well-length does not vary much within this dataset. In other cases, I normalize drilling time by well-length to account for differences where they are more likely to have a notable impact on outcomes. I also make the assumption that staying within the target zone is not an issue with any of the wells—a reasonable assumption since there is no reason to believe that any of these wells were botched.

4.1.2 Conversion of drilling report data into operation sequences

Given this reporting structure, the drilling activities of each well can be approximated using a sequence of these operation codes. These sequences will ultimately form the basis for comparing wells' drilling activity to each other for similarity in 4.2.2. It is helpful to disaggregate these sequences into different "drilling modules," and this is straight-forward to do because the reporting format uses "phase codes" associated with the section of well under construction. The four drilling modules for these wells are *1) drilling the surface hole*—drilling the first section of wellbore, *2) casing the surface hole*—casing and cementing the first section of the well, *3) drilling the production hole*—drilling the second (and in this case final) section of wellbore, *4) casing the production hole*—casing and cementing the second section of the well.

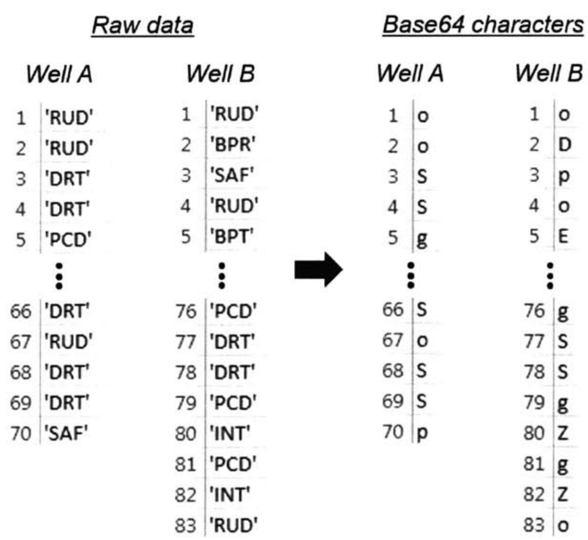


Figure 4-1: Sequences of operation codes (raw data) mapped to Base64 characters.

In preparation for comparison, I used a program in MATLAB to separate operation codes into the drilling modules of each well. In the dataset for this analysis the drilling operation codes were three letters in length, so for ease of implementation, I used a key to convert each drilling operation code into a Base64 character. The sequential procedures for each well are represented by a vector of symbols, which can be mapped back to the original operation codes and the identifying information of the well. An illustration of this encoding of operations for each well into a string is shown in Figure 4-1, which uses the drilling the production hole module from two wells for illustration. Descriptions of the operation codes in Figure 4-1 are included in Appendix A in Table A.5.

4.2 Evaluating operational variability

Standardization of drilling procedures is a key driver toward reduction over time of the unit cost of production in unconventional oil and gas fields. This may be aspirational, in the sense of efforts to modularize well designs and planned sequences of activities across similar wells. It may also be used to refer to the similarity of actual procedures between different wells, which is difficult to quantify. In this section, I present an original method for quantifying this dimension of learning and present some interesting results obtained from applying it to the field data described in the previous section.

4.2.1 Approximate string matching and operational dissimilarity

In order to measure standardization across a development campaign, and how procedures vary over time, I make use of approximate string matching (ASM). ASM is an algorithmic technique used for the pairwise comparison of character strings that represent qualitative measurements, steps or symbols in a sequential system—such as nucleotides in DNA or RNA sequences—to quantify the degree of similarity [76]. ASM had never been applied to petroleum production systems before we proposed it at the 2014 SPE ATCE in Amsterdam [119]. In the past it has been applied to short strings, such as in online search queries and word processor spell-checkers, and also with long strings, such as in determining the similarity of sequences of DNA in the field of bioinformatics [32, 76]. I use this technique to develop a measure for operational variability, the first objective metric for evaluating and comparing the rate and consistency of standardization across different wells, sections of wells, and rigs.

The purpose of the ASM technique is to quantify the dissimilarity between two chains of symbols. In order to do this, the algorithm determines the minimum number of single-character edit operations necessary to change one string to another. This minimum number of edit operations is often called a "string distance," but I will refer to it formally as the *operational dissimilarity*, since we are comparing sequences of operations. A higher operational dissimilarity corresponds to a greater variation in the field operations between two wells. ASM can be implemented using several algorithms that allow for different edit operations. In this analysis, I chose to use the Damerau-Levenshtein algorithm[1], which is used for comparison of DNA sequences in bioinformatics—the most similar application to our own [32, 101, 76]. This variant of ASM includes the potential edits described in Appendix A in Table A.6.

The Damerau-Levenshtein ASM algorithm first creates a blank matrix for comparing two strings of elements. The matrix size is $(m+1) \times (n+1)$, where m and n are the lengths of the respective strings (or number of operations). The first column and row are used to compare each string with a "blank" reference string and are numbered according to character position, which is the minimum number of edits between each of the strings and the blank string. This matrix with the initial numbering is shown in Figure 4-2.

		o	D	p	o	E	•••	g	S	S	g	Z	g	Z	o
	0	1	2	3	4	5		76	77	78	79	80	81	82	83
o	1														
o	2														
S	3														
S	4														
g	5														
⋮							⋱								
S	66														
o	67														
S	68														
S	69														
p	70														

Figure 4-2: The initial comparison matrix is generated for the strings being compared. Many cells have been omitted here and are represented with ellipsis.

A dynamic program then proceeds through each position, moving left-to-right and then top-to-bottom. At each position, the algorithm finds and stores the minimum number of edits associated with changing the first substring to the second substring, where each substring is

[1] More precisely, I use the restricted edit form of Damerau-Levenshtein, which is computationally more efficient and does not allow substrings to be edited more than once. The results of this version are nearly identical with the full Damerau-Levenshtein algorithm.

defined as the characters up to and including the current position (i.e. to the left and above the current position in the matrix). In order to do this, the algorithm combines the editing associated with the character combination at the current position with the previous stored values. The potential edits and the process for evaluating the new associated values relative to the previous position's minimum edit values are summarized in Appendix A in Table A.7. The minimum total value out of all of these potential edit operations is the value that will be stored in the cell for the positions currently under consideration.

In this manner, the algorithm effectively stores the minimum number of differences as it works across the two strings, and the bottom right cell holds the resulting minimum edits, or operational dissimilarity, between the first string of operation codes and the second. Figure 4-3 provides a detailed example of this.

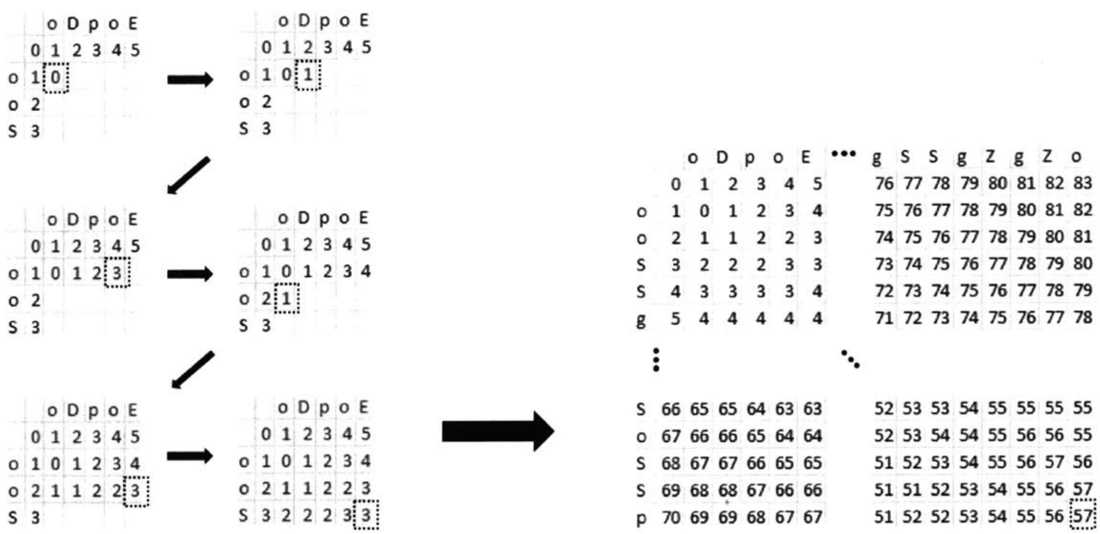


Figure 4-3: On the left, a series of substring comparison steps are shown (some steps are excluded here). The gray box indicates the two substrings being compared and the red box is the minimum edit operations at that position. On the right is an overview of the matrix after the entire strings have been compared with the overall operational dissimilarity in the bottom right red box.

Using this ASM algorithm, I systematically compared the well drilling modules in the analyzed dataset discussed here in order to find the operational dissimilarity for each pairing of wells. However, given that we are considering sequences of different lengths, one concern with this metric of operational dissimilarity is that it is an absolute measure of the total number of edits. Longer strings (i.e. wells with more operational steps) can incorrectly appear more different from each other than shorter strings that are compared with each other—the greater number of potential differences present in a longer string tends to exaggerate the difference by skewing the operational dissimilarity without regard for the overall length of strings being compared. This is particularly a problem since we are interested in looking at specific drilling modules, which are different lengths (such as longer modules like drilling the production hole and shorter modules like casing the surface hole) and want to be able to compare trends between these. What we really need to measure is how different (or similar) two sequences are relative to the maximum number of differences. To develop this metric, we normalize the operational dissimilarity by dividing it by the maximum number of potential differences, which is the length of the longer string being compared. This provides a normalized measure of dissimilarity, ranging from zero, if the two strings are identical, to one, if there is no similarity between the two strings and every character must be either substituted, inserted, or deleted.

After carrying out ASM on a set of drilling modules from wells, we need a way to visualize the normalized operational dissimilarity between each pair of wells. One approach to this is using a color plot. Figure 4-4 provides an example of this visualization using a small sample of wells. The left and bottom axes are labeled with well numbers (ordered sequentially by date), increasing from top to bottom and left to right, as is standard in matrices. Note that diagonally across, there is a line of cells with a value of zero, since a well compared to itself has an operational dissimilarity of zero. Additionally, the plot is symmetrical about this line, since approximate string matching is commutative. Temporally, since the wells on each axis are ordered by date, the top left is a comparison of the oldest wells to each other and the bottom right is a comparison of the most recent wells to each other. Of particular interest to us is the "gradient," or gradual change in color toward lower values which can be seen in this particular example moving from older wells to newer wells. This indicates that wells became more standardized and similar to each other over time.

Figure 4-5 is a color plot of the normalized operational dissimilarity, calculated by using ASM, for the production hole drilling module across nearly 200 wells drilled sequentially in this campaign. The darker shades along the left and top of the plot indicate that earlier wells have a great amount of dissimilarity both with each other (upper left corner) and with

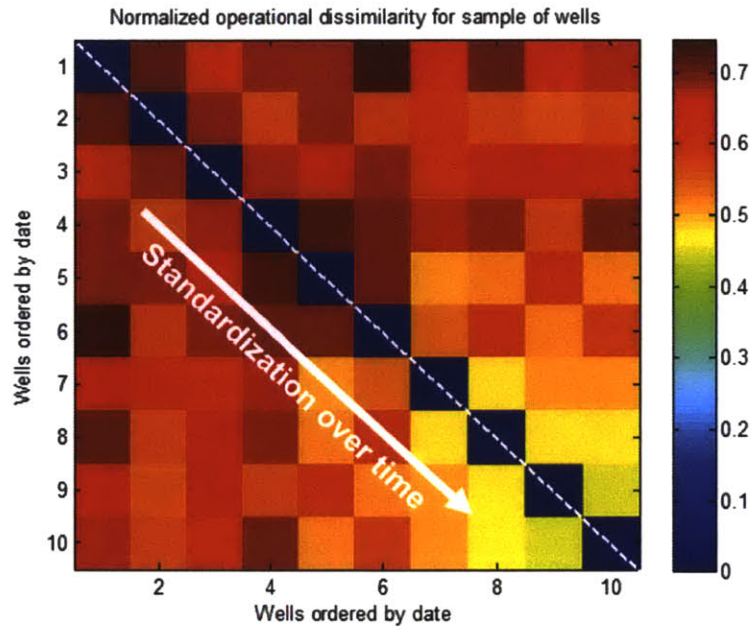


Figure 4-4: A color plot can be used to visualize the results of ASM for a series of wells. This color plot is similar to viewing a discretized surface plot from above. The plot here is not of actual data, and is intended only to be illustrative.

later wells (along the upper and left side). On the other hand, later wells (lower right corner of the plot) appear to be more similar to each other.

4.2.2 Demonstration of operational variability

Although it is useful to visualize all of the information generated from the ASM in color plots, it is easier to inspect trends by using a single metric for each well. Additionally, when considering standardization in a given field, we are less interested in how each well compares to every other well in the field—to understand standardization we really want to understand the nature of procedural variation in temporally close wells. In order to measure this, we formally define operational variability as the mean of the normalized operational dissimilarity of the previous ten wells compared with the current well[2]. This provides a simple metric for individual wells indicating how procedurally standardized each module in them is.

Referring back to Figure 4-5, the section of data used for calculating operational variability is outlined with dotted lines. A plot of operational variability for drilling the production

[2]A scope different than the previous ten wells can be chosen, but there is an inherent trade-off between being able to apply the technique earlier on and introducing too much noise with a shorter scope.

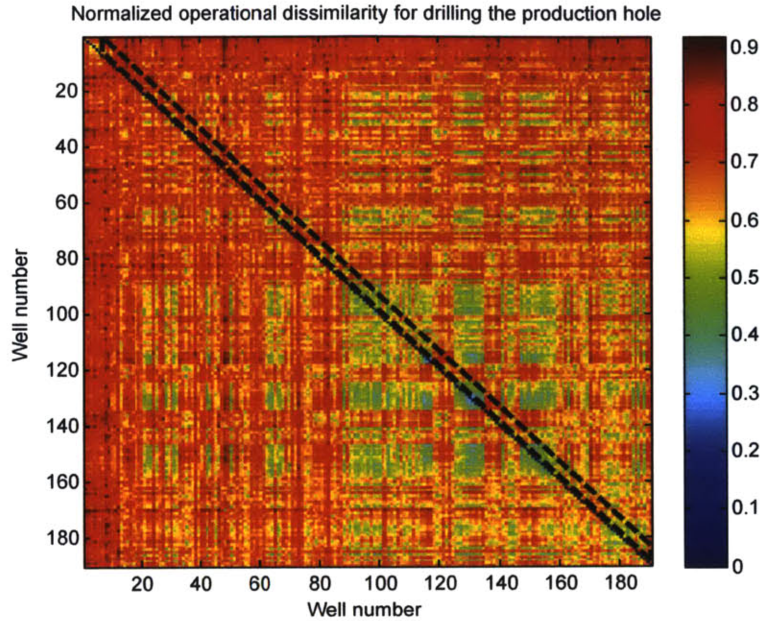


Figure 4-5: Color plot of normalized operational dissimilarity for drilling the production hole.

hole versus chronological well number is shown in Figure 4-6. There is a strong trend toward greater standardization across the drilling campaign within this module.

The trends of standardization are not identical across all drilling modules and some interesting insights can be drawn from comparisons. We can apply a linear fit to operational variability in a module to produce a trend line we will call a "standardization curve," because of the similarity of this concept to the learning curve. Figure 4-7 shows a comparison of standardization curves in different drilling modules. Note that there are nearly parallel trends for both of the drilling-based modules as well as for both casing-based modules. It is natural that the drilling-based modules experience higher levels of standardization, or reduction in operational variability. There is simply a greater variety of activities that can be carried out in the drilling procedures than for casing, which is a more straight-forward activity. Drilling-based activities have more experimentation early on, associated with finding the best techniques. Once these best practices have been identified the drilling procedures are rapidly standardized in order to improve efficiency. Also, observe that modules related to the production hole—which is longer and more complex to both drill and case because of the presence of a curved section and long, horizontal section—have higher operational

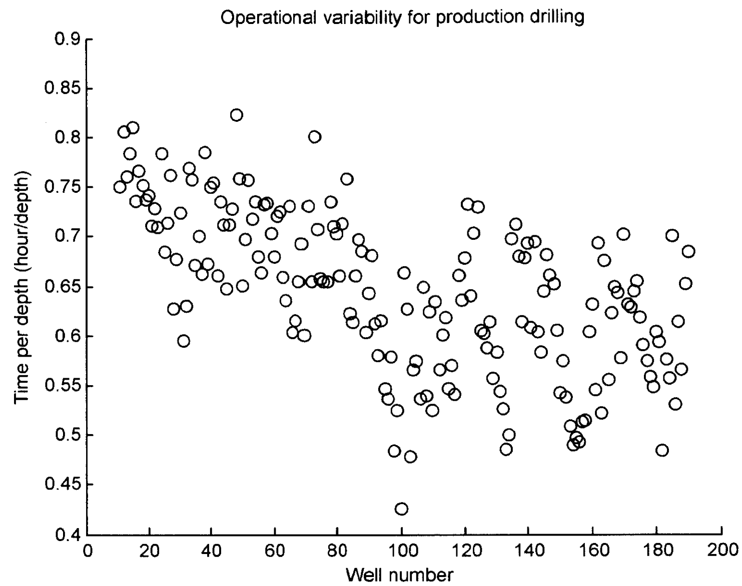


Figure 4-6: Operational variability captures the operational dissimilarity of a well relative to the other most recently drilled wells (in this case using a scope of the ten previous wells).

variability than their surface hole counterparts. This illustrates how this metric can be used to understand the nature of learning in different parts of wells. This type of insight should guide management toward focusing on modules with greater opportunity for standardization improvements. In some ways, standardization may be most prevalent and also most important for more complex sequences of activities in the drilling process.

It is worth considering a potential relationship between standardization (i.e. the reduction of operational variability shown in Figure 4-6) and time-based performance improvement, as is frequently discussed with the learning curve (see Section 2.3.1). Learning curves will be developed and applied to this drilling data in the next section, but for now the normalized time and chronological well number will be plotted without a fitted curve simply to show general shifts in performance over time and how operational variability can provide further illumination on change brought about through learning. Figure 4-8 shows time for drilling the production hole in each well normalized by the length of this section of well[3]. This can be compared to the plot in Figure 4-6 showing operational variability in this drilling

[3]Here we normalize the length, although it is not necessary, as wells are fairly similar in length. Later, in Section 4.3, this normalization is not used because when considering the entire well construction process, slight variations in total well length have little impact. It is generally recommended that both non-normalized and normalized time are examined and an evaluation of the relevance of well length differences is made from this.

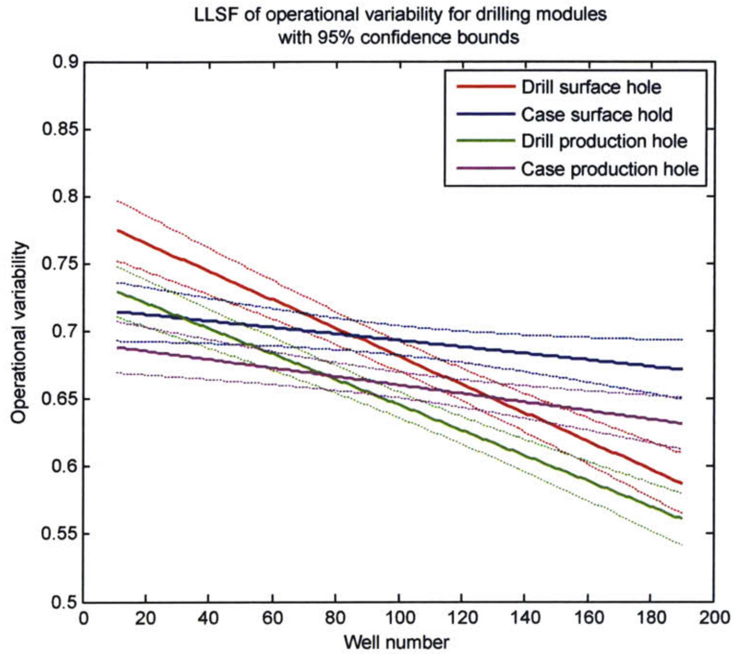


Figure 4-7: A comparison of operational variability trends in different drilling modules. Solid lines are linear least squares fits (LLSF) and corresponding dashed lines are 95% confidence bounds established using non-simultaneous functional bounds [109].

module for each well. In both cases, learning is reflected in the general downward progression of the data, toward faster drilling and more standardized, consistent sequences of drilling procedures. Both of these findings are somewhat expected based on literature discussed in Section 2.3.3. It is unclear at this point whether there is a causal relationship between these trends in either direction but some correlation is apparent. This corollary relationship is interesting because each metric is measuring fundamentally different aspects of how the drilling of wells has changed over time: the length of time a well takes to drill and how similar operationally a well is to other recently drilled wells. Because the operational variability has been normalized by the the number of operations, the length of time spent drilling has no direct impact on the standardization curve, indicating that these are two separate trends taking place alongside each other. Although it has been theorized that standardization is an important aspect driving time performance improvement in the drilling of unconventional fields, prior to this work, standardization had never been quantified alongside time-based improvement for comparison of the two phenomena.

Rather than treating it as a separate indicator, we can also use the metric of operational

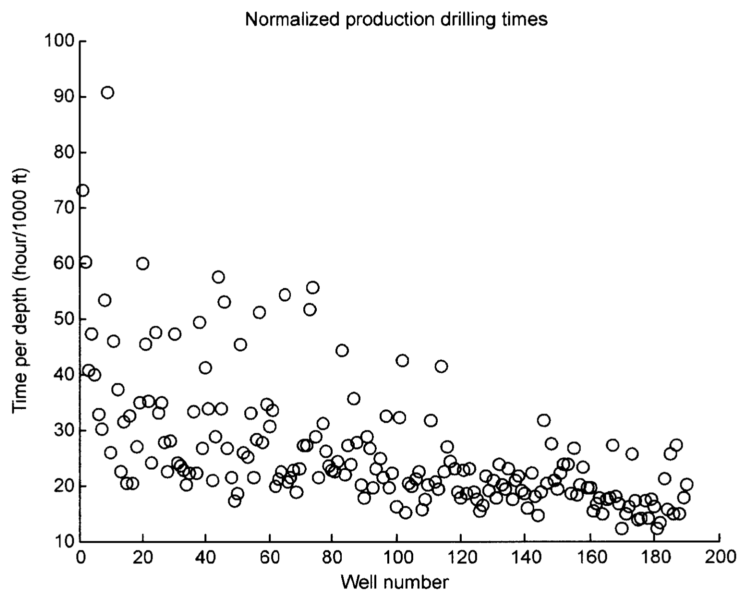


Figure 4-8: Normalized time for drilling the production hole. Compare the downward trend with that shown for operational variability in 4-6.

variability to enhance a plot of time-based improvement. In this manner, we can use a single plot to provide even more information about how operations have changed over time, both in their duration and relative sequencing. To do this, a "bubble" plot is used to plot normalized times of drilling the production hole with variable marker sizes. The marker sizes are determined by the operational variability of the production hole drilling procedures in each well. Normalized times for the first ten wells are excluded from the plot because they are outside the scope of measurement for operational variability. This makes it possible to view relative changes in operational variability and the trend of standardization as a dimension of the learning process.

An interesting application for this combined plotting of time-based performance and operational variability is analyzing and comparing learning within specific drilling rigs or crews. The drilling campaign under analysis actually consists of several rigs operating throughout. In addition to looking at aggregate performance improvements and standardization across the entire field development campaign (as we have been doing so far), we can also disaggregate learning effects for the different rigs. Figure 4-10 shows normalized production hole drilling time for each rig, from the eleventh to final well drilled by the rig (in order to limit plotting to wells which have a corresponding operational variability measure). This type of

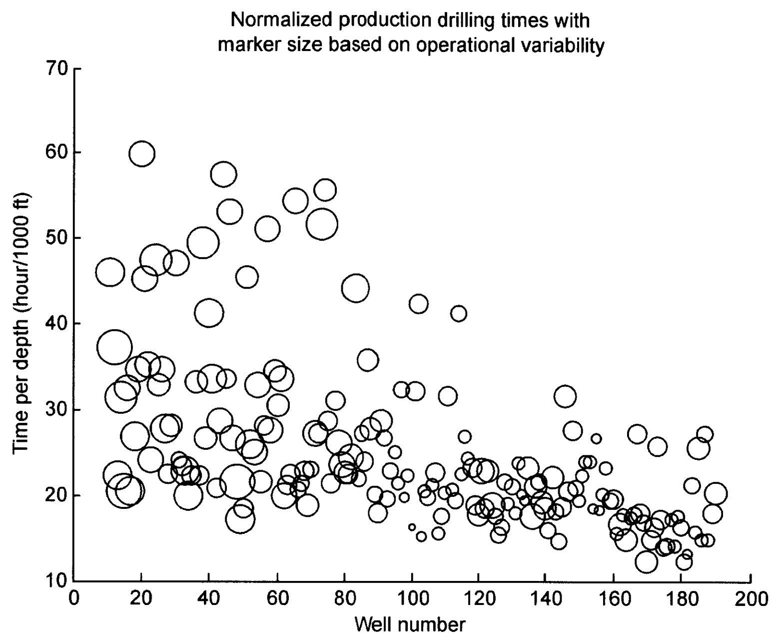


Figure 4-9: Operational variability and time-based performance combined into one "bubble" plot with marker size based on operational variability.

plot can provide insight into how rigs should best be deployed within fields. For instance, Rig 3, which operated for only a short time, showed extraordinary procedural consistency, indicating an ability to implement fairly standardized procedures immediately. In contrast, Rigs 1, 2, and 4, deployed early in the campaign, had high operational variability as well as substantial scatter in time-based performance at the start. This may be attributable to planned experimentation and ambiguous engineering plans. Over the course of the campaign though, they all demonstrated an ability to internalize learning and achieve good operational consistency and performance improvements. Rig 5, on the other hand, was brought in late during the campaign and had a large amount of operational variability and inconsistency in time-based performance. This may indicate an inability to follow standard operating procedures and incorporate broader field learnings or may be due to an oversight by management at bringing this rig up to speed with best practices. It would require further assessment to determine if management is the culprit, but if the Rig 5 crew itself is mostly to blame, then we can conclude that Rig 3 is a more reliable rig to bring into a field in which some standard operating procedures have already been established.

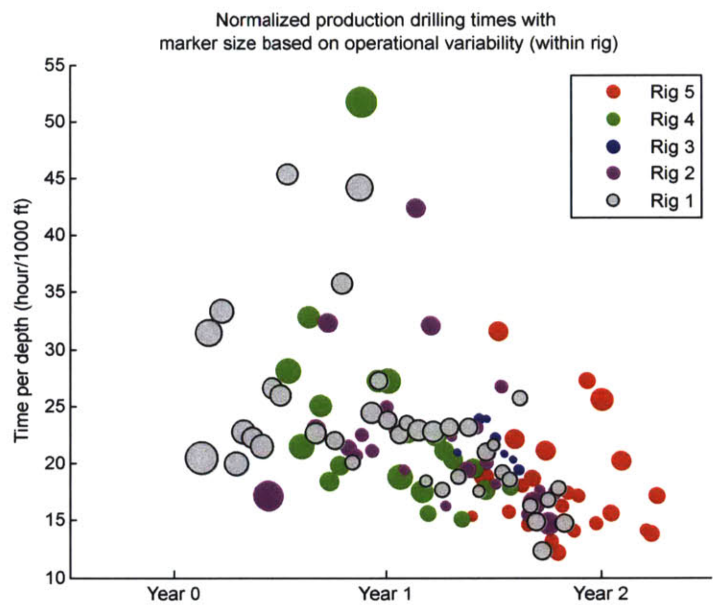


Figure 4-10: Operational variability can be used to provide additional insight into differences between drilling rig performance. Operational variability is measured within the wells drilled by each rig and only the eleventh to final well is plotted. See Figures B-20 to B-24 in Appendix B for each rigs data shown independently.

4.2.3 Enabling automated and data-driven learning

ASM and the metric of operational variability can be used to measure standardization in drilling field operations, which is an established concept though one that has not been rigorously measurable before this research. The results here indicate some connection between operational variability and improvement in drilling time performance. This is a useful first step in deconstructing the inputs to the learning curve and can be used to better understand the role of standardization in reducing drilling time. The ability to measure and record standardization can inform the design of incentives for engineers and field workers in order to balance the different roles of standardization and experimentation. Additionally, the measurement of standardization is useful for planning, implementing, and evaluating drilling campaigns. For instance, ASM can be used to call attention to parts of a well that could benefit from further standardization (and vice versa).

Going forward, this technique can be made more operational and the learning curve can be better understood by developing a systematic means for measuring experimentation, or "continuous improvement." The challenge of doing this, due to natural variability in time-based performance, is discussed in the next section. The combination of improved understanding of continuous improvement and standardization with the ability to measure these will lead to an improved understanding of the different levers that drive the learning curve in drilling. This can also help engineers and managers to balance the two objectives.

Ultimately, these methods can play a critical role in future "smart" drilling rigs and advanced drilling campaign management software, enabling rapid adaptation and greater overall drilling efficiency. By automating measurement of specific dimensions of learning like standardization, data analytics and machine learning techniques can be rapidly deployed with incoming streams of data to assist in directing operations, evaluating well designs and making new drilling plans. This will enable the oil and gas industry to overcome "data paralysis" and make better use of abundant field data. Although further work is needed for the development of additional tools and frameworks for learning, the tools presented here for measuring standardization are a fundamental early step toward improving drilling planning and the management of data-driven learning in unconventional drilling.

4.3 Performance variability in the learning curve

4.3.1 The issue with learning curve noise

Continuous improvement is the central process of TQM (total quality management) generating improvements in effectiveness and efficiency in a range of industries [155]. Here, I use it to refer more generally to any intentional learning activity during drilling, which is intended to drive improvements in performance and shift activity favorably along the learning curve. For instance, a new drilling fluid might be introduced during a challenging section of drilling a well in an effort to improve drilling efficiency. As identified in Section 2.3 though, introducing changes is not enough, and if done unsystematically can actually harm the learning process. It is also necessary to measure outcomes, infer causal effect, and if warranted, enshrine the new practice as a protocol (such as an SOP).

Zangwill and Kantor [155] developed an approach for evaluating progress made relative to the learning curve—an important examination given the interest in better managing the learning curve. The authors develop a differential equation form of the learning curve that can incorporate both the power-law (Equation 2.1) and exponential (Equation 2.2) forms of the learning curve in literature by using different assumptions. From this more general model, they describe a finite difference approach to estimate progress on the learning curve (i.e. the rate of learning implied by a continuous improvement) defined as,

$$E(q) = -\Delta M/(\Delta q N(q)) \tag{4.1}$$

In Equation 4.1, $E(q)$ is the effectiveness of the continuous improvement, q is the production number (i.e. the cumulative well count) and Δq is the number of items processed in the period of interest (while the continuous improvement is being implemented. $N(q)$ is the "waste" remaining in the system and ΔM is the measured improvement in the cost or time metric. Figure 4-11 illustrates these quantities on a sample learning curve.

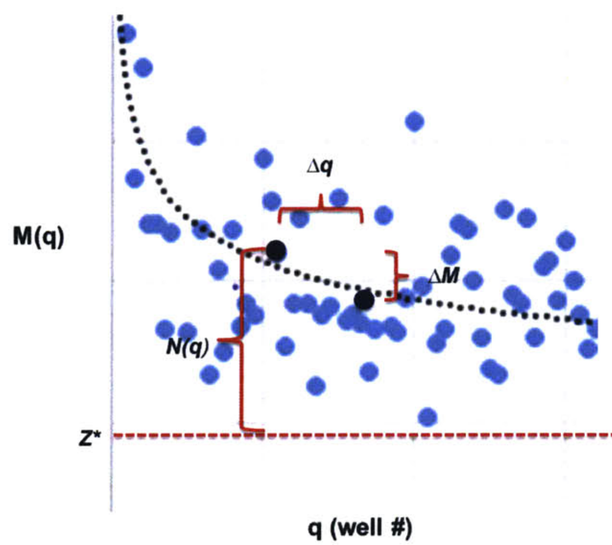


Figure 4-11: Illustration of the variables used by Zangwill and Kantor [155] to evaluate effectiveness.

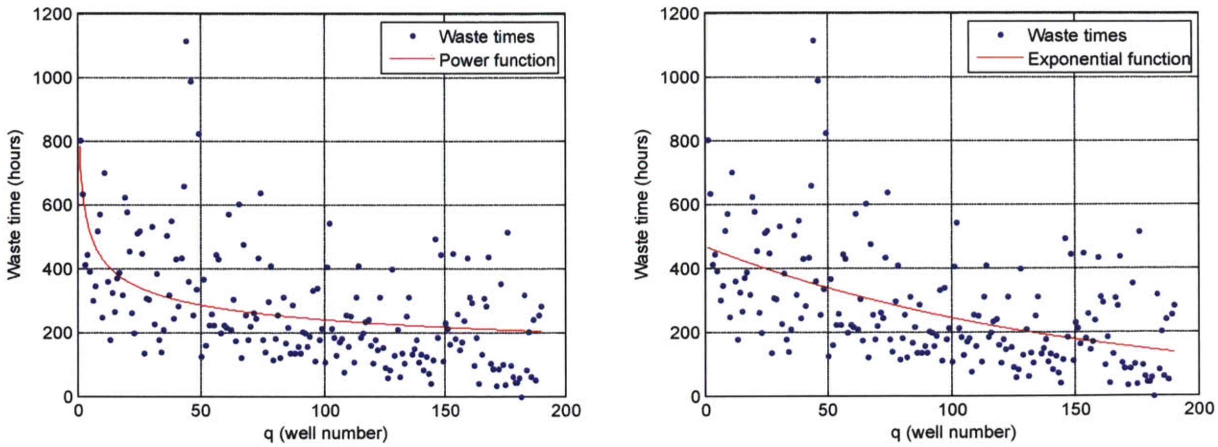
A major challenge in reliably estimating ΔM is the abundant noise, or scatter, of data points distributed around the learning curve. The difference in M between two points captures some change along the learning curve but also an unknown error term associated with each point. Zangwill and Kantor attempt to address this concern by suggesting that M be measured at the level most directly associated with the change to reduce error. There are two issues with this though. First, in a process like drilling there is significant interconnectedness of different procedures making it difficult to measure a direct output of an activity at times. Second, Zangwill and Kantor presume that if the effect of an activity can be measured directly it will be a deterministic representation of the causal impact of a continuous improvement. This is a poor assumption in a situation like drilling a well, which is fraught with uncertainty about the subsurface. Any measure of an outcome should be viewed as a statistical sampling of the effect in question; an exact replication of activities might yield a slightly different outcome in the field.

Additionally, there are situations, such as developing learning-based incentives or evaluating particular crews and rigs, in which it would be desirable to measure learning within an overall learning curve, rather than for a subset activity in the learning curve. As an example of this, an analysis I carried out found that a collection of different rigs all made learning progress while drilling several wells on a drilling pad, but when the rigs were moved to a new drilling pad, they each experienced a decline in efficiency with the first well on the pad. This form of "forgetting" is an important consideration in the learning process and has been previously discussed in the context of drilling rigs [17]. Importantly, some rigs were better at internalizing learning and had lower rates of forgetting. This difference in progress retention allowed the less forgetful rigs to, over time, outperform other rigs that had higher rates of learning within pads. Although the specific results of this analysis have not been included in this thesis, this anecdote illustrates the value of considering continuous improvement effectiveness as part of the overall learning curve.

4.3.2 Characterizing the variability of noise

When considering the variability of performance, it is less relevant to consider absolute values, than the quantity of waste remaining in the system. This is identical to the measured performance but with an assumed constant technical limit subtracted from the measured effect[4] [155, 3]. The assumption of a technical limit is not trivial and may change over time, but for the purposes of evaluating performance, waste in the system is usually relatively large

[4]The technical limit is included in the exponential law (see Equation 2.2) so this term is set to zero during the curve fitting of waste time here.



(a) Power-law learning curve

(b) Exponential learning curve

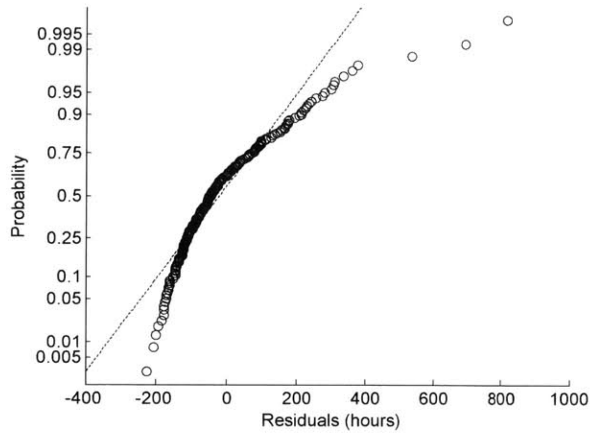
Figure 4-12: The power-law and exponential functions, the two most common forms of the learning curve, were fitted to the waste times for wells in a drilling campaign using ordinary least squares nonlinear regression.

early on when the technical limit is most uncertain, making the accuracy of early estimates of technical limits less important [155]. By the time the remaining waste becomes small enough to make position of the technical limit a concern, more reliable estimates can be made of its quantity (and management of continuous improvement is less economically important at this stage anyway).

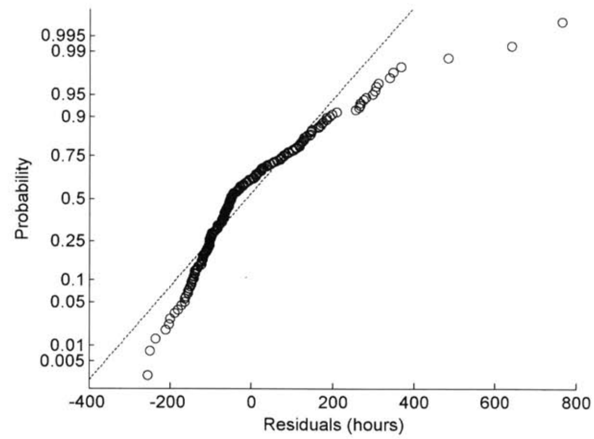
Given that I have drilling data over the course of a development campaign, I will simply assume a technical limit equal to the minimum observed value. In situations where less data is available, this would not be appropriate and the technical limit must be defined based on technical analysis, analogues, and best estimates that are revised as more data becomes available.

Nonlinear regression was used to fit both the power and exponential forms of the learning curve to the waste times. The results of this are shown in Figure 4-12.

It is important to understand the nature of performance variability within the learning curve. As in Chapter 3, probability plots provide a useful tool for this evaluation. In this instance, I use them to evaluate residuals from the curve fits. Visually inspecting Figure 4-12, it is fairly apparent that the noise is not evenly distributed and has an upper skew to it. Nevertheless, I will begin by considering a normal approximation for the noise (which is unlikely to be appropriate given this skew) and then examine a lognormal distribution fit to



(a) Residuals for power-law fit

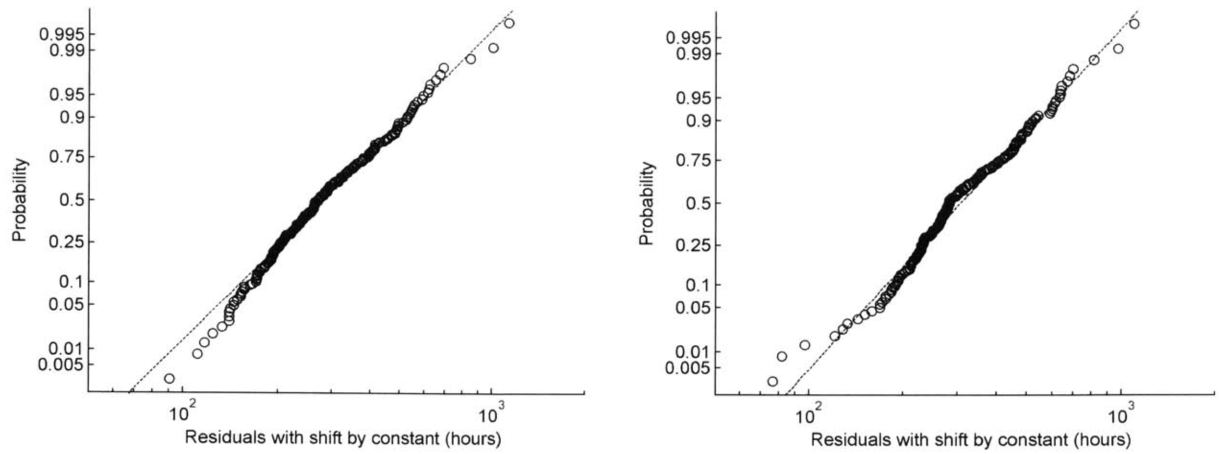


(b) Residuals for exponential fit

Figure 4-13: Normal probability plots for residuals of power and exponential learning curve fits.

the residuals. Figure 4-13 shows normal probability plots for the residuals of both the power and exponential curve fitting. This distribution provides a very poor description of noise in both curve types.

Evaluating a lognormal shape is not straightforward because residuals are both positive and negative (for those below the fitted curve) and a lognormal distribution by definition can only include positive values. Therefore, the data points need to first be shifted by some constant so that the distribution shape can be compared a lognormal one. Such a constant, though inherently arbitrary is not trivial to select. It is not enough to shift the distribution such that the lowest observed point is at (or very close to) zero. In a collection of around 200 samples, even from a true lognormal distribution, there is unlikely to be a value occurring relatively close to zero. To determine an appropriate shift constant for a sample size n, I first calculate the inverse CDF for what would statistically be the lowest expected value (i.e. $F = 1/n$). Of course, the lowest sample could occur below or above this value but it seems a reasonable starting point for the lowest data point if a lognormal distribution is a good approximation. I then shift the distribution such that the minimum value observation coincides with this expected lowest observation. Because the inverse CDF of the lognormal distribution depends on the shift constant selected, an iterative process in MATLAB was used to obtain this. A lognormal probability plot applied to both sets of residuals is shown



(a) Residuals for power-law fit

(b) Residuals for exponential fit

Figure 4-14: Lognormal probability plots for residuals (with constant shift) of power and exponential learning curve fits.

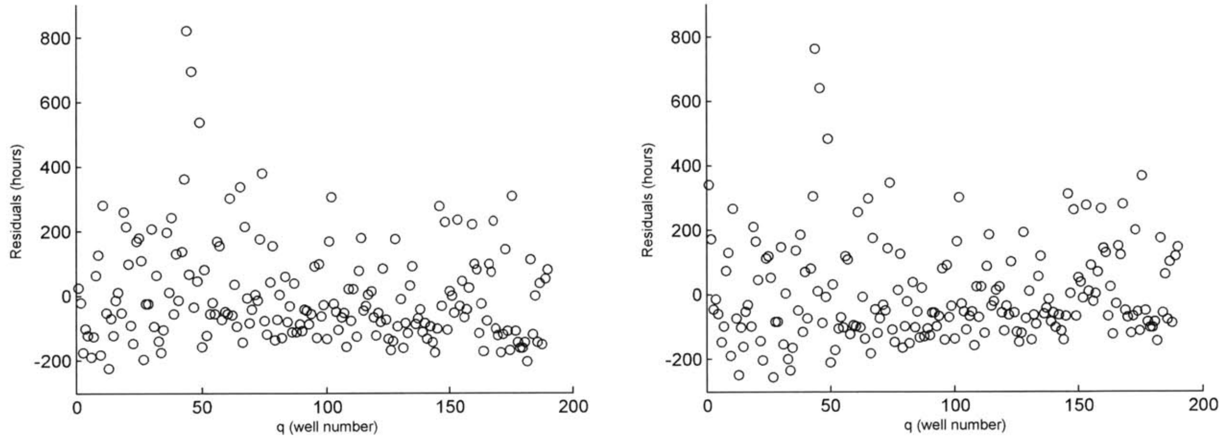
in Figure 4-14. This model for the residuals is quite good, suggesting that a lognormal distribution is a good approximation for the performance variability in a learning curve.

It is also worth considering how the distribution of learning curve noise might change throughout the learning curve. To determine if variance is fairly constant throughout the learning curve (i.e. homoscedastic), I used a graphical and formal test for heteroscedasticity in the data. From a plot of the residuals relative to q, the production unit number, the residual appears consistent and randomly arranged over the course of all wells (Figure 4-15). An Engle test was also applied in MATLAB and confirmed that for both sets of residuals, heteroscedasticity is not present.

I also consider the consistency of the distribution shape throughout the learning curve. To do this, I break the sequence of wells into three roughly equal-sized groups. Figure 4-16 shows probability plots for residuals in each fit for the first group of wells. Figure 4-17 shows the same for the second group of wells, and Figure 4-18 for the third, and final, group of wells. Lognormality is apparent for all three of these categories.

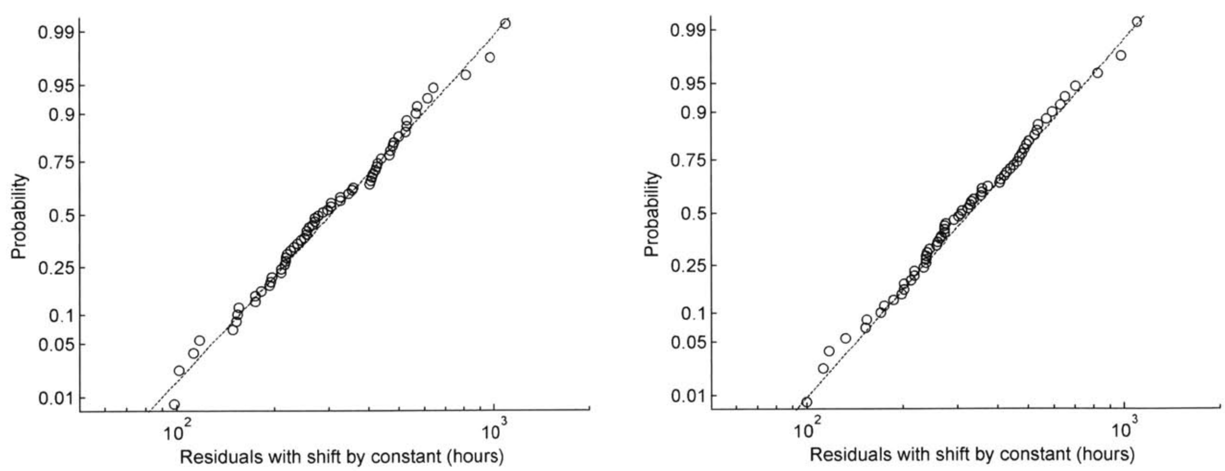
4.3.3 Implications of drilling performance variability

As was the case in Chapter 3, lognormality is more an insightful observation and convenient model than a surprising result. Complex tasks often have right-skewed distributions of



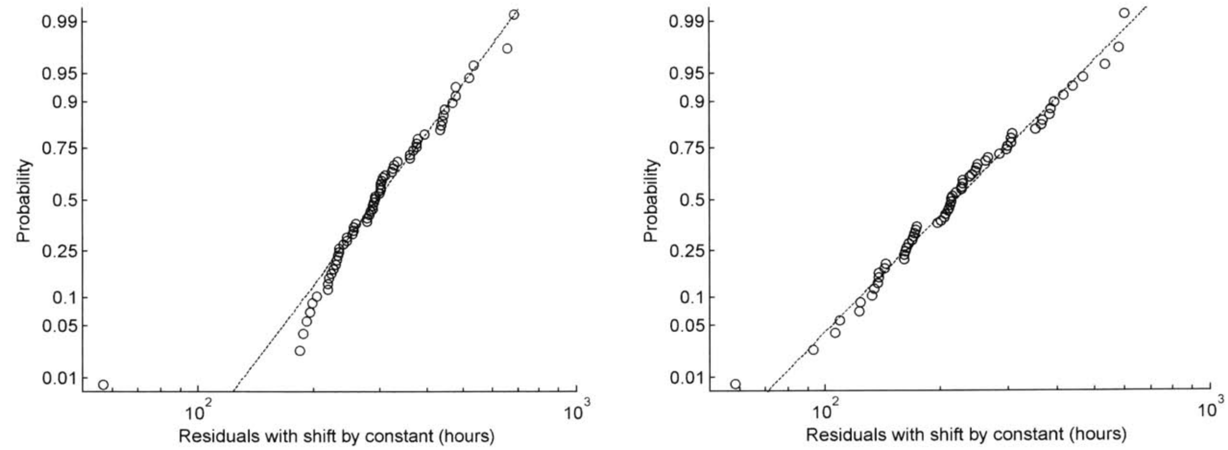
(a) Residuals for power-law fit (b) Residuals for exponential fit

Figure 4-15: A plot of residuals in sequential order suggests homoscedasticity.



(a) Residuals for power-law fit (b) Residuals for exponential fit

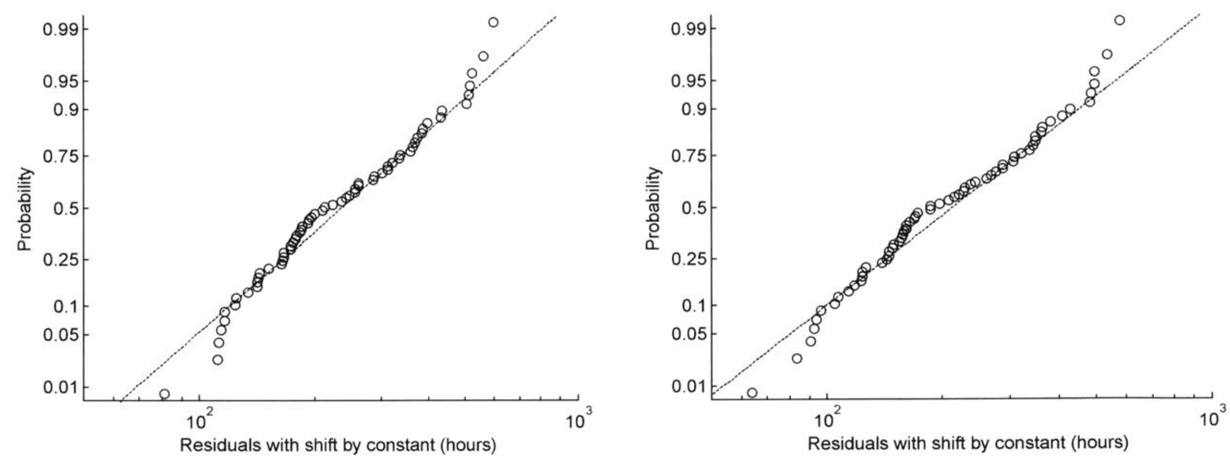
Figure 4-16: Lognormal probability plots for residuals (with constant shift) of power and exponential learning curve fits for first group (out of three groups) of wells.



(a) Residuals for power-law fit

(b) Residuals for exponential fit

Figure 4-17: Lognormal probability plots for residuals (with constant shift) of power and exponential learning curve fits for second group (out of three groups) of wells.



(a) Residuals for power-law fit

(b) Residuals for exponential fit

Figure 4-18: Lognormal probability plots for residuals (with constant shift) of power and exponential learning curve fits for third group (out of three groups) of wells.

efficiency due to the multiplicative effects of any errors [88]. This is likely to be the case in drilling because mistakes may interact with each other and have compounding effects. Lognormality also often accompanies systems that are positive-constrained. Interestingly, here the data faces a non-zero lower constraint in the form of a technical limit, meaning that the waste time in the system, rather than the actual time, is lognormally distributed. This point emphasizes the importance when characterizing variability in a system of considering physical constraints and appropriate data transformations that facilitate the identification of common patterns.

Though the approach and assumptions used here to characterize performance variability are similar to those used with productivity in Chapter 3, it is important to consider the limited extent of the data and place less credence in the findings. To form broad conclusions about patterns of performance variability in drilling operations, it would be necessary to investigate further the presence of learning curve noise in other drilling campaigns and further refine the procedure for shifting the data. Additionally, it would be worthwhile to investigate performance variability for specific sets of procedures to determine if there are similar patterns across all activities, or if certain activities are more variable and others more consistent. Still the analysis here illustrates the importance of considering patterns of variability through the use of empirical models of data (in this case learning curves), data transformation, and graphical methods of data analysis. Extreme accuracy in the characterization of variability may be less important than the selection of a convenient and reasonable model, such as the lognormal distribution here, that enables the use of other statistical methods. This naturally depends on the modeling goals.

Characterizing variability in performance is an important consideration and can be used to analyze and compare the effectiveness of rigs, crews, and management in a rigorous manner. The assumption of a reasonable distribution can also be used to better understand the technical limit, which can be related to some very low probability value at the low end of the distribution of waste times.

Furthermore, it is apparent from these findings that ordinary least squares regression is not appropriate for fitting learning curves, because the assumption that errors are normally distributed fails. Other approaches, such as robust fitting or using a log-linear approach may be more reliable. Consideration of the variability within the curve could even be used to develop a Bayesian model of learning that takes into account natural variability, uncertainty about the learning parameter, and uncertainty about the technical limit. Such a model would be an extremely powerful way to rigorously capture learning dynamics while avoiding some of the pitfalls of curve fitting—particularly early on—and would make it possible to incorporate more information into the process. This is a very promising area for future work, although it is beyond the scope of this thesis.

Most importantly, when evaluating continuous improvements and inferring effects of changes made to the drilling process, it is not enough to look at absolute changes. Improvement must be measured statistically and with reference to the learning curve. Attempts to design experiments and evaluate the effectiveness of changes will otherwise be misguided and will underestimate the number of trials needed to estimate a causal impact with any confidence.

Chapter 5

Closing remarks

Thus far, development of shale gas and tight oil resources has proceeded at a rapid pace without an adequate understanding of the underlying drivers of drilling and production variability in fields. If this continues, it is clear that large gains in development efficiency and effectiveness will be left on the table. The pressure from lower oil prices and increasing scrutiny on the long-term resource potential marks a changing landscape for companies engaged in shale gas and tight oil production. No longer can they rely on eager investors to underwrite unpredictable and inconsistent results. Characterization of performance variability that has thus far been intrinsic to drilling and production in these fields is the first critical step toward disentangling aspects of uncertainty that can be mitigated through better practices or anticipated through applied statistics and data analytics. Only when these challenges are properly understood and the resulting patterns of variability recognized, can the unconventional development process be made into a "conventional" development process that is both streamlined and reliable.

The focus of this concluding chapter will be on broader implications of the findings in this thesis and future areas that should be addressed.

5.1 Statistical importance of considering variability

It should be clear from the analysis in this thesis that it is important and worthwhile to not only incorporate uncertainty into forecasts and models, but also to choose a reasonable distribution for doing so. In Section 4.3, I showed that a simple and widely accepted statistical approach like least squares regression fitting of learning curves may be misguided because error is not normally distributed—a key assumption of the approach which is unfortunately often not questioned. In natural systems and systems in which values can only take on positive values (or are constrained by some other lower bound, like a technical limit) there is strong precedent to look for a lognormal distribution shape[1]. Although, it is often under-appreciated and viewed as simply an off-shoot of the more popular normal distribution, many in the scientific community regard the lognormal distribution as equally fundamental and more useful than the normal distribution in many situations. The ease with which log-transformations can be carried out with basic computational or spreadsheet software means properties described by this distribution can leverage the abundant statistical approaches designed for normal distributions[2].

Historically, there is some bias in industry toward consistent overestimation of reserves, and the use of a distribution like the lognormal one, which probabilistically yields more low values than high values, can help to reign in this tendency [131]. No distribution will perfectly capture real world behavior and thus when using such a model it is also important to keep in mind any systematic deviations. In Section 3.2.3, deviations from lognormality at the extreme ends of the empirical distribution were discussed, and the deviations found reinforce the need to manage industry expectations, because the amount of very low-performing wells exceeds, and amount of very high-performing wells falls short of, what is expected even with this skewed distribution.

There are statistical implications to the prevalence of the lognormal distribution in shale gas and tight oil well productivity, especially related to the nature of sampling from this distribution. With a skewed distribution for productivity, like the lognormal one found in this thesis, an individual well is more likely to fall below than above the population mean for a

[1]Occasionally, other less empirically-based distributions are used to represent variability of geological properties, such as the triangular distribution, which has been used by USGS [116]. This can lead to overestimation as compared to use of the lognormal distribution, so it is important to consider whether use of such a distribution is appropriate for the system [131].

[2]There are some situations, such as when there are a very small number of samples, in which log-transformation of a lognormal distribution can introduce some bias. Skeweness often leads to tradeoffs between bias and error. A good reference on this topic is a paper by Blackwood on the lognormal distribution in environmental systems [11].

field. This makes the production risk higher for drilling one well than for drilling many wells. In fact the sample mean increases with the number of samples, making mean a misleading measure of central tendency unless a very large number of wells are drilled [131]. Median well performance, which is a better estimate of the "typical outcome" (rather than "total outcome" across many wells), is notably lower than mean well performance in a field [11]. In Chapter 3, Table 3.1 showed that within counties, the mean often differs from the median by greater than 10%. This underlies the importance of considering more than one measure of central tendency and spread when considering variability in shale and tight plays. Furthermore, this nuance should be taken into consideration when assessing the environmental footprint of developing these resources because no single, static metric will capture the distribution of impacts fully.

5.2 Improving decision-making in unconventionals

Developing reasonable assumptions about probability distributions of outcomes can enable greater incorporation of statistical approaches into decision-making in unconventionals. For instance, in Section 3.3 I demonstrated the usefulness of a Bayesian approach to inferring the parameters of the productivity distribution within specific areas of a play early in development. This example only skims the surface of the predictive potential of a more fully developed spatial Bayesian model of unconventional field productivity. Additionally, based on the lognormality found for learning curve noise in Section 4.3, a Bayesian method can be developed for evaluating continuous improvements and probabilistically predicting progress on the learning curve. There are advantages to Bayesian statistical approaches over those based entirely on classical statistics, as was discussed in Section 3.3. Foremost though is the ability to expand simple models to include additional information, such as through hierarchical models, and eventually develop machine learning techniques based on them.

In addition to direct statistical and data mining techniques, it is important to base strategies on an awareness of the nature of performance variability. In the old paradigm of conventional oil and gas development, the resources lowest on the cost-curve could generally be systematically identified and produced first. However, in unconventionals the wide variability in production rates, even within sweet-spot areas, means that thus far, beyond aiming to be in higher quality acreage, it is impossible to avoid sampling from the entire supply curve for an area simultaneously. Additionally, the drilling process—which benefits from improvements in efficiency over time due to learning—remains hampered by a poor understanding of how to manage natural and intentional (i.e. experimentally-driven) variability in drilling activity. There is a need to counter the prevailing "factory drilling" mentality with a more formalized understanding of uncertainty; the notion that these are "engineering" plays that simply need to be "manufactured" understates the importance of performance variability. If anything, these are *statistical plays*, and development potential can only be fully realized with breakthroughs in the ability to leverage field data into planning and decision-making.

Learning in unconventional fields requires experimentation with different methods, but as Section 2.3 discussed, this must be carried out in a systematic manner with comprehension of natural and explainable variability. This is equally true for drilling, in which continuous improvements require statistical inference of effects, and completions, in which many hydraulic fracturing parameters are evaluated for their effect on well productivity. Outcomes really need to be thought of as a combination of both an effect variable and an unknown random variable. Experimental design and related methods can utilize knowledge about observed variance in order to statistically disentangle these effects. Design of experiments can and should be driven by the value of information since they are taking place in a development context. For example, when it is possible and worthwhile, large individual changes rather than multiple, small changes to a system will enable inferences to be reached with greater confidence in fewer trials. "Tampering"—or making constant changes to drilling and completions processes while failing to consider an underlying randomness to outcomes—may actually undermine learning efforts by limiting the ability for standardization to lead to greater efficiency of implementation. Some coordination may be required because several teams involved in a well's design and construction may be pursuing different initiatives. In such a case, each group has a desire to see their own efforts prove successful, but it is difficult to determine how much any one of the initiatives can be attributed with an outcome. There may actually be a bias in management toward over-experimentation because those who have had successful experiments in the past are often promoted to these roles and they may have an unbalanced and overconfident perspective on experimentation [108, 103].

Industry culture plays an important role in decision-making and should be revisited in light of these findings. There is a bias in industry toward wanting to drill more wells, which is partly reinforced by financial incentives and social pressures within companies [131]. This culture is actually a holdover from the conventional mindset, in which drilling was viewed as the only way to really find a very productive well or prospect, even if this actually destroys value in the process [131]. This "hero's journey" mentality (also referred to as the "prospector myth") is a belief that persistence will pay off in the end, and that the odds can be beat [131]. Like a gambler at a slot machine, failures are tolerated because each one is assumed to bring the player one step closer to an enormous payout that will make up for all of the losses. This kind of wasteful and dangerous mindset has been identified in the industry, even prior to the shale gas and tight oil development of recent years [131]. I simply wish to re-emphasize here that, if anything, appreciating the "table odds" is even more important with unconventionals, and part of the importance of formally describing variability of performance in these systems is to move toward a more rational approach to decision-making.

Characterization of the statistical behavior in these systems can supply the foundation needed for developing additional statistical and data mining techniques to improve decision-making in unconventional development. This will enable further incorporation of abundant field data into the planning and engineering of fields. The work presented in this thesis is an important step toward distinguishing the "signal" from the "noise," in performance-related field data but further work will be needed to accomplish this. Drilling operations and initial production data have been analyzed to demonstrate the wide range of opportunities across the development spectrum. However, it is important to also consider how hydraulic fracturing varies both intentionally (varied designs) and stochastically (in terms of unpredictability of fracture growth, for instance). Additional production history data should be analyzed to identify trends such as more rapid decline in some situations, evidence of well interference, and other important patterns. Geological data can also be incorporated into models like the one for well productivity here, in order to better understand the factors leading to differences in outcomes.

5.3 Paving the rocky path forward for shale

If unconventional oil and gas is to play a major role domestically and globally in the coming decades, we will need a concerted research effort into the underlying physics that govern shale production. The characterization in this thesis of well-to-well production variability should aid this effort by providing a statistical basis for evaluating the role of different geological and completion parameters and stochastically modeling subsurface behavior. For instance, detection of well interference, or shared pressure drawdown area, may be possible by measuring statistical differences in wells drilled at different times within a certain spatial proximity.

Despite the apparent stochasticity of production to date, with better scientific understanding, improved techniques for extraction, and more optimal decision-making, it may still prove possible to shift the productivity distribution toward fewer poor-performing wells and more high-performing wells. Inclusion of additional variables may eliminate some of the productivity uncertainty, as there are surely properties beyond well length that can control for some of the observed variability. At the very least, well placement can be improved beyond the current approach of widespread drilling of evenly-spaced wells across an entire play, and increased density drilling in sweet-spot areas. The ability to extract more resource with fewer wells will not only improve the economics of this resource but will also reduce the environmental footprint associated with development.

Economical sustained production from shale gas and tight oil plays in North America will require well productivity gains and the reduction in development costs to keep up with (or outstrip) the pace at which the best acreage in fields is exploited. This allows production to be sustained through more effective development of the naturally less productive parts of a play. Understanding the balance that can be struck between these forces is the key challenge for those attempting to assess the long-term productive potential of these resources. Reliable resource evaluations that account for production variability across fields and systematic changes in development costs will help policymakers, public utilities, and the private sector make informed decisions about long-term energy infrastructure investments.

If the economic benefits of unconventionals—and public health benefits of switching electricity generation from coal to shale gas—are to be realized outside of North America it is important to codify learning from the development of these resources in North America. This will enable better identification of promising resource plays and the best drilling areas within them, based on a strong statistical understanding of variability within plays. It will also ensure that learning from drilling in North American unconventionals is transferred to these other areas, and that the best techniques for managing learning and experimentation are deployed. Development of shale and tight resources outside of North America faces greater economic hurdles than it did in the U.S. and Canada, and it will be necessary to minimize and remove inefficiencies very early on if these efforts are to be successful.

Of course, looming over all of this today is the weak global oil price, presenting a new hurdle to unconventional resource development both abroad and in North America as well. A year ago, despite all of the uncertainties associated with drilling and production in unconventionals, strong oil prices were taken as a given. Now, this has changed and future price trajectory is unclear. This makes it more important than ever to understand these other sources of uncertainty and find methods and strategies to mitigate or anticipate them. The inefficiencies of yesterday allow us to frame the right questions today, and with inspiration and insight, find a better solution for tomorrow.

Appendix A

Tables

The plays used in Chapter 3 include Barnett, Bakken, Marcellus, Eagle Ford, and Haynesville. In order to generate the database of wells for the study the filtering criteria in Table A.1 was used for each play. General details about the plays are summarized in Table A.2, including the location, geological age, extent, depth, average thickness, total organic carbon, porosity, and technically recoverable resources. This information has been assimilated from a range of references and is only intended to be descriptive of the plays. Many of the values still have a high degree of uncertainty associated with them.

Tables A.3 and A.4 present additional results from other plays to support the findings of consistent absolute variability in different Barnett counties (Table 3.1) and vintages (Table 3.2).

Tables A.5 to A.7 describe some example drilling procedures and the steps involved in calculating operational dissimilarity using ASM, as discussed in Chapter 4.

Table A.1: Filtering criteria used for wells in each play.

Play	Barnett	Bakken	Marcellus	Eagle Ford (Gas)	Haynesville
State(s)	Texas	North Dakota	Pennsylvania	Texas	Texas, Louisiana
Number of wells	9,138	5,483	1,584	2,159	2,221
Production type	Gas	Oil; Oil and gas	Gas	Gas	Gas
Well type	Horizontal	Horizontal	Horizontal	Horizontal	Horizontal
Status	Active	Active	Active	Active	Active
Earliest start date	2005	2009	2009	2010	2009
Depths	At least 1,000 feet	At least 1,000 feet	At least 1,500 feet	At least 1,000 feet	At least 1,000 feet
Perforation length	At least 1,000 feet	At least 1,000 feet	At least 1,500 feet	At least 1,000 feet	At least 1,000 feet
Peak rate	Highest rate in first 3 months	Highest rate in first 6 months	Highest rate in first 3 months	Highest rate in first 3 months	Highest rate in first 3 months
Production history	At least 3 months of production	At least 6 months of production	At least 3 months of production	At least 3 months of production	At least 3 months of production

Table A.2: General properties of the plays included in the characterization of productivity variability.

Play	Barnett	Bakken	Marcellus	Eagle Ford (Gas)	Haynesville
Location	North Texas [30]	North Dakota, South Dakota, and Montana [56]	Ohio, Pennsylvania, New York, and West Virginia [30]	South Texas [30]	East Texas and northwest Louisiana [30]
Geological age	Mississippian [30]	Devonian-Mississippian [151]	Devonian [30]	Cretaceous [30]	Jurassic [30]
Extent, sq mi	6,500 [48]	6,500 [48]	95,000 [48]	3,300 [48]	9,000 [48]
Depth, ft	7,500 [48]	4,500-7,500 [48]	4,000-8,500 [48]	7,000 [48]	10,500-13,500 [48]
Average Thickness, ft	300 [48]	22 [48]	125 [48]	200 [48]	250 [48]
Total organic carbon (TOC), %	4.2 [30]	11 [140]	4.8 [30]	3.8 [30]	2.4 [30]
Porosity, %	5 [120]	8 [48]	8 [48]	9 [48]	8.5 [48]
Technically recoverable resources[1]	100 MMBO; 26,700 BCFG; 1,100 MMBNGL [127]	7,400 MMBO; 6,700 BCFG; 500 MMBNGL [56]	84,200 BCFG; 3,400 MMBNGL [26]	900 MMBO; 51,900 BCFG; 2,000 MMBNGL [39]	60,700 BCFG [39]

[1]MMBO≡ million barrels of oil. BCFG ≡ billion cubic feet of gas. MMBNGL ≡ million barrels of natural gas liquids.

Table A.3: Summary statistics for specific productivity of counties in all plays. mscf/mo/ft is thousand standard cubic feet per month per foot. bbl/mo/ft is barrels per month per foot.

Play	**County**	**P90-P10**	**P50 (mscf/mo/ft)**	**Mean (mscf/mo/ft)**
Marcellus	Bradford	3.45	24.84	28.18
Marcellus	Lycoming	3.38	30.20	33.28
Marcellus	Susquehanna	6.67	55.79	61.72
Marcellus	Tioga	3.71	20.39	23.04
Marcellus	Washington	3.81	19.40	25.42
Eagle Ford	Dewitt	2.83	23.01	26.45
Eagle Ford	Dimmit	3.16	8.77	9.43
Eagle Ford	Karnes	2.77	20.18	21.34
Eagle Ford	Lasalle	4.16	18.85	19.81
Eagle Ford	Webb	3.70	20.77	22.87
Haynesville	Caddo	1.83	51.97	54.89
Haynesville	De Soto	2.58	65.88	67.89
Haynesville	Panola	2.38	40.59	43.61
Haynesville	Red River	3.00	86.26	89.30
Haynesville	Sabine	2.25	60.12	62.52
Play	**County**	**P90-P10**	**P50 (bbl/mo/ft)**	**Mean (bbl/mo/ft)**
Bakken	Dunn	4.38	1.62	1.78
Bakken	McKenzie	3.35	1.81	1.98
Bakken	Mountrail	4.33	1.75	2.04
Bakken	Williams	3.44	1.35	1.47
Bakken	Divide	3.04	0.96	1.04

Table A.4: Spread of productivity by vintage in all plays.

Play	**Vintage**	**P90-P10**
Marcellus	2010	4.45
Marcellus	2011	4.56
Marcellus	2012	7.03
Marcellus	2013	7.39
Bakken	2009	6.14
Bakken	2010	4.47
Bakken	2011	4.12
Bakken	2012	4.20
Bakken	2013	3.94
Eagle Ford	2010	4.41
Eagle Ford	2011	4.36
Eagle Ford	2012	4.25
Eagle Ford	2013	5.39
Haynesville	2010	3.12
Haynesville	2011	2.81
Haynesville	2012	2.18

Table A.5: Descriptions of operations included in Figure 4-1.

Operation	**Description**	**Operation code**
Rig up/down	On-site equipment and materials preparation for any operation	RUD
Drill string trip	Run-in-hole and pull-out-of-hole with assembly	DRT
Pump/ circulate/ displace	Pumping fluid into the well down the string or annulus when not carrying out other operations such as drilling, coring, hole opening, milling, reaming, or washing down, working on pipe, pressure testing or cementing	PCD
Safety related	Covers non-routine safety activities that are not specifically covered by another operation code	SAF
BOP install/ remove	Includes all work in assembling/removing BOPs or diverter	BPR
BOP test	All activities associated with the function and pressure testing of BOPs	BPT
Inflow test	Pressure testing using the formation pressure as a source of pressure	INT

Table A.6: Description and examples of the edit operations within ASM and corresponding procedural changes they identify.

Edit type	**Description of drilling procedural change**	**Example of procedural change between two wells**
Insertion of a character	An operation was inserted between two existing operations.	Break up drilling sequence with circulation of bottoms up
Deletion of a character	An existing operation was removed from the drilling procedure.	Slide drill continuously rather than breaking up slide with period of rotary drilling
Substitution of one character for another	An existing operation was replaced with a different operation in a procedure.	Drill with same bottom-hole-assembly (BHA) instead of tripping to change BHA
Transposition of two adjacent characters	Two existing sequential operations in a procedure had their order reversed.	Pump lost-circulation mud immediately before landing curve instead of immediately after landing curve

Table A.7: Potential edits, conditions restricting their use, and the method for finding the new value in each cell. *Cost* is equal to 0 if the current characters being compared are identical, and 1 otherwise, which allows for the option of making no change.

Edit type	**Conditions for this edit**	**New value from edit**
Insertion	N/A	Value in the cell to the left + 1
Deletion	N/A	Value in the cell above + 1
Substitution	N/A	Value in the cell diagonal (left and above) + *cost*
Transposition	Must not be the first character in either string; Current and previous character in one string should be reverse of current and previous character in the other string	Value in the cell diagonal two away (up two and left two) + *cost*

Appendix B

Figures

Figures B-1 to B-19 are included in this Appendix to reinforce and provide greater breadth to the findings of Section 3.2.2 of the thesis. The lognormality of productivity is demonstrated across a range of plays, categories of wells, and a different metric of initial productivity.

Figures B-20 to B-24 expand upon the contents of Figure 4-10.

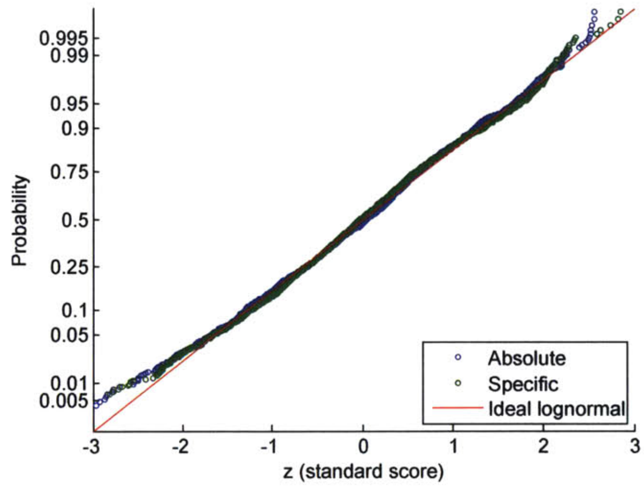


Figure B-1: Probability plot comparing absolute and specific productivity in the Marcellus to an ideal lognormal distribution.

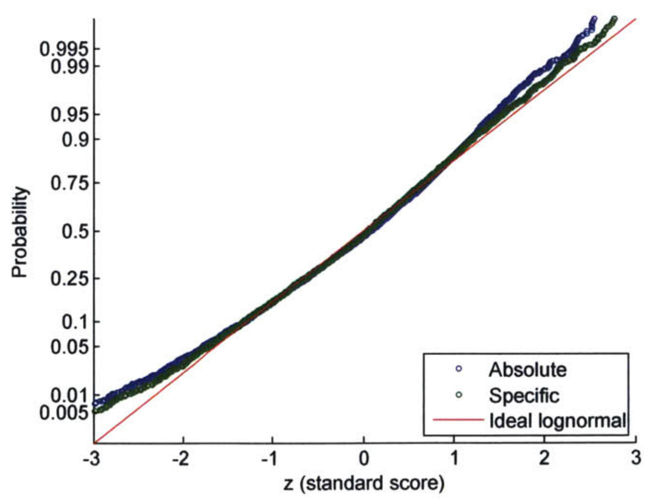


Figure B-2: Probability plot comparing absolute and specific productivity in the Bakken to an ideal lognormal distribution.

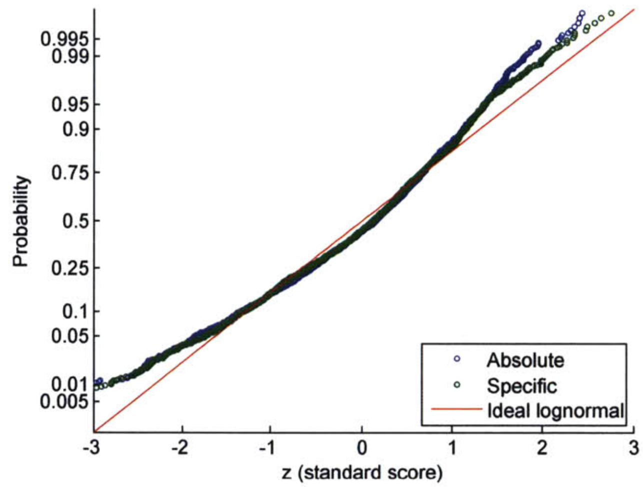


Figure B-3: Probability plot comparing absolute and specific productivity in the Eagle Ford to an ideal lognormal distribution.

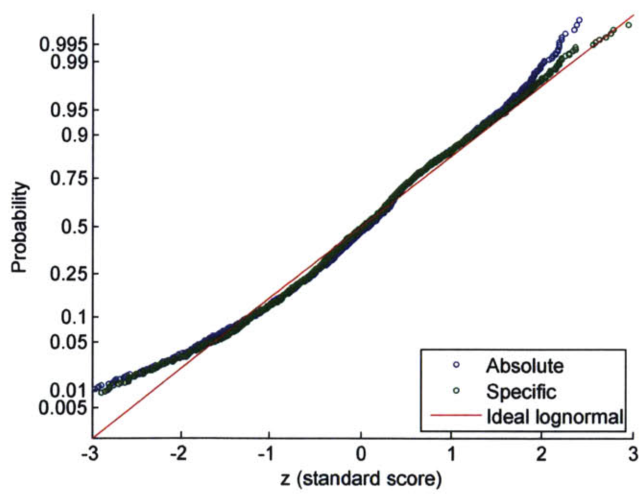


Figure B-4: Probability plot comparing absolute and specific productivity in the Haynesville to an ideal lognormal distribution.

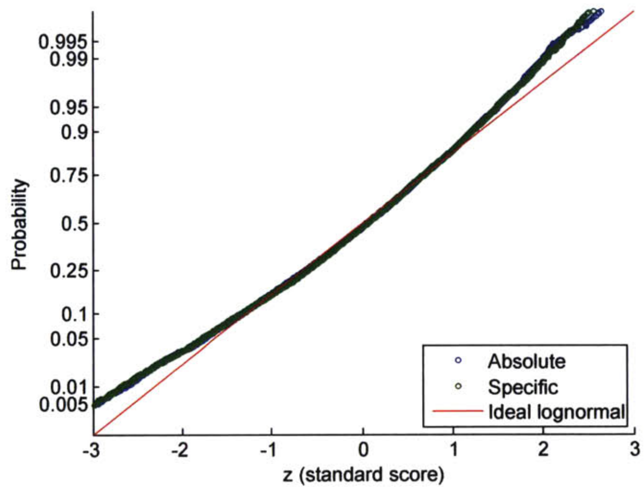
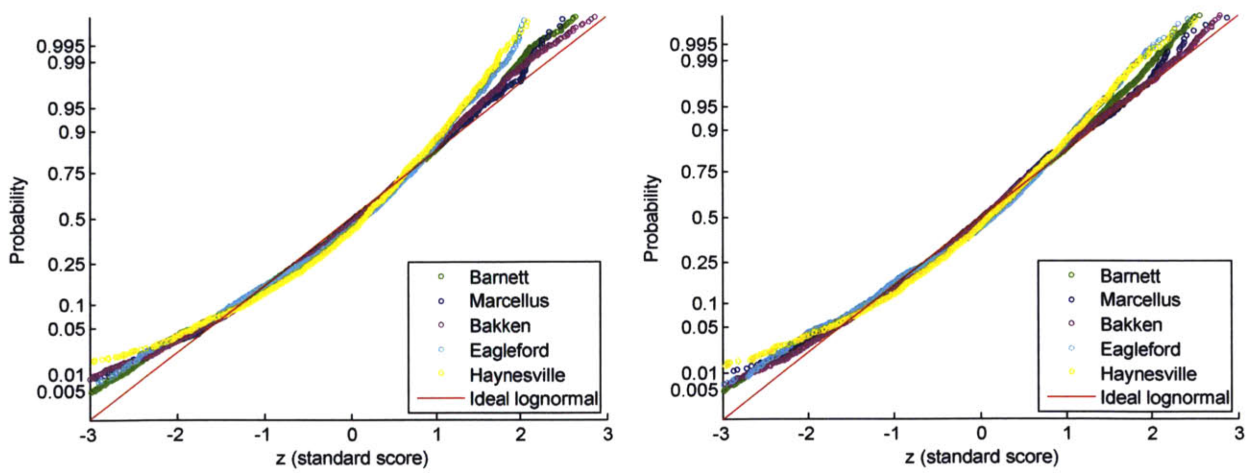


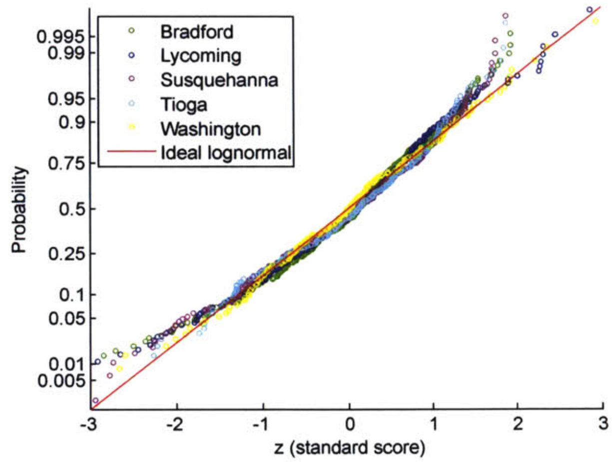
Figure B-5: Probability plot comparing absolute and specific productivity in the Barnett (based on first 12 months of production) to an ideal lognormal distribution.



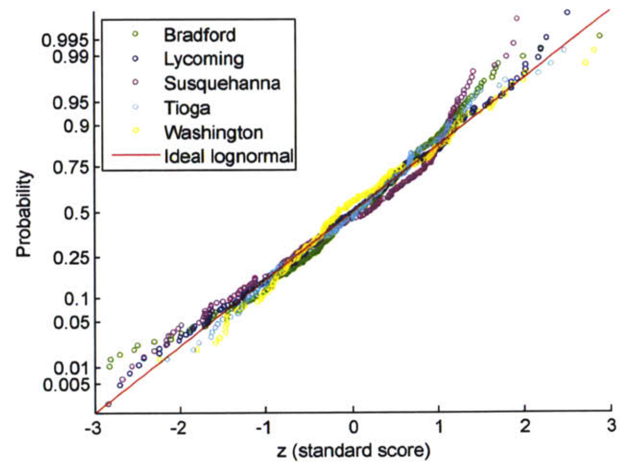
(a) Absolute productivity

(b) Specific productivity

Figure B-6: Probability plots comparing absolute and specific productivity of all plays (based on first 12 months of production) to an ideal lognormal distribution.

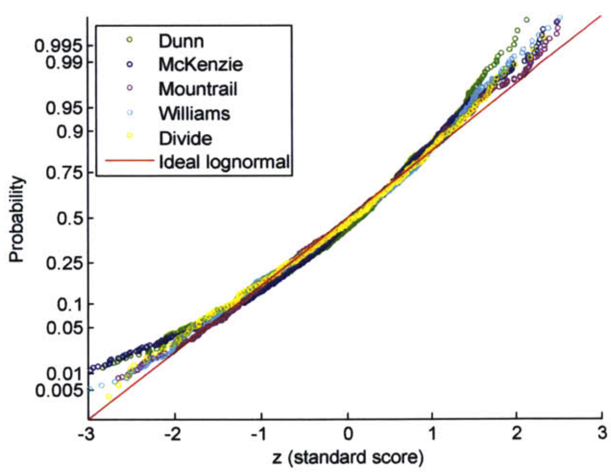


(a) Absolute productivity

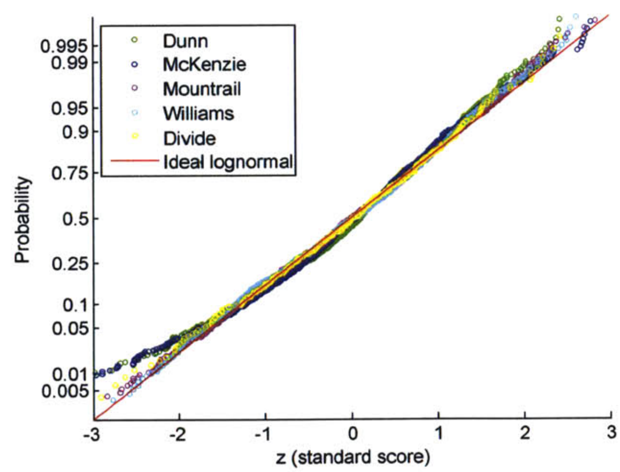


(b) Specific productivity

Figure B-7: Probability plots comparing absolute and specific productivity of well ensembles in different Marcellus counties to an ideal lognormal distribution.

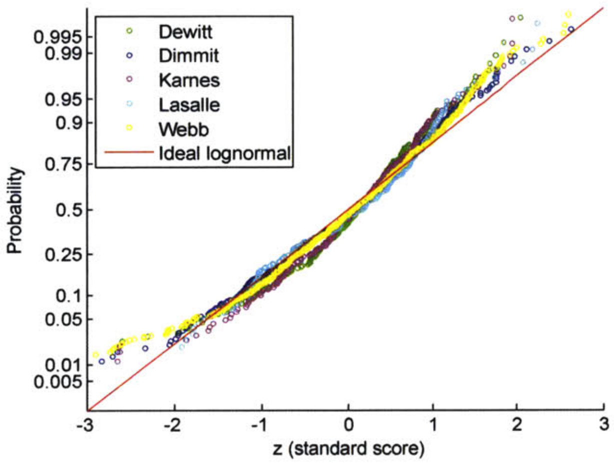


(a) Absolute productivity



(b) Specific productivity

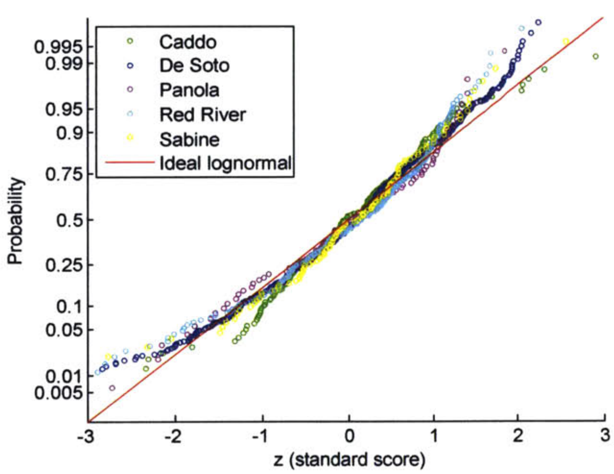
Figure B-8: Probability plots comparing absolute and specific productivity of well ensembles in different Bakken counties to an ideal lognormal distribution.



(a) Absolute productivity

(b) Specific productivity

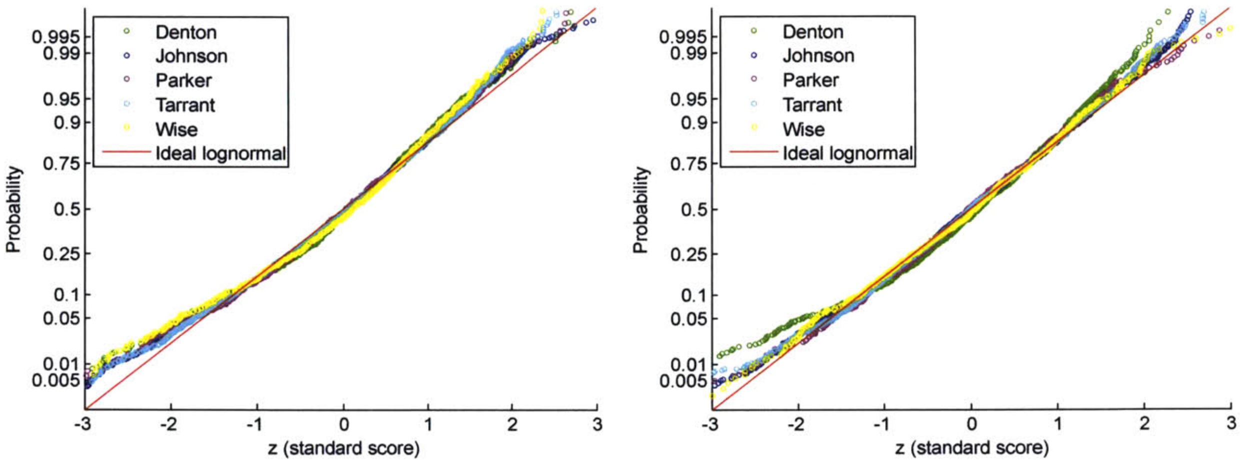
Figure B-9: Probability plots comparing absolute and specific productivity of well ensembles in different Eagle Ford counties to an ideal lognormal distribution.



(a) Absolute productivity

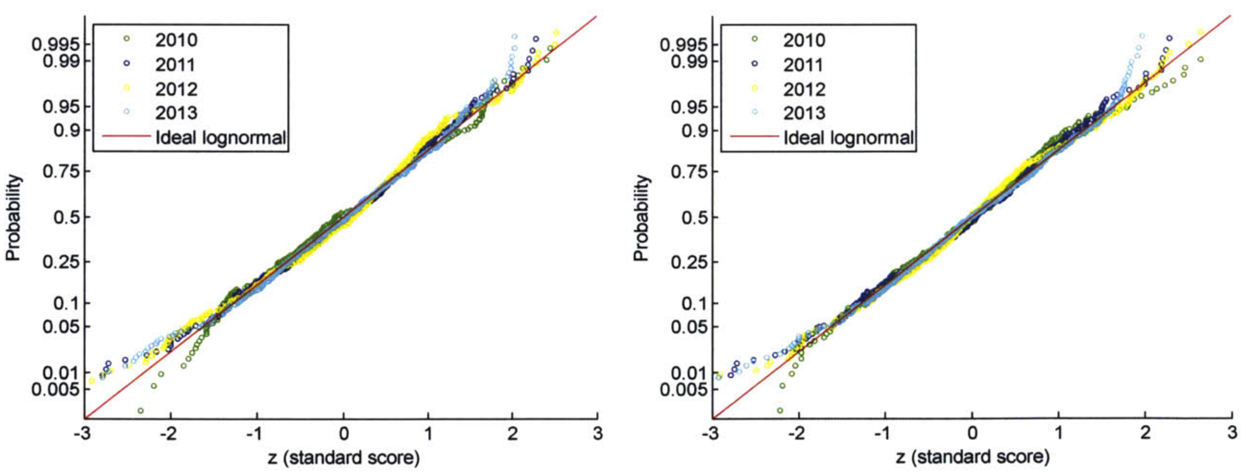
(b) Specific productivity

Figure B-10: Probability plots comparing absolute and specific productivity of well ensembles in different Haynesville counties to an ideal lognormal distribution.



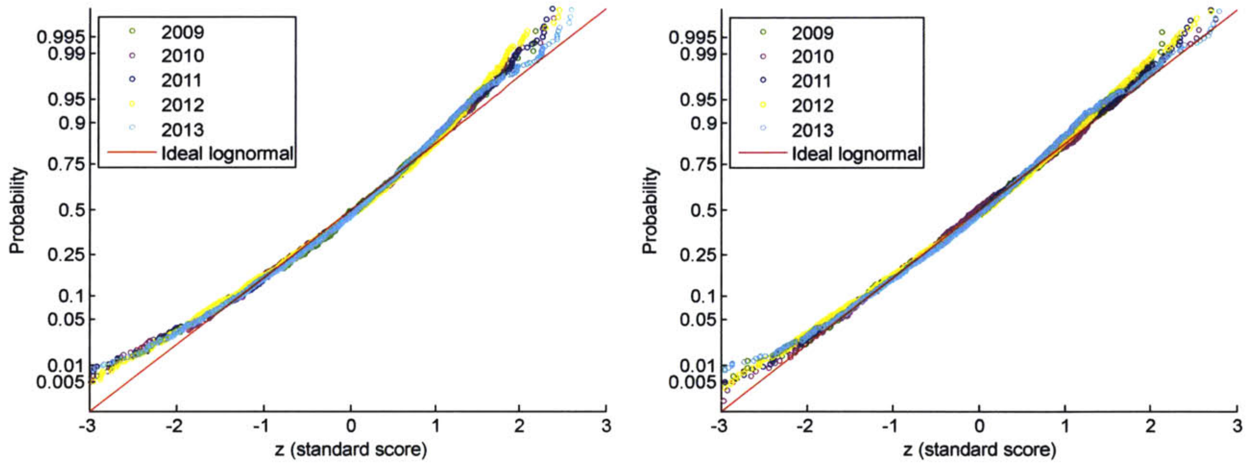
(a) Absolute productivity (b) Specific productivity

Figure B-11: Probability plots comparing absolute and specific productivity of well ensembles in different Barnett counties (based on first 12 months of production) to an ideal lognormal distribution.



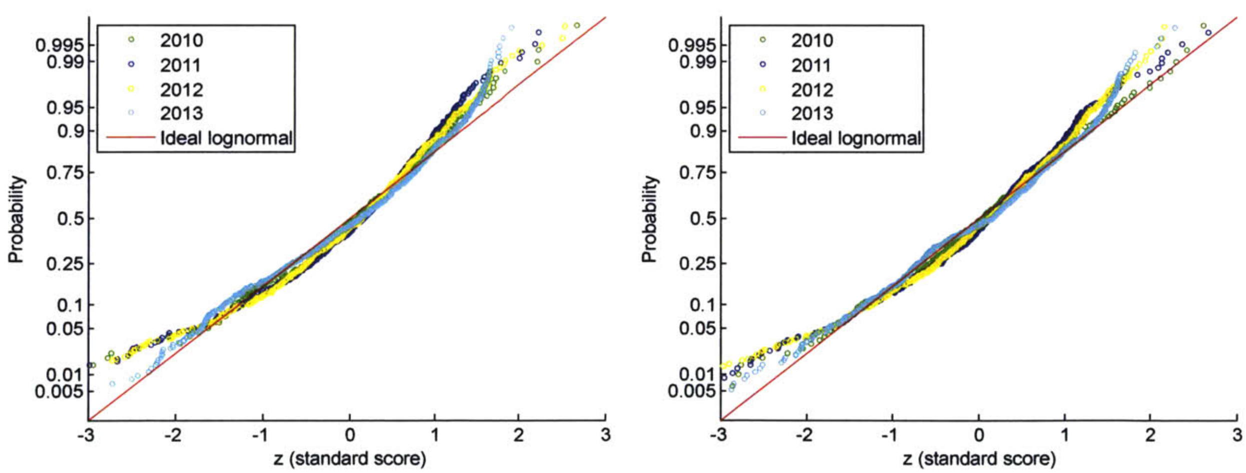
(a) Absolute productivity (b) Specific productivity

Figure B-12: Probability plots comparing absolute and specific productivity of different vintages in the Marcellus to an ideal lognormal distribution.



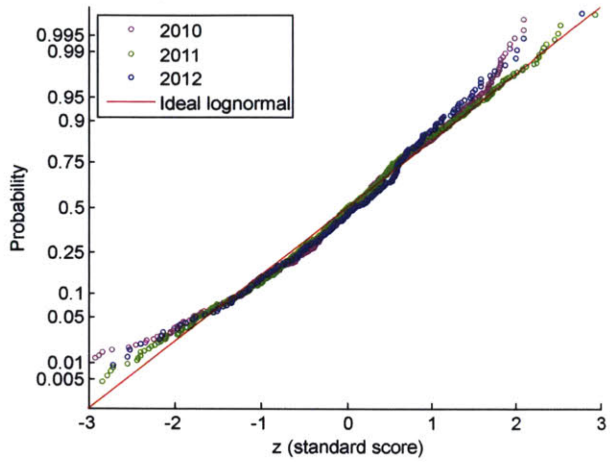
(a) Absolute productivity (b) Specific productivity

Figure B-13: Probability plots comparing absolute and specific productivity of different vintages in the Bakken to an ideal lognormal distribution.



(a) Absolute productivity (b) Specific productivity

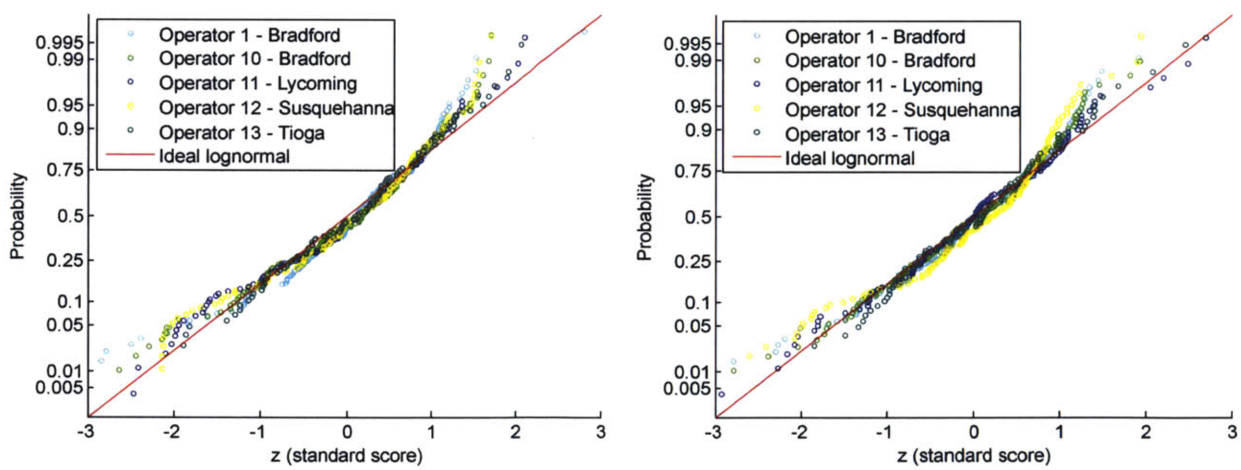
Figure B-14: Probability plots comparing absolute and specific productivity of different vintages in the Eagle Ford to an ideal lognormal distribution.



(a) Absolute productivity

(b) Specific productivity

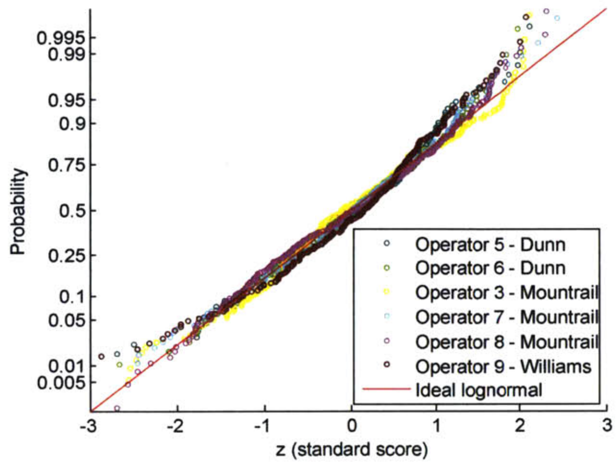
Figure B-15: Probability plots comparing absolute and specific productivity of different vintages in the Haynesville to an ideal lognormal distribution.



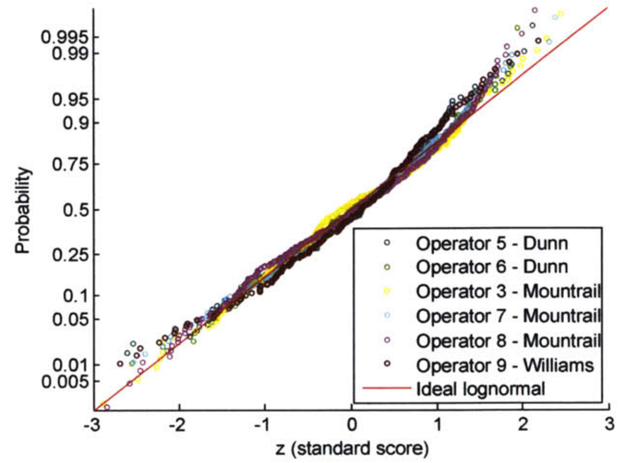
(a) Absolute productivity

(b) Specific productivity

Figure B-16: Probability plots comparing absolute and specific productivity of different operating company well portfolios in Marcellus counties to an ideal lognormal distribution.

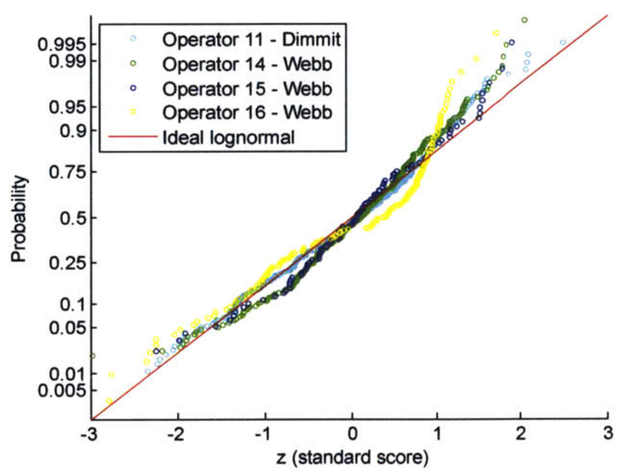


(a) Absolute productivity

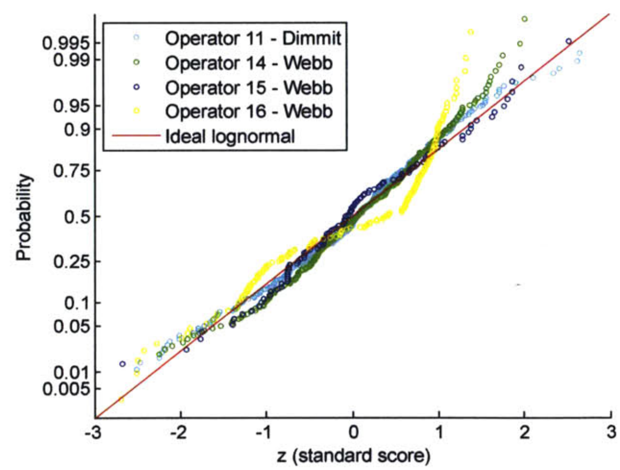


(b) Specific productivity

Figure B-17: Probability plots comparing absolute and specific productivity of different operating company well portfolios in Bakken counties to an ideal lognormal distribution.

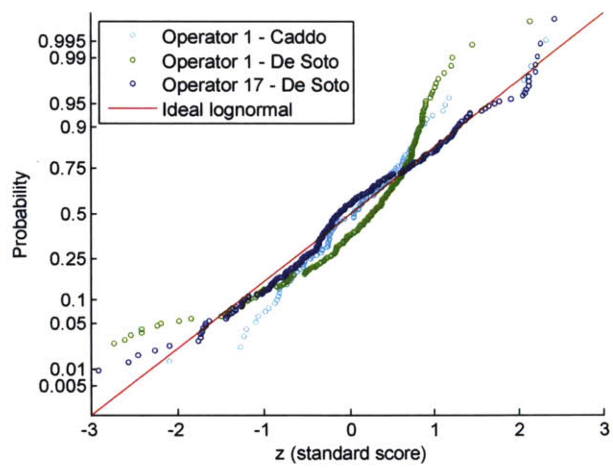


(a) Absolute productivity

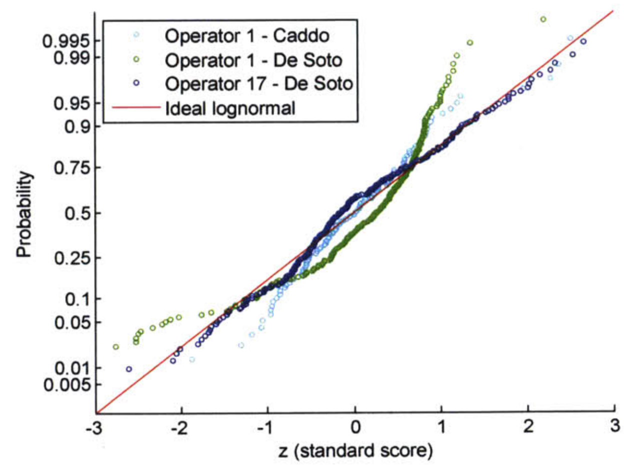


(b) Specific productivity

Figure B-18: Probability plots comparing absolute and specific productivity of different operating company well portfolios in Eagle Ford counties to an ideal lognormal distribution.



(a) Absolute productivity



(b) Specific productivity

Figure B-19: Probability plots comparing absolute and specific productivity of different operating company well portfolios in Haynesville counties to an ideal lognormal distribution.

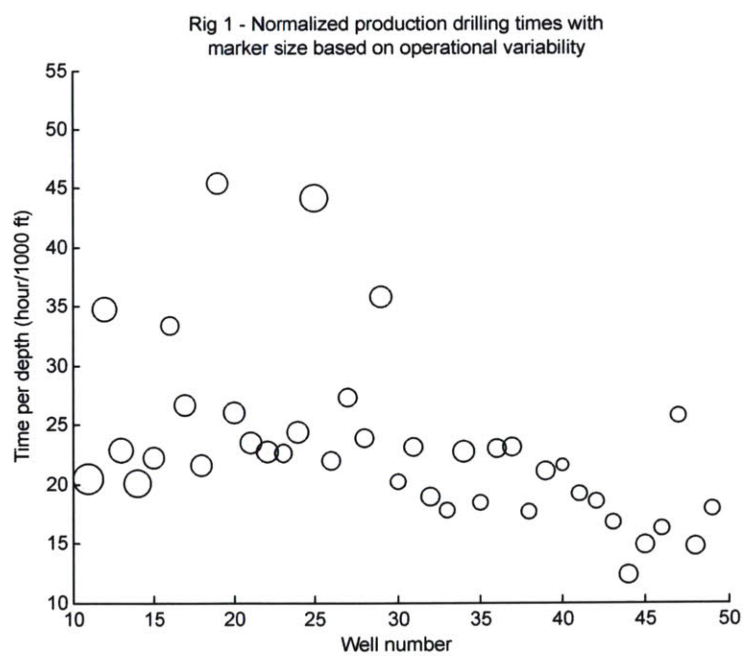


Figure B-20: Operational variability and time based performance for rig 1-drilling the production hole.

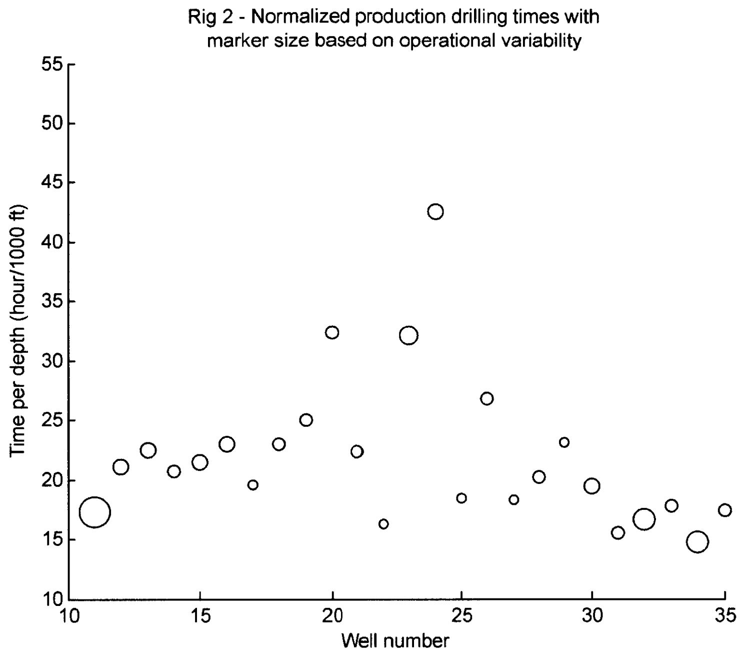


Figure B-21: Operational variability and time based performance for rig 2-drilling the production hole.

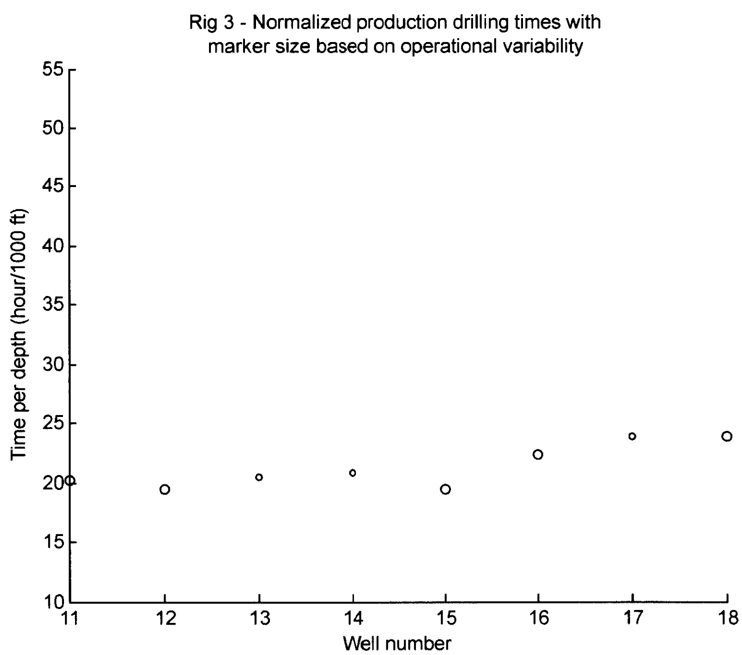


Figure B-22: Operational variability and time based performance for rig 3-drilling the production hole.

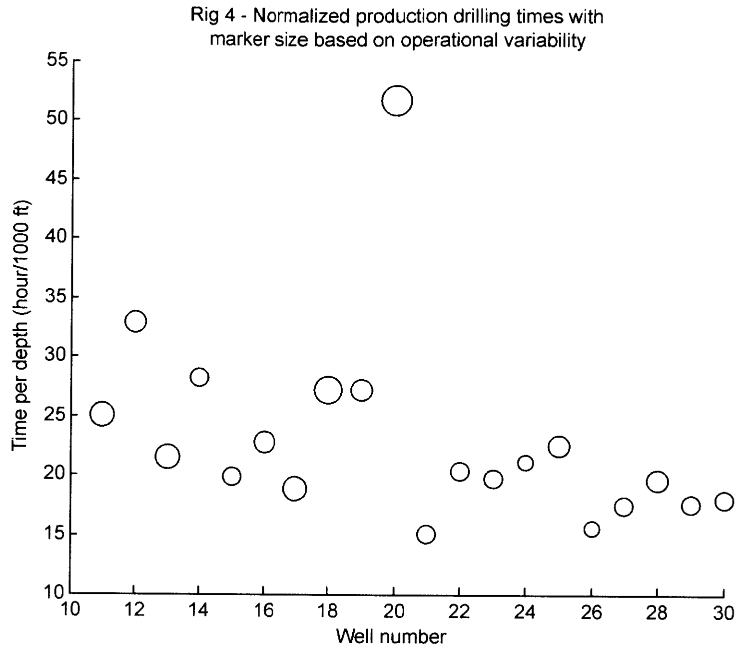


Figure B-23: Operational variability and time based performance for rig 4-drilling the production hole.

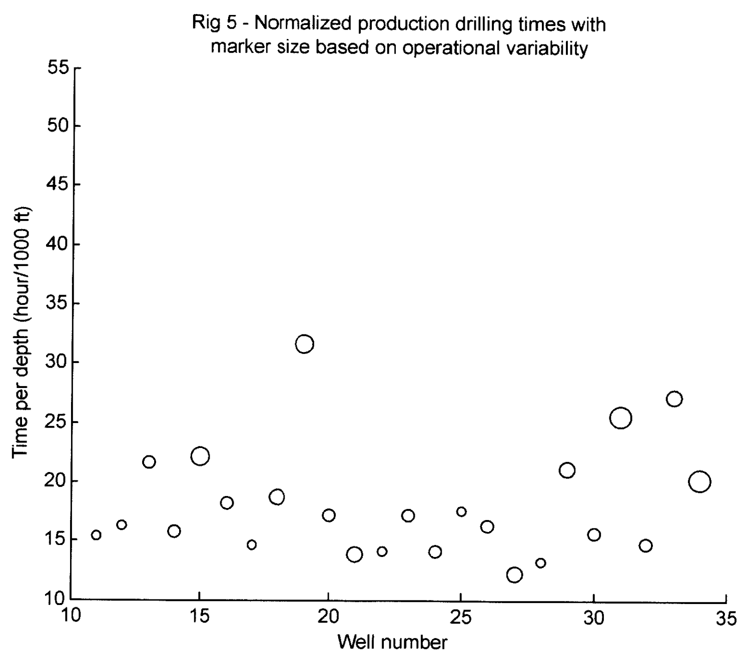


Figure B-24: Operational variability and time based performance for rig 5-drilling the production hole.

Bibliography

[1] C. Adams. Oil price fall threatens $1tn of projects. *Financial Times*, 15 Dec. 2014.

[2] John Aitchison and James A.C. Brown. *The lognormal distribution with special reference to its uses in economics.* Cambridge University Press, London, 1957.

[3] Ketil Andersen, Eric Edgar Maidla, Gerhard Thonhauser, Buddy King, and Per Arild Sjowall. Case history: Automated performance measurement of crews and drilling equipment. *Society of Petroleum Engineers/International Association of Drilling Contractors Drilling Conference and Exhibition*, 2009.

[4] G. Andreatta, G.M. Kaufman, R.G. McCrossan, and R.M. Procter. *The shape of Lloydminster oil and gas deposit attribute data*, pages 411–431. Quantitative Analysis of Mineral and Energy Resources. Springer, 1988.

[5] J. Arps and T. Roberts. Economics of drilling for cretaceous oil and gas on the east flank of the denver-julesberg basin. *American Association of Petroleum Geologists Bulletin*, 42(11):2549–2566, 1958.

[6] J.J. Arps. Analysis of decline curves. *Transactions of the American Institute of Mining Engineers*, 160:228–247, 1945.

[7] E.D. Attanasi and L.J. Drew. Lognormal field size distributions as a consequence of economic truncation. *Mathematical Geology*, 17(4):335–351, 1985.

[8] C.N. Zou Z. Yang S.Z. Tao X.J. Yuan R.K. Zhu L.H. Hou S.T. Wu L. Sun G.S. Zhang B. Bai L. Wang X.H. Gao Z.L. Pang. Continuous hydrocarbon accumulation over a large area as a distinguishing characteristic of unconventional petroleum: The Ordos Basin, North-Central China. *Earth-Science Reviews*, 126:358–369, 2013.

[9] World Bank. Understanding the plunge in oil prices: sources and implications. Report, World Bank, 2015.

[10] Dimitri P. Bertsekas and John T. Tsitsiklis. *Bayesian Statistical Inference*, book section 8. Athena Scientific, Belmont, Massachusetts, 2 edition, 2008.

[11] Larry G. Blackwood. The lognormal distribution, environmental data, and radiological monitoring. *Environmental Monitoring and Assessment*, 21:193–210, 1992.

[12] Roger Bohn and Michael A. Lapre. Accelerated learning by experimentation. *Social Science Research Network*, 2010.

[13] Paul Bommer. *A primer of oilwell drilling: A basic text of oil and gas drilling.* The University of Texas, 7 edition, 2008.

[14] Kent A. Bowker. Barnett Shale gas production, Fort Worth Basin: Issues and discussion. *American Association of Petroleum Geologists Bulletin*, 91(4):523–533, 2007.

[15] Paul G. Bradley and Gordon M. Kaufman. Reward and uncertainty in exploration programs. *American Association of Petroleum Geologists Bulletin*, 54(12):2472–2472, 1970.

[16] A. Brandt, G. Heath, E. Kort, F. O'Sullivan, G. Petron, S.M. Jordaan, P. Tans, J. Wilcox, A.M. Gopstein, D. Arent, S. Wofsy, N.J. Brown, R. Bradley, G.D. Stucky, D. Eardley, and R. Harriss. Methane leaks from North American natural gas systems. *Science*, 343:733–35, 2014.

[17] J.F. Brett and K.K. Millheim. The drilling performance curve: a yardstick for judging drilling performance. *Society of Petroleum Engineers Annual Technical Conference and Exhibition*, 1986.

[18] G. Brown and J.W. Saunders. Lognormal genesis. *Journal of Applied Probability*, 18:542–547, 1981.

[19] Michael R. Brule. Big data in exploration and production: Real-time adaptive analytics and data-flow architecture. *Society of Petroleum Engineers Digital Energy Conference*, 2013.

[20] Kent Burkholder, Ellen Miriam Coopersmith, and Jan Schulze. Appraisal excellence in unconventional reservoirs. *Society of Petroleum Engineers Canadian Unconventional Resources Conference*, 2012.

[21] Bunyamin Can and C.S. Kabir. Probabilistic production forecasting for unconventional reservoirs with stretched exponential production decline model. *Society of Petroleum Engineers Reservoir Evaluation and Engineering*, 15(1):41, 2012.

[22] John M. Chambers, William S. Cleveland, Beat Kleiner, and Paul A. Tukey. *Graphical Methods For Data Analysis.* Wadsorth International Group and Duxbury Press, Boston, 1983.

[23] H.K. Cho, K.P. Bowman, and G.R. North. A comparison of gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission. *Journal of Applied Meteorology*, 43:1586–1597, 2004.

[24] Jan Cienski. Eni joins shale gas exodus from Poland. *Financial times*, 15 Jan. 2014.

[25] C.R. Clarkson, J.L. Jensen, and S. Chipperfield. Unconventional gas reservoir evaluation: What do we have to consider? *Journal of Natural Gas Science and Engineering*, 8:9–33, 2012.

[26] J.L. Coleman, R.C. Milici, T.A. Cook, R.R. Charpentier, Mark Kirschbaum, T.R. Klett, R.M. Pollastro, and C.J. Schenk. Assessment of undiscovered oil and gas resources of the Devonian Marcellus Shale of the Appalachian Basin Province, 2011. Report 2011-3092, U.S. Geological Survey Fact Sheet, 2011.

[27] E. Crooks. US shale: What lies beneath. *Financial Times*, 26 Aug. 2014.

[28] R.A. Crovelli. A comparison of analytic and simulation methods for petroleum play analysis and aggregation. Report 86-97, U.S. Geological Survey, 1986.

[29] R.A. Crovelli. Probability theory versus simulation of petroleum potential in play analysis. *Annals of Operations Research*, 8:363–381, 1987.

[30] Mark E. Curtis, Carl H. Sondergeld, Raymond J. Ambrose, and Chandra S. Rai. Microstructural investigation of gas shales in two and three dimensions using nanometer-scale resolution imaging. *American Association of Petroleum Geologists Bulletin*, 96(4):665–677, 2012.

[31] Laurie P Dake. *Fundamentals of reservoir engineering.* Elsevier, New York, 1983.

[32] Fred J. Damerau. A technique for computer detection and correction of spelling errors. *Communications of the Association for Computing Machinery*, 7(3):171–176, 1964.

[33] John P. de Wardt. Lean drilling-introducing the application of automotive lean manufacturing techniques to well construction. *Society of Petroleum Engineers/International Association of Drilling Contractors Drilling Conference and Exhibition*, 1994.

[34] John P de Wardt. Manufacturing wells: Myth or magic. *Society of Petroleum Engineers/International Association of Drilling Contractors Drilling Conference and Exhibition*, 2012.

[35] Andres Mauricio Del Busto Pinzon. *Key Economic Drivers Impacting Eagle Ford Development from Resource to Reserves.* Texas A&M University, 2013.

[36] John Deutch. The good news about gas-the natural gas revolution and its consequences. *Foreign Affairs*, 90:82, 2011.

[37] K. Ditzel, J. Plewes, and B. Broxson. US manufacturing and LNG exports: economic contributions to the US economy and impacts on US natural gas prices. Report, Charles River Associates for Dow Chemical Company, Feb. 25 2013.

[38] L.J. Drew and J.C. Griffith. Size, shape and arrangement of some oilfields in the USA. *Symposium on Computer Applications in the Mineral Industries*, 1965.

[39] R.F. Dubiel, P.D. Warwick, Sharon Swanson, Lauri Burke, L.R.H. Biewick, R.R. Charpentier, J.L. Coleman, T.A. Cook, Kris Dennen, Colin Doolan, Catherine Enomoto, P.C. Hackley, A.W. Karlsen, T.R. Klett, S.A. Kinney, M.D. Lewan, Matt Merrill, Krystal Pearson, O.N. Pearson, J.K. Pitman, R.M. Pollastro, E.L. Rowan, C.J. Schenk, and

Brett Valentine. Assessment of undiscovered oil and gas resources in Jurassic and Cretaceous strata of the Gulf Coast, 2010. Report 2011-3020, U.S. Geological Survey Fact Sheet, 2011.

[40] L.L. Eberhardt and R.O. Gilbert. Gamma and lognormal distributions as models in studying food-chain kinetics. Report BNWL-1747, Battelle, Pacific Northwest Laboratories, 1973.

[41] M.J. Economides and K.G. Nolte. *Reservoir Stimulation*. J. Wiley, Chichester, 3 edition, 2000.

[42] The Economist. An interview with George Mitchell: The industry can no longer simply focus on the benefits of shale gas, 1 Aug. 2013.

[43] The Economist. The oil price: Dead cat rally, 10 Mar. 2015.

[44] The Economist. The petrostate of America, 15 Feb. 2014.

[45] The Economist. Energy prices: Pump aligning, 17 Jan 2015.

[46] The Economist. Cheaper oil: Winners and losers, 25 Oct 2014.

[47] Morgan R. Edwards and Jessika E. Trancik. Climate impacts of energy technologies depend on emissions timing. *Nature Climate Change*, 4(5):347–352, 2014.

[48] EIA. Review of emerging resources: U.S. shale gas and shale oil plays. Report, United States, Energy Information Administration, 2011.

[49] EIA. Technically recoverable shale oil and shale gas resources: an assessment of 137 shale formations in 41 countries outside the united states. Report, United States, Energy Information Administration, 2013.

[50] EIA. Annual energy outlook 2014 with projections to 2040. Report, United States, Energy Information Administration, 2014.

[51] Jill Feblowitz. The big deal about big data in upstream oil and gas. Report, Hitachi Data Systems, 2012.

[52] M.J. Fetkovich, A.M. Works, T.S. Thrasher, and M.D. Bradley. Depletion performance of layered reservoirs without crossflow. *Society of Petroleum Engineers Formation Evaluation*, 5(03):310–318, 1990.

[53] B. Forbes, J. Ehlert, and H. Wilczynski. The flexible factory: The next step in unconventional gas development. Report, Schlumberger Business Consulting, 2009.

[54] C.N. Fredd, J.L. Daniels, and J.D. Baihly. $40 billion learning curve: Leveraging lessons to minimize the overall investment in unconventional plays. *Society of Petroleum Engineers Middle East Unconventional Resources Conference*, 2015.

[55] Hans-Christian Freitag. New prices, challenges, and opportunities in the next wave of unconventional resource development. *Journal of Petroleum Technology*, 2015.

[56] S.B. Gaswirth, K.R. Marra, T.A. Cook, R.R. Charpentier, D.L. Gautier, D.K. Higley, T.R. Klett, M.D. Lewan, P.G. Lillis, C.J. Schenk, M.E. Tennyson, and K.J. Whidden. Assessment of undiscovered oil resources in the Bakken and Three Forks Formations, Williston Basin Province, Montana, North Dakota, and South Dakota, 2013. Report 2013-3013, U.S. Geological Survey Fact Sheet, 2013.

[57] Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis.* Texts in Statistical Science. Chapman & Hall/CRC, 2 edition, 2004.

[58] Marion Gerson. The techniques and uses of probability plotting. *Journal of the Royal Statistical Society*, 24(4):235–257, 1975.

[59] R. Gold. Fracking gives U.S. energy boom plenty of room to run. *Wall Street Journal*, 14 Sep. 2014.

[60] Xinglai Gong, Raul Alberto Gonzalez, Duane McVay, and Jeffrey D. Hart. Bayesian probabilistic decline curve analysis quantifies shale gas reserves uncertainty. *Society of Petroleum Engineers Canadian Unconventional Resources Conference*, 2011.

[61] Raul Alberto Gonzalez. Probabilistic decline curve analysis reliably quantifies uncertainty in shale gas reserves regardless of stage of depletion. *Society of Petroleum Engineers Eastern Regional Meeting*, 2012.

[62] Quanxin Guo, Lujun Ji, Vusal Rajabov, James Friedheim, Christin Portella, and Rhonna Wu. Shale gas drilling experience and lessons learned from Eagle Ford. In *Society of Petroleum Engineers Americas Unconventional Resources Conference.*

[63] Stephen K. Hacker, Bertrand Jouslin de Noray, and Cindy Johnston. Standardization versus improvement: Approaches for changing work process performance. Report, European Quality, 2001.

[64] Brent W. Hale and William M. Cobb. Barnett shale: A resource play-locally random and regionally complex. *Society of Petroleum Engineers Eastern Regional Meeting*, 2010.

[65] Russell Hall, Robin Bertram, Gary Gonzenbach, Jim Gouveia, Brent Hale, Paul Lupardus, Paul McDonald, Bill Vail, and Marshall Watson. Guidelines for the practical evaluation of undeveloped reserves in resource plays. *Society of Petroleum Evaluation Engineers Monograph Series*, 3, 2010.

[66] Ursula Hammes, H. Scott Hamlin, and Thomas E. Ewing. Geologic analysis of the Upper Jurassic Haynesville Shale in East Texas and west Louisiana. *American Association of Petroleum Geologists Bulletin*, 95(10):1643–1666, 2011.

[67] William J Haskett. Unconventional type curves: Useful, or sirens of destruction? *Society of Petroleum Engineers Annual Technical Conference and Exhibition*, 2011.

[68] William J. Haskett and P. Jeffrey Brown. Recurrent issues in the evaluation of unconventional resources. *American Association of Petroleum Geologists Search and Discovery*, 40674:12–15, 2011.

[69] Bo Hedberg and Sten Jönsson. Designing semi-confusing information systems for organizations in changing environments. *Accounting, Organizations and Society*, 3(1):47–64, 1978.

[70] Robert A. Hefner III. The United States of gas. *Foreign Affairs*, 93(3):9–14, 2014.

[71] Keith Holdaway. *Harness Oil and Gas Big Data with Analytics: Optimize Exploration and Production with Data-Driven Models.* John Wiley & Sons, Inc., Hoboken, 2014.

[72] Robert W Howarth, Anthony Ingraffea, and Terry Engelder. Natural gas: Should fracking stop? *Nature*, 477(7364):271–275, 2011.

[73] HPDI. HPDI production database, 2014.

[74] J. David Hughes. Energy: A reality check on the shale revolution. *Nature*, 494(7437):307–308, 2013.

[75] Neil Hume. Bhp takes $2.84bn writedown on shale gas. *Financial Times*, 3 Aug. 2012.

[76] S. Ignacimuthu. *Basic Bioinformatics.* Alpha Science International Ltd., Oxford, 2 edition, 2013.

[77] Chi U. Ikoku. Application of learning curve models to oil and gas well drilling. *Society of Petroleum Engineers California Regional Meeting*, 1978.

[78] S. Ikonnikova, J. Browning, S. Horvath, and S. Tinker. Well recovery, drainage area, and future drill-well inventory: empirical study of the Barnett shale gas play. *Society of Petroleum Engineers Reservoir Evaluation & Engineering*, 17(4):484–496, 2014.

[79] D. Ilk, J.A. Rushing, A.D. Perego, and T.A. Blasingame. Exponential vs. hyperbolic decline in tight gas sands-understanding the origin and implications for reserve estimates using Arps' decline curves. *Society of Petroleum Engineers Annual Technical Conference and Exhibition*, 2008.

[80] Mason Inman. The fracking fallacy. *Nature*, 516(7529):28–30, 2014.

[81] IHS Global Insight. America's new energy future: The unconventional oil and gas revolution and the US economy. Report, 2012.

[82] C. Jablanowski, A. Ettehad, B. Ogunyomi, and I. Srour. Integrating learning curves in probabilistic well-construction estimates. *Society of Petroleum Engineers Drilling & Completions*, 3, 2011.

[83] R. Jackson, A. Vengosh, J. Carey, R. Davies, T. Darrah, F. O'Sullivan, and G. Petron. The environmental costs and benefits of fracking. *Annual review of Environment and Resources*, 2014.

[84] H. Jacoby, F. O'Sullivan, and S. Paltsev. The influence of shale gas on U.S. energy and environmental policy. *Economics of Energy & Environmental Policy*, 1:37–51, 2012.

[85] Daniel M. Jarvie, Ronald J. Hill, Tim E. Ruble, and Richard M. Pollastro. Unconventional shale-gas systems: The Mississippian Barnett Shale of north-central Texas as one model for thermogenic shale-gas assessment. *American Association of Petroleum Geologists Bulletin*, 91(4):475–499, 2007.

[86] Steffen Jenner and Alberto J. Lamadrid. Shale gas vs. coal: Policy implication from environmental impact comparisons of shale gas, conventional gas, and coal on air, water, and land in the United States. *Energy Policy*, (53):442–453, 2013.

[87] Paul L. Joskow. Natural gas: from shortages to abundance in the united states. *The American Economic Review*, 103(3):338–343, 2013.

[88] Boyan Jovanovic and Yaw Nyarko. A bayesian learning model fitted to a variety of empirical learning curves. *Brookings Papers on Economic Activity. Microeconomics*, 1995:247–305, 1995.

[89] Mark J. Kaiser and Yunke Yu. Drilling and completion cost in the Louisiana Haynesville Shale. *Natural Resources Research*, 24(1):5–31, 2015.

[90] Gordon M. Kaufman. *Statistical decision and related techniques in oil and gas exploration.* Prentice-Hall, 1963.

[91] Gordon M Kaufman. Statistical issues in the assessment of undiscovered oil and gas resources. *The Energy Journal*, 14(1):183–215, 1993.

[92] Gordon M. Kaufman, Y. Balcer, and D. Kruyt. *A probabilistic model of oil and gas discovery*, pages 113–142. Methods of Estimating the Volume of Undiscovered Oil and Gas Resources. American Association of Petroleum Geologists, 1975.

[93] Truman Lee Kelley. *Fundamentals of Statistics.* Harvard University Press, 1947.

[94] Daniel H. Kim. The link between individual and organizational learning. *Sloan Management Review*, 35(1):37–50, 1993.

[95] D.G. Krige. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52(6):119–139, 1951.

[96] D.G. Krige. One the departure of ore value distributions from the lognormal model in South African gold mines. *Journal of the South African Institute of Mining and Metallurgy*, 61(4), 1960.

[97] D.G. Krige. A study of gold and uranium distribution patterns in the Klerksdorp gold field. *Geoexploration*, 4:43–53, 1966.

[98] John K. Kruschke. *Doing Bayesian Data Analysis.* Elsevier Science, 2 edition, 2015.

[99] Michael A. Lapre and Ingrid M. Nembhard. *Inside the Organizational Learning Curve: Understanding the Organizational Learning Process.* Now Publishers Inc, 2011.

[100] J. Lee and R. Sidle. Gas-reserve estimation in resource plays. *Society of Petroleum Engineers Economics & Management*, 2:86–91, 2010.

[101] Vladimir I. Levenshtein. Binary codes capable of correcting deletions, insertions and reversals. *Soviet physics doklady*, 10:707, 1966.

[102] Michael Levi. *The Power Surge: Energy, Opportunity, and the Battle for America's Future.* Oxford University Press, New York, 2013.

[103] Barbara Levitt and James G. March. Organizational learning. *Annual review of sociology*, 14(1):319–338, 1988.

[104] Ferdinand K. Levy. Adaptation in the production process. *Management Science*, 11(6):B136–B154, 1965.

[105] Eckhard Limpert, Werner A. Stahel, and Markus Abbt. Log-normal distributions across the sciences: keys and clues. *BioScience*, 51(5):341–352, 2001.

[106] Susan Lund. Game changers: Five opportunities for US growth and renewal. Report, McKinsey, 2013.

[107] J.P. March, J.G. Olsen. The uncertainty of the past: organizational learning under ambiguity. *European Journal of Political Economy*, 3:147–171, 1975.

[108] Zur March, James G Shapira. Managerial perspectives on risk and risk taking. *Management Science*, 33(11):1404–1418, 1987.

[109] MathWorks. Confidence and prediction bounds. *MATLAB documentation*, 2015.

[110] MathWorks. Probability plots (probplot). *MATLAB Documentation*, 2015.

[111] Shawn Maxwell. Microseismic location uncertainty. *Canadian Society of Exploration Geophysicists Recorder*, 34(4):41–46, 2009.

[112] Shawn Maxwell. Microseismic: Growth born from success. *Society of Exploration Geophysicists: The Leading Edge*, 29(3):338–343, 2010.

[113] Joseph B. Mazzola and Kevin F. McCardle. A bayesian approach to managing learning-curve uncertainty. *Management Science*, 42(5):680–692, 1996.

[114] William D. McCain. *The properties of petroleum fluids.* PennWell Books, Tulsa, 1990.

[115] Steven McGinn. BHP Billiton confirms Eagle Ford condensate exports. *World Oil*, 2014.

[116] Christophe McGlade, Jamie Speirs, and Steve Sorrell. Methods of estimating shale gas resources-comparison, evaluation and implications. *Energy*, 59:116–125, 2013.

[117] J. Meisner and J. Demirmen. The creaming method: A bayesian procedure to forecast future oil and gas discoveries in mature exploration provinces. *Journal of the Royal Statistics Society Series A*, 144:1–31, 1981.

[118] Rubert G. Miller Jr. *Beyond ANOVA: Basics of applied statistics.* Chapman & Hall/CRC Texts in Statistical Science. Chapman & Hall/CRC, 1997.

[119] Justin B. Montgomery and Francis M. O'Sullivan. Measuring drilling standardization using approximate string matching. *Society of Petroleum Engineers Annual Technical Conference and Exhibition*, 2014.

[120] Scott L. Montgomery, Daniel M. Jarvie, Kent A. Bowker, and Richard M. Pollastro. Mississippian Barnett Shale, Fort Worth basin, north-central Texas: Gas-shale play with multi-trillion cubic foot potential. *American Association of Petroleum Geologists Bulletin*, 89(2):155–175, 2005.

[121] Amit Shankar Mukherjee and Michael A. Lapre. Knowledge driven quality improvement. *Management Science*, 44(11):S35–S49, 1998.

[122] S. Mustafiz. State-of-the-art petroleum reservoir simulation. *Petroleum Science and Technology*, 26(10-11):1303–1329, 2008.

[123] OGJ. Shell reshuffles us shale assets in two major deals. *Oil & Gas Journal*, 14 Aug. 2014.

[124] William G. Ouchi. Markets, bureaucracies, and clans. *Administrative Science Quarterly*, 25(1):129–141, 1980.

[125] Tad W. Patzek, Frank Male, and Michael Marder. Gas production in the Barnett Shale obeys a simple scaling theory. *Proceedings of the National Academy of Science*, 110(49):19731–19736, 2013.

[126] PGC. Potential supply of natural gas in the united states. Report, Potential Gas Committee Biennial Report, 2013.

[127] Richard M. Pollastro, Ronald J. Hill, Thomas A. Ahlbrandt, Ronald R. Charpentier, Troy A. Cook, Timothy R. Klett, Mitchell E. Henry, and Christopher J. Schenk. Assessment of undiscovered oil and gas resources of the Bend Arch-Fort Worth Basin Province of north-central Texas and southwestern Oklahoma, 2003. Report 2004-3022, U.S. Geological Survey Fact Sheet, 2004.

[128] S. E. Prensky. A survey of recent developments and emerging technology in well logging and rock characterization. *The Log Analyst*, 35(2):15–45, 1994.

[129] Reuters. China finds shale gas challenging, halves 2020 output target, 7 Aug. 2014.

[130] Howard Rogers. Shale gas-the unfolding story. *Oxford Review of Economic Policy*, 27(1):117–143, 2011.

[131] P.R. Rose. *Risk Analysis and Management of Petroleum Exploration Ventures.* Methods in Exploration. American Association of Petroleum Geologists, Tulsa, 2001.

[132] J.A. Rushing, A.D. Perego, R.B. Sullivan, and T.A. Blasingame. Estimating reserves in tight gas sands at HP/HT reservoir conditions: use and misuse of an Arps decline curve methodology. *Society of Petroleum Engineers Annual Technical Conference and Exhibition*, 2007.

[133] SAS. Analytic innovations address new challenges in the oil and gas industry. Report, SAS Institute, 2014.

[134] Schlumberger. Oilfield glossary, 2015.

[135] R.L. Schmoyer, J.J. Beauchamp, C.C. Brandt, and F.O. Hoffman Jr. Difficulties with the lognormal model in mean estimation and testing. *Environmental and Ecological Statistics*, 3:81–97, 1996.

[136] Daniel P. Schrag. Is shale gas good for climate change? *Daedalus*, 141(2):72–80, 2012.

[137] J.H. Schuenemeyer and L.J. Drew. A procedure to estimate the parent population of the size of oil and gas fields as revealed by a study of economic truncation. *Mathematical Geology*, 15(1):145–161, 1983.

[138] Arcangelo Sena, Gabino Castillo, Kevin Chesser, Simon Voisey, Jorge Estrada, Juan Carcuz, Emilio Carmona, and Peggy Hodgkins. Seismic reservoir characterization in resource shale plays: Stress analysis and sweet spot discrimination. *Society of Exploration Geophysicists: The Leading Edge*, 30(7):758–764, 2011.

[139] Michael Shellenberger, Ted Nordhaus, Alex Trembath, and Jesse Jenkins. Where the shale gas revolution came from: Government's role in the development of hydraulic fracturing in shale. Report, Breakthrough Institute, 2012.

[140] Stephen A. Sonnenberg. TOC and pyrolysis data for the Bakken Shales, Williston Basin, North Dakota and Montana. *American Association of Petroleum Geologists Bulletin*, 2011.

[141] Didier Sornette. *Critical phenomena in natural sciences: chaos, fractals, selforganization and disorder: concepts and tools.* Synergistics. Springer, Berlin and New York, 2 edition, 2006.

[142] R. Strickland, D. Purvis, and T. Blasingame. Practical aspects of reserves determinations for shale gas. *Society of Petroleum Engineers North American Unconventional Gas Conference and Exhibition*, 2011.

[143] S.B. Suslick and D.J. Schiozer. Risk analysis applied to petroleum exploration and production: an overview. *Journal of Petroleum Science and Engineering*, 44(1):1–9, 2004.

[144] New York Times. Documents: Leaked industry e-mails and reports, Accessed 18 Aug. 2014.

[145] The Economist Intelligence Unit. US: Failing in shale. Report, 20 Nov. 2013.

[146] Peter P Valkó. Assigning value to stimulation in the barnett shale: a simultaneous analysis of 7000 plus production histories and well completion records. *Society of Petroleum Engineers Hydraulic Fracturing Technology Conference*, 2009.

[147] Rainer Van Den Bosch and Antonio Paiva. Benchmarking unconventional well performance predictions. *Society of Petroleum Engineers/European Association of Geoscientists & Engineers European Unconventional Resources Conference & Exhibition-From Potential to Production*, 2012.

[148] Eric van Oort, James Griffith, and Barry Schneider. How to accelerate drilling learning curves. *Society of Petroleum Engineers/International Association of Drilling Contractors Drilling Conference and Exhibition*, 2011.

[149] A. Vengosh, R. Jackson, N. Warner, T. Darrah, and A. Kondash. A critical review of the risks to water resources from unconventional shale gas development and hydraulic fracturing in the united states. *Environmental Science & Technology*, 48(15):8334–8348, 2014.

[150] George Stephen Wattley. Are the oil recoveries distributed lognormally? if not, there may be information in the differences. *Society of Petroleum Engineers Latin American & Caribbean Petroleum Engineering Conference*, 2007.

[151] Rick L. Webster. *Petroleum Source Rocks and Stratigraphy of the Bakken Formation in North Dakota*. American Association of Petroleum Geologists Bulletin, 1984.

[152] Rick Wilkinson. *Speaking oil & gas*. BHP Billiton Petroleum, Perth, 2006.

[153] T.P. Wright. Factors affecting the costs of airplanes. *Journal of Aeronautical Science*, 3(122-128), 1936.

[154] Daniel Yergin. *The quest: energy, security, and the remaking of the modern world*. Penguin, New York, 2011.

[155] Willard I. Zangwill and Paul B. Kantor. Toward a theory of continuous improvement and the learning curve. *Management Science*, 44(7):910–920, 1998.

[156] M. Zoback. Managing the seismic risk posed by wastewater disposal. *Earth*, 57:38–43, 2012.