UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

PRODUCTION DATA ANALYSIS AND FORECASTING: BAYESIAN PROBABILISTIC METHODS TO QUANTIFY UNCERTAINTY, ANALYZE PRODUCTION PERFORMANCE TRENDS IN STACK

A THESIS

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

MASTER OF SCIENCE

By

NAGA SANTHOSH ADAVENI

Norman, Oklahoma

2020

PRODUCTION DATA ANALYSIS AND FORECASTING: BAYESIAN PROBABILISTIC METHODS TO QUANTIFY UNCERTAINTY, ANALYZE PRODUCTION PERFORMANCE TRENDS IN STACK

A THESIS APPROVED FOR THE

GALLOGLY COLLEGE OF ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Talayeh Razzaghi, Chair

Dr. Deepak Devegowda

Dr. Charles Nicholson

Acknowledgements

I would like to express my sincere gratitude to Dr. Deepak Devegowda for trusting me and guiding me on this master's thesis. The perspective that professor provided, and all the petroleum engineering training has reflected in the quality of this work.

I would like to thank Dr. Talayeh Razzaghi and Dr. Charles Nicholson for agreeing to chair and co-chair my thesis committee. Their input has been invaluable.

With full-time employment, I would not have been able to complete this degree without the support of my family. I cannot thank my wife Sravanthi and my child Samhith enough for the sacrifices they have had to make, as I took away the family time with them for the last two years. Thank you to my parents Adaveni Mani Mohan and Adaveni Srivani for their unconditional love, support, and motivation. The support of my family is the only thing that drives me across the finish line.

I would like to thank DSA Program Directors Dr. Sridhar Radhakrishnan and Dr. Randa Shehab, staff and faculty for their support and efforts to make this program a seamless experience for working professionals.

I would like to thank Devon Energy for providing with the production and completions data and for allowing me to publish this research. I sincerely thank Patrick Kamann and Johan Daal for listening to my proposal, trusting me, and helping me to obtain the necessary approvals. I would like to thank the Subsurface Technology Group, in particular Jake Dickson and Emily Lair, for all the support they have provided. I would like to thank Subrahmanyam Nittala, Managing Director at Accenture, who has been a tremendous motivator for me. His advice and words of wisdom have enlightened me to strive for this degree.

Abstract

In this work, a Bayesian approach of probabilistic estimation for decline curve analysis in unconventional reservoirs is presented. The primary objectives of this study are the quantification of the uncertainty for production forecasting and do a parent-child analysis for wells from the same play.

MCMC-based Metropolis algorithm is used for sampling from the proposal distributions to generate posterior distributions for the decline curve parameters. This sampling technique is applied for three models: Arps, Duong, and power law exponential models. Prior and likelihood distributions are established for the three models based on our understating of the data and the models. Forecast estimates are generated using multiple intervals of initial production data to understand how the sampling algorithm generates better estimates with increasing amount of training data.

282 oil and gas wells Meramec STACK unconventional play are used in this work to quantify the production forecasting uncertainty. Results show that the MCMC-based approach was able to establish uncertainty bounds, matching MAP estimates for cumulative production. Based on the amount of production data available and the nature of the flow, the model that fits best can vary. Using the estimated decline curve parameters, parent-child well comparison analysis is done to understand the changing production dynamics in the Meramec STACK play.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In this chapter, I provide a general overview of the aspects of the oil and gas industry that is necessary to facilitate an easier understanding of this research. Post that, I introduce the decline curve models. In Section 1.1, I present the concepts of conventional and unconventional reservoirs, vertical and horizontal wells. In section 1.2, I present the concepts of transient and boundary dominated flow for wells. In section 1.3, I will discuss how the amount of oil or gas production of a well, captured at a daily or monthly frequency, is important for understanding the production trends. In Section 1.4, I will further explain the procedures to extrapolate the production trend, to understand the future production and thereby the economic potential of the well. In the following sections, I introduce the concepts of decline curve analysis (DCA) and various industry standard models for decline curves. In chronological order from Section 1.5 through 1.9, I review and present the workings of the DCA models. The Arps model [1945], a standard model for forecasting production for conventional reservoirs, is presented first followed by various newer models that are developed for unconventional reservoirs. In Section 1.10, I discuss the need for understanding the forecasting uncertainty irrespective of the method used for predicting production. Section 1.11 details the profile of wells in the STACK area chosen for this analysis.

1.1 Reservoir

A reservoir is a subsurface pool of oil or gas contained in porous or fractured rock formations. There are two types of reservoirs - conventional and unconventional. Conventional resources are located under impermeable rock formations called a caprock, which allow for trapping of hydrocarbons. Reservoir and fluid characteristics of conventional reservoirs typically permit oil or natural gas to flow readily into wells. They are developed using vertical wellbores and produce oil and natural gas at economic flow rates with minimal or no stimulation.

Unconventional resources are trapped by low permeability and low porosity rocks. Oil or gas cannot be extracted at economical rates and require the use of more sophisticated means to extract oil and gas economically.

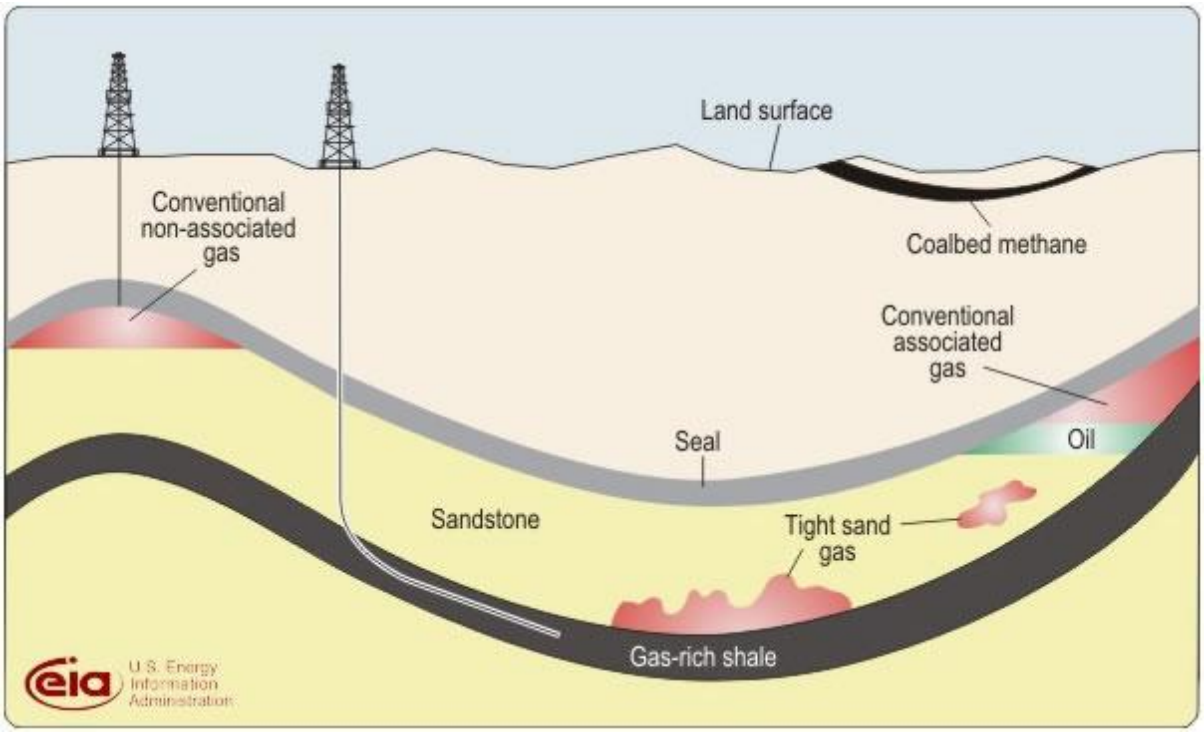

Figure 1.1 Conventional and unconventional resources (US EIA).

Figure 1.1 shows a conventional and unconventional play in the same schematic, with a vertical well producing from the conventional non-associated gas reservoir and a horizontal well producing from an unconventional gas-rich shale, shows the use of horizontal wells to develop a shale resource and, shows various unconventional resources.

Stimulation is a treatment performed to restore or enhance the productivity of a well. Stimulation in shale oil and gas reservoirs is done by hydraulic fracturing treatments. Figure 1.2 shows a four-stage completion for a horizontal well with four fractures per stage. Fracturing creates a highly conductive path for fluids between the reservoir and wellbore. Horizontal wells in very-low-permeability formations such as shales are typically hydraulically fractured with 10 or more stages starting at the “toe” of the well and working back to the “heel” where the well bends up to the surface. The combination of horizontal wells and hydraulic fracturing has rendered production from unconventional shales economic and viable.

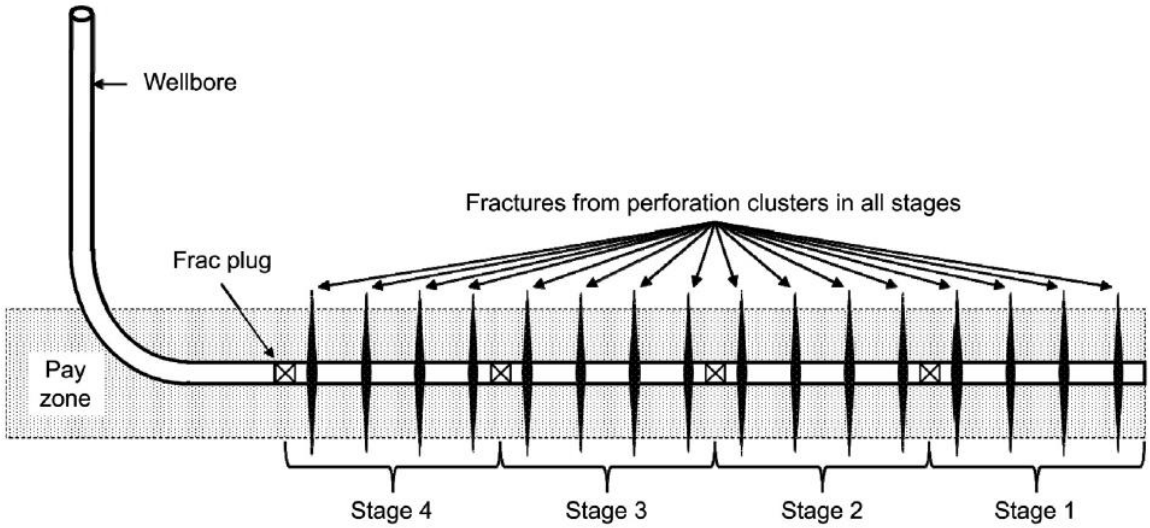

Figure 1.2 Multi-stage horizontal well (Guo et al. 2017). The black vertical lines represent hydraulic fractures. In the figure above there are four stages of fractures with four fractures per stage.

1.2 Reservoir flow

Flow in a reservoir is characterized as being one of the two types: transient flow and boundary dominated flow (BDF). During the initial transient flow period, pressure transients migrate outward from the well without encountering any boundaries. This happens when a well is placed in a region where there has been no prior oil and gas production, so the well only 'sees' the original reservoir pressure. In a system like this, the pressure transients created by the producing well move outward, draining larger and larger reservoir volumes. This is the transient flow period. Eventually, of course, the pressure transients intersect pressure transients originating from other wells in the vicinity or a physical boundary. The onset of this effect is called boundary dominated flow.

Figure 1.3 shows the cross-section of a reservoir into which a horizontal well has been drilled. The cross-section shows three fractures of the horizontal well. The three cross-sectional pictures show the pressure profiles after 5, 10 and 60 years of production. The varying pressure in the reservoir is represented by different colors, with dark red indicating the highest pressure and dark blue the lowest pressure in the reservoir. The pressure around the fractures continues to decrease as the reservoir drains. After 5 years of production, the pressure transients move outwards, but have not reached all the boundaries. After 60 years of production, the pressure transients have reached all the

boundaries and the pressure is decreasing at the boundary indicating boundary dominated flow for the well.

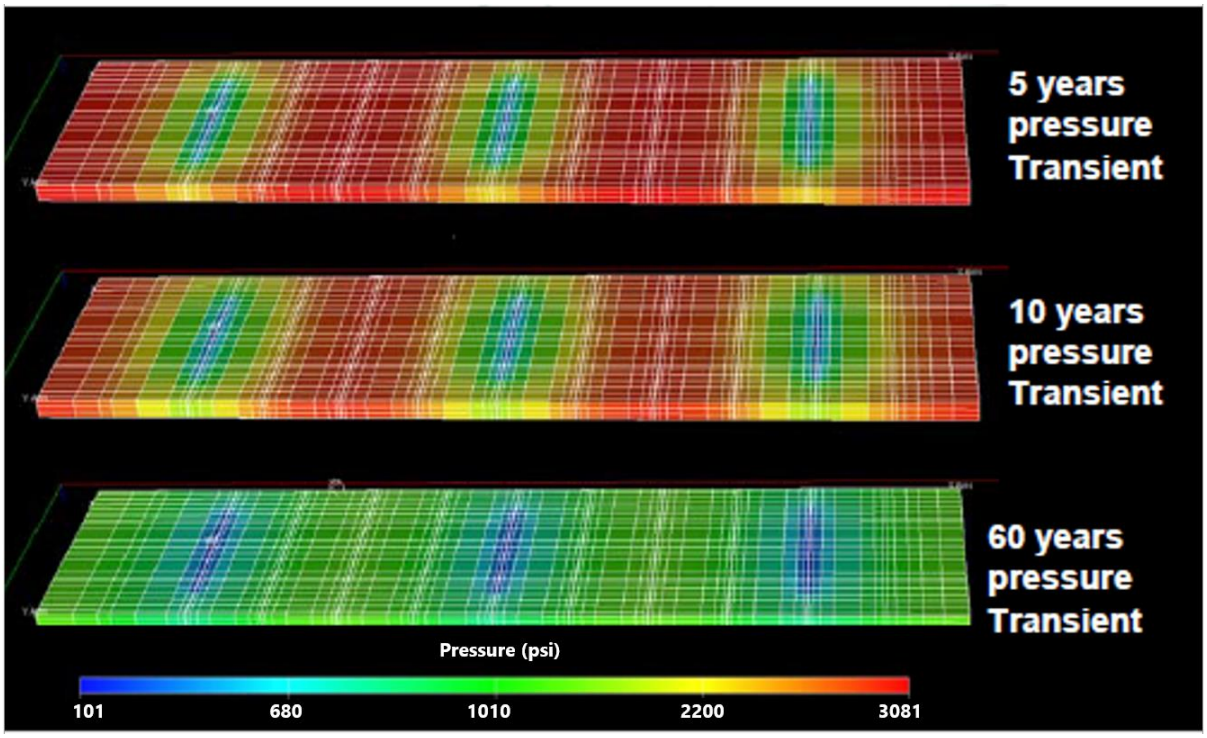

Figure 1.3 Cross-section view of a reservoir with a horizontal well. Figure shows a single stage with three fractures. The three cross-section images show the pressure profiles after 5, 10 and 60 years of production (Chen. 2016).

1.3 Production data

Production can begin as soon as wells are completed, and surface facilities are installed. Well production rates are produced fluid per unit of time. These production rates are recorded on a daily or monthly basis. Figure 1.4 shows daily rates and cumulative production for an oil well as a function of time. Figure 1.5 shows daily rates and cumulative production for a gas well over time.

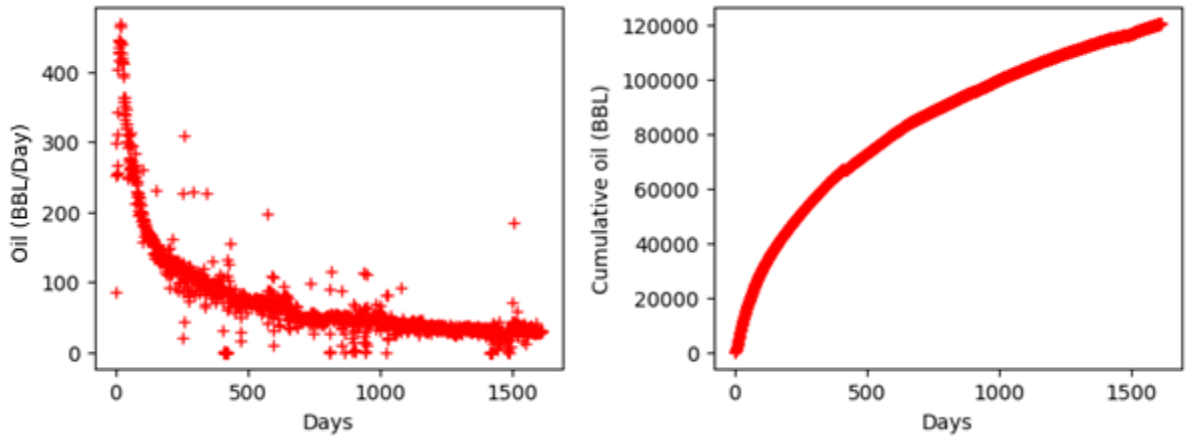

Figure 1.4 Daily oil rate (left) and cumulative oil production (right) for an oil well.

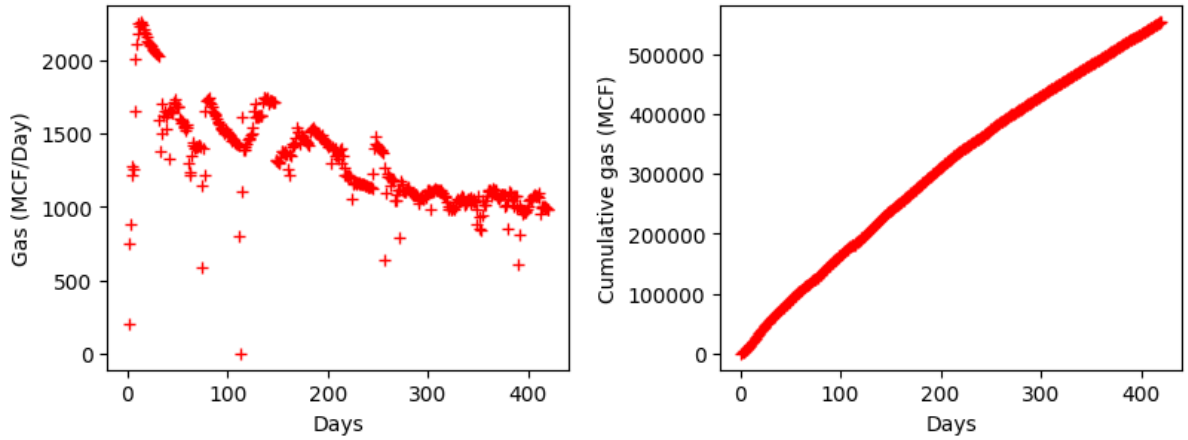

Figure 1.5 Daily gas rate (left) and cumulative gas production (right) for a gas well.

1.4 Decline curve analysis

One of the biggest challenges in the oil and gas industry is forecasting production trends. This is typically done using a workflow called decline curve analysis (DCA). Production data is plotted as in Figure 1.4 and Figure 1.5, and the trends are extrapolated to forecast future production, obtained after analyzing historical production data. These trends are then used to determine future oil and gas production, to determine if some form of intervention is necessary to assess future economic viability and estimated ultimate recovery (EUR) of a well. If the forecasted production rates are low, the operator may decide to intervene in the form of drilling more wells or stimulating existing wells.

Decline curve analysis (DCA) methods, in a variety of forms, have been used in the petroleum industry for more than fifty years to analyze production data and forecast reserves. These models forecast the future production data using empirical rate-time

equations with parameters. The unknown parameters in the decline curves are determined by fitting the decline curves to historical data. Future production is estimated by extrapolating the decline curve to a specified final rate. The final rate is usually determined as the lowest rate that is still economically viable. For this reason, the specified final rate is called economic rate or abandonment rate. Reserves are the difference between cumulative production at abandonment and current cumulative production. I will be presenting various decline curve models in Sections 1.5 – 1.9. Figure 1.6 shows the projected daily oil rate for an oil well using Arps (1945) hyperbolic decline model. Using the Arps decline model (1945), the production has been forecasted for 880 days based on the initial production data for 440 days.

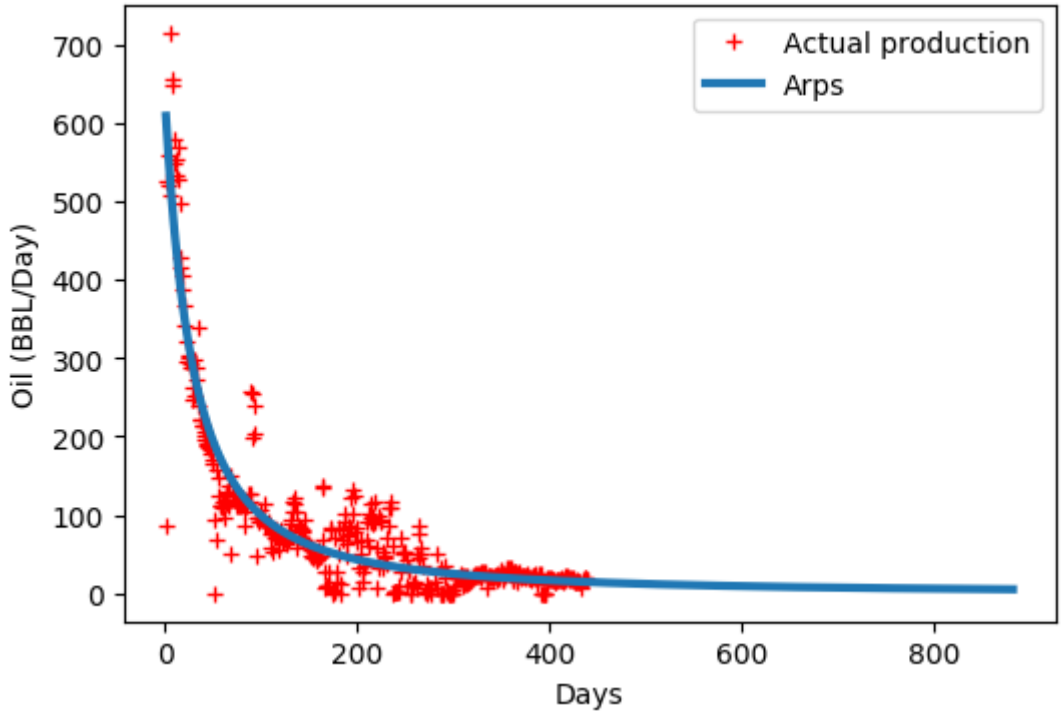

Figure 1.6 Projected production rate forecasted using Arps (1945) hyperbolic decline model.

To summarize, decline curve analysis is important for two reasons: for predicting future production and thereby estimating the ultimate recovery for oil and gas wells. These metrics help understand if the development is economical. In 1945, J.J. Arps proposed decline models when there was no formalism to predict future production and reserve estimates (Arps, 1945). While the Arps model has been applied successfully for conventional wells, their extension to unconventional wells is not straightforward (Duong, 2011).

For conventional reservoirs, due to high reservoir permeability, the transient flow lasts a few days. The original Arps model is valid only for boundary-dominated flow. In unconventional shale reservoirs where the matrix permeability is extremely low, there is a chance that the flow will not become boundary dominated during its life (Lee et al. 2010). Unfortunately, this implies that the Arps model cannot be applied to decline curve analyses for unconventional reservoirs. The application of Arps to the transient flow period in unconventional wells can result in expected ultimate recovery (EUR) predictions to be unrealistically high.

Since 2008, several new decline-curve models have been introduced to estimate unconventional reservoirs. Ilk et al. (2008) introduced the power-law decline curve to model the decrease in the decline exponent b with time. Valko et al. (2010) introduced the stretched exponential production decline curve and used it to quantify the uncertainty in field-production forecasts. Duong (2011) proposed a new model for unconventional reservoirs with very low permeability, and the shape of this curve is suited for wells that exhibit long periods of transient flow. Daal et al. (2019) recently presented conceptual decline curve models which conform to the long-term transient flow behavior observed in unconventional reservoirs. In the following sections, I present a detailed summary of these models' decline curve analysis, their application, different parameters, and their purpose.

1.5 Arps model

Arps (1945) proposed that the production drop over a given constant interval is a percentage of the preceding production rate. This production drop fraction as a percentage per month value is called the loss ratio.

$$a = \frac{q}{\frac{dq}{dt}} \qquad \text{.....(1)}$$

In Eq. 1, a represents loss ratio, q represents production rate, and t represents time. If the drop in the production rate per unit of time is proportional to the production rate, we obtain a specific form of the production decline called an exponential decline.

$$q_t = q_i * exp(-D_i * t) \qquad \text{.....(2)}$$

In Eq. 2, q_t represents production rate at time t. q_i represents the rate at t=1. D_i represents the initial decline rate which is the reciprocal of the loss ratio. Subsequent observations determined that the exponential decline is conservative and led to the development of other decline curves such as the hyperbolic decline as represented in Eq. 3. Note that the exponential decline is obtained directly from Eq. 3 in the limit of b tending to 0. If b is equal to 1, we obtain what is known as the harmonic decline.

$$q_t = \frac{q_i}{(1 + bD_i t)^{1/b}} \qquad \text{.....(3)}$$

In Eq. 3, q_i, q_t, D_i represent the same values as in Eq.2, b represents the hyperbolic decline constant. Arps model is strictly valid for boundary-dominated flow (BDF) and hence has been applied successfully for conventional reservoirs. Its application to unconventional reservoirs is only valid when transient flow has ended. In any case, the parameters to be estimated for calculating production forecast using Arps model are initial rate (q_i), decline rate (D_i) and hyperbolic decline constant (b).

Figure 1.7 shows a regression matching of daily rates for an oil well with the calculated production using the least square fit match using Arps model.

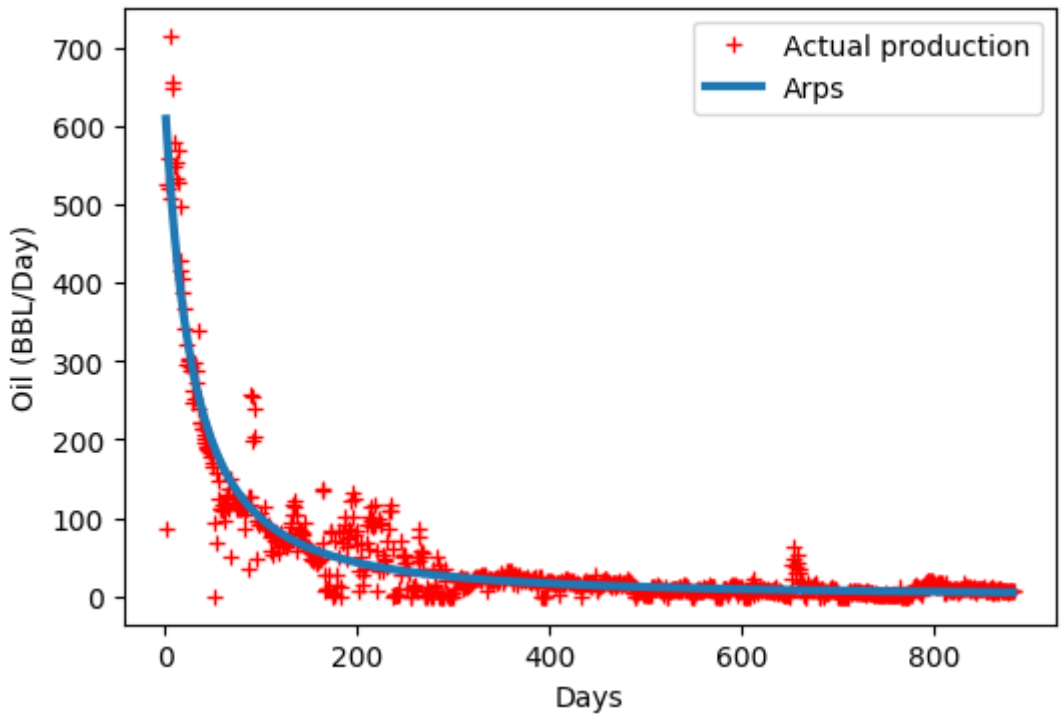

Figure 1.7 Calculated production matching versus actual production for oil well using Arps model.

1.6 Power law model

In practice, for any reservoir, before the onset of BDF, the b-value is always higher than 1. Extrapolating short-term declines to predict long-term production, results in extremely high overestimation of the well EUR. Ilk et al. (2008) present the need for an appropriate model for reserves extrapolation in tight reservoirs. A new power law loss ratio is developed which has more generality than the hyperbolic rate decline relation for tight gas/shale gas wells. As a non-hyperbolic approach to reserves estimates, Ilk et al. (2008) develop a method which employs a different functional form for the decline D parameter as given by:

$$D = D_\infty + D_1 t^{-(1-n)} \quad \text{.....(4)}$$

In Eq. 4, D_1 and D_∞ stand for the decline constant intercept at t =1 and t = ∞ respectively, n is the time exponent, D is the decline at time t. Eq. 4 is the power law loss-ratio formulation. Its interpretation is that the loss ratio can be approximated by a decaying power law function with a constant behavior at large time (D_∞ being a constant). This model is flexible enough to model transient, transition and BDF in many cases, but at long times the relation reduces to the traditional exponential decline relation since the contribution of the power law term is relatively smaller. Substituting Eq. 4 into the loss ratio Eq. 1 and integrating yields the below:

$$q = \hat{q}_i \, exp\,[\,-D_\infty t - \hat{D}_i t^n] \quad \text{.....(5)}$$

In Eq. 5, q stands for Rate at time t, $\hat{q}_i$ stands for Rate intercept, D_1 and D_∞ stand for the decline constant intercept at t =1 and t = ∞ respectively. $\hat{D}_i$ is D_1 divided by n where n is the time exponent. Figure 1.8 shows how Arps model compares to the power law model (Ilk et al. 2008). Arps hyperbolic model shows a constant value for production rate q and decline value D at early times whereas those values tend to be non-constant for the power law model. At late times, power law model shows exponential rate decline for q and a constant D value which takes the value of the parameter D_∞. The unknown parameters that need to be estimated for power law model are the decline rates at t =1 and t = ∞, D_1 and D_∞ respectively, Rate intercept ($\hat{q}_i$) and the time exponent (n).

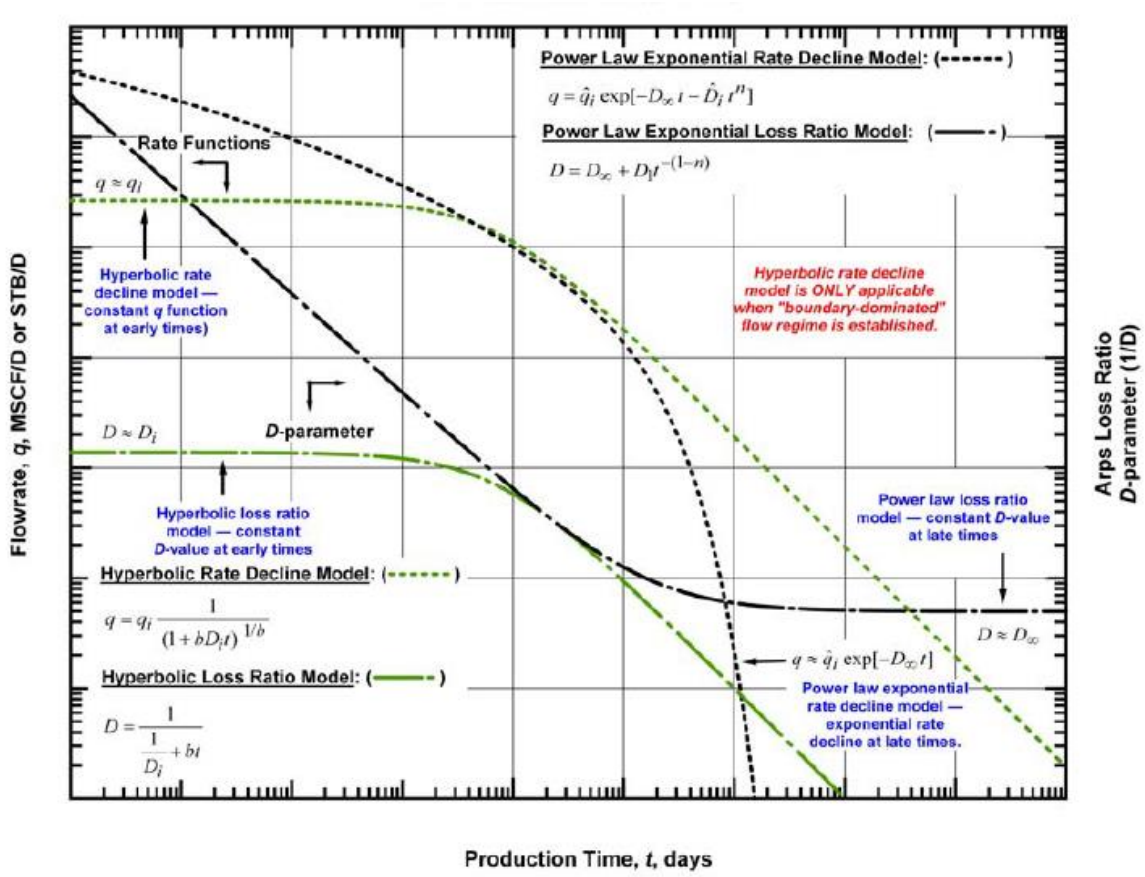

Figure 1.8 Rate and loss ratio comparison of Arps hyperbolic and power law model (Ilk et al. 2008). The green curves represent the production data and decline rate parameter using the Arps equation. The black curves show the rates and decline rate observed are subsequently fitted with a power law model.

1.7 Stretched exponential production decline model

Valko et al. (2010) propose a model that relies on parameter processing for a large group of wells using a concept called 'group-data controlled forecast'. A natural interpretation of the stretched exponential decay of a quantity is that it is generated by a sum(integral) of pure exponential decays with a 'fat-tailed' probability distribution of the time constants. The stretched exponential production decline (SEPD) model assumes that the actual production decline is determined by a great number of contributing volumes individually in exponential decay in pseudo-steady state, but with a specific distribution of characteristic time constants. The distribution is determined by a parameter pair (n, τ). τ is the median of the characteristic time constants and n is the exponent parameter for the SEPD model. The nearer the value of n is to zero, the larger is the tail of the distribution, or in other words, the elementary volumes have very large time constants. SEPD equation is as shown below:

$$q = q_0 exp[-(t/\tau)^n] \quad \text{.....(6)}$$

In Eq. 6, q indicates the rate at time t and q_0 indicates the rate at time t=0. Although the SEPD model has been applied often, it was specifically developed for the operating conditions associated with Barnett Shale. No claim is made that it applies to all plays with a similar consistency and hence its applicability remains untested. The unknown parameters that need to be estimated for Stretched exponential production decline model are initial rate (q_0), median characteristic time constant (τ) and the exponent parameter (n).

1.8 Duong model

Duong (2014) propose an alternative approach to estimate EUR from wells in which fracture flow is dominant and matrix contribution is negligible. For fracture flows at a constant flowing bottom hole pressure, a log-log plot of rate, q divided by cumulative production, G_p versus time will yield a straight line with a unit slope regardless of fracture type as shown in Figure 1.9. In practice, a slope of greater than unity is normally observed because of actual field operations, data approximation, and flow-regime changes. A rate/time or cumulative production/time relationship can be established based on the intercept and slope values of this log-log plot and initial gas or oil rate. Results show that this alternative approach is easier to use, gives a reliable EUR and can be used to replace the traditional decline methods for unconventional reservoirs.

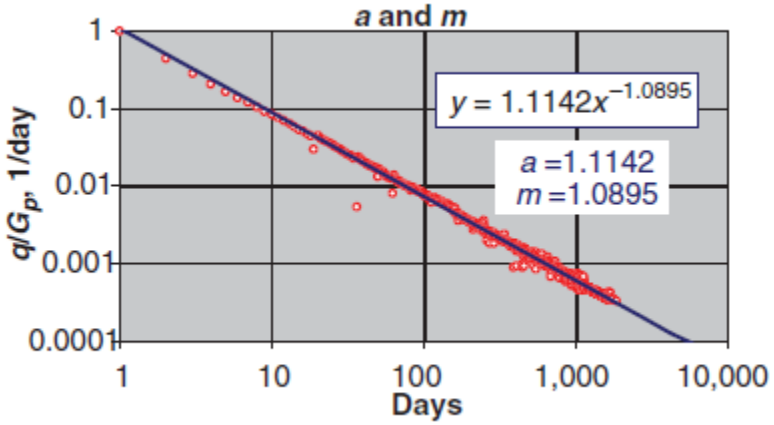

Figure 1.9 A log-log plot of q/G_p vs time for a shale gas well (Duong 2014).

Duong (2014) used this finding to derive an empirical decline model for cumulative production G_p as proposed in Eq. 7 that is based on a long-term linear flow in many wells in tight and shale gas reservoirs. This applies to unconventional reservoirs with very low permeability. The shape of the curve is well-suited even for long periods of transient flow. Like other unconventional methods, this method will also predict a finite EUR and is a very conservative approach.

$$G_p = \frac{q_1}{a} e^{\frac{a}{1-m}(t^{1-m}-1)} \qquad \text{.....(7)}$$

In Eq. 7, q_1 is the theoretical rate at t = 1. Estimates of a and m are derived from the intercept on the y-axis and slope from the log-log plot as shown in Figure 1.9. The ratio of daily rate q and cumulative production G_p is defined by Eq. 8 below:

$$\frac{q}{G_p} = at^{-m} \qquad \text{.....(8)}$$

The daily flow rate is derived from Eq. 7 and Eq. 8 as shown below:

$$q = q_1 t^{-m} e^{\frac{a}{1-m}(t^{1-m}-1)} \qquad \text{.....(9)}$$

The unknown parameters that need to be estimated for this model are the initial rate (q_1), slope and intercept from the plot in Figure 1.9, m and a respectively.

1.9 Variable power law exponential model

Daal et al. (2019) present conceptual decline curve models which conform to the long-term transient flow behavior in unconventional reservoirs. These models are validated by matching production rate, loss ratios and the b-factors. As depicted in Figure 1.10, these models are built on the concept that the loss ratio and the b-factor are not continuous functions of time.

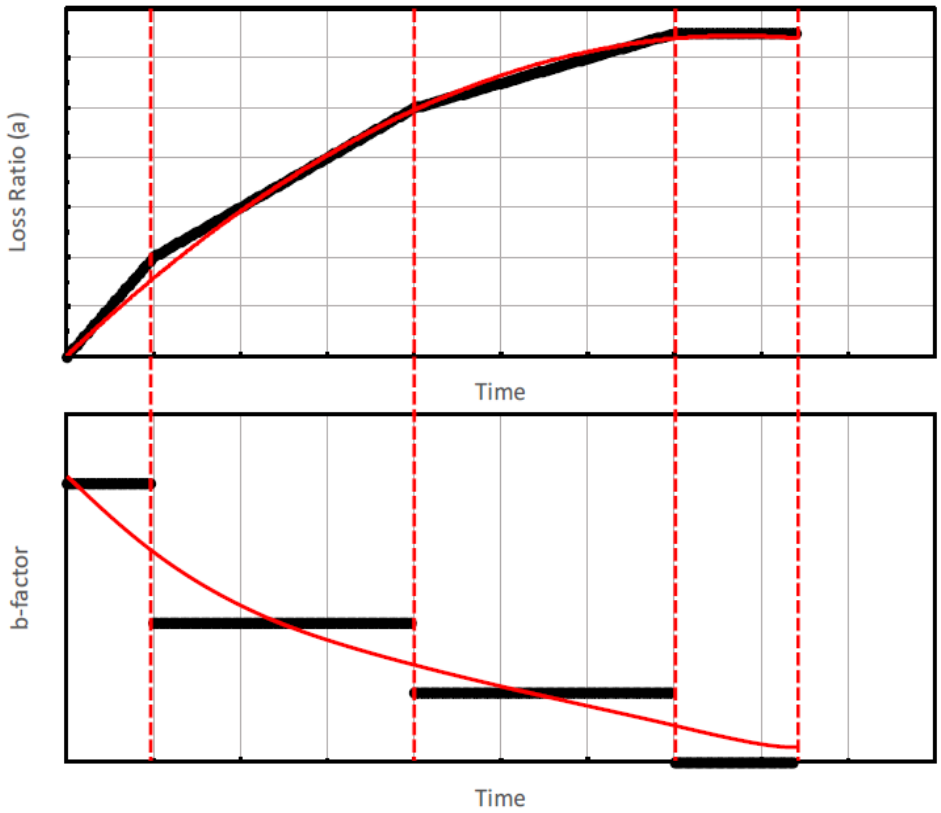

Figure 1.10 Conceptual model proposed by Daal et al. (2019).

Since pinpointing the duration of the flow regimes is challenging, the model develops generalized expressions for b-factor as expressed in Eq. 10.

$$b(t) = \frac{d}{dt}[a] = f(t, \beta_1, \beta_2, \ldots \beta_n) \qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(10)$$

The work by Daal et al. (2019) proposes two models: power law exponential model and variable power law exponential model. These models are built on the observations of quadratic changes in loss ratio and power law changes in inverse loss ratio, respectively. Variable power law exponential model is an extension of the SEPD model (Valko et al., 2010) where the primary difference is that the extension accounts for curvature in the inverse loss ratio plot which SEPD represents as a straight line.

Using production data from Barnett-Shale, the proposed models are compared to existing models and show a unique level of flexibility in fitting production data. The power law exponential fits the power law behavior of inverse loss ratio through time. This corresponds to the inverse loss ratio following a straight-line relationship on a log-log plot. Eq. 11 represents the power law exponential relationship between the current

rate (q) and initial rate (q_i) using decline curve parameters α_0 and β_0. These parameters are empirically approximated based on loss ratios for a given well.

$$q = q_i exp\left[\frac{\alpha_0 t^{1-\beta_0}}{\beta_0 - 1}\right] \quad \text{.....(11)}$$

Variable power law exponential is a decline curve for observation of variable power law behavior of inverse loss ratio through time. This corresponds to the inverse loss ratio following a curved line relationship on a log-log plot. Eq. 12 represents the power law exponential relationship between the current rate and initial rate based on a predefined set number of decline curve parameters α_j and β_j.

$$q = q_i exp\left[\sum \frac{\alpha_j t^{1-\beta_j}}{\beta_j - 1}\right] \quad \text{.....(12)}$$

These decline curve equations are derived with an emphasis to match not only rate-time historical production, but also b-factor and the loss ratio. The unknown parameters that need to be estimated for these models are the initial rate (q_i) and decline curve parameters α_j and β_j.

1.10 Uncertainty in production estimates

There have been several modifications to the traditional Arps model over the past decade, that account for specific scenarios such as elongated transient periods in low permeability reservoirs and the, power law behavior of the loss ratio.

As with any regression model, the results provide fixed parameter estimates with a single prediction of rate/cumulative production/EUR. In general, the authors do not address the uncertainty associated with the parameters, do not consider the noise in the measurement (the production data), nor do they provide production forecasts with uncertainty. Figure 1.11 and Figure 1.12 show the daily production and cumulative production matching for an oil well using Arps (1945), power law exponential (Daal et al. 2019) and Duong (2014) models. Each model with different sets of parameters obtained using regression show different estimates for daily production and cumulative production. This plot underlines the importance of calculating uncertainty estimates which provides a distribution for future production of a well.

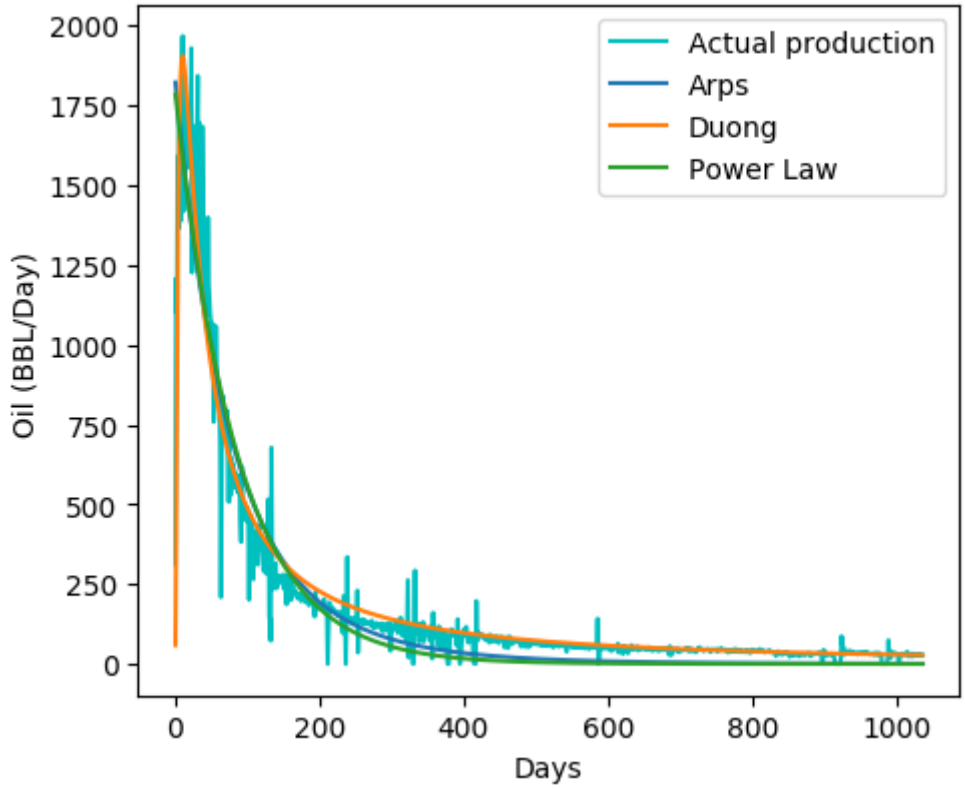

Figure 1.11 Production matching for an oil well using various decline models.

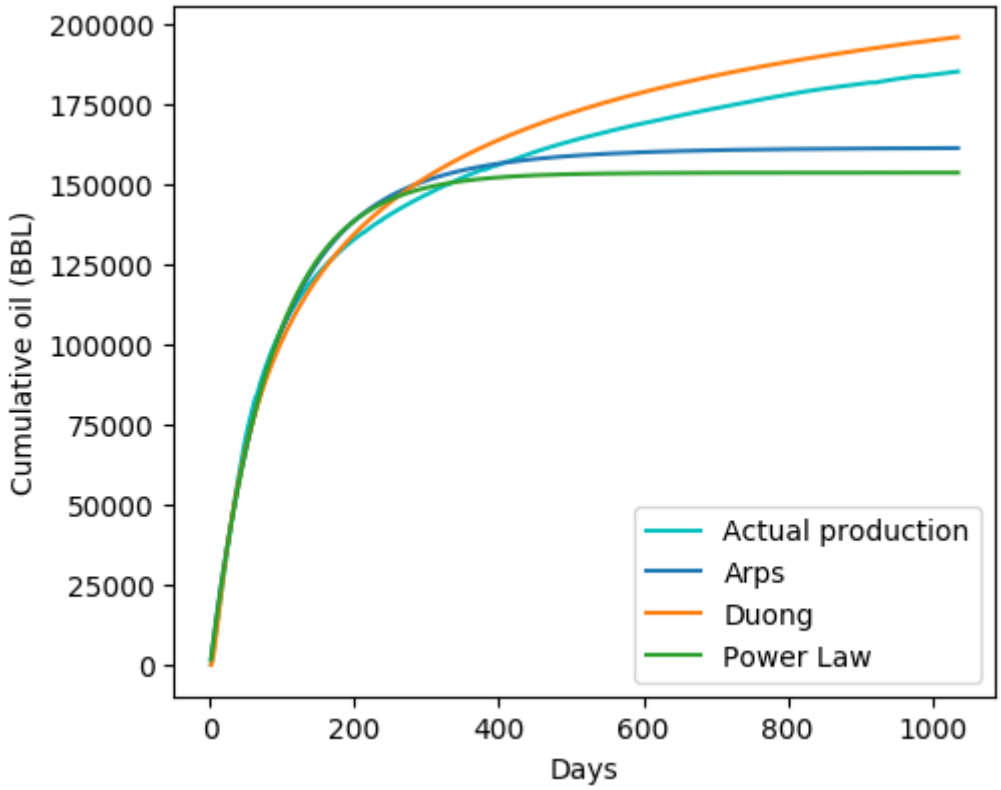

Figure 1.12 Cumulative production matching for an oil well using various decline models.

Although we have numerous methods to analyze production data from unconventional shale wells, we do not have proper methods to quantify uncertainty in the forecasts (McVay et al. 2014). This becomes important when attempting to calculate the economic value of a development project, to make decisions about drilling more or fewer wells or to design surface facilities.

There have been a few attempts to embrace the uncertainty resulting from noise in the data. Jochen and Spivey (1996) and Cheng et al. (2010) developed bootstrap methods that can generate probabilistic decline forecasts and thereby quantify reserves uncertainty. The modified bootstrap method (MBM) (Cheng et al. 2010) is shown to provide estimates of uncertainty in cumulative production that envelope the true value for several field case studies.

McVay et al. (2014) proposed a solution that uses a Bayesian methodology to approximate Arps decline curve parameters. Their approach is based on the Markov chain Monte Carlo (MCMC) method for parameter estimation and was tested on 197 wells. This was the first paper to provide a comprehensive treatment of uncertainty in forecasts using samples from the MCMC-derived posterior distribution. The major drawback of this approach was that it was not applied to and tested with several of the more recent unconventional shale well models.

This thesis documents the application and validity of the MCMC approach to decline curve parameter estimation for several different decline curve models when applied to production data from unconventional wells. Additionally, there has been no comprehensive analyses of parent and child well performance in terms of decline curve parameters. I address this shortcoming in my thesis. In the next section, I provide a brief description of the field data used in this thesis.

1.11 Field characteristics

The purpose of this research is to ascertain the production performance using uncertainty estimates for wells in a single play. In this thesis, I use data from the Mississippian Meramec formation in the Sooner Trend Anadarko Canadian and Kingfisher (STACK) play, which is an unconventional target in the Anadarko basin in west central Oklahoma. It is one of the most productive tight-oil systems in the Anadarko basin

(Almasoodi et al. 2020). Since 2017, the Meramec has been a high-interest area for oil and gas production (Haustveit et al. 2017), but the recent drop in oil prices in 2020 has severely dampened enthusiasm for additional drilling operations. The Meramec STACK play is a multi-layered tight oil reservoir and most development is centered in the Kingfisher and Canadian counties in Oklahoma (Li et al. 2020). A map of the Devon Energy company acreage is shown in Figure 1.13.

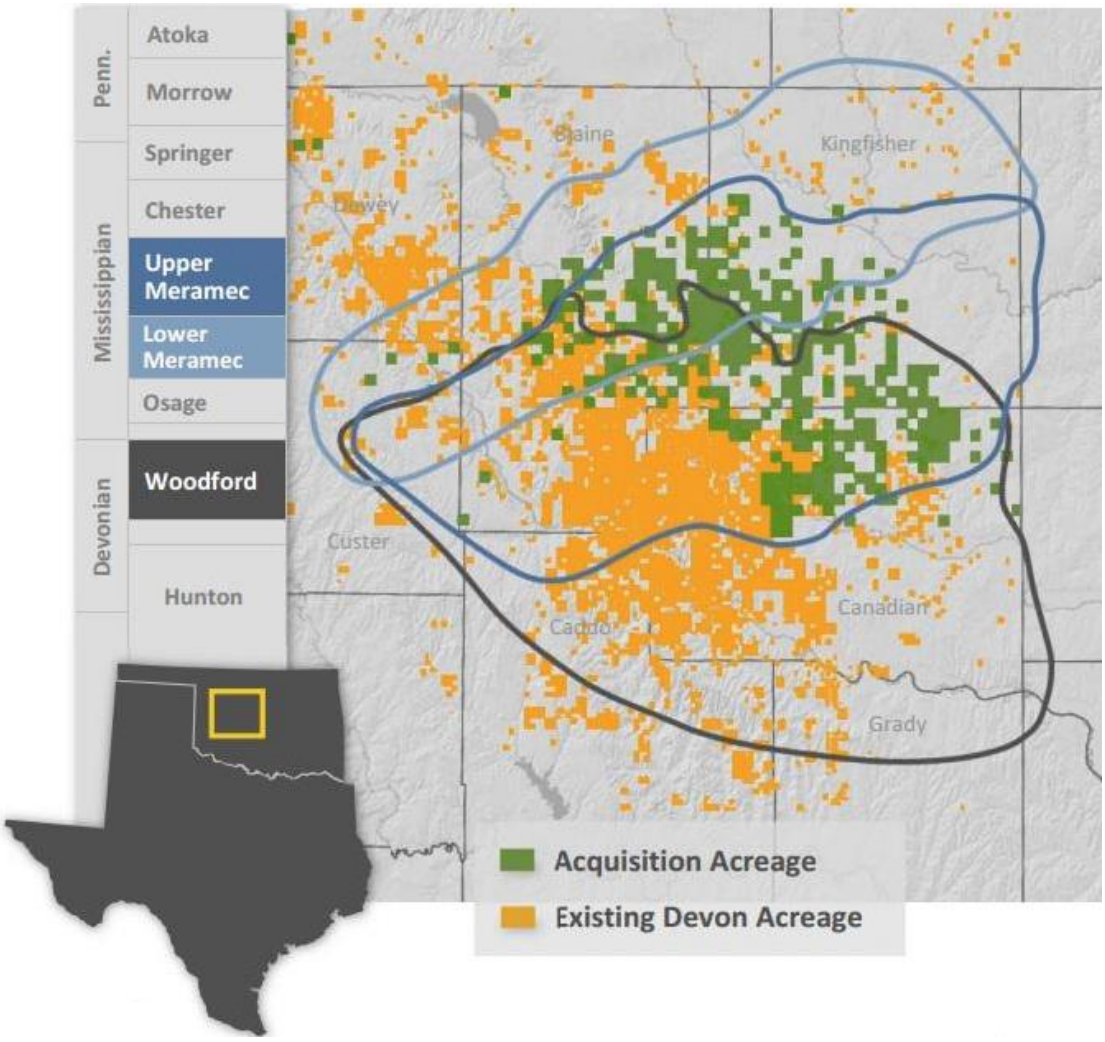

Figure 1.13 Meramec STACK play (Devon energy, 2015).

The wells that came online in the past one year are excluded while performing the uncertainty analysis. 268 oil and gas wells have been identified in the Meramec field for this research. Figure 1.14 shows the number of wells that came online every year since October 2014. The first producing well was drilled in October 2014 and starting with the second quarter of 2018, there has been a prolific increase of infill drilling in the area. Infill drilling method refers to the drilling of additional wells in a field to recover additional oil.

It has the effect of decreasing the spacing between wells and increases the probability of well interference.

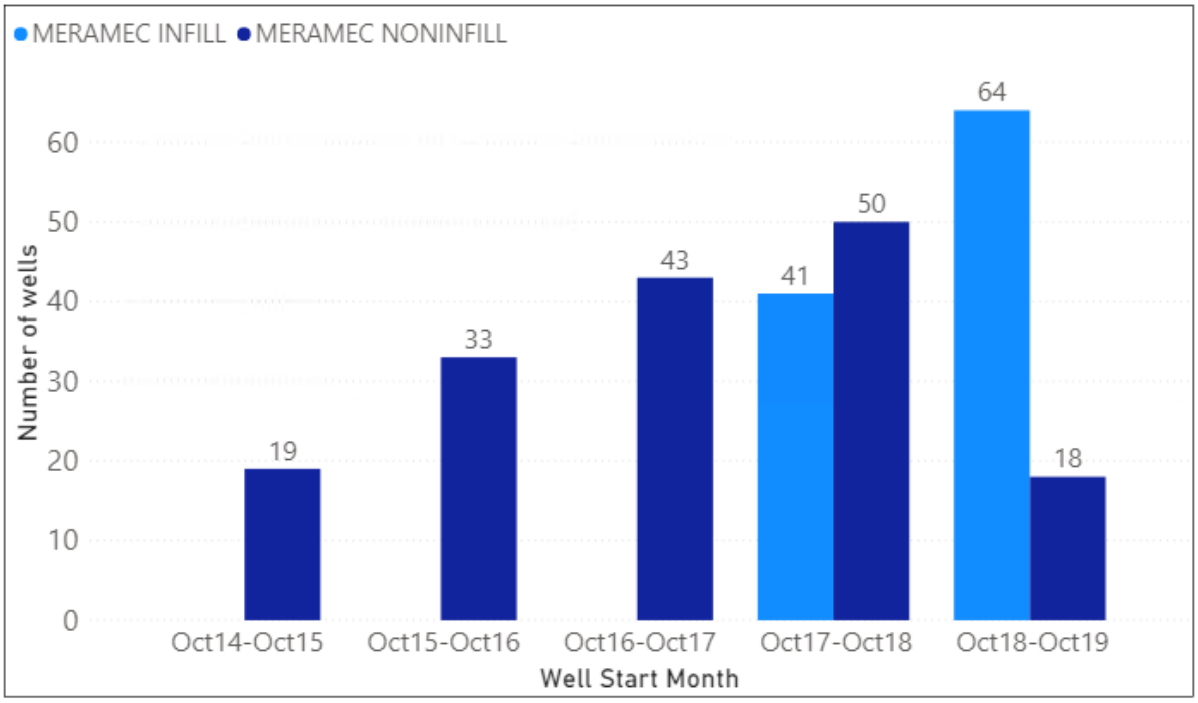

Figure 1.14 Meramec wells operated by Devon energy.

1.12 Organization of the Thesis

This thesis is organized into four chapters. Chapter 1 discusses the introduction to the decline curve models and their application for wells based on various classifications. Chapter 2 explains the Mathematical foundation of this research and explain about Bayesian methods, MCMC and the samplers that made the computation of random variable estimates so effective. It also includes a simple demonstration of how a simple regression fit can be obtained by specifying a model using PyMC3 (2016). Key estimates like maximum a posteriori (MAP), posterior probability distribution, are explained. Chapter 3 provides a summary of the parameter distributions generated using Arps (1945), power law exponential (Daal et al. 2019) and Duong (2014) models for the chosen wells from STACK. It will also include the summary of metrics R-squared, RMSE and MAE for the posterior estimates of the sample wells. Posterior distribution graphs for sample wells using three decline models are shown. An analysis of change in the initial rates and the decline rates over time is shown vs. number of frac states, lateral length and the amount of proppant used. Based on our understanding of the decline curve values for wells in this field, we present analysis on how infill development and the completion

design has impacted the production in the field. Finally, Chapter 4 concludes the thesis and presents the scope of future work.

Chapter 2

Mathematical foundation

In this chapter, I introduce the probabilistic framework which is the basis for Bayes theorem. In the following sections, I explain the advent of Markov chain Monte Carlo (MCMC) methods and provide a detailed explanation on four sampling methods. I end this chapter by providing a simple example of Bayesian linear regression to demonstrate the application of MCMC to parameter estimation and uncertainty quantification.

From a Bayesian perspective, probability quantifies the degree of confidence we have in the estimate of a parameter. A frequentist approach treats unknown parameters as fixed values, thereby no probabilities (prior) can be assigned to them.

2.1 Bayes theorem

Consider a situation with some noisy measurements, y and a possible model with parameters, θ that generated that data. We would like to assess the confidence we have that the model with parameters, θ generated the noisy measurements. If this model is unlikely, we would like to find the most probable set of model parameters that generated the noisy observations or measurements.

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \qquad \text{.....(13)}$$

This is expressed by Bayes' theorem as shown in Eq. 13. The term, $p(\theta|y)$ is denoted as the posterior distribution and reflects the confidence we have that a model with parameters, θ generated the data, y. The maximum of this distribution is called the Maximum A Posteriori estimate of θ also known as the MAP estimate.

To compute the posterior probability of the model parameters, we need to define $p(\theta)$ which is the prior distribution of the model parameters before even looking at the data, y. $p(y|\theta)$ is the likelihood function and, expresses the plausibility of the data given the parameters, θ. $p(y)$ is called the marginal likelihood. Also known as evidence, it is the probability of observing the data averaged over all possible values of parameters and is generally not computed because Bayes' theorem is a proportionality if marginal likelihood is ignored as expressed in Eq. 14.

$$p(\theta|y) \propto p(y|\theta)p(\theta) \quad \text{.....(14)}$$

To find the MAP estimate of the model parameters, we do not need to know the proportionality constant. While the analytic form of the posterior distribution can occasionally be computed, it makes sense to use sampling methods to compute the posterior, especially in high-dimensional spaces. In the next section, I describe a few of these sampling approaches.

2.2 Markov chain Monte Carlo (MCMC)

MCMC uses Monte Carlo integration using Markov chains to integrate over high-dimensional probability distributions to infer model parameters. Monte Carlo integration draws samples and then creates sample averages to approximate expectations. Markov chain Monte Carlo draws samples by constructing a Markov chain for a long time. There are different samplers to construct these chains. All the newer samplers including Gibbs Sampler (Geman et al. 1984) are special cases of the general framework of Metropolis et al. (1953) and Hastings (1970).

2.2.1 Monte Carlo integration

Metropolis and Ulam (1949) first published a document on Monte Carlo simulation. This paper introduces Monte Carlo particle methods which form the basis for modern methods such as bootstrap filters, condensation, and survival of the fittest algorithms. The features of the posterior distribution – moments, quantiles, highest posterior density regions, MAP estimates are all legitimate for Bayesian inference and they can be expressed in terms of posterior expectations of functions of the parameter θ. Let X be a vector of random variables with a posterior distribution p(x). The posterior expectation for f(X) is given by the equation:

$$E[f(X)] = \frac{\int f(x)p(x)dx}{\int p(x)dx} \quad \text{.....(15)}$$

Monte Carlo integration evaluates E[f(X)] by drawing samples $\{X_t, t = 1, \dots, n\}$ from p(x) and then approximating as shown in Eq. (16) below:

$$E[f(X)] \approx \frac{1}{n}\sum_{t=1}^{n} f(x_t) \qquad \text{.....(16)}$$

The population mean of f(X) is estimated by a sample mean and when the samples are independent, the law of large numbers ensures that the approximation is accurate by increasing the sample size n. One way of generating the samples x_t is by using a Markov Chain which makes it Markov chain Monte Carlo.

2.2.2 Markov chain

Markov chain is a sequence for which a set of random variables (X_1, X_2, ... , X_t) are identified such that at each time t >= 0, the next state X_t is sampled from a distribution $P(X_t|X_{t-1})$ which depends only on the previous state of the chain and does not depend on the prior history of the chain (X_1, X_2, ... , X_{t-2}). A sequence of random variables (X_1, X_2, ... , X_t) on a discrete state space is called a first order Markov Chain if:

$$p(X_t = x_t \mid X_{t-1} = x_{t-1}, \ldots, X_1 = x_1) = p(X_t = x_t \mid X_{t-1} = x_{t-1}) \qquad \text{.....(17)}$$

In other words, the distribution of X_t given the whole history of the process is the same as the distribution of X_t given just the most recent value, X_{t-1}. Therefore, the more steps that are included, the more closely the sample distribution matches actual distribution. The chain will gradually forget its initial state and the probability function will converge to a unique stationary distribution. We discuss various sampling methods in the upcoming sections.

2.2.3 Rejection sampling

In the 1940s, Ulam and Neumann developed many Monte Carlo algorithms, including importance sampling and rejection sampling. Assume that we wish to generate samples from a distribution p(x) that is not one of the standard distributions. For example, it may be a product of several different other distributions. We define a proposal distribution, q(x) and a constant M such that $p(x) \leq Mq(x)$ for $x \in X$. A value $x^{(i)}$ sampled from q(X) will be accepted if it meets condition in Eq. 18

$$u < \frac{p(x^{(i)})}{Mq(x^{(i)})} \qquad \text{.....(18)}$$

u in Eq. 18 is a random value drawn from a uniform distribution with the range (0,1). This process is repeated until a predetermined number of iterations are reached. The resulting accepted values are then distributed according to p(x). Figure 2.1 illustrates this concept of rejection sampling.

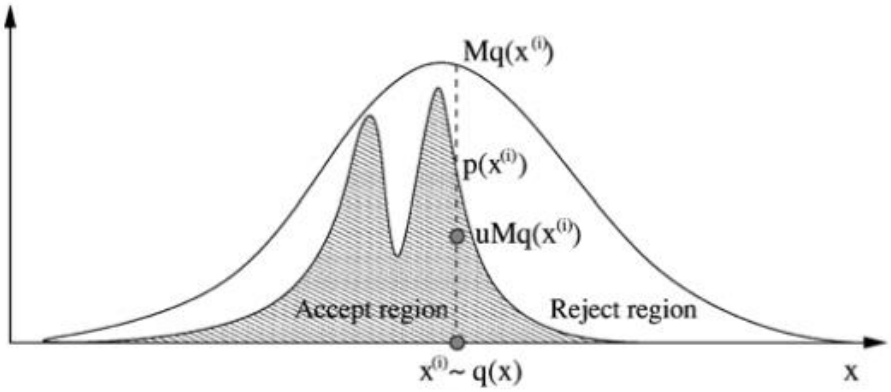

Figure 2.1 Rejection sampling [Andrieu et al., 2003].

This simple method suffers from a severe limitation since it is not always possible to bound p(X)/q(X) with a reasonable constant M.

2.2.4 Importance sampling

Importance sampling helps approximate expectations from distributions especially when the distribution has a complex, non-standard form. As with rejection sampling, we draw samples from a proposal distribution, q(x). Assuming we want to estimate the mean of the function, f(x), with respect to probability distribution, p(x). Using this principle, the possible Monte Carlo simulation for the integral sum for f(x) is shown in Eq. 19.

$$I = \sum_{i=1}^{N} f(x_i) w(x_i) \qquad \text{.....(19)}$$

w(x) in Eq. 19 is the importance weight and is obtained using Eq. 20.

$$w(x) \triangleq \frac{p(x)}{q(x)} \qquad \text{.....(20)}$$

p(x) denotes the posterior density and q(x) is the proposal distribution. The proposal distribution which minimizes the variance of the estimator integral must be chosen.

2.2.5 Metropolis-Hastings sampling

Metropolis-Hastings (MH) algorithm is the most popular MCMC method (Hastings, 1970; Metropolis et al., 1953). An MH step of a distribution p(X) and the proposal distribution q(X) involves sampling a candidate value x_i given the current value x, which is $q(x_i \mid x)$. The Markov chain then moves towards x_i with acceptance probability defined by Eq. 21.

$$A(x, x_i) = min\left\{1, \frac{p(x_i)q(x|x_i)}{p(x)q(x_i|x)}\right\} \qquad \text{.....(21)}$$

Similar to rejection sampling, a value 'u' is drawn from a uniform distribution with the range (0,1). If the acceptance probability is greater than 'u', the Markov chain moves towards x_i with the acceptance probability value of $A(x, x_i)$, otherwise it remains at x for that iteration. This process is repeated until the total number of iterations for sampling are exhausted. Figure 2.2 indicates how a Gaussian proposal distribution better approximates a bimodal target distribution as the number of iterations increase from 100 to 5000. The proposal normal distribution has a sampled mean and a variance of 100. For this example, the target distribution is identified by Eq. 22:

$$p(x) = 0.3 * e^{-0.2*x^2} + 0.7 * e^{(-0.2(x-10)^2)} \qquad \text{.....(22)}$$

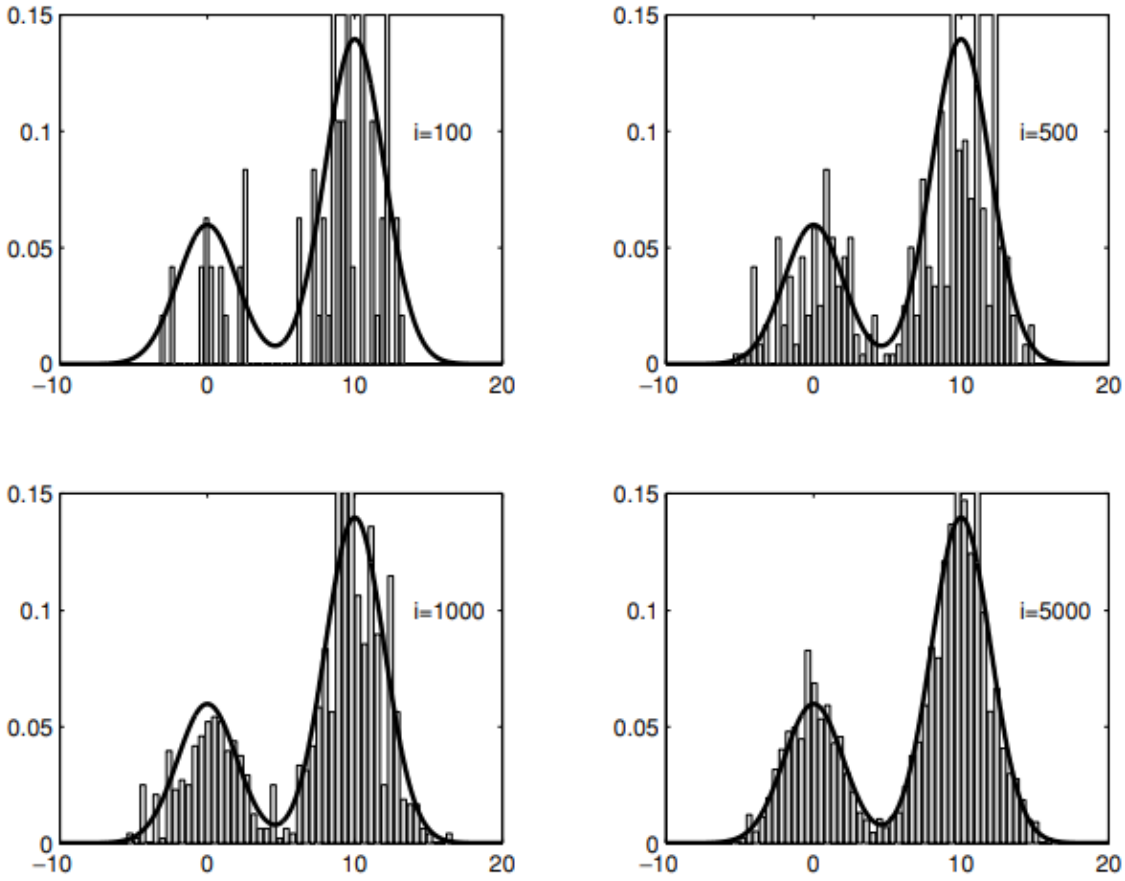

Figure 2.2 Target distribution and histogram of the MCMC samples for different iterations [Andrieu et al., 2003].

2.2.6 Gibbs sampling

Donald and Stuart Geman (1984) showed how the Metropolis Hastings algorithm could be adapted to the high-dimensional problems that arise in Bayesian statistics. This algorithm was named after physicist Josiah Gibbs. Gibbs sampler generates posterior samples for each variable by sampling from its conditional distribution whereas the remaining variables are fixed to their current values. For D random variables X_1, X_2, X_3…, X_D, below algorithm shows the sampling mechanism:

Step 1: Initialize initial values $x^{(0)}$ for the random variables from proposal distribution q

Step 2: for I = 1, 2, … N:

$$x_1^{(i)} \sim p(X_1=x_1 | X_2=x_2^{(i-1)}, X_3=x_3^{(i-1)}, \dots, X_D=x_D^{(i-1)})$$

$$x_2^{(i)} \sim p(X_2=x_2 | X_1=x_1^{(i-1)}, X_3=x_3^{(i-1)}, \dots, X_D=x_D^{(i-1)})$$

.

.

.

$x_D^{(i)} \sim p(X_D=x_D | X_1=x_1^{(i-1)}, X_2=x_2^{(i-1)}, \dots, X_{D-1}=x_{D-1}^{(i-1)})$

This process continues until the proposal distribution matches the actual posterior distribution. Using this algorithm, we simulate samples by sweeping through all the posterior conditionals, one random variable at a time. Since they are initialized using random variables, the early samples may not represent the actual posterior distribution. With many iterations, MCMC converges to the posterior eventually.

To summarize, importance sampling works for scenarios where it is easy to calculate the proposal distribution and easy to specify a constant value as the threshold for the proposal to posterior distribution ratio. Rejection sampling on the other hand is for scenarios where it is difficult to produce samples from the population to model rare events. Gibbs sampling breaks the curse of dimensionality by producing low dimensional conditionals. Metropolis sampling creates a Markov chain based on an acceptance-rejection step. Gibbs sampling can be considered a special case of Metropolis algorithm with an acceptance probability of one. Given the volatility in the well performance and production data, Metropolis sampler is used for generating the posterior distributions for all decline curve models.

2.3 Posterior analysis

The result of a Bayesian analysis is a posterior distribution which is a set of plausible values generated based on prior, data and the likelihood (model). Posterior spread is proportional to the parameter uncertainty and a larger spread indicates a higher degree of uncertainty.

A commonly used method to summarize posterior spread is to use a highest-posterior density (HPD) interval. If 95% HPD for the parameter is [2-5], it means that according to the given prior and model, the distribution of the variable in question is between 2 and 5 with a probability of 95%. In common terms, this interval is a credible interval. Performing a Bayesian analysis enables the possibility of having a probability distribution of a parameter. This is not possible with the frequentist approach of

confidence intervals since the parameters are fixed by design and the confidence interval either contains or does not contain the true value of the parameter.

The mode value of the posterior distribution is called maximum a posteriori (MAP). It is generally found using the numerical optimization methods. The posterior predictive distribution is sampled to calculate outcome production values. The computed cumulative production values can further be classified into P10, P50 and P90 estimates by performing percentile calculations.

2.4 Bayesian example using PyMC3

PyMC3 (Salvatier et al., 2016) is a package that enables probabilistic programming in Python. Though the base code for this package is written using Python, the computationally demanding parts are written using NumPy and Theano. Theano compiles Python code to C code.

For this demonstration, I use a simple linear equation of the form $y = \alpha + \beta x$. The first step is to assume values for intercept and slope of the linear equation and calculate the outcome value. The second step is to add random noise to the value. Figure 2.3 shows the true regression line in thick black and the points generated by adding random noise as blue dots.

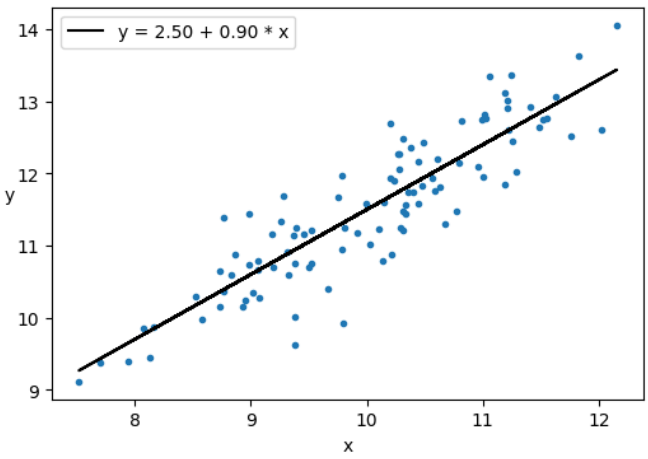

Figure 2.3: Actual data (straight-line) with added error (blue dots).

The second step is the model definition where prior distributions are assigned to the parameters. The model defines the intercept α as a normal distribution with a mean value of 0 and a standard deviation of 10. The slope β is also defined as a normal

distribution with a mean value of 0 and a standard deviation of 1. Likelihood is defined as a normal distribution whose mean value is deterministic value generated by the original linear equation. Half-Cauchy distribution is used for standard deviation of the likelihood. This distribution with heavy tails suits model parameters with a higher uncertainty. The model now has prior distributions defined for all parameters and a likelihood function defined as normal distribution. Figure 2.4 shows the posterior trace as a frequency distribution plot as well as a line chart over all iterations. The posterior trace is generated over four sampling chains of 2000 iterations and each chain is shown as a different color line. The trace shows good convergence proving the choice of priors to be correct.

Posterior trace values are plugged into the original linear equation to obtain the range of outcome values shown in Figure 2.5. All outcomes plotted together explain the amount of uncertainty associated with the predicted parameter values.

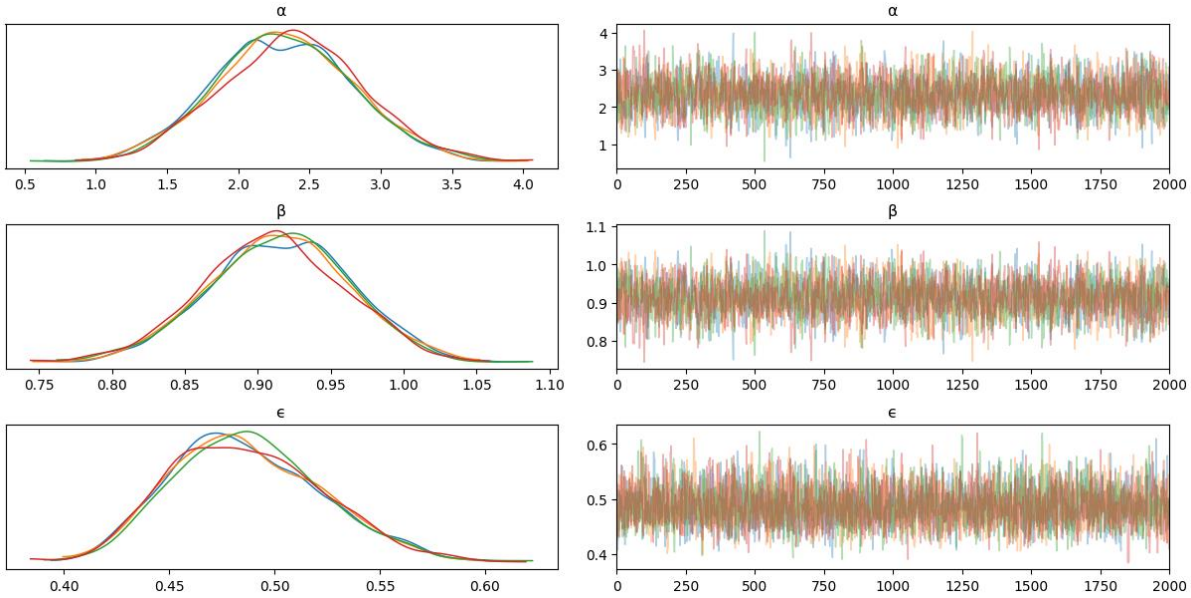

Figure 2.4: Posterior trace of three parameters.

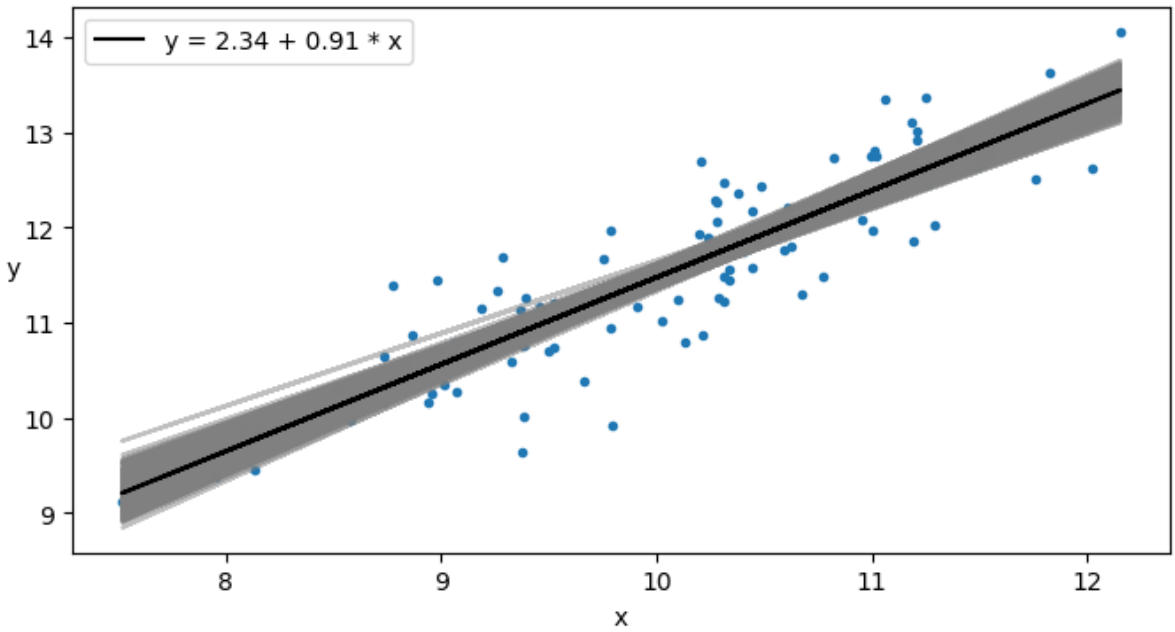

Figure 2.5: Predictions generated from posterior trace.

The thick black line in Figure 2.5 is a calculated outcome using the mean values from posterior trace, blue dots are the actual data used to train the model and the remaining gray lines represent all possible outcomes.

Chapter 3

MCMC Application and Results

In this chapter, I will present a detailed analysis and summary of decline curve parameter estimation and uncertainty estimates generated for 268 wells in the STACK play. In the first section, I discuss the data gathered for the purpose of this research. In the second section, I will discuss the parameter convergence pattern and the posterior distribution of the parameters generated for an oil well using the Arps model. In the third section, I will present the results for few sample oil and gas wells using three DCA models – Arps (1945), Duong (2014) and power law exponential (Daal et al. 2019) models. In the fourth section, I will present the summary statistics of the parameters approximated for three models.

3.1 Data

This section describes the various production and completion data attributes collected for this thesis.

Rates: The daily production rates for oil, gas, and water for all wells. The earliest well came online in 2014 and the last well came online in 2019.

Downtime: Wells undergo downtime occasionally due to various operational reasons that may be intentional or non-intentional. Chemical treatment, offset activities, equipment maintenance necessitates intentional downtime while the involuntary reasons include equipment failure. For this research, the number of daily downtime hours and the reason for downtime is capture for all wells.

Stimulation: A treatment performed to restore or enhance the productivity of a well. Stimulation in shale gas reservoirs is done by hydraulic fracturing treatments. Fracturing creates a highly conductive flow path between the reservoir and wellbore.

Frac fluid: A fluid injected into a well as part of the stimulation operation.

Proppant: Sized particles mixed with frac fluid to hold fractures propped open after a hydraulic fracturing treatment.

Stim stages or frac stages: The number of intervals that are stimulated using hydraulic fracturing for a horizontal wellbore.

Lateral length: The length of the horizontal wellbore.

Infill drilling: The drilling of additional wells in a field to recover additional oil. It has the effect of decreasing the spacing between wells and increases the probability of well interference.

3.2 Prior and Likelihood function

In this section, I will present the priors and likelihood functions as probabilistic distributions to obtain the posterior distributions for the decline curve parameters. The prior distribution essentially quantifies the users' degree of confidence in a parameter value or set of parameter values. The likelihood function, for a given model, expresses the probability that the measurements obtained were generated by the model. For this research, I use non-informative or uniform priors.

In the Arps (1945) model, the value of b and log values of initial rate q_i and, decline rate D_i are modeled as uniform distributions with the ranges specified below:

$$-4.61 < \log(q_i) < 13.8$$

$$-2.3 < \log(D_i) < 3.91$$

$$0 < b < 2$$

The upper- and lower-limits for $\log(q_i)$ are chosen based on an examination of initial rates across the field. The range for $\log(D_i)$ is chosen to account for abnormal or normal daily decline rates. The upper limit of 2 for hyperbolic b parameter is chosen because b tends to be greater than 1 for unconventional plays.

For the Duong (2014) model, I first use a curve fitting procedure to arrive at the initial estimates for log value of the theoretical rate at day 1, slope and intercept of the log-log plot between the ratio of daily-rate and cumulative production q/G_p versus time, $\log(q_1)$, m, and a respectively, and define uniform distributions around these initial estimates as prior distributions.

For the power law exponential (Daal et al. 2019) model, I first use a curve fitting procedure to arrive at the initial estimates for the log value of the initial rate q_i and the two dimensionless decline curve parameters, α and β, log(q_i), log(α), and log(β) respectively, and define uniform distributions around this value to be used as prior distributions.

The likelihood function can be modeled as a normal distribution if the noise in the data is assumed to be from a zero-mean, normal distribution (McVay et al. 2014).

$$p(y|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left(\frac{-\left(y_{abs} - y_{pred}\right)^2}{2\sigma^2}\right) \qquad \text{.....(23)}$$

Eq. 23 shows the likelihood modeled as a normal distribution for a noise standard deviation of σ (BBL or MCF) in the production data. In Eq. 23, y_{abs} and y_{pred} are the observed and predicted values of flow rates, respectively. In this research, θ represents the decline curve parameters for Arps (1945), Duong (2014) and power law exponential (Daal et al. 2019) models respectively, as shown below:

$$\theta = [log(q_i), log(D_i), b]^T$$
$$\theta = [log(q_1), m, a]^T$$
$$\theta = [log(q_i), log(\alpha), log(\beta)]^T$$

For a well producing for n days, y_{pred} = (y_1, y_2, … ,y_n) indicates the 'n' sequential observations of production. The likelihood function for the 'n' sequential observations becomes:

$$p(y\,|\theta) \propto exp\left[\left(y_{pred} - y_{abs}\right)^T C_y^{-1}\left(y_{pred} - y_{abs}\right)\right] \qquad \text{.....(24)}$$

C_y represents the variance between the n observations as shown below:

$$C_y = \begin{bmatrix} \sigma_{y_1}^2 & \sigma_{y_1y_2}^2 & \dots & \sigma_{y_1y_n}^2 \\ \sigma_{y_2y_1}^2 & \sigma_{y_2}^2 & \dots & \sigma_{y_2y_n}^2 \\ & \cdot & & \\ & \cdot & & \\ & \cdot & & \\ \sigma_{y_ny_1}^2 & \sigma_{y_ny_2}^2 & \dots & \sigma_{y_n}^2 \end{bmatrix} \qquad \text{.....(25)}$$

The observations are independent of each other, which implies the variances between different observations can be assumed as zero. Assuming the remainder variances are equal denoted by σ_y^2, the matrix in Eq. 25 can be represented as:

$$C_y = \begin{bmatrix} \sigma_y^2 & 0 & \dots & 0 \\ 0 & \sigma_y^2 & \dots & 0 \\ & & \cdot & \\ & & \cdot & \\ & & \cdot & \\ 0 & 0 & \dots & \sigma_y^2 \end{bmatrix} \qquad \text{……..(26)}$$

I use Metropolis algorithm for sampling. Using a Markov chain, we can approximate the distribution of a parameter θ_s given the most recent value θ_{s-1}. While constructing the Markov chain, we draw samples from a proposal distribution 'q' since the posterior distribution is unknown. There is a probability 'α' that that the candidate $\theta_{proposal}$ drawn from the proposal distribution is accepted and a probability (1- α) that it is not accepted. Ratio of the posterior probability of $\theta_{proposal}$ to θ_{s-1} is given by Eq .25 (McVay et al. 2014).

$$\text{Posterior probability ratio} = \frac{p(\theta_{proposal}|y)}{p(\theta_{s-1}|y)} \qquad \text{……..(25)}$$

The Markov chain generated by use of the Metropolis algorithm will converge to the desired posterior distribution when the acceptance ratio 'α' equals the minimum of numeric value one and ratio of the normalized posterior probability of $\theta_{proposal}$ to θ_{s-1} is given by Eq. 26 (McVay et al. 2014).

$$\text{Normalized posterior probability ratio} = \frac{\frac{p(\theta_{proposal}|y)}{q(\theta_{proposal}|\theta_{s-1})}}{\frac{p(\theta_{s-1}|y)}{q(\theta_{s-1}|\theta_{proposal})}} \qquad \text{……(26)}$$

Eq. 26 is derived from Eq. 19 discussed in Section 2.2.5. The reason to normalize proposal distribution is to keep the Markov chain independent of the proposal distribution (McVay et al. 2014).

As defined in Eq. 12 in Section 2.1, posterior distribution is proportional to the product of the likelihood and the prior distribution.

$$p(\theta|y) \propto p(y|\theta)p(\theta) \qquad \text{……..(27)}$$

The maximum number of distinct posterior parameter estimates generated for a well per model are limited to 2000.

3.3 MCMC Convergence

In this section, I will present the results of parameter estimation using the Arps model for a randomly selected oil well (Well 1). The well has been continually producing since May 2018 with no offset activities and therefore is a good candidate to apply and test the MCMC method for parameter estimation. I chose four iterations of MCMC and generate over 20000 successive parameter samples for each iteration for a total of 80000 samples. Figure 3.1 shows the Arps posterior parameter histograms using the initial 20% (176 days) of the production data. Figures 3.2, 3.3 and 3.4 show corresponding figures when the initial 30% (264 days), 40% (352 days) and 50% (441 days) of the production data. The histograms show progressively narrower distributions with the availability of more data and a corresponding decrease in the uncertainty of the parameter estimates. The mode frequency has increased from close to 40,000 to 60,000 for the parameters' q_i and b.

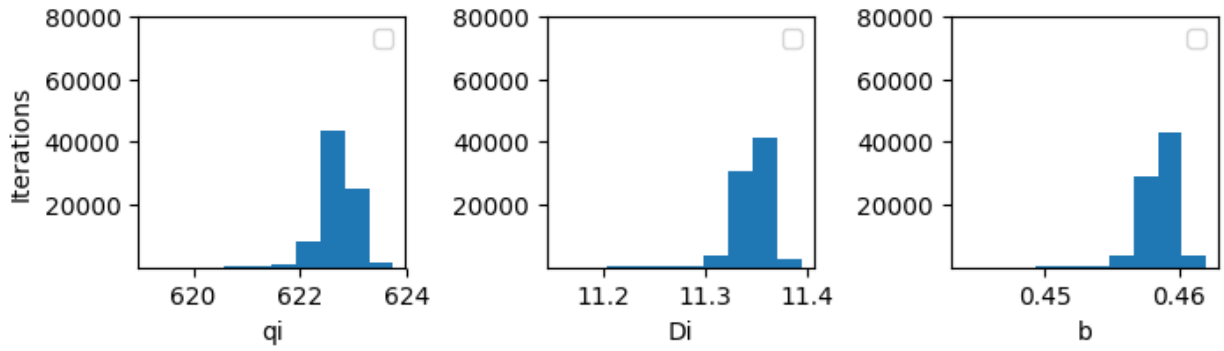

Figure 3.1 Posterior parameter distribution for the model with 20% of data used for training.

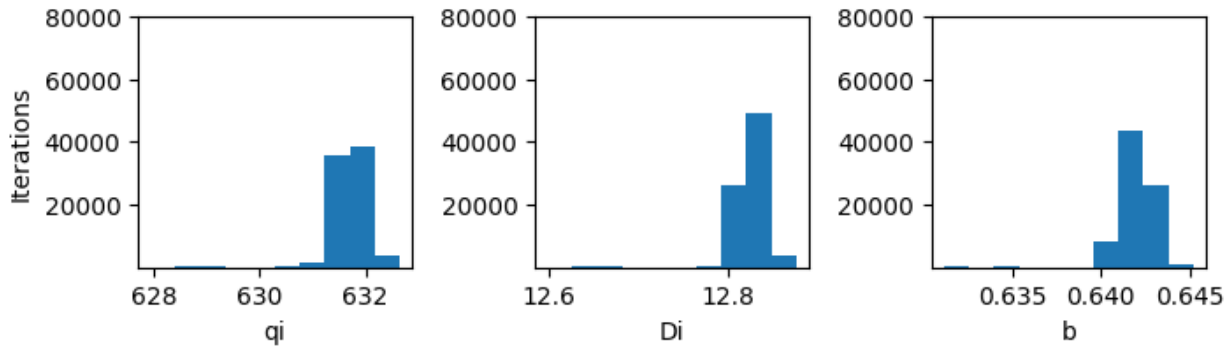

Figure 3.2 Posterior parameter distribution for the model with 30% of data used for training.

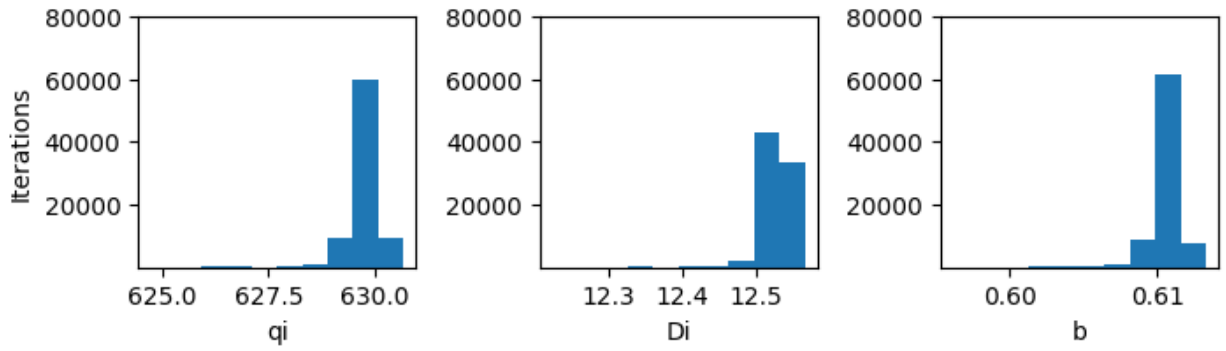

Figure 3.3 Posterior parameter distribution for the model with 40% of data used for training.

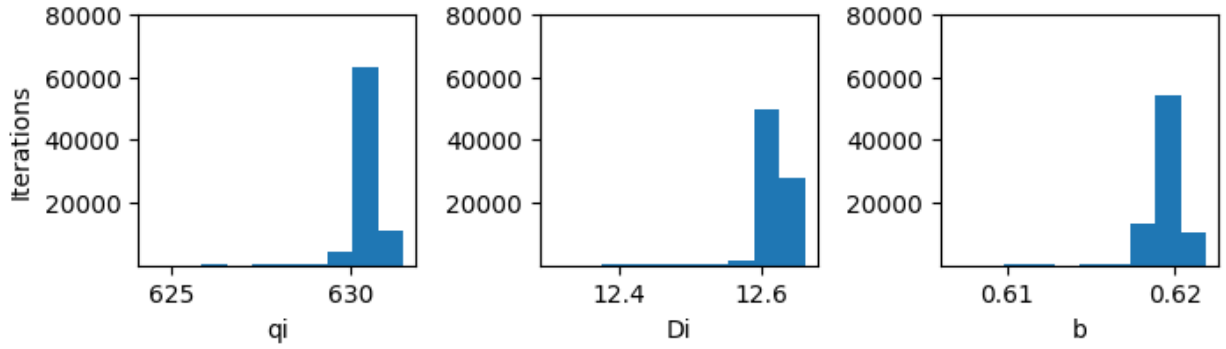

Figure 3.4 Posterior parameter distribution for the model with 50% of data used for training.

Figure 3.5 shows the actual versus predicted daily and cumulative production for Well1. The cumulative production values are calculated using the posterior parameter distribution obtained by training the model with 20% of the initial production data. For the same model, Figure 3.6 shows the parameter distribution for Well1. This is a posterior trace plot which plots the trace values as sampled using the model definition. The values shown on the plot are $\log(q_i)$, $\log(D_i)$ and b. The four draws, each with 20000 iterations, are shown in four different colors.

Figure 3.7 shows the actual versus predicted daily and cumulative production for Well1. The cumulative production values are calculated using the posterior parameter distribution obtained by training the model with 30% of the initial production data. For the same model, Figure 3.8 shows the posterior parameter distribution. The comparison between Figure 3.5 and Figure 3.7 shows the improved match between the calculated and the actual cumulative production. Figures 3.9, 3.10 show similar plots obtained by training the model with 40% of the initial production data. Figures 3.11, 3.12 show similar plots obtained by training the model with 50% of the initial production data. There is little spread associated with the forecasts and relatively less uncertainty for Well1.

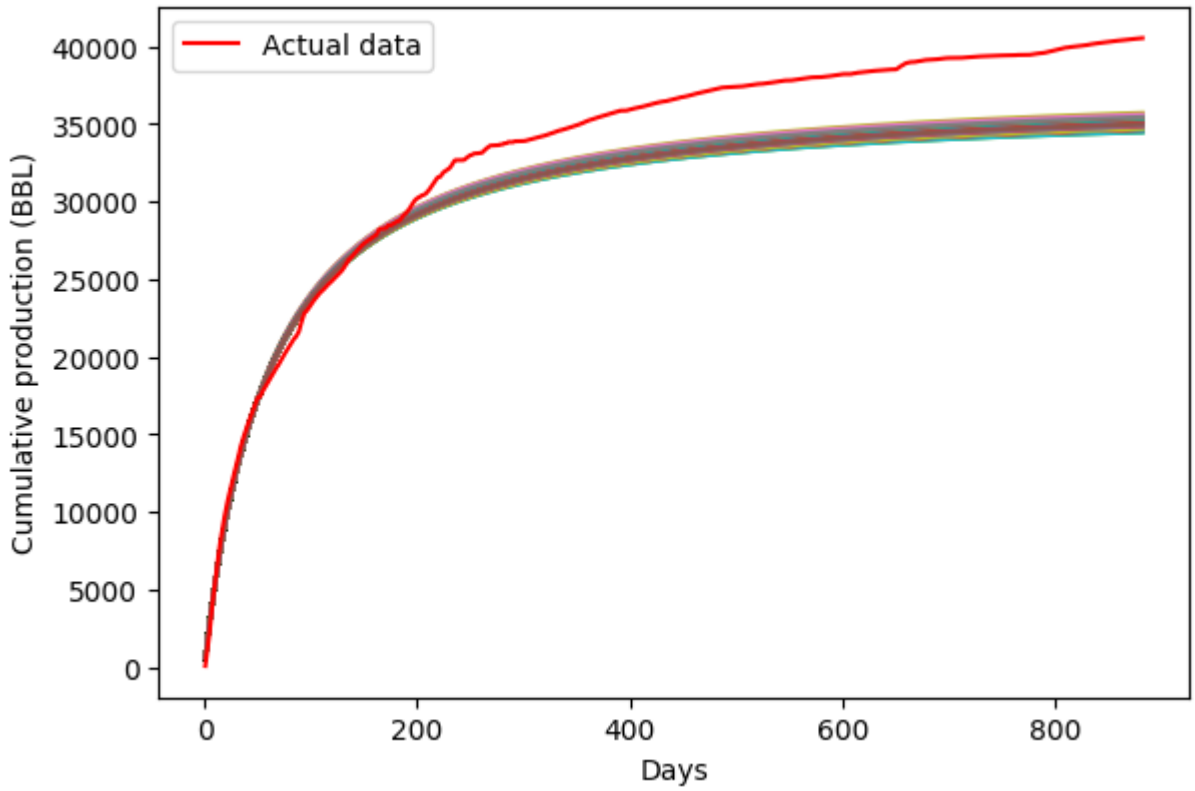

Figure 3.5 Cumulative production calculated using posterior parameter distribution for the model trained with 20% of the initial production data.

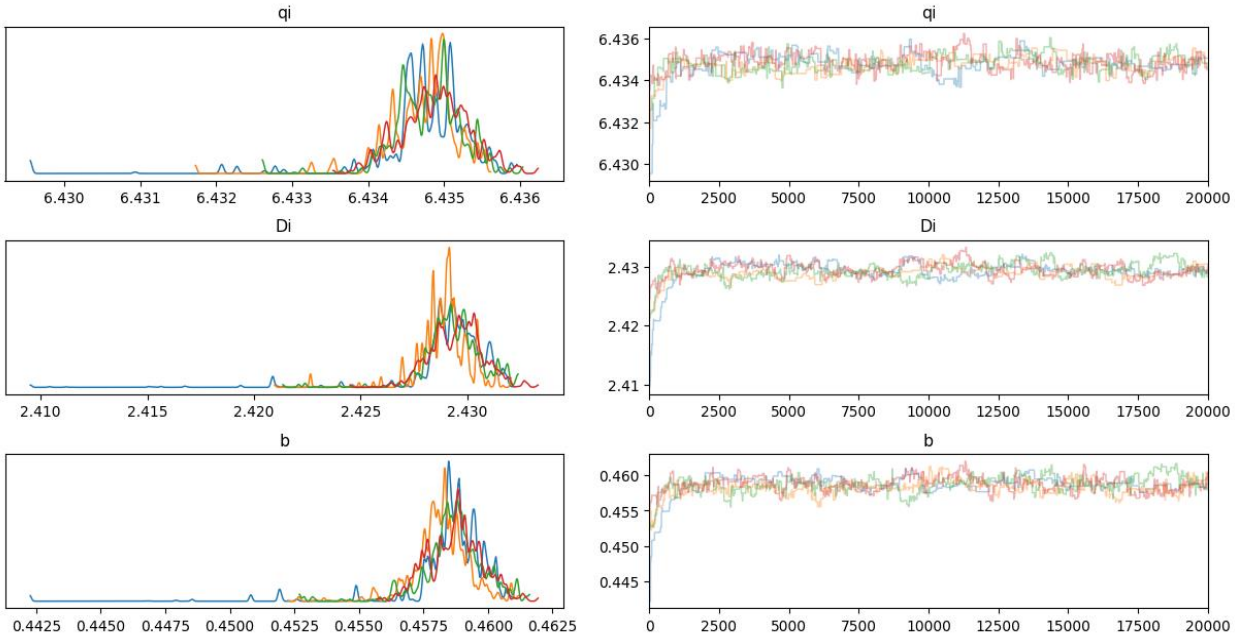

Figure 3.6 Posterior parameter distribution for Well1 with 20% of initial production data used for training.

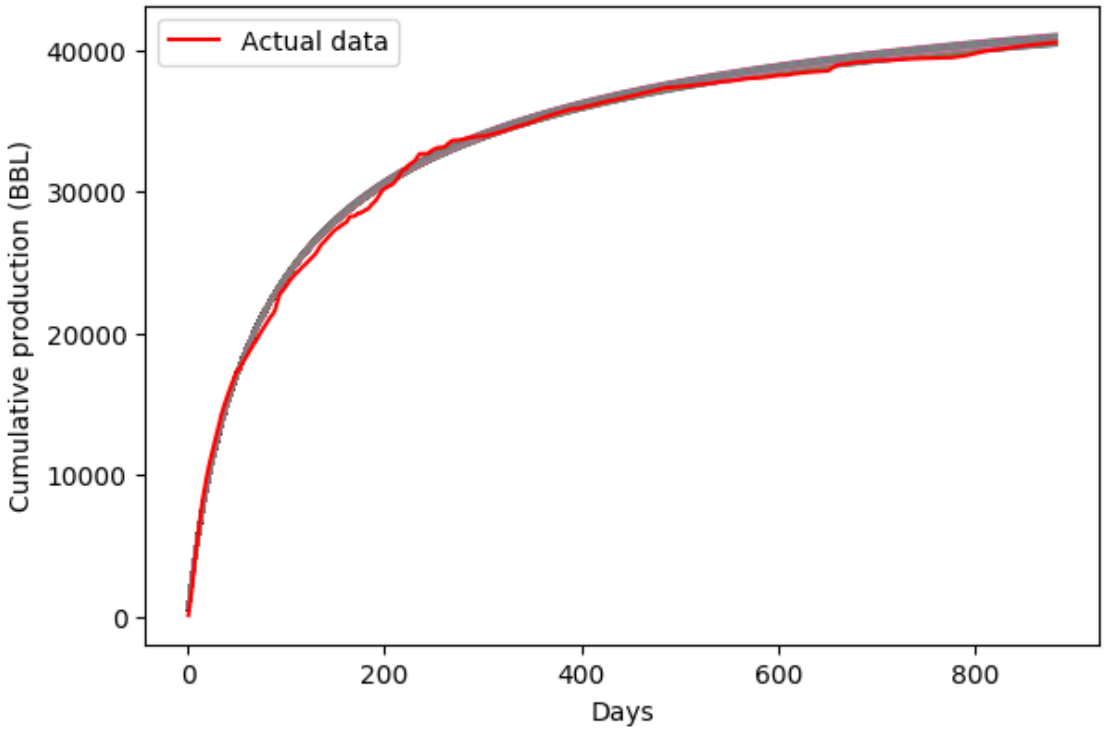

Figure 3.7 Cumulative production calculated using posterior parameter distribution for the model trained with 30% of the initial production data.

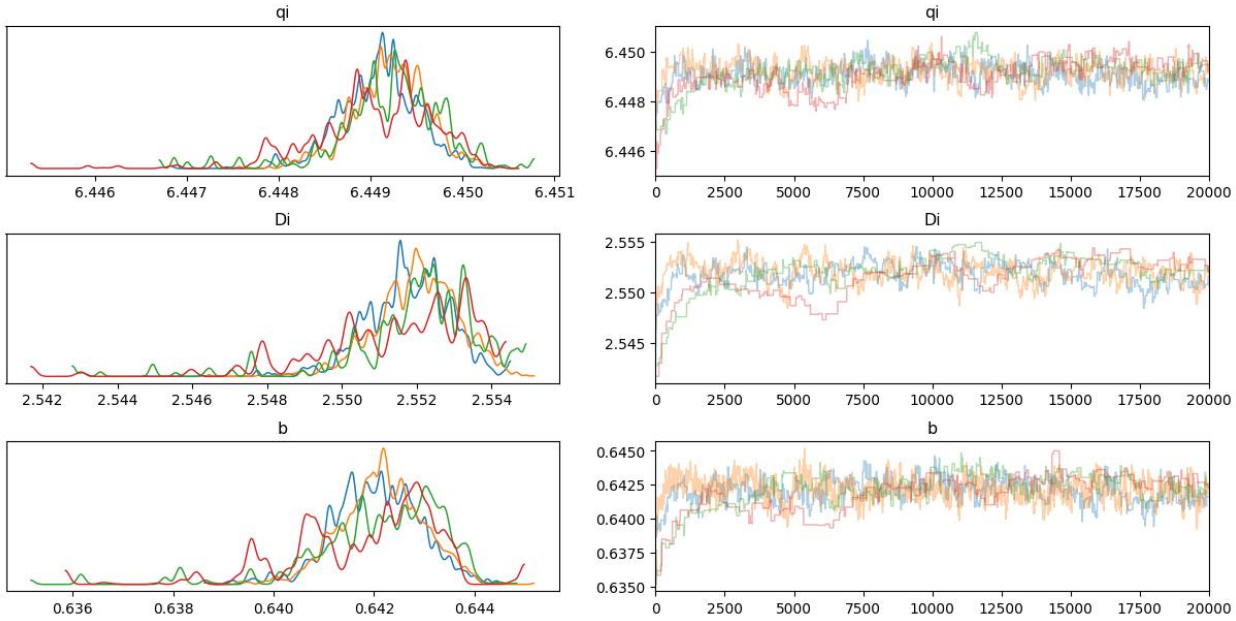

Figure 3.8 Posterior parameter distribution for Well1 with 30% of initial production data used for training.

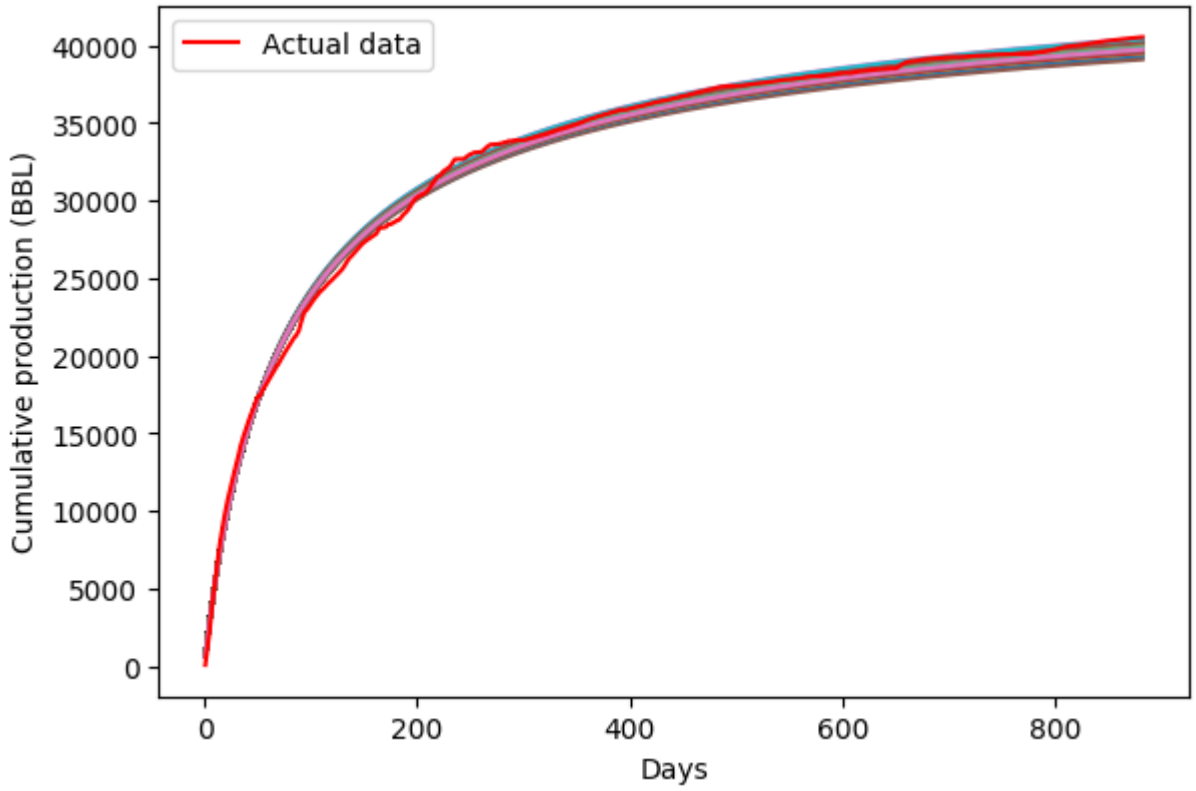

Figure 3.9 Cumulative production calculated using posterior parameter distribution for the model trained with 40% of the initial production data.

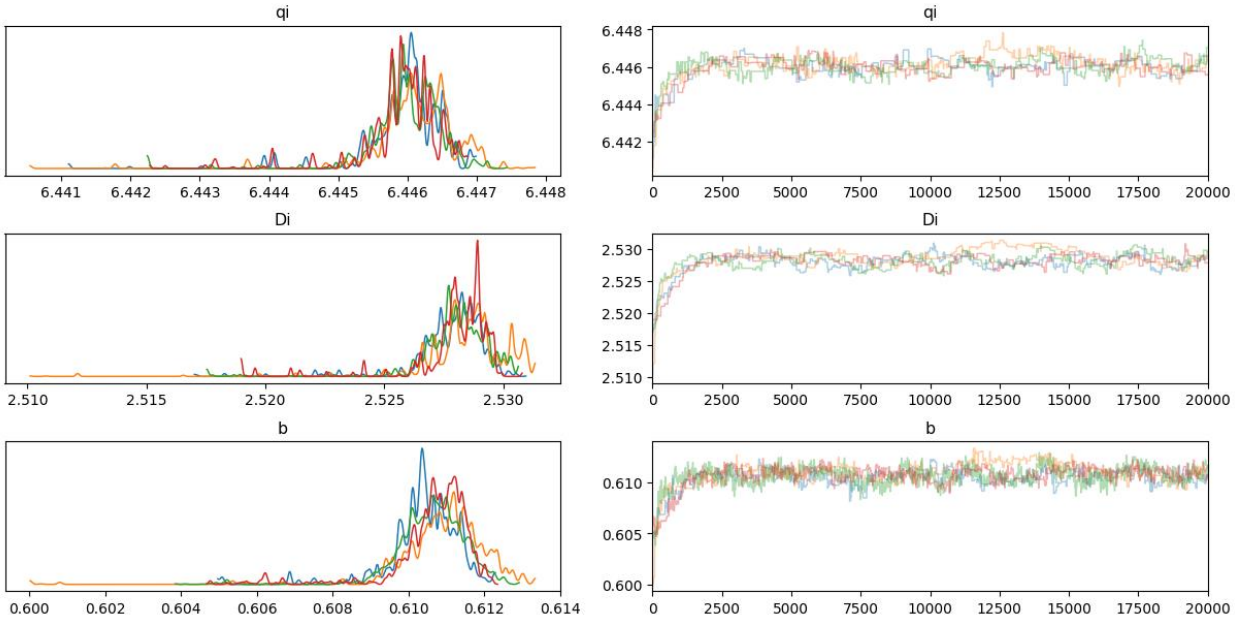

Figure 3.10 Posterior parameter distribution for Well1 with 40% of initial production data used for training.

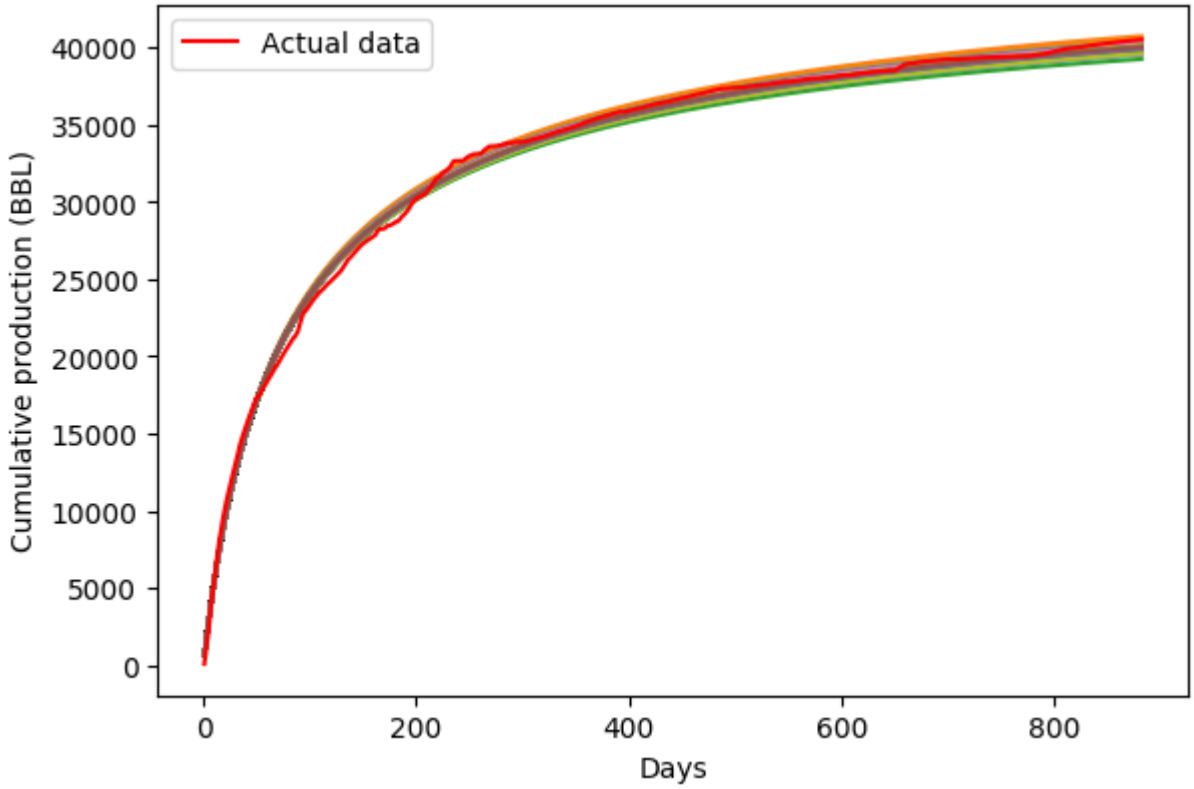

Figure 3.11 Cumulative production calculated using posterior parameter distribution for the model trained with 50% of the initial production data.

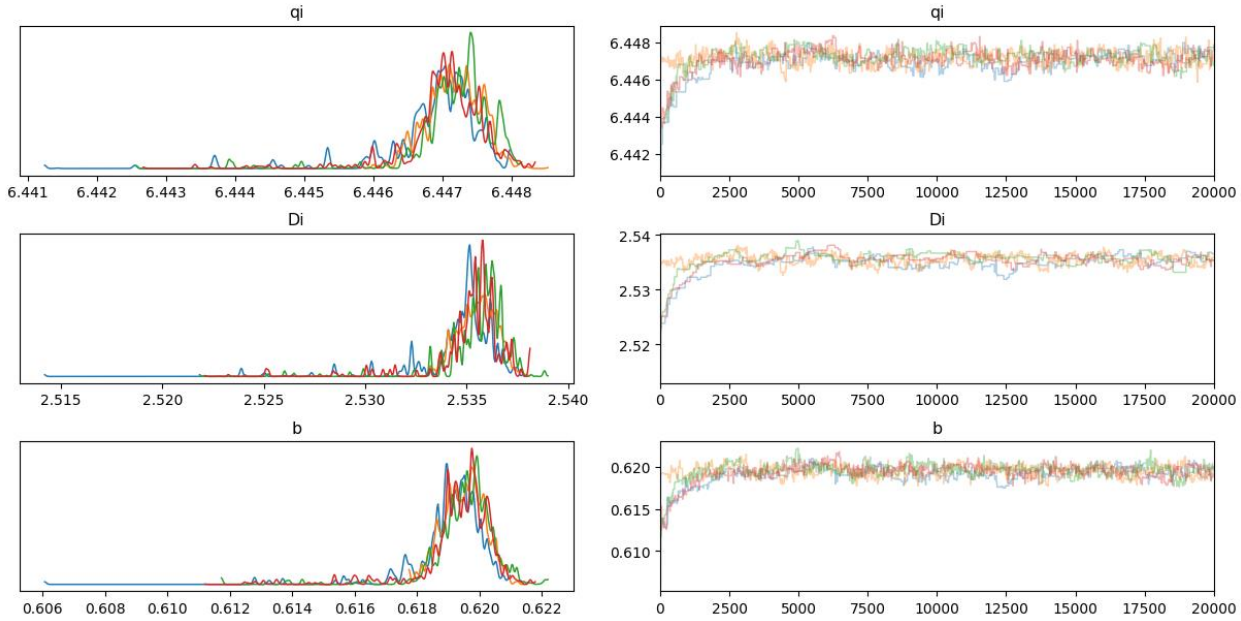

Figure 3.12 Posterior parameter distribution for Well1 with 50% of initial production data used for training.

Figure 3.13 shows the predicted cumulative production and the match to this data for MAP estimates obtained by training the model with 20% (176 days), 30% (264 days), 40% (352 days), 50% (441 days) of initial production data. To assess the impact of changing amounts of initial production data on MCMC parameter estimation, I vary the amount of production data from 176 days to 264 days to 352 days to 441 days and compare the performance of the models given by the MAP parameter estimates. With just 20% of the data, we clearly see that we are underestimating the true production. As more data becomes available, the fidelity of the MAP estimates improves.

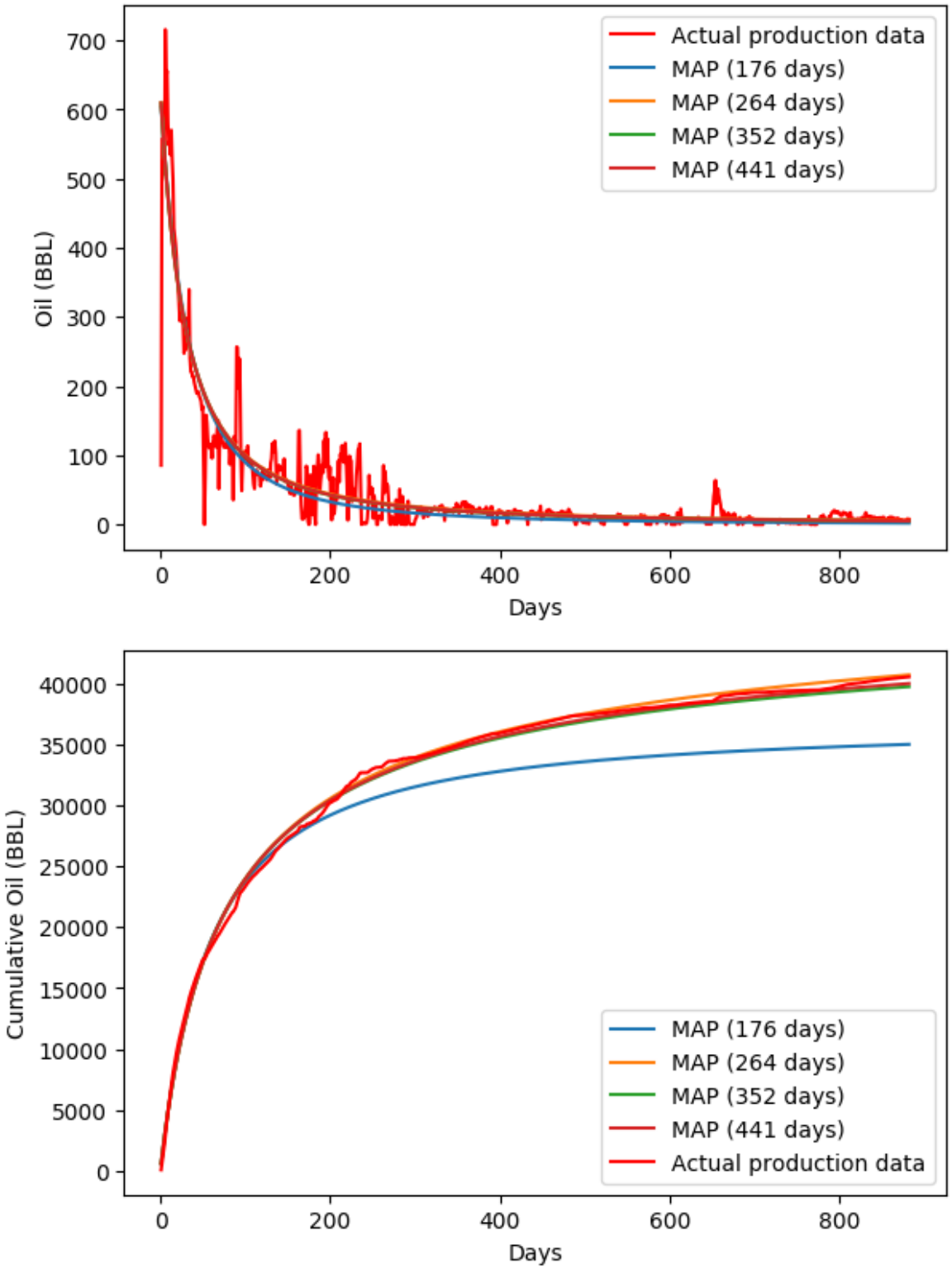

Figure 3.13 Calculated cumulative production and the actual production generated using various training intervals for Well 1 using Arps model.

3.4 Field-wide results

268 wells from Meramec field in STACK are analyzed using the MCMC-based parameter estimation with the PyMC3 (2016) package. The models considered in this work are Arps (1945), power law exponential (Daal et al. 2019) and Duong (2014) decline curve models. For each well, the posterior distributions and the MAP estimates are calculated. MAP estimates are generated for various training intervals, posterior distributions are generated by using 50% of the initial production data for training.

I do not apply any smoothing to the production data neither are wells with a significant amount of downtime excluded from the dataset. However, if there is significant downtime for a well due to events like offset activities, that well will be considered a candidate for multi-segment forecasting. For multi-segment forecasting, I fit a decline curve prior to the shutdown and successive, separate decline curves for each period of long downtime periods to account for changing completions, artificial lift method or nearby well activity.

3.4.1 Metrics

Four metrics are captured for simulated production data from the set of parameter estimates. They are R-Squared, Root mean square error (RMSE), Mean absolute error (MAE) and the percentage difference between the true data and the model-predicted production.

The coefficient of determination or R-squared is a measure that determines the proportion of variance and therefore is a goodness of fit indicator showing how well the model fits the data. It is quantified as the regression sum-of-squares divided by the total sum-of-squares or one minus the mean square error between the actual and predicted values divided by the variance in the dependent variables.

RMSE represents difference between the predicted and the observed values.

MAE measures the average magnitude of the errors in a set of predictions, without considering their direction. It is the average of the absolute difference between the predicted values and observed value. All individual differences are weighted equally in the average.

RMSE gives a relatively larger weight to big errors since the errors are squared before they are averaged. RMSE is a better indicator when large errors are undesirable. But given the volatile nature of the transient period and the large errors that occur during that period, MAE is captured in addition to RMSE for a better understanding of the quality of the models. The last metric, difference percentage is a percentage of the difference between calculated and actual cumulative production divided by the actual cumulative production.

3.4.2 Oil well production matching using the Arps decline curve

Using the MCMC approach for parameter estimation, I show the results of the match to daily rate and cumulative production in Figure 3.14 for a single oil well. Analysis on how the actual cumulative production matches with the MAP, P10 and P90 estimates for all the wells is shown after the single well analysis.

50% of the production data (582 days) is used for training the model. From the posterior parameter distribution, 401 distinct parameter estimates are generated. For the production calculated using these 401 estimates, R-squared, RMSE and MAE are captured. Table 3.1 shows the average values for the 401 estimates for Well2.

Avg. R2	Avg. RMSE	Avg. MAE
0.91	39.53	18.44

Table 3.1 Metrics summary for Well 2 posterior analysis.

Table 3.2 shows the summary of the 401 production estimates captured. All values shown are shown in MMBLS (millions of barrels).

Max. Cumulative	Min. Cumulative	P90	P50	P10	Actual
211.09	185.35	193.21	196.08	201.26	201.12

Table 3.2 Production estimates summary in MMBLS for Well 2 posterior analysis.

Table 3.3 shows the data associated with the MAP estimate for Well 3. MAP estimate underestimates the cumulative oil production by 2.49%. Figure 3.15 shows the actual daily and cumulative production in comparison to the calculated daily and cumulative production based on the MAP estimate.

q_i	D_i	b	Estimate	R2	RMSE	MAE	Cumulative (MMBLS)	Difference %
737.29	3.87	1.14	MAP (582 days)	0.91	39.41	18.25	196.1	2.49

Table 3.3 MAP estimate summary for Well 2.

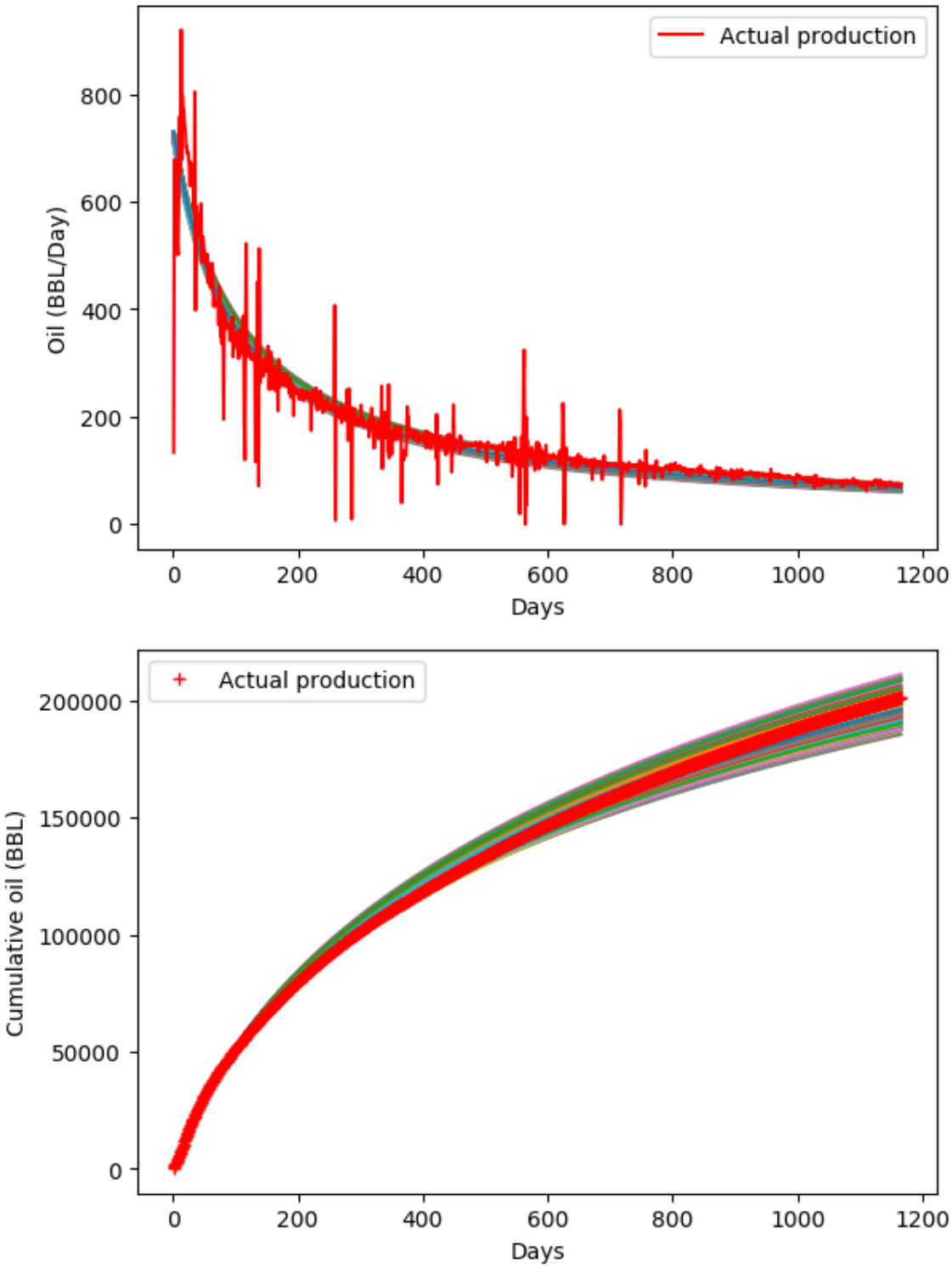

Figure 3.14 Estimates for forecasted daily production date and cumulative oil production for Well 2 using the Arps model, with the parameter estimate samples from their posterior distributions.

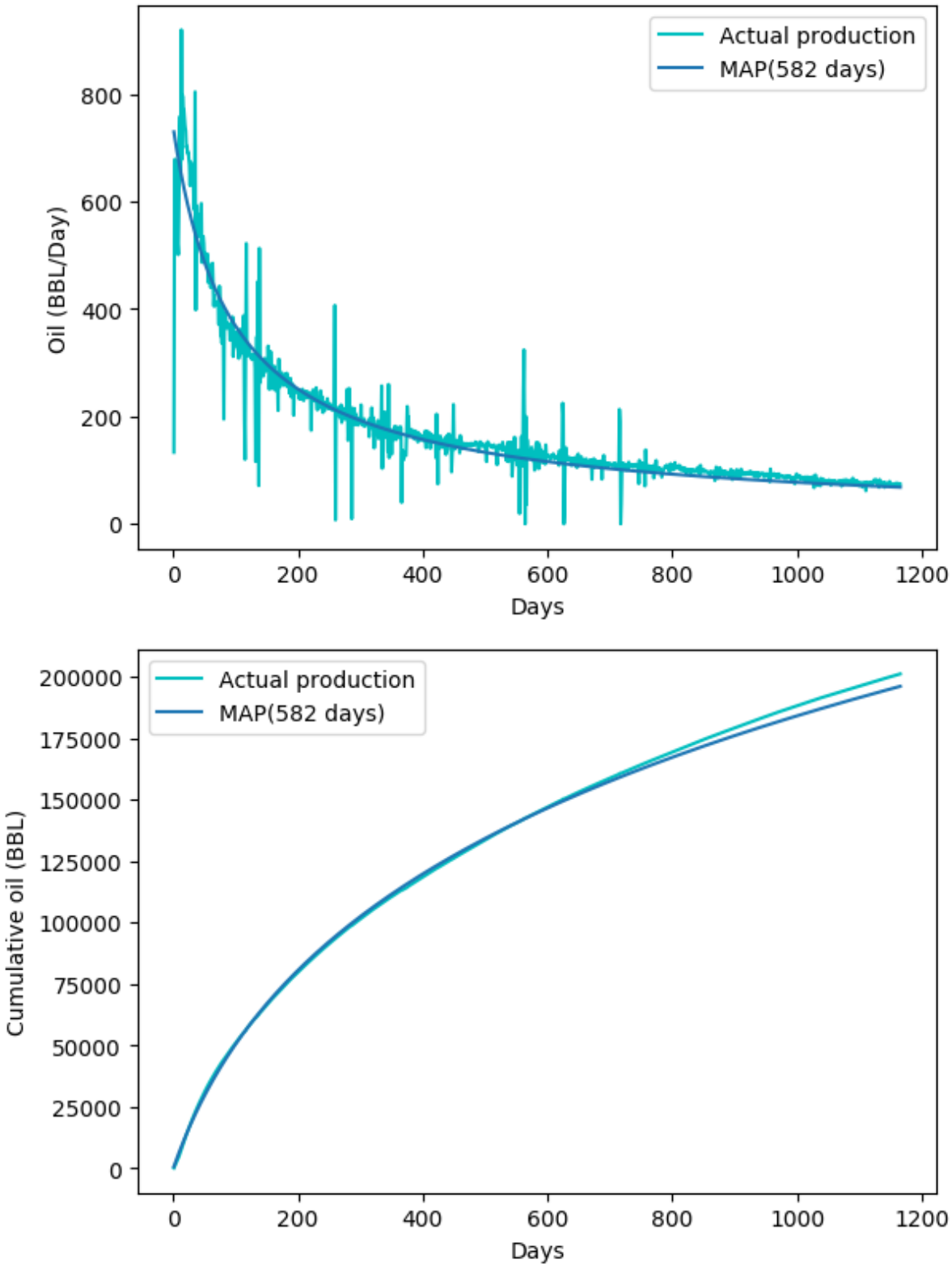

Figure 3.15 Estimates for forecasted daily production date and cumulative oil production for Well 2 using the Arps model, with the MAP estimate.

I will now present how the MCMC approach for parameter estimation was able to generate the MAP estimates for 265 oil wells. Figure 3.16 shows the correlation between the cumulative actual production and the cumulative production generated using MAP estimates. The linear regression line shows a strong relationship with an R-squared value of 0.979 which indicates good validity of the MAP estimates.

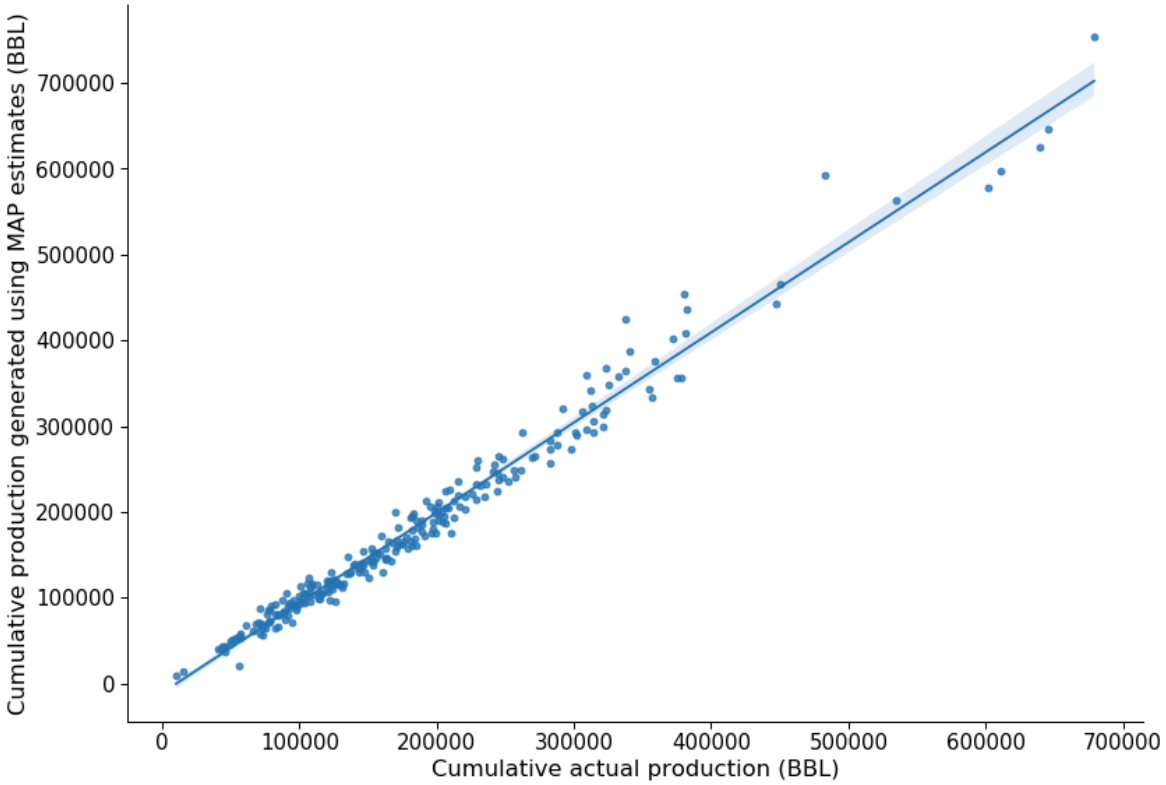

Figure 3.16 Regression line between the cumulative actual production and the cumulative production generated using MAP estimates for oil wells using Arps model.

Figure 3.17 shows the frequency plot of wells based on the condition if the cumulative actual production lies in the range of [P90 – P10] values of the posterior distribution. The x-axis has labels with intervals of 12 months starting Oct 2014. Figure 3.18 shows a similar frequency plot based on whether the actual cumulative production falls within [-10%, 10%] range of the MAP estimate. The percentage of wells failing outside the range of [P10-P90] and the MAP +/- 10% estimates increases as amount of production data available for training the model decreases.

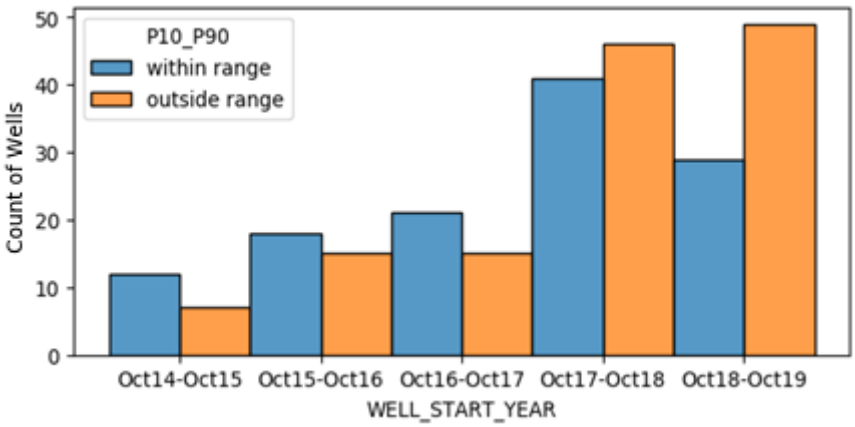

Figure 3.17 Histogram plot showing the distribution of oil wells falling within and outside the range of [P90-P10] estimates, while using Arps model.

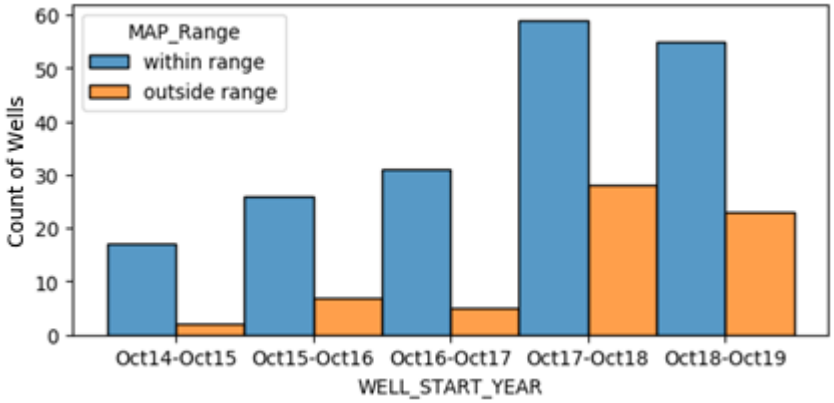

Figure 3.18 Histogram plot showing the distribution of oil wells falling within and outside the range of [-10%, 10%] of the MAP estimate, while using Arps model.

3.4.3 Gas well production matching using the Arps decline curve

Using the MCMC approach for parameter estimation, I show the results of the match to daily rate and cumulative production in Figure 3.19 for a single gas well. Analysis on how the actual cumulative production matches with the MAP, P10 and P90 estimates for all the wells is shown after the single well analysis.

50% of the production data (207 days) is used for training the model. From the posterior parameter distribution, 281 distinct parameter estimates are generated. For the production calculated using these 281 estimates, R-squared, RMSE and MAE are captured. Table 3.4 shows the average values for these metrics.

Avg. R2	Avg. RMSE	Avg. MAE
0.86	184.23	104.6

Table 3.4 Metrics summary for Well 3 posterior analysis.

Table 3.5 shows the summary of the 281 production estimates captured. All values shown are shown in MMCF (Million cubic feet).

Max. Cumulative	Min. Cumulative	P90	P50	P10	Actual
512.15	452.66	472.66	478.15	487.02	475.84

Table 3.5 Production estimates summary in MMCF for Well 3 posterior analysis.

Table 3.6 shows the data associated with the MAP estimate for Well 3. MAP estimate overestimates the cumulative oil production by 0.57%. Figure 3.20 shows the actual daily and cumulative production in comparison to the calculated daily and cumulative production based on the MAP estimate.

q_i	D_i	b	Estimate	R2	RMSE	MAE	Cumulative (MMCF)	Difference %
2864.6	5.73	1.82	MAP (207 days)	0.87	183.3	102.5	478.57	-0.57

Table 3.6 MAP estimate summary for Well 3.

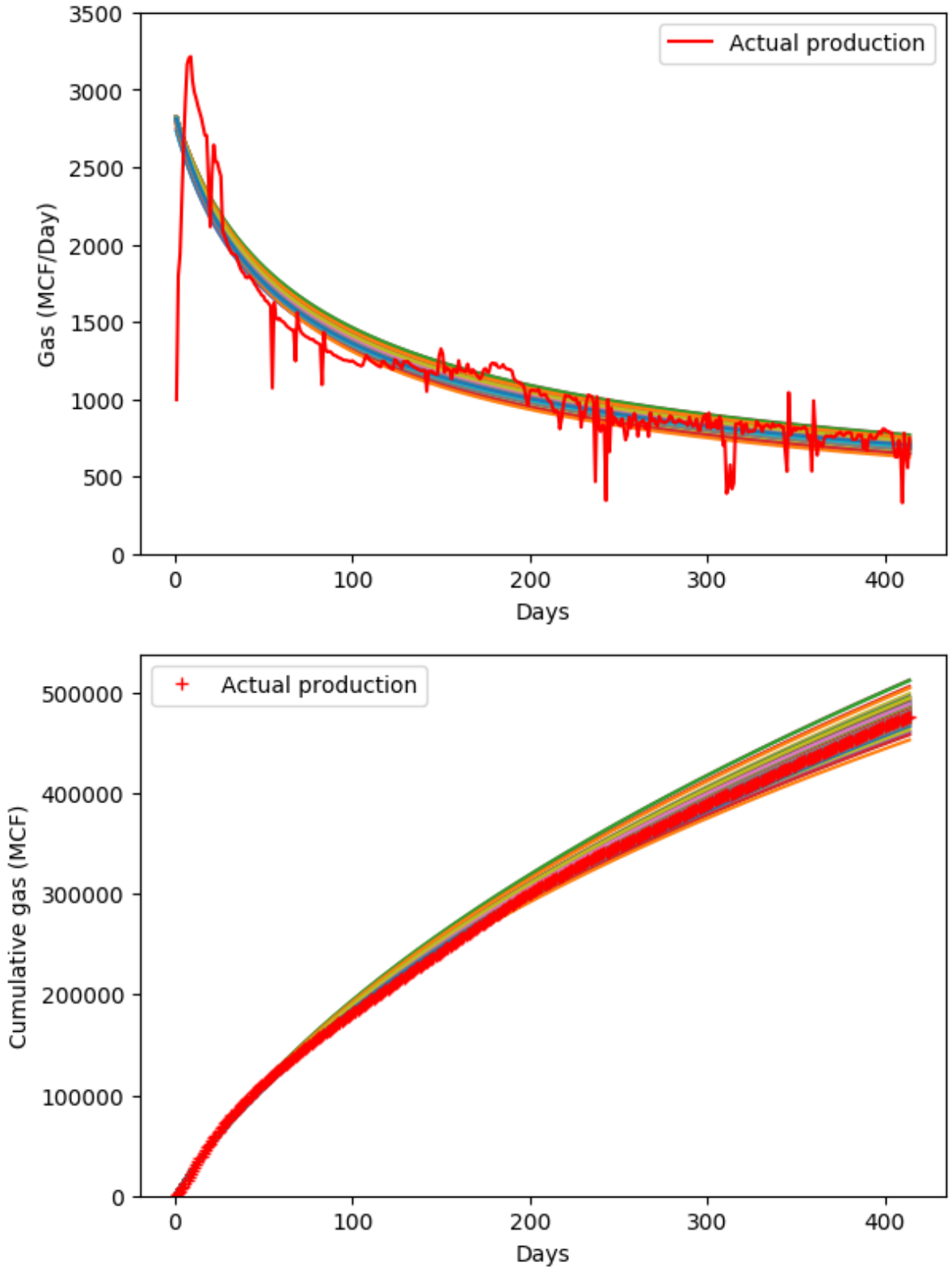

Figure 3.19 Estimates for forecasted daily production date and cumulative gas production for Well 3 using the Arps model, with the parameter estimate samples from their posterior distributions.

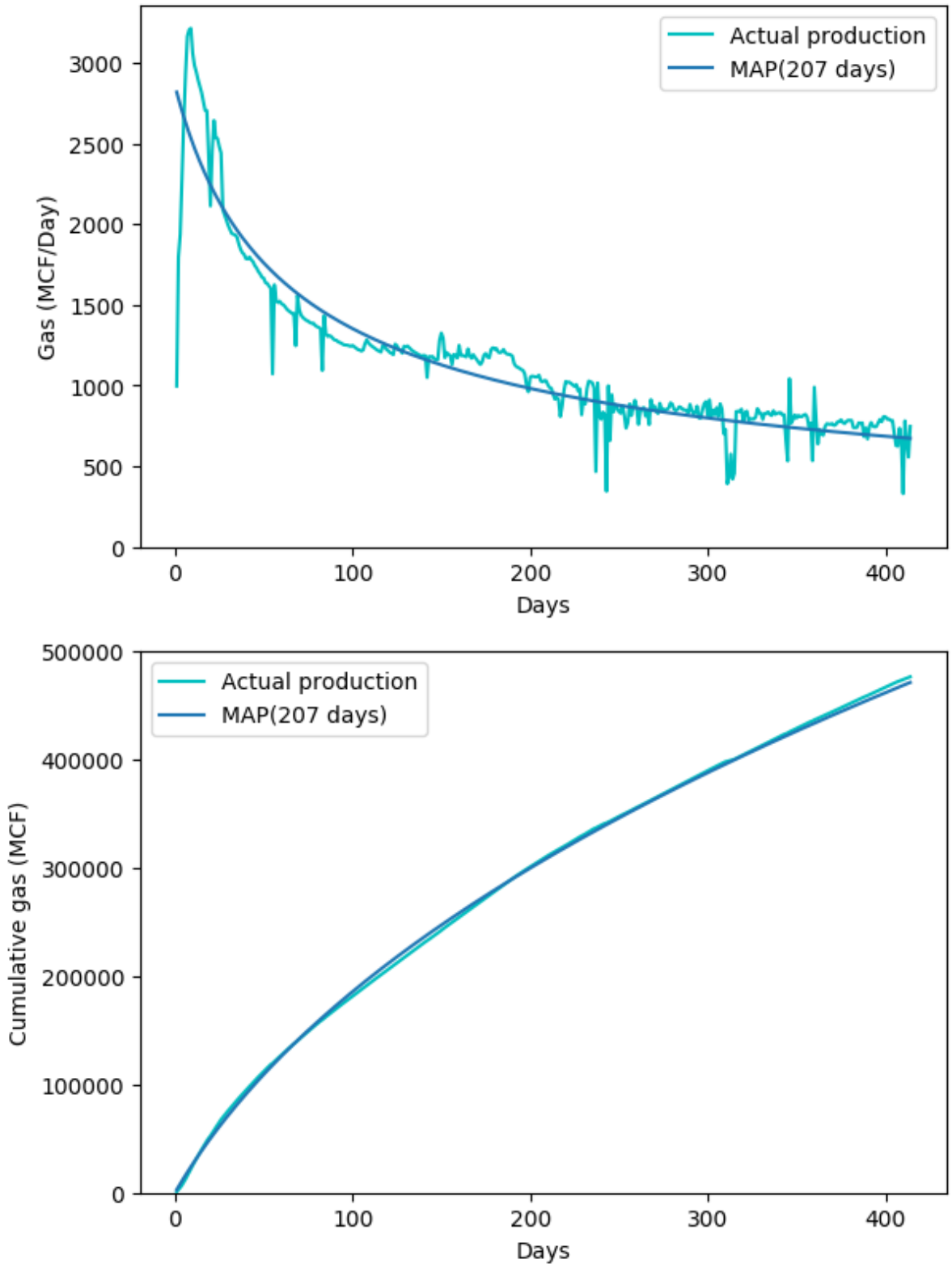

Figure 3.20 Estimates for forecasted daily production date and cumulative gas production for Well 3 using the Arps model, with the MAP estimate.

I will now present how the MCMC approach for parameter estimation was able to generate the MAP estimates for 247 gas wells. Figure 3.21 shows the correlation between the cumulative actual production and the cumulative production generated using MAP estimates. The linear regression line shows a strong relationship with an R-squared value of 0.967 which indicates good validity of the MAP estimates.

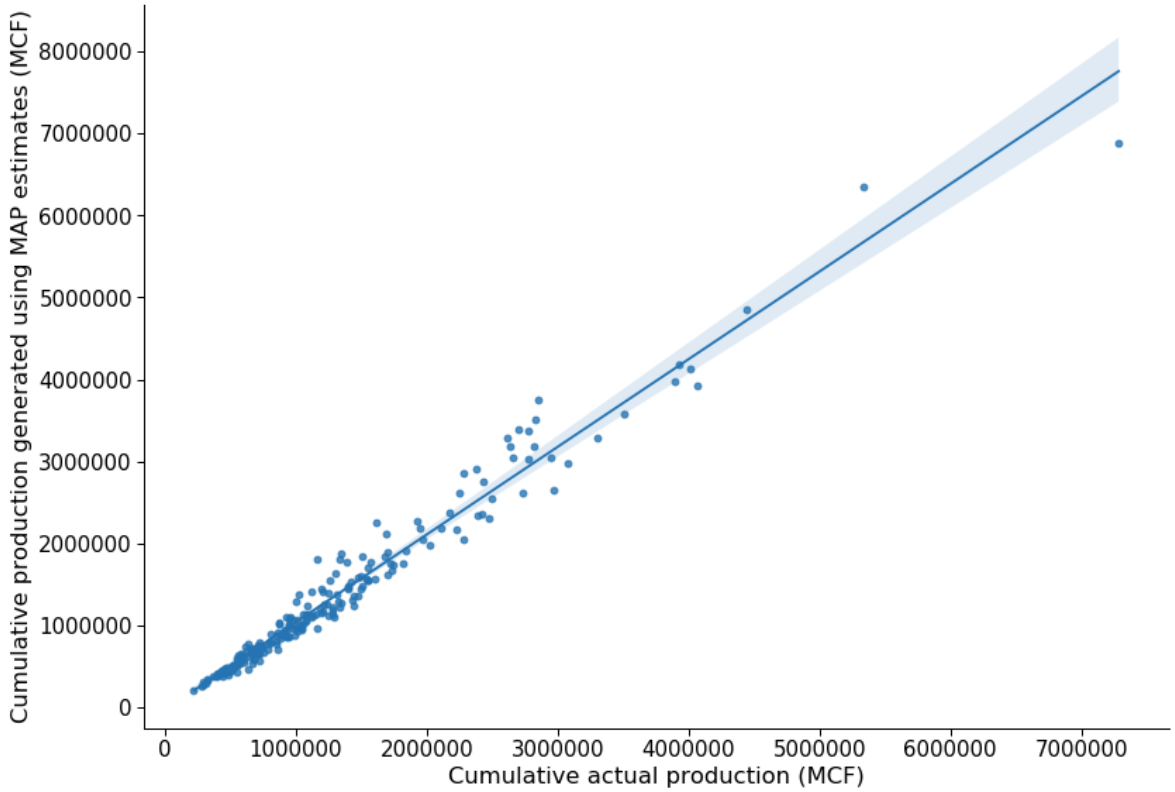

Figure 3.21 Regression line between the cumulative actual production and the cumulative production generated using MAP estimates for gas wells using Arps model.

Figure 3.22 shows the frequency plot of wells based on the condition if the cumulative actual production lies in the range of [P90 – P10] values of the posterior distribution. The x-axis has labels with intervals of 12 months starting Oct 2014. Figure 3.23 shows a similar frequency plot based on whether the actual cumulative production falls within [-10%, 10%] range of the MAP estimate. The frequency plot in Figure 3.22 does not show a similar pattern as seen with the oil wells using Arps for the [P10-P90] estimates. The frequency plot in Figure 3.23 indicates a similar trend as seen with MAP estimates for oil wells using Arps.

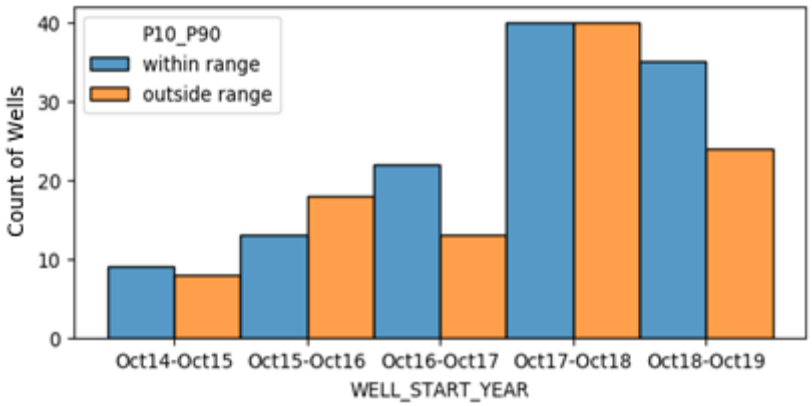

Figure 3.22 Histogram plot showing the distribution of gas wells falling within and outside the range of [P90-P10] estimates, while using Arps model.

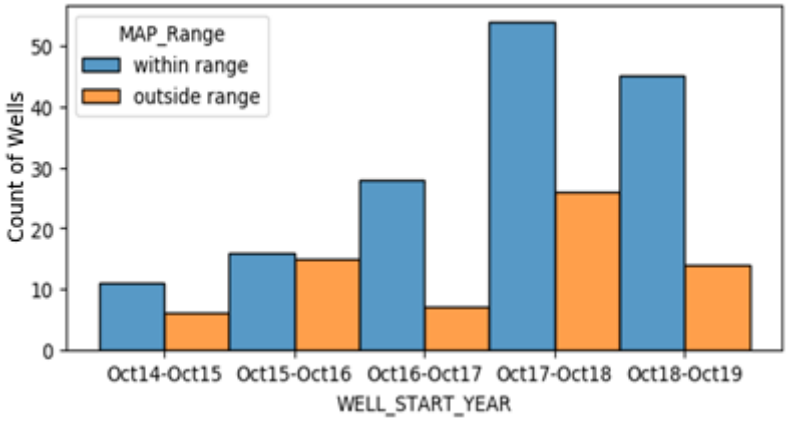

Figure 3.23 Histogram plot showing the distribution of gas wells falling within and outside the range of [-10%, 10%] of the MAP estimate, while using Arps model.

3.4.4. Oil well production matching using the power law exponential decline curve

Using the MCMC approach for parameter estimation, I show the results of the match to daily rate and cumulative production in Figure 3.24 for a single oil well using the power law exponential decline curve. Analysis on how the actual cumulative production matches with the MAP, P10 and P90 estimates for all the wells is shown after the single well analysis.

50% of the production data (441 days) is used for training the model. From the posterior parameter distribution, 2000 distinct parameter estimates are generated. For the production calculated using these 2000 estimates, R-squared, RMSE and MAE are captured. Table 3.7 shows the average values for these metrics captured for Well 4.

Avg. R2	Avg. RMSE	Avg. MAE
0.85	34.44	17.56

Table 3.7 Metrics summary for Well 4 posterior analysis.

Table 3.8 shows the summary of the 2000 production estimates captured. All values shown are shown in MMBLS (millions of barrels).

Max. Cumulative	Min. Cumulative	P90	P50	P10	Actual
42.73	28.65	35.18	35.77	36.29	40.53

Table 3.8 Production estimates summary in MMBLS for Well 4 posterior analysis.

Table 3.9 shows the data associated with the MAP estimate for Well 4. MAP estimate underestimates the cumulative oil production by 11.75%. Figure 3.25 shows the actual daily and cumulative production in comparison to the calculated daily and cumulative production based on the MAP estimate.

q_i	α	β	Estimate	R2	RMSE	MAE	Cumulative (MMBLS)	Difference %
683.81	0.07	0.37	MAP (441 days)	0.85	34.39	17.55	35.76	11.75

Table 3.9 MAP estimate summary for Well 4.

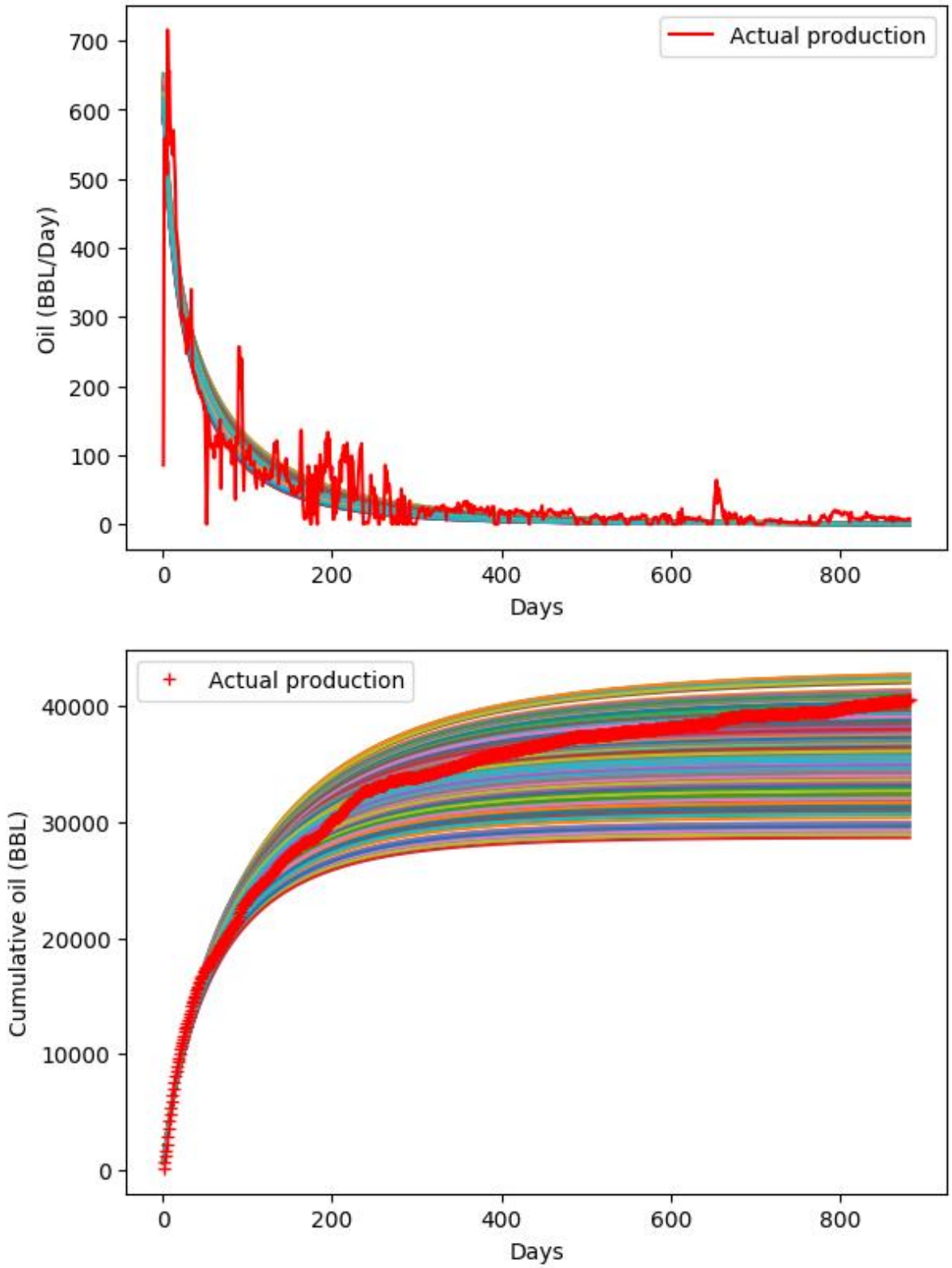

Figure 3.24 Estimates for forecasted daily production date and cumulative oil production for Well 4 using the power law exponential model, with the parameter estimate samples from their posterior distributions.

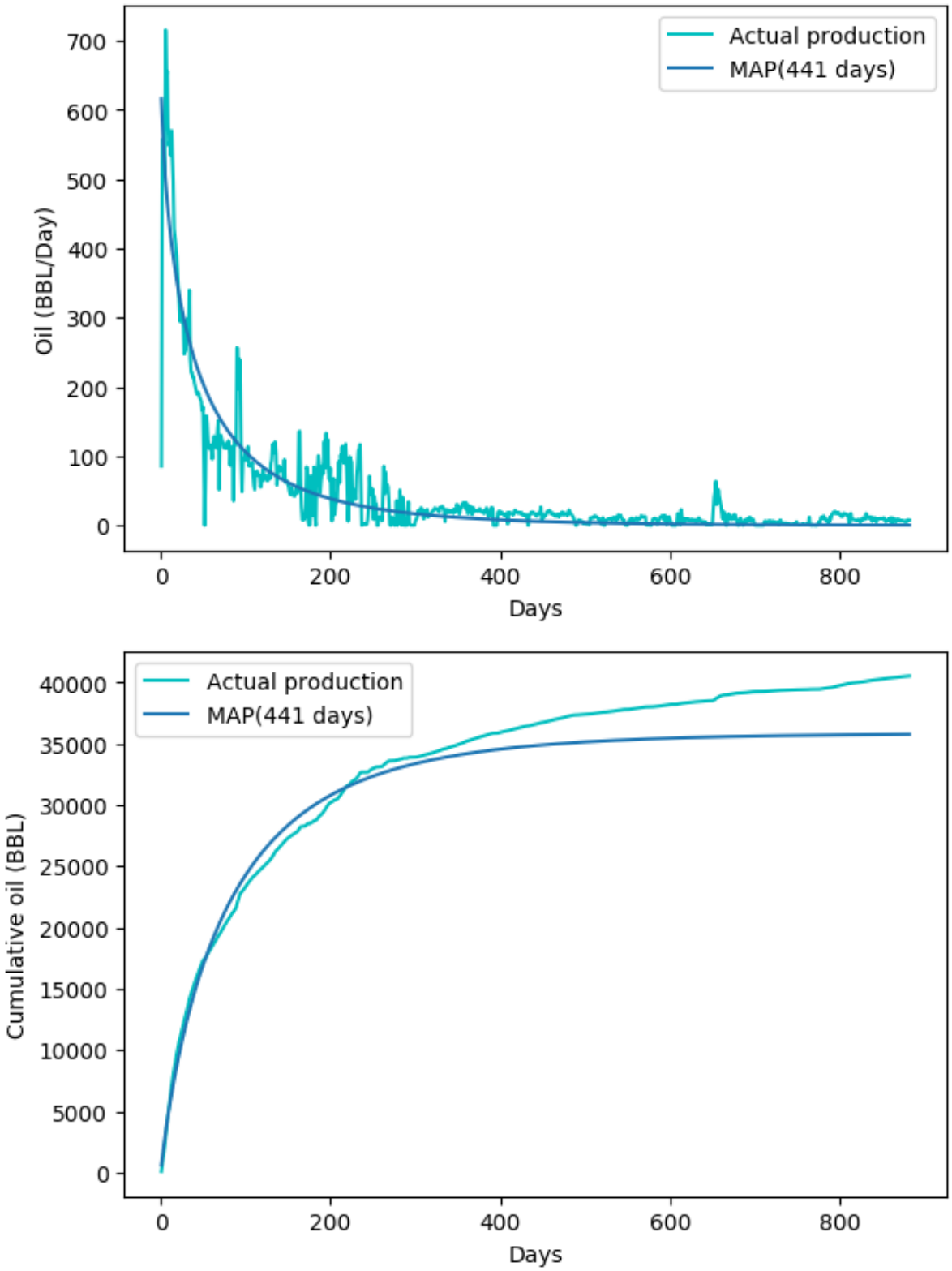

Figure 3.25 Estimates for forecasted daily production date and cumulative oil production for Well 4 using the power law exponential model, with the MAP estimate.

The underestimation of production during BDF has resulted in the significant drop of cumulative production from around day 250 as shown in Figure 3.25.

I will now present how the MCMC approach for parameter estimation was able to generate the MAP estimates for 181 oil wells using the power law exponential model. Figure 3.26 shows the correlation between the cumulative actual production and the cumulative production generated using MAP estimates. The linear regression line shows a strong relationship with an R-squared value of 0.972 which indicates good validity of the MAP estimates.

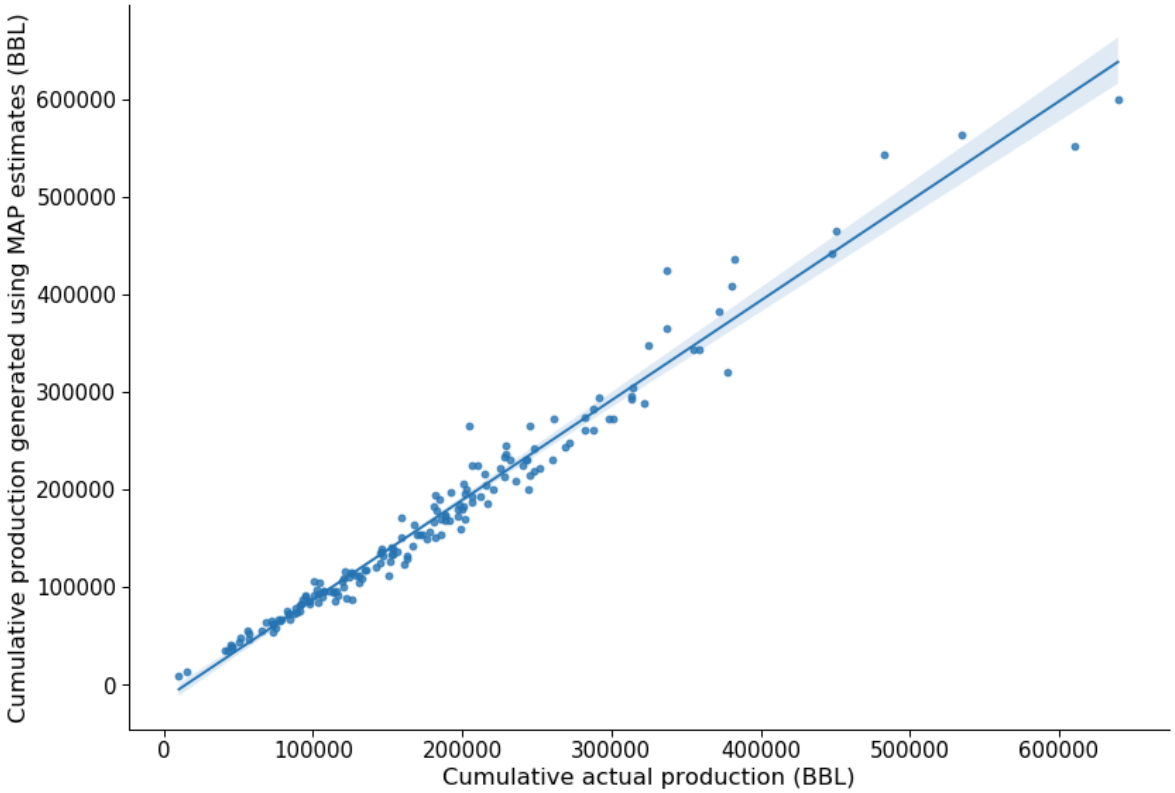

Figure 3.26 Regression line between the cumulative actual production and the cumulative production generated using MAP estimates for oil wells using power law exponential model.

Figure 3.27 shows the frequency plot of wells based on the condition if the cumulative actual production lies in the range of [P90 – P10] values of the posterior distribution. The x-axis has labels with intervals of 12 months starting Oct 2014. Figure 3.28 shows a similar frequency plot based on whether the actual cumulative production falls within [-10%, 10%] range of the MAP estimate. The percentage of wells failing outside the range of [P10-P90] and the MAP +/- 10% estimates increases as amount of production data available for training the model decreases.

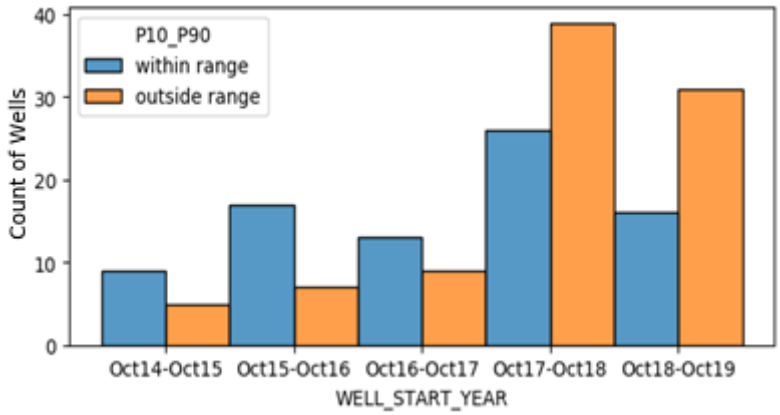

Figure 3.27 Histogram plot showing the distribution of oil wells falling within and outside the range of [P90-P10] estimates, while using power law exponential model.

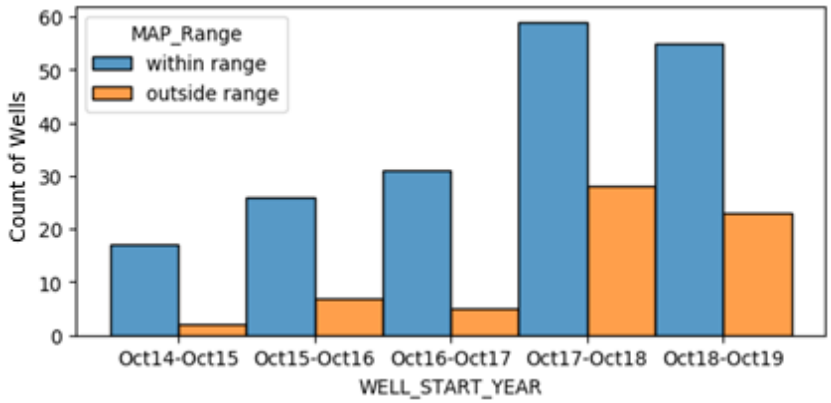

Figure 3.28 Histogram plot showing the distribution of oil wells falling within and outside the range of [-10%, 10%] of the MAP estimate, while using power law exponential model.

3.4.5. Gas well production matching using the power law exponential decline curve

Using the MCMC approach for parameter estimation, I show the results of the match to daily rate and cumulative production in Figure 3.29 for a single gas well using the power law exponential decline curve. Analysis on how the actual cumulative production matches with the MAP, P10 and P90 estimates for all the wells is shown after the single well analysis.

50% of the production data (819 days) is used for training the model. From the posterior parameter distribution, 2000 distinct parameter estimates are generated. For the production calculated using these 2000 estimates, R-squared, RMSE and MAE are captured. Table 3.10 shows the average values for these metrics captured for Well 4.

Avg. R2	Avg. RMSE	Avg. MAE
0.79	458.3	313.6

Table 3.10 Metrics summary for Well 5 posterior analysis.

Table 3.11 shows the summary of the 2000 production estimates captured. All values shown are shown in MMCF (million cubic feet).

Max. Cumulative	Min. Cumulative	P90	P50	P10	Actual
4171.4	2020.41	2731.12	3100.89	3528.54	3307.17

Table 3.11 Production estimates summary in MMCF for Well 5 posterior analysis.

Table 3.12 shows the MAP estimate details for Well 5. MAP estimate underestimates the cumulative gas production by 6.37%.

q_i	α	β	Estimate	R2	RMSE	MAE	Cumulative (MMCF)	Difference %
4689.2	0.003	0.13	MAP (819 days)	0.83	422.01	268.83	3096.6	6.37

Table 3.12 MAP estimate summary for Well 5.

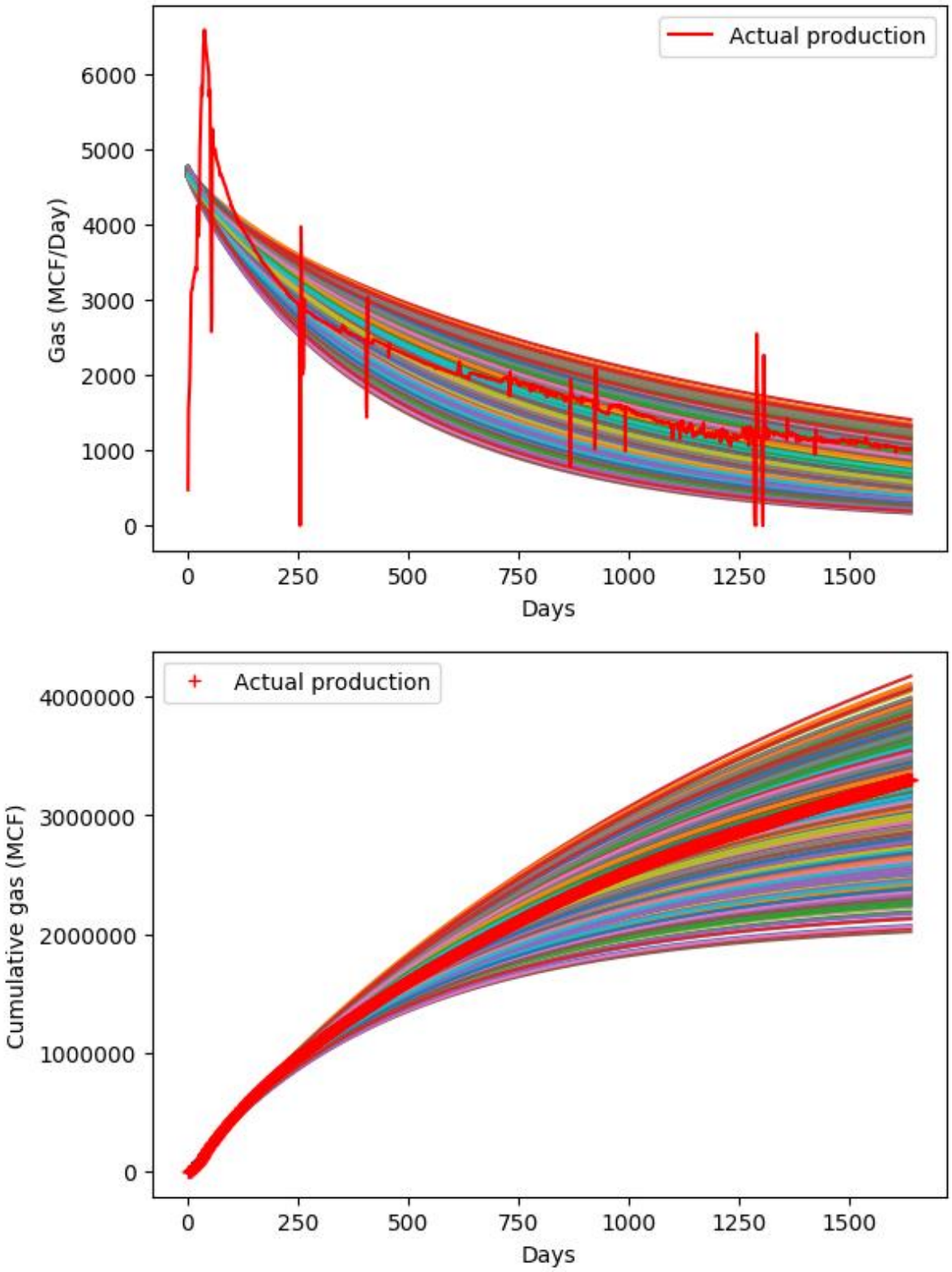

Figure 3.29 Estimates for forecasted daily production date and cumulative gas production for Well 5 using the power law exponential model, with the parameter estimate samples from their posterior distributions.

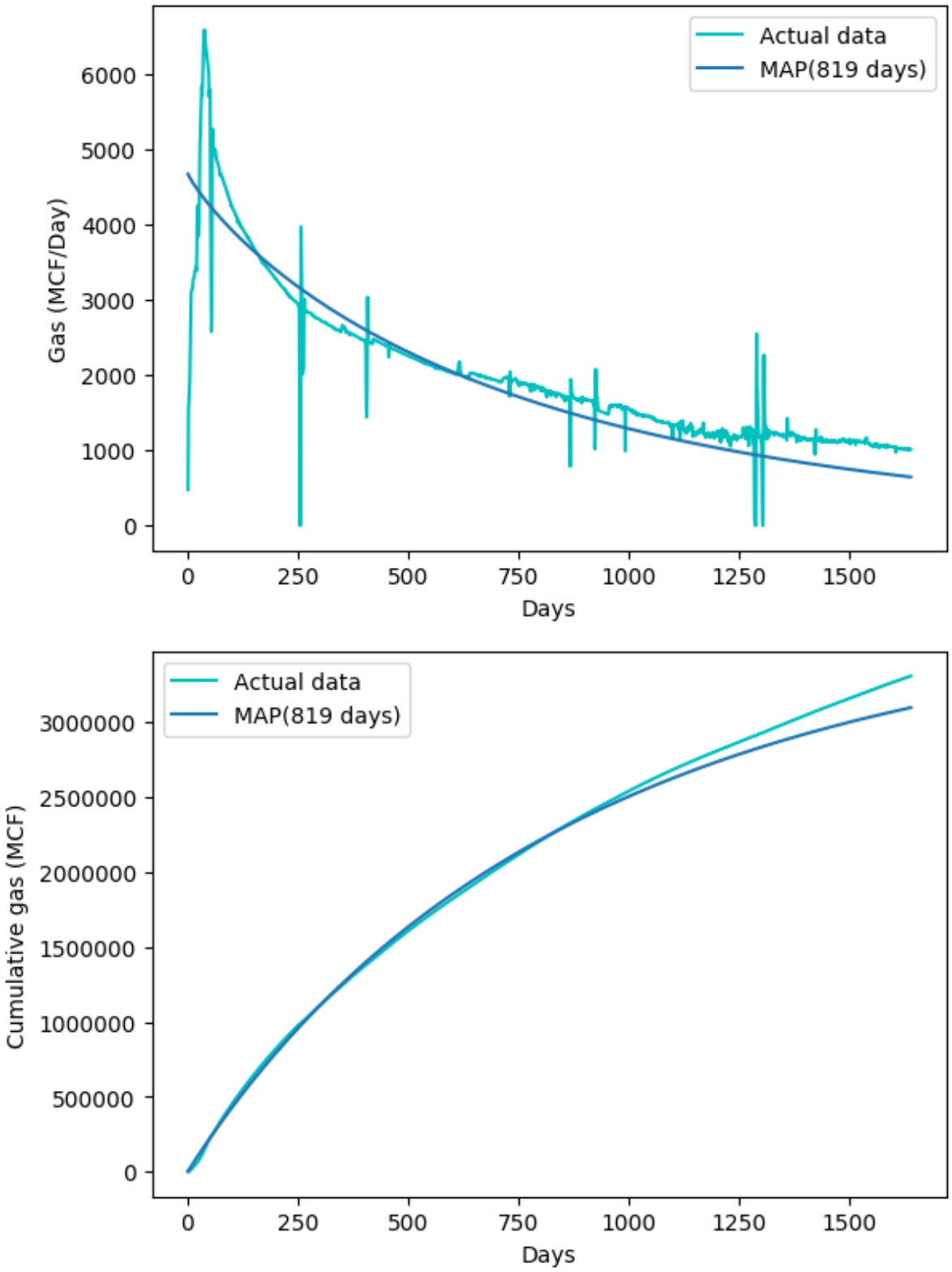

Figure 3.30 Estimates for forecasted daily production date and cumulative gas production for Well 5 using the power law exponential model, with the MAP estimate.

Figure 3.30 shows the actual daily and cumulative production in comparison to the calculated daily and cumulative production based on the MAP estimate.

I will now present how the MCMC approach for parameter estimation was able to generate the MAP estimates for 113 gas wells using the power law exponential model. Figure 3.31 shows the correlation between the cumulative actual production and the cumulative production generated using MAP estimates. The linear regression line shows a strong relationship with an R-squared value of 0.958 which indicates good validity of the MAP estimates.

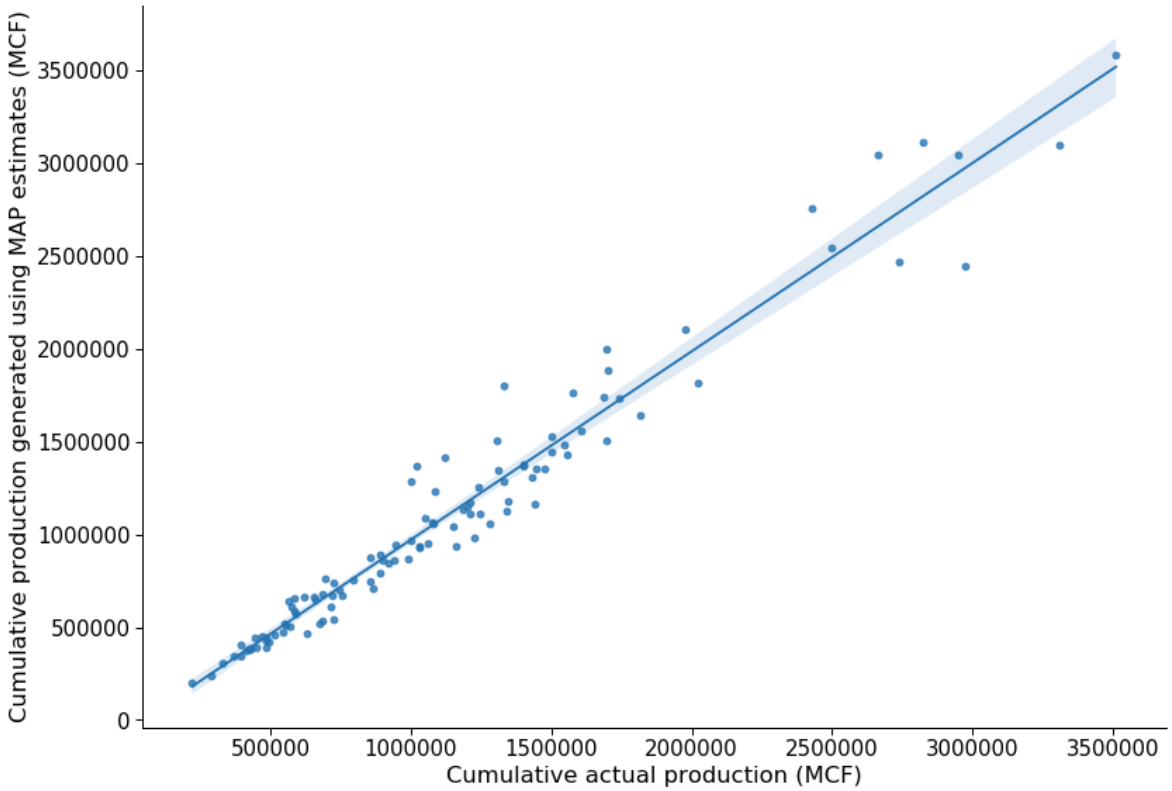

Figure 3.31 Regression line between the cumulative actual production and the cumulative production generated using MAP estimates for gas wells using power law exponential model.

Figure 3.32 shows the frequency plot of wells based on the condition if the cumulative actual production lies in the range of [P90 – P10] values of the posterior distribution. The x-axis has labels with intervals of 12 months starting Oct 2014. Figure 3.33 shows a similar frequency plot based on whether the actual cumulative production falls within [-10%, 10%] range of the MAP estimate. For all durations, we see higher number of wells failing with in the [P10-P90] range as opposed to the trends seen so far.

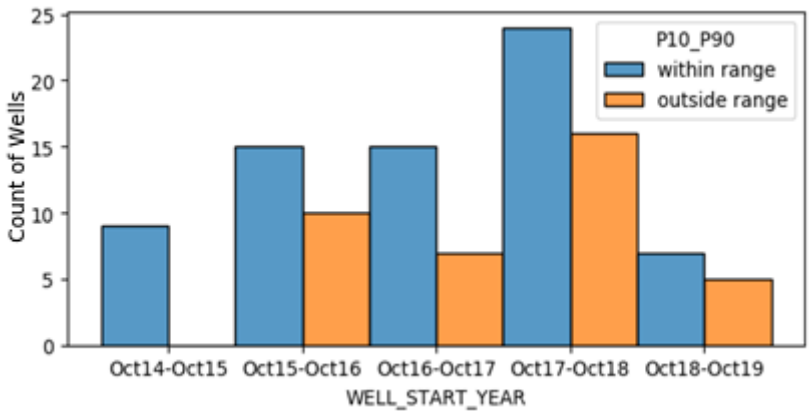

Figure 3.32 Histogram plot showing the distribution of gas wells falling within and outside the range of [P90-P10] estimates, while using power law exponential model.

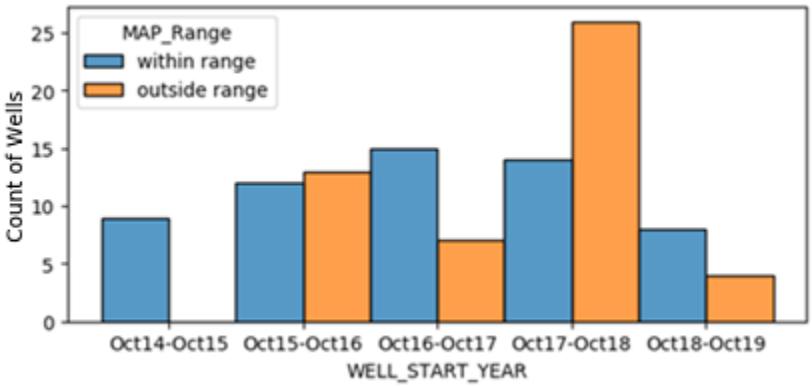

Figure 3.33 Histogram plot showing the distribution of gas wells falling within and outside the range of [-10%, 10%] of the MAP estimate, while using power law exponential model.

3.4.6 Oil well production matching using the Duong decline curve

Using the MCMC approach for parameter estimation, I show the results of the match to daily rate and cumulative production in Figure 3.34 for a single oil well using the Duong decline curve. Analysis on how the actual cumulative production matches with the MAP, P10 and P90 estimates for all the wells is shown after the single well analysis.

50% of the production data (429 days) is used for training the model. From the posterior parameter distribution, 1190 distinct parameter estimates are generated. For the production calculated using these 1190 estimates, R-squared, RMSE and MAE are captured. Table 3.13 shows the average values for these metrics captured for Well 4.

Avg. R2	Avg. RMSE	Avg. MAE
0.8	76.44	40.29

Table 3.13 Metrics summary for Well 6 posterior analysis.

Table 3.14 shows the summary of the 1190 production estimates captured. All values shown are shown in MMBLS (millions of barrels).

Max. Cumulative	Min. Cumulative	P90	P50	P10	Actual
104.6	76.5	82.9	89.8	96.6	83.9

Table 3.14 Production estimates summary in MMBLS for Well 6 posterior analysis.

Table 3.15 shows the data associated with the MAP estimate for Well 6. MAP estimate overestimates the cumulative oil production by 7.19%. Figure 3.35 shows the actual daily and cumulative production in comparison to the calculated daily and cumulative production based on the MAP estimate.

Q_i	m	a	Estimate	R2	RMSE	MAE	Cumulative (MMBLS)	Difference %
149.64	1.42	3.43	MAP (429 days)	0.8	75.87	39.87	89.93	-7.19

Table 3.15 MAP estimate summary for Well 6.

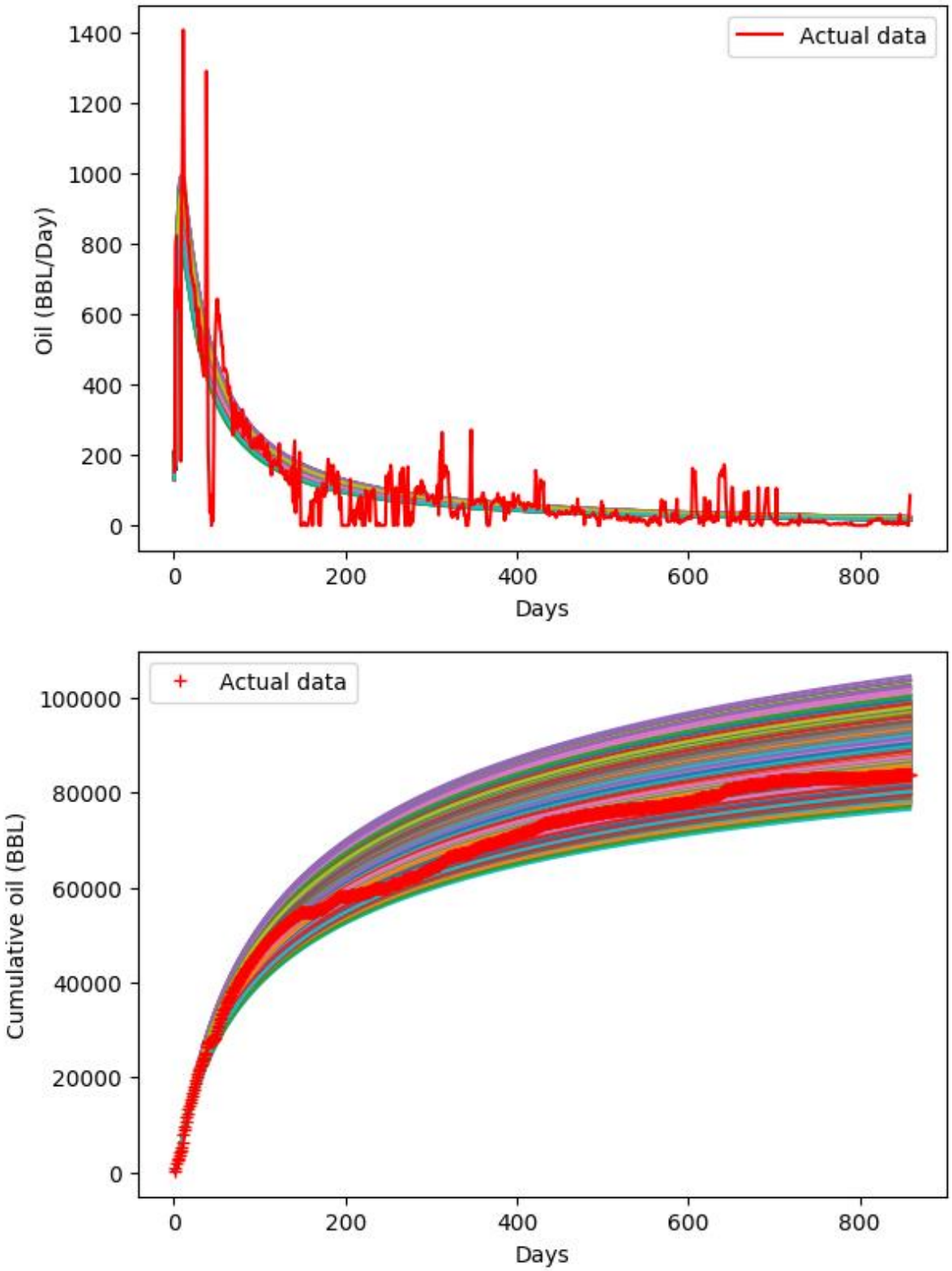

Figure 3.34 Estimates for forecasted daily production date and cumulative oil production for Well 6 using the Duong model, with the parameter estimate samples from their posterior distributions.

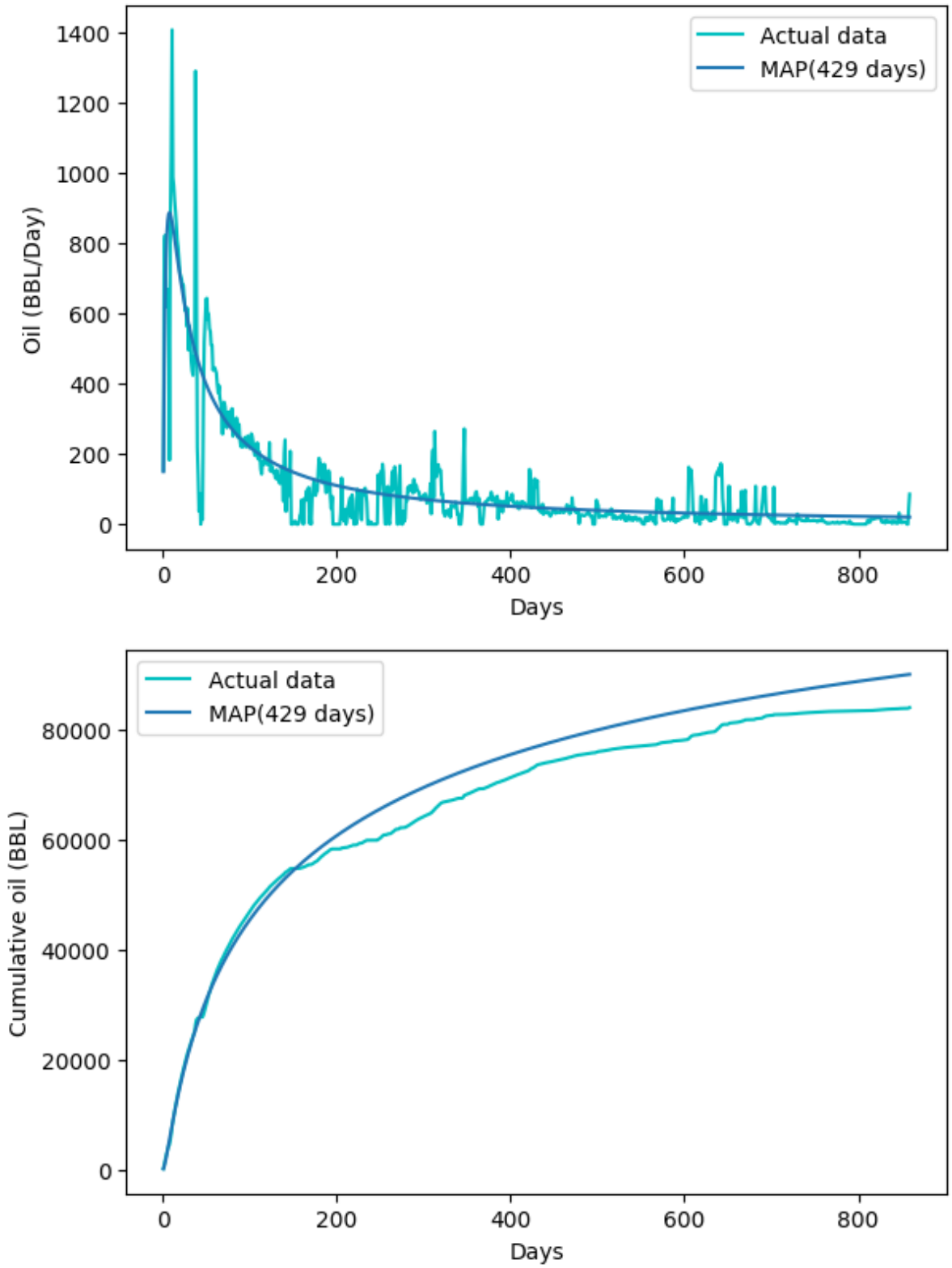

Figure 3.35 Estimates for forecasted daily production date and cumulative oil production for Well 6 using the Duong model, with the MAP estimate.

I will now present how the MCMC approach for parameter estimation was able to generate the MAP estimates for 169 oil wells using the Duong model. Figure 3.36 shows the correlation between the cumulative actual production and the cumulative production generated using MAP estimates. The linear regression line shows a strong relationship with an R-squared value of 0.954 which indicates good validity of the MAP estimates.

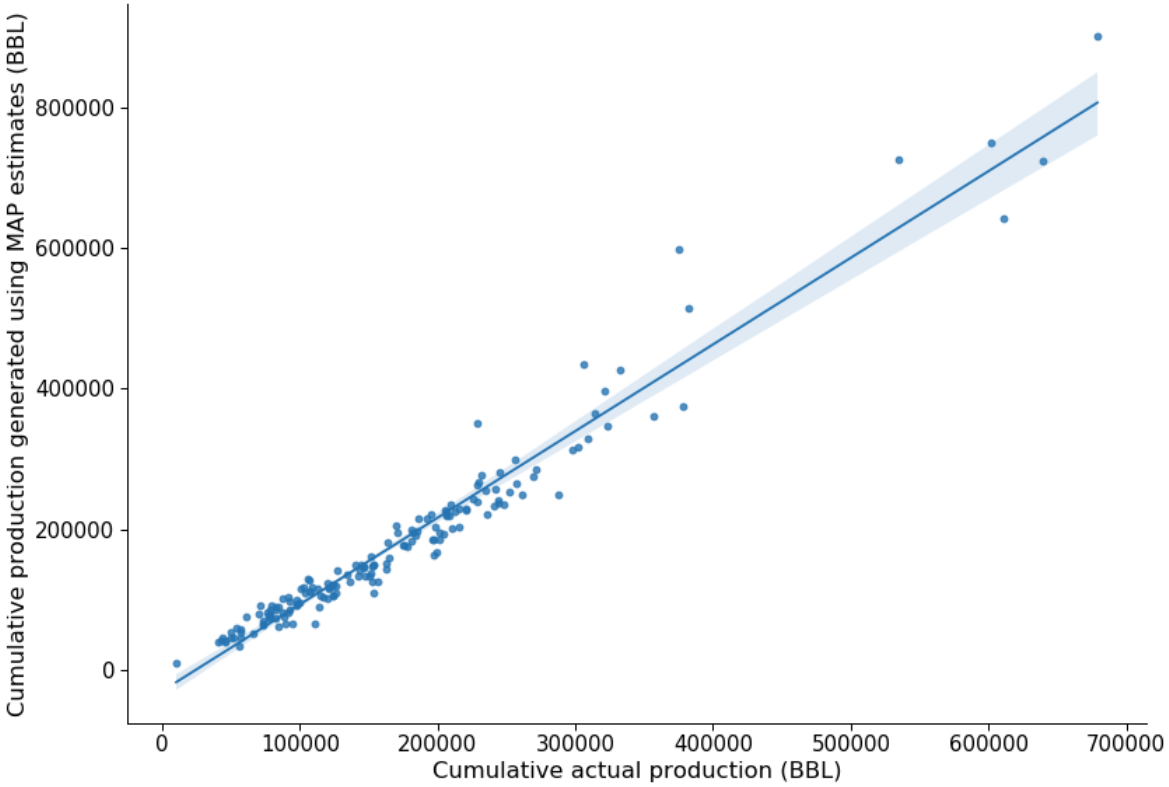

Figure 3.36 Regression line between the cumulative actual production and the cumulative production generated using MAP estimates for oil wells using Duong model.

Figure 3.37 shows the frequency plot of wells based on the condition if the cumulative actual production lies in the range of [P90 – P10] values of the posterior distribution. The x-axis has labels with intervals of 12 months starting Oct 2014. Figure 3.38 shows a similar frequency plot based on whether the actual cumulative production falls within [-10%, 10%] range of the MAP estimate. The percentage of wells failing outside the range of [P10-P90] and the MAP +/- 10% estimates increases as amount of production data available for training the model decreases.

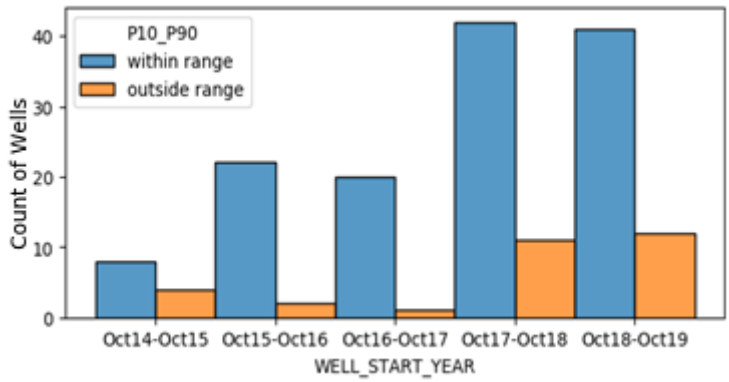

Figure 3.37 Histogram plot showing the distribution of oil wells falling within and outside the range of [P90-P10] estimates, while using Duong model.

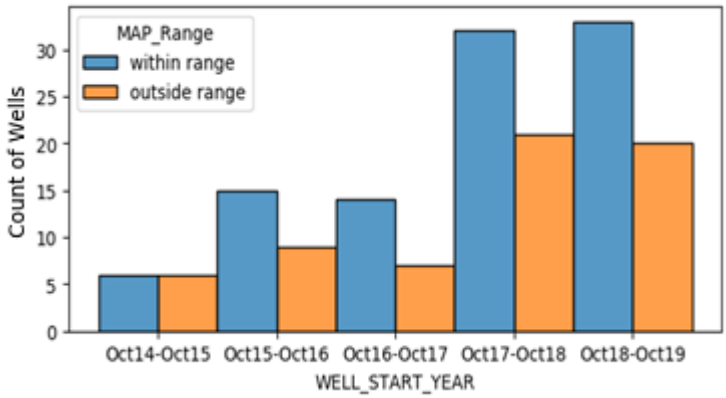

Figure 3.38 Histogram plot showing the distribution of oil wells falling within and outside the range of [-10%, 10%] of the MAP estimate, while using Duong model.

3.4.7 Gas well production matching using the Duong decline curve

Using the MCMC approach for parameter estimation, I show the results of the match to daily rate and cumulative production in Figure 3.39 for a single gas well using the Duong decline curve. Analysis on how the actual cumulative production matches with the MAP, P10 and P90 estimates for all the wells is shown after the single well analysis.

50% of the production data (210 days) is used for training the model. From the posterior parameter distribution, 450 distinct parameter estimates are generated. For the production calculated using these 450 estimates, R-squared, RMSE and MAE are captured. Table 3.16 shows the average values for these metrics captured for Well 7.

Avg. R2	Avg. RMSE	Avg. MAE
0.33	264.48	190.19

Table 3.16 Metrics summary for Well 7 posterior analysis.

Table 3.17 shows the summary of the 450 production estimates captured. All values shown are shown in MMCF (million cubic feet).

Max. Cumulative	Min. Cumulative	P90	P50	P10	Actual
635.57	453.97	495.35	568.19	618.22	554.8

Table 3.17 Production estimates summary in MMCF for Well 7 posterior analysis.

Q_i	m	a	Estimate	R2	RMSE	MAE	Cumulative (MMCF)	Difference %
823.36	1.13	1.68	MAP (210 days)	0.64	194.07	134.83	569.63	-2.67

Table 3.18 MAP estimate summary for Well 7.

Table 3.18 shows the data associated with the MAP estimate for Well 7. MAP estimate overestimates the cumulative gas production by 2.67%. Figure 3.40 shows the actual daily and cumulative production in comparison to the calculated daily and cumulative production based on the MAP estimate.

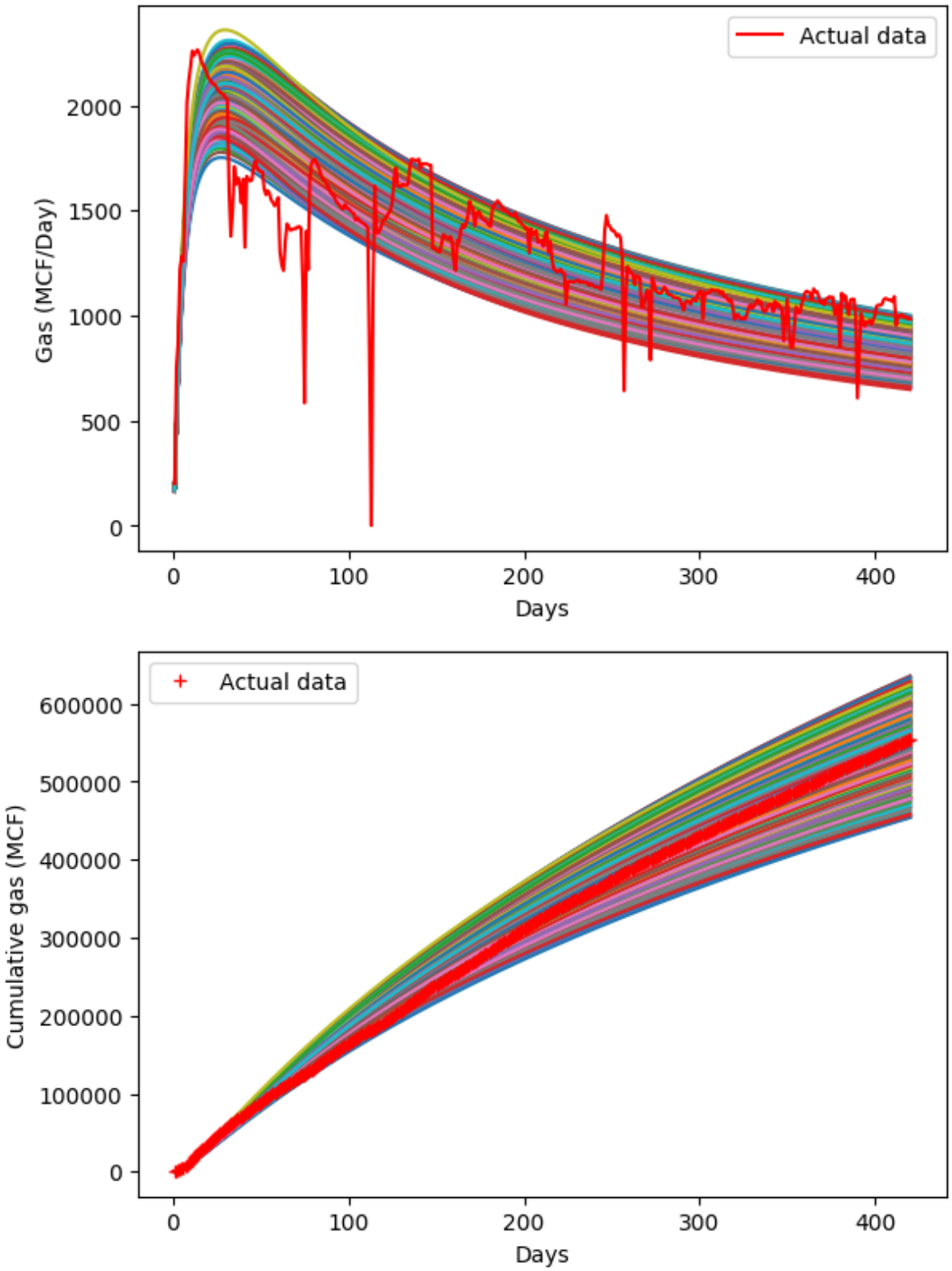

Figure 3.39 Estimates for forecasted daily production date and cumulative gas production for Well 7 using the Duong model, with the parameter estimate samples from their posterior distributions.

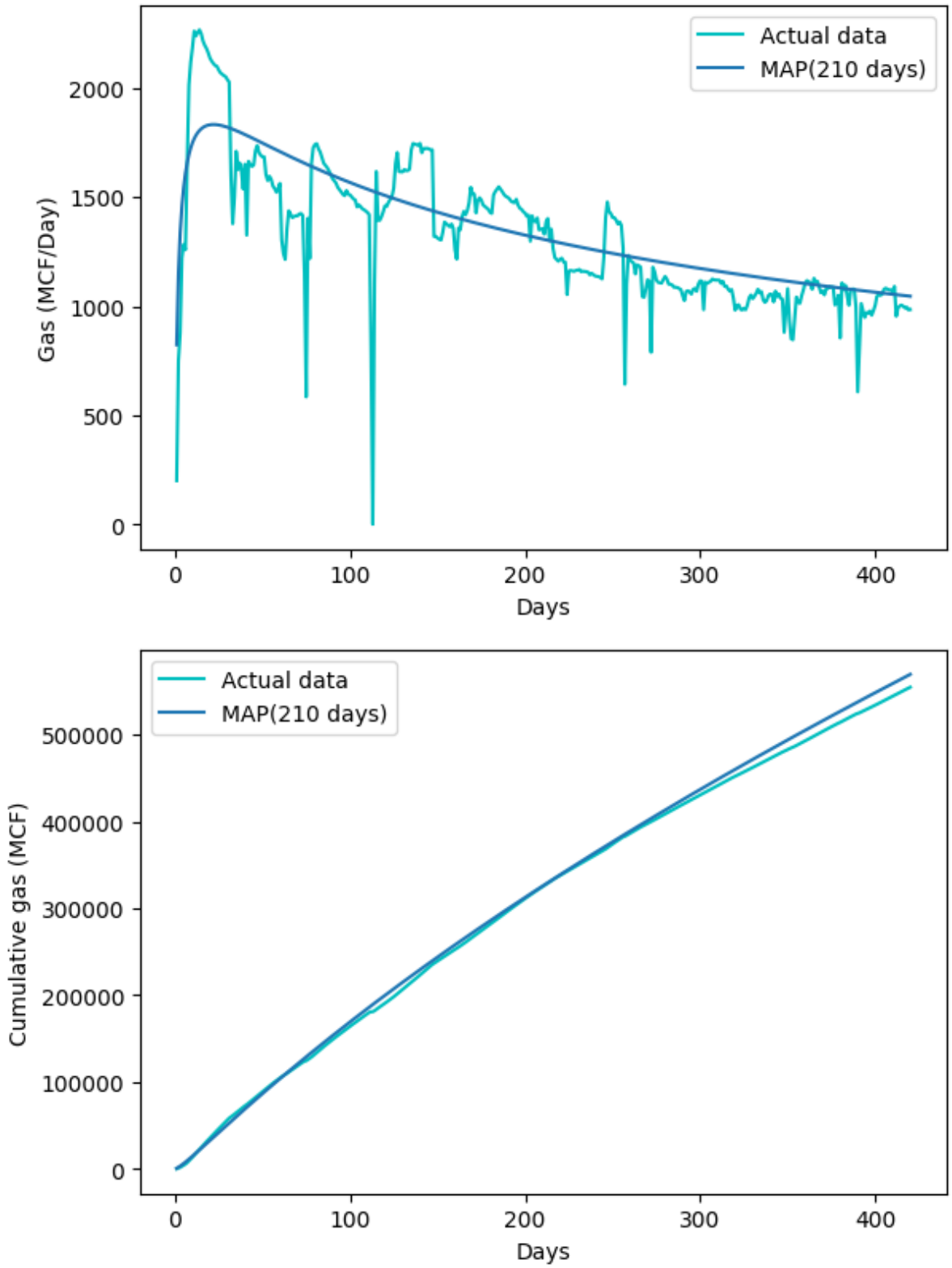

Figure 3.40 Estimates for forecasted daily production date and cumulative gas production for Well 7 using the Duong model, with the MAP estimate.

This well has a little over one year of production data with most of the duration being in the transient flow period. Duong model closely matches the production using MAP estimate indicating the effectiveness of this model for wells with prolonged with transient flow periods.

I will now present how the MCMC approach for parameter estimation was able to generate the MAP estimates for 101 oil wells using the Duong model. Figure 3.41 shows the correlation between the cumulative actual production and the cumulative production generated using MAP estimates. The linear regression line shows a strong relationship with an R-squared value of 0.947 which indicates good validity of the MAP estimates.

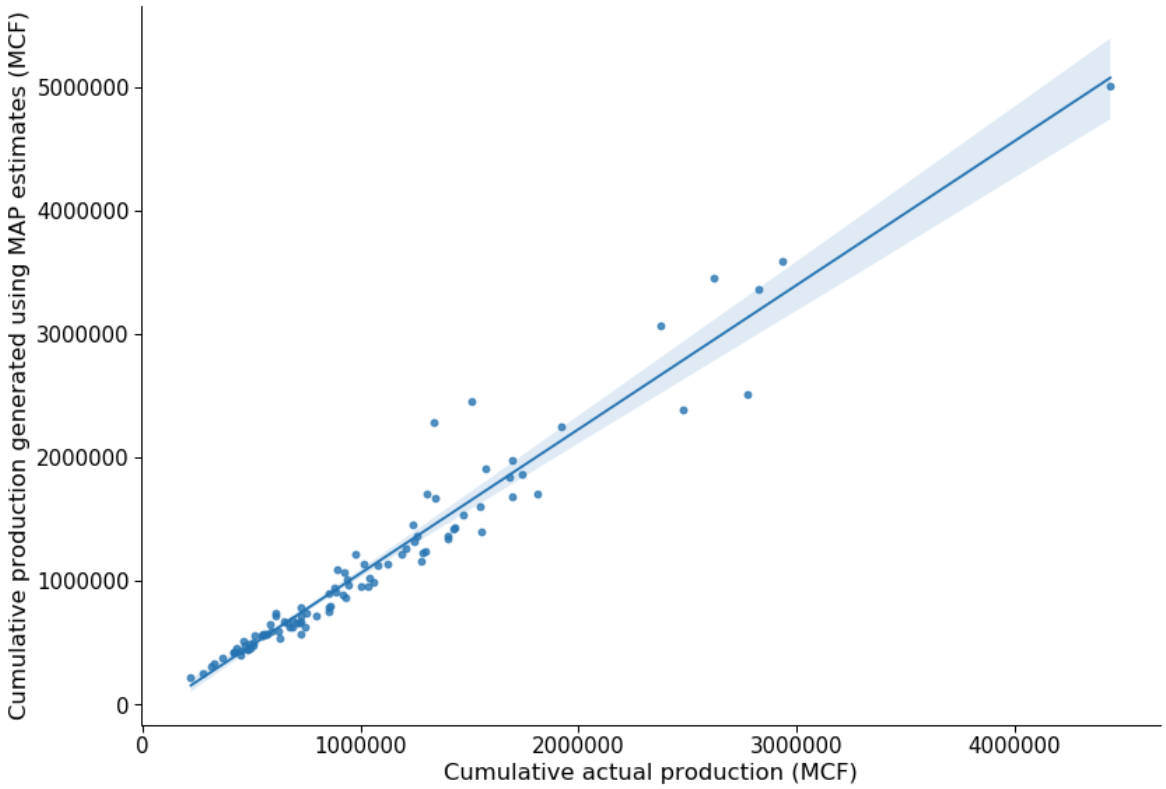

Figure 3.41 Regression line between the cumulative actual production and the cumulative production generated using MAP estimates for gas wells using Duong model.

Figure 3.42 shows the frequency plot of wells based on the condition if the cumulative actual production lies in the range of [P90 – P10] values of the posterior distribution. The x-axis has labels with intervals of 12 months starting Oct 2014. Figure 3.43 shows a similar frequency plot based on whether the actual cumulative production falls within [-10%, 10%] range of the MAP estimate. As seen earlier, gas wells do not follow the same trends as the oil wells for ranges. Duong model for gas wells shows the

highest percentage of wells to have the actual cumulative production in the [P10-P90] range.

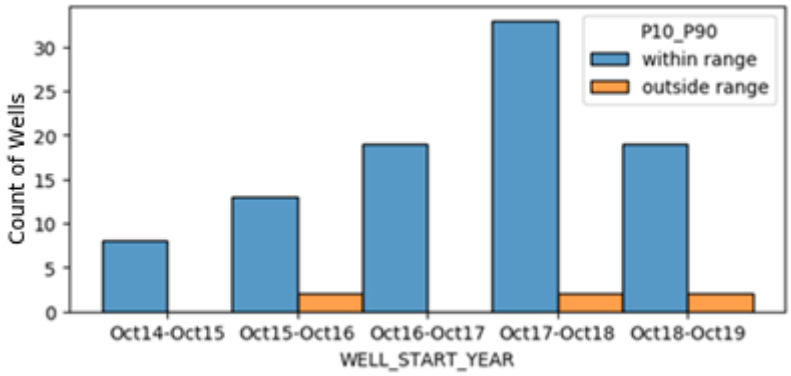

Figure 3.42 Histogram plot showing the distribution of gas wells falling within and outside the range of [P90-P10] estimates, while using Duong model.

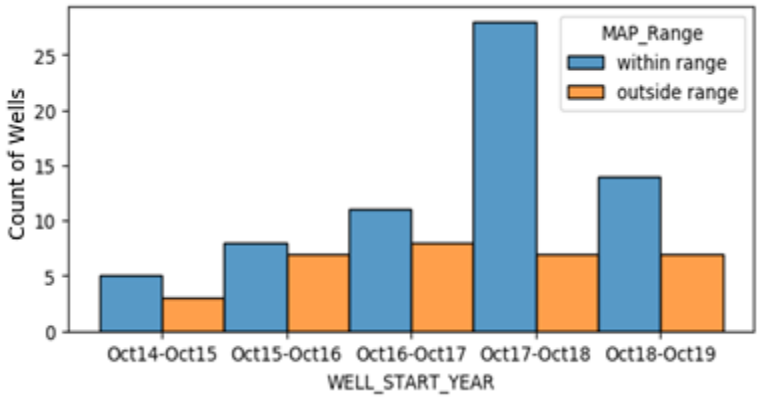

Figure 3.43 Histogram plot showing the distribution of gas wells falling within and outside the range of [-10%, 10%] of the MAP estimate, while using Duong model.

3.4.8 Multi-segment forecast for an oil well using the Arps decline curve model

Segment being a general word, the reference made here indicates a production interval that does not include any events with a downtime to the well which has a potential to alter the production post the downtime. Those events include offset well drilling or any other completions activity. The need for multi-segment forecast arises with the fact that the production rates need not necessarily produce at the earlier predicted rates. This implies that the for a well with a single downtime event, two segments of production exist which are before and after the downtime. For Well 8 as shown in Figure 3.44, this occurs at Day 570. The graph shows that the MAP estimate is underpredicting the production post Day 880.

Figure 3.44 shows the same well with a single MAP estimate from the model trained using 50% of the production data over both the segments. Two segments are trained using Arps as likelihood as if they are both different wells each starting on Day 1 and then merged back to show the results for Well 8. For the duration of the second segment, there is better matching when compared with the single forecast in Figure 3.45.

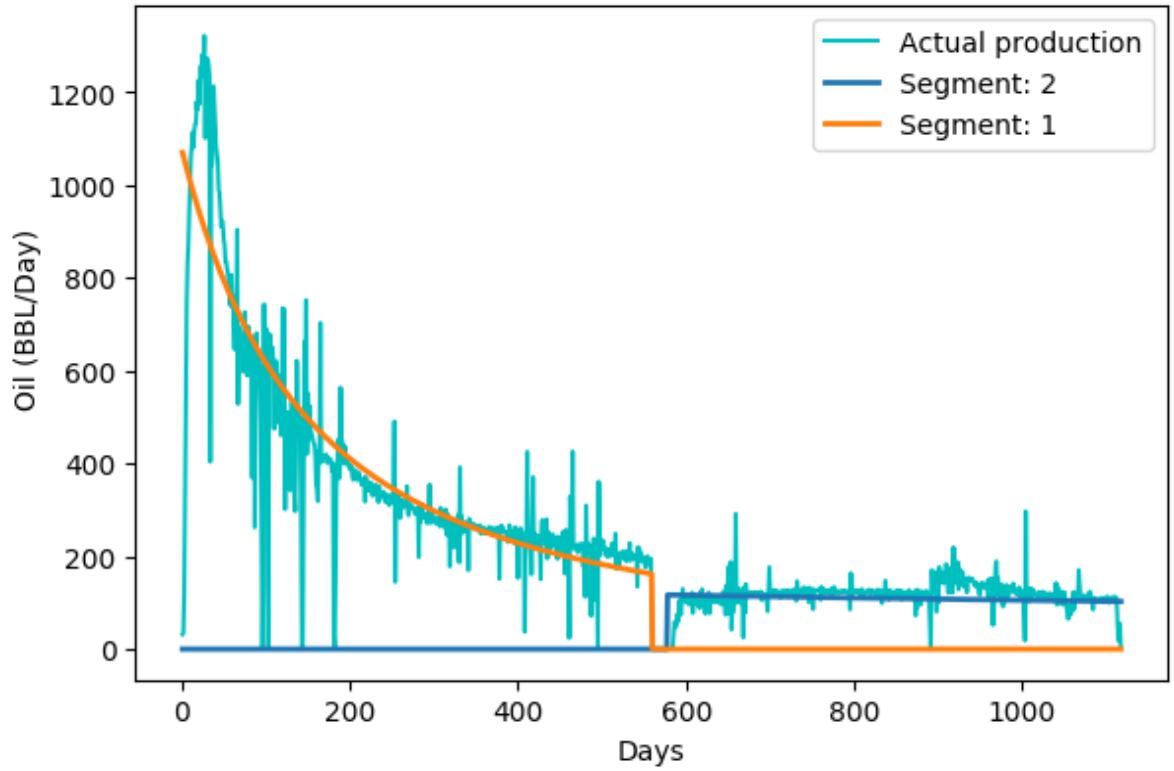

Figure 3.44 Estimates for forecasted daily production rate for Well 8 using the Arps model, with different MAP estimates for two segments.

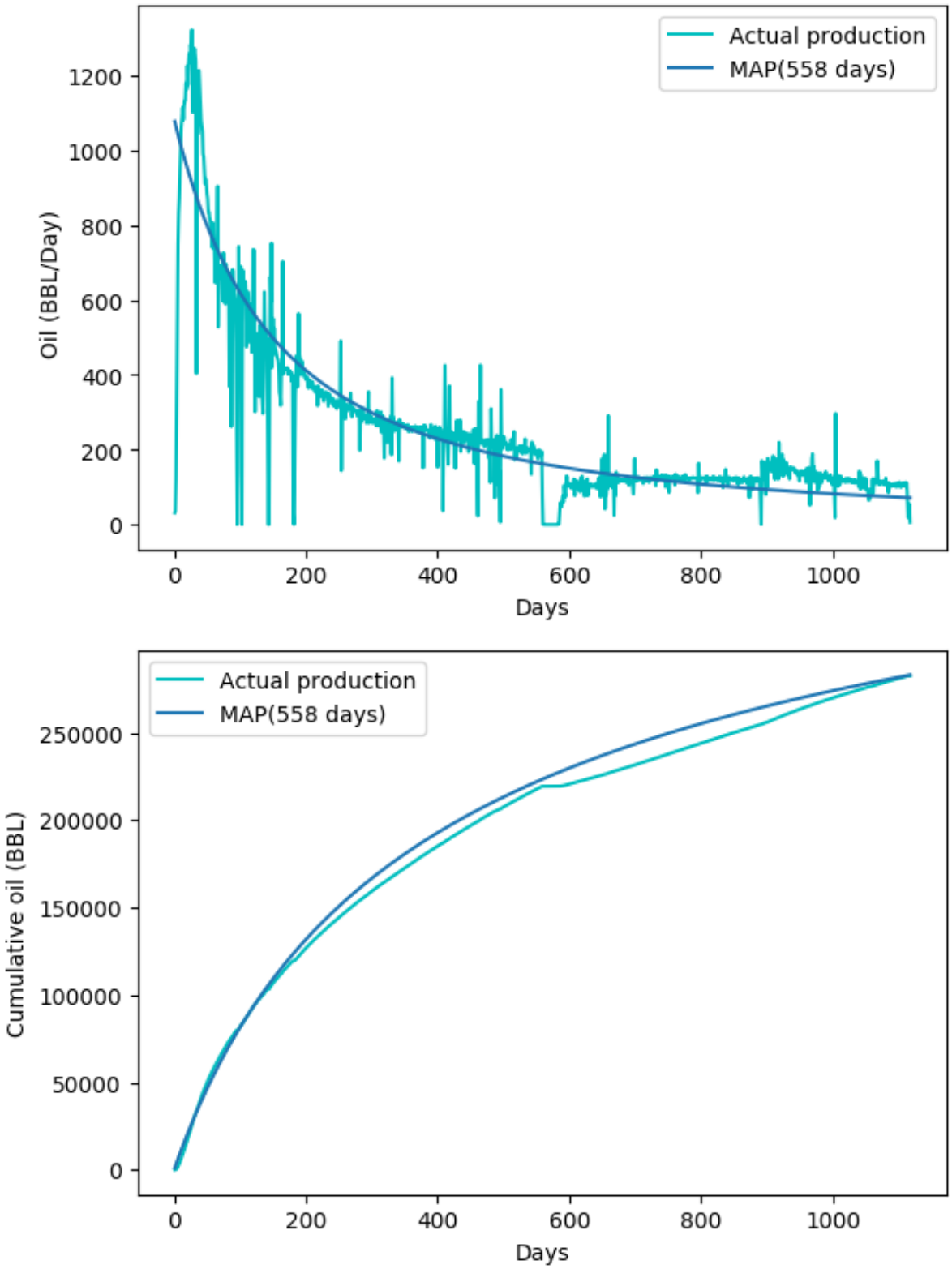

Figure 3.45 Estimates for forecasted daily production rate and cumulative oil production for Well 8 using the Arps model, with the MAP estimate.

3.4.9 MCMC summary

Figure 3.46 shows the number of wells for which I was able to generate the decline curve parameters using the three models – Arps, Duong, and power law exponential.

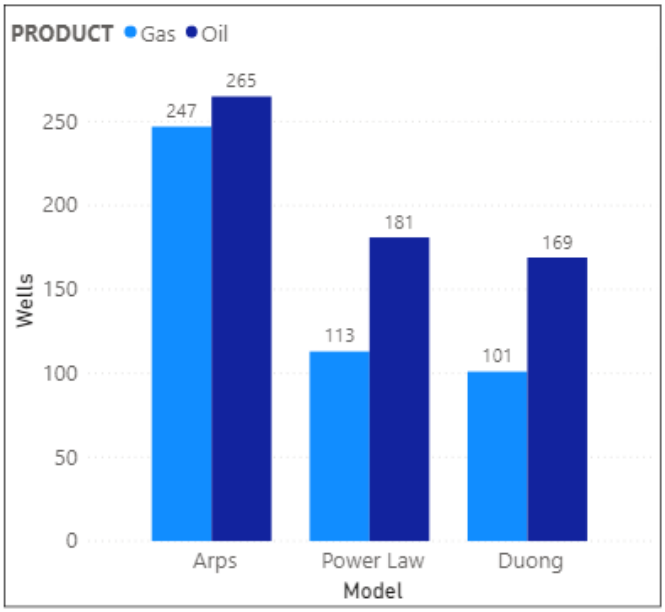

Figure 3.46 Classification of the total number of wells based on the product and decline curve model for which I was able to estimate the decline curve parameters.

The reason for the number of gas wells being less than the oil wells is because of the higher volatility associated with the gas rates when compared with the oil rates. When multiple patterns of production occur due to high volatility, models will converge to bad estimates. Arps generated the estimates for the highest number of wells. Duong models show best matching when the ratio of daily rate and cumulative production plotted vs time follows a straight line on a log-log plot. Power law exponential model shows best matching when the loss ratio exhibits the power law behavior. 121 of the total 268 available wells had significant downtime during their lifetime which could have impacted the single segment forecast for those wells by not producing a posterior estimate that generates plausible production estimates. The R-squared values for the linear regression plots between the cumulative actual production and cumulative predicted production calculated using MAP estimates for the three models are listed below:

	Arps	Power law exponential	Duong
Oil	0.979	0.972	0.954
Gas	0.967	0.958	0.947

Table 3.19 R-squared values for linear regression plots.

3.5 Well analyses for the Meramec play

In this section, I present, how over years, changing completion design has impacted the normalized initial rates. I also present how infill development has impacted the parent well production.

3.5.1 Infill versus non-infill wells performance

Completion design is defined by a set of attributes: amount of proppant used, lateral length of the completion, number of stimulation or hydraulic fracturing stages. Normalized initial rate is the actual initial rate divided by the amount of proppant used per feet of lateral length. Eq. 26 shows the definition for the normalized initial rate:

$$Qi\_Normalized = \frac{Q_i}{Lateral\ length} \qquad \text{.....(26)}$$

Figure 3.47 shows an increasing trend of Qi_Normalized until Oct 2017 – Oct 2018 and then following a declining trend. Infill development has started in the second quarter of 2018 and there has been no new non-infill development in the Meramec play since the first quarter of 2019.

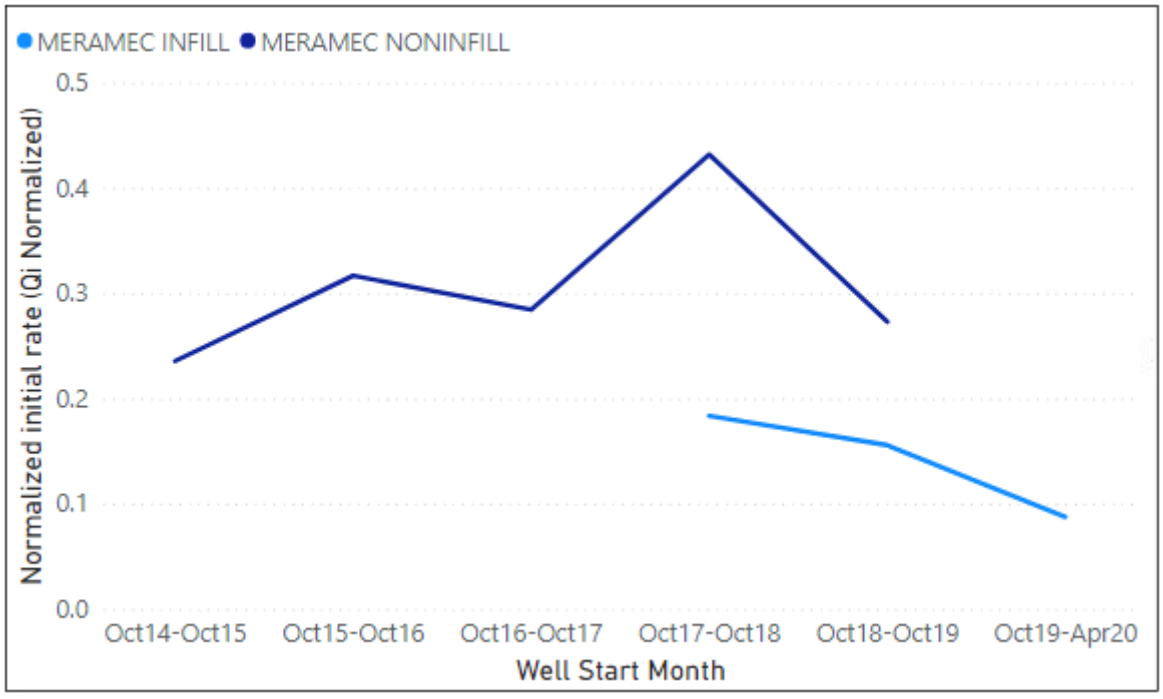

Figure 3.47 Average of normalized initial rates (qi) for infill and non-infill wells.

Figure 3.48 and Figure 3.50 show a clear increasing trend of the lateral length and the amount of proppant used for completing the wells. Figure 3.49 shows the changing pattern for the number of stimulation stages between infill and non-infill wells with no clear trend.

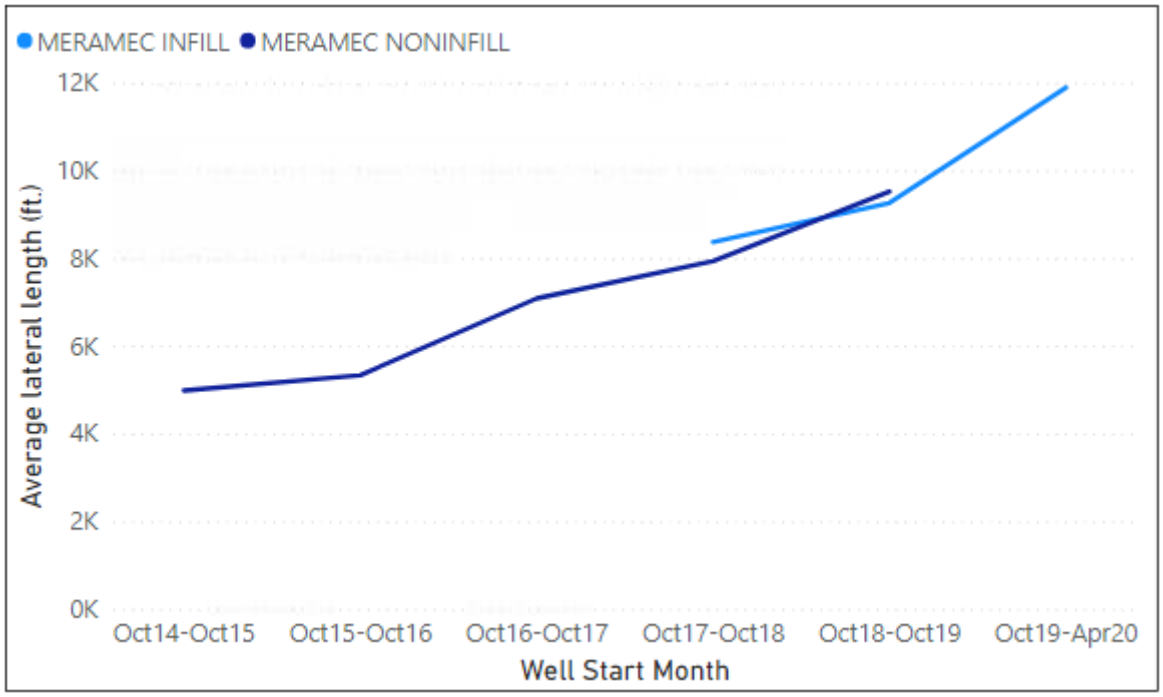

Figure 3.48 Average of lateral length over years for infill and non-infill wells.

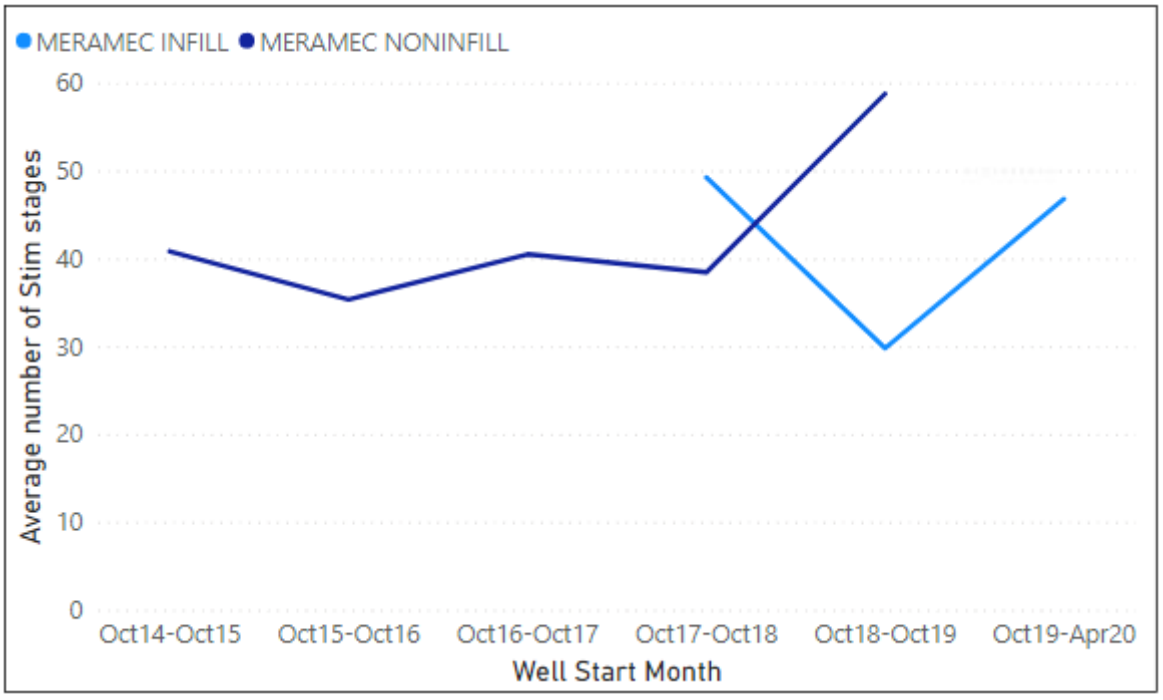

Figure 3.49 Average number of stimulation stages over years for infill and non-infill wells.

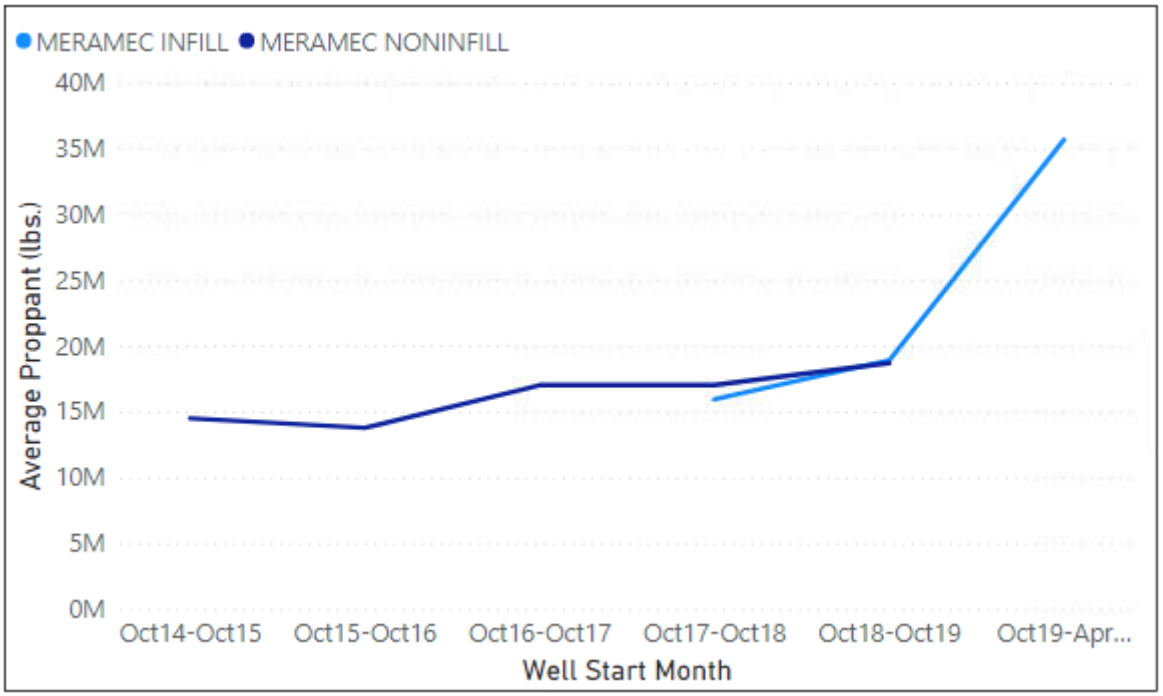

Figure 3.50 Average amount of proppant used over years for infill and non-infill wells.

3.5.2 Parent and child well analysis

For this analysis, I have picked five wells within a radius of 2000 feet, three of which came online in the second quarter of 2018 and the two that came online in the second quarter of 2020, I compare the values of the MAP estimates Q_i_Normalized, D_i and b. All three values are significantly less for the newer wells even with higher lateral length, proppant and the number of stimulation stages used. Figures 3.51, 3.52 and 3.53 show these values and their declining trend for the newer wells. The three non-infill wells have come online in the second quarter of 2018 and the two infill wells have come online in the second quarter of 2020. Figures 3.54, 3.55 and 3.56 show an increasing trend for the number of stages, lateral length and proppant used for completing the well. The five wells have a true vertical depth in range of [9300 ft – 9650 ft]. Figure 3.56 shows that the newer infill wells are in an exponential decline state given the Arps decline exponent 'b' is close to the value zero.

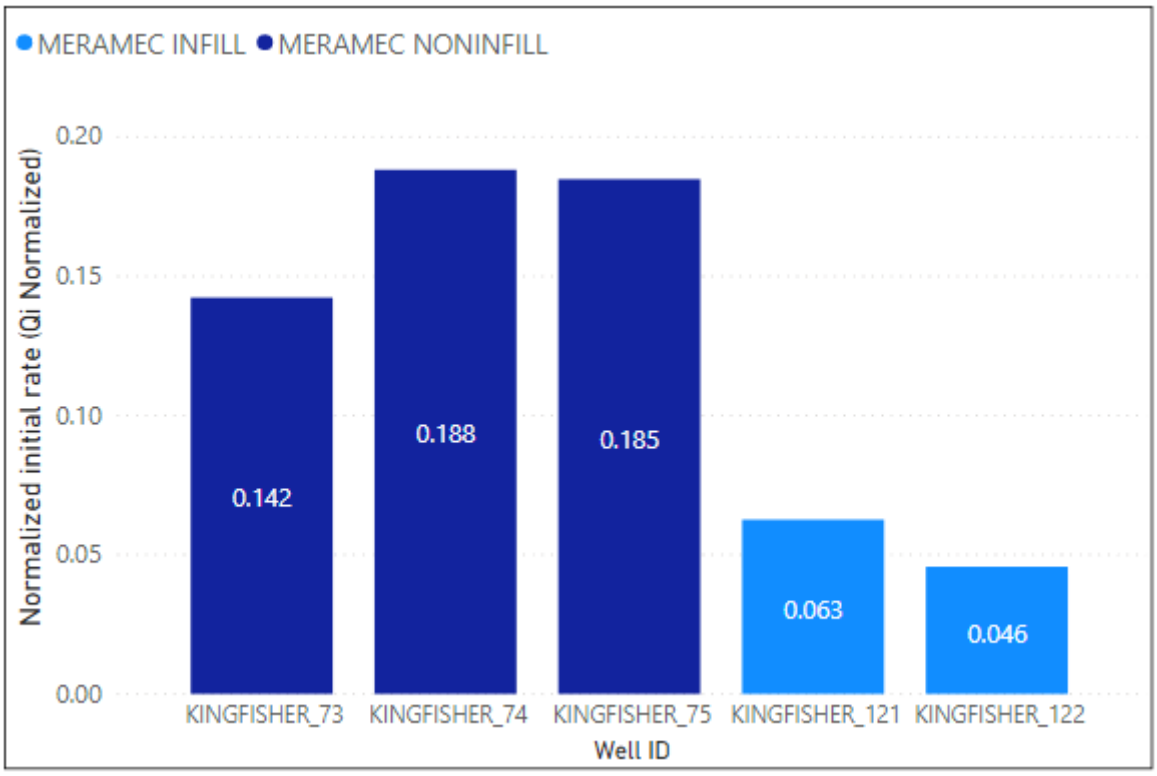

Figure 3.51 Arps normalized initial rate (q_i) for the Kingfisher county oil wells.

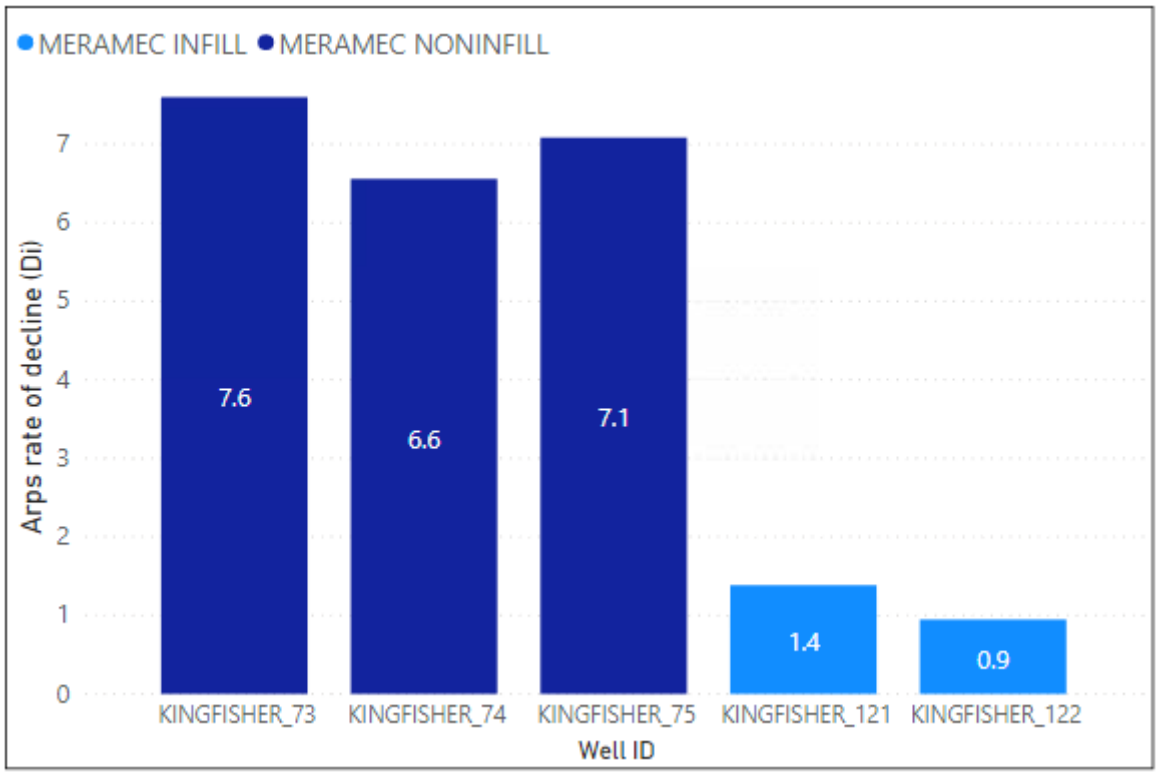

Figure 3.52 Arps rate of decline (D_i) for the five Kingfisher county oil wells.

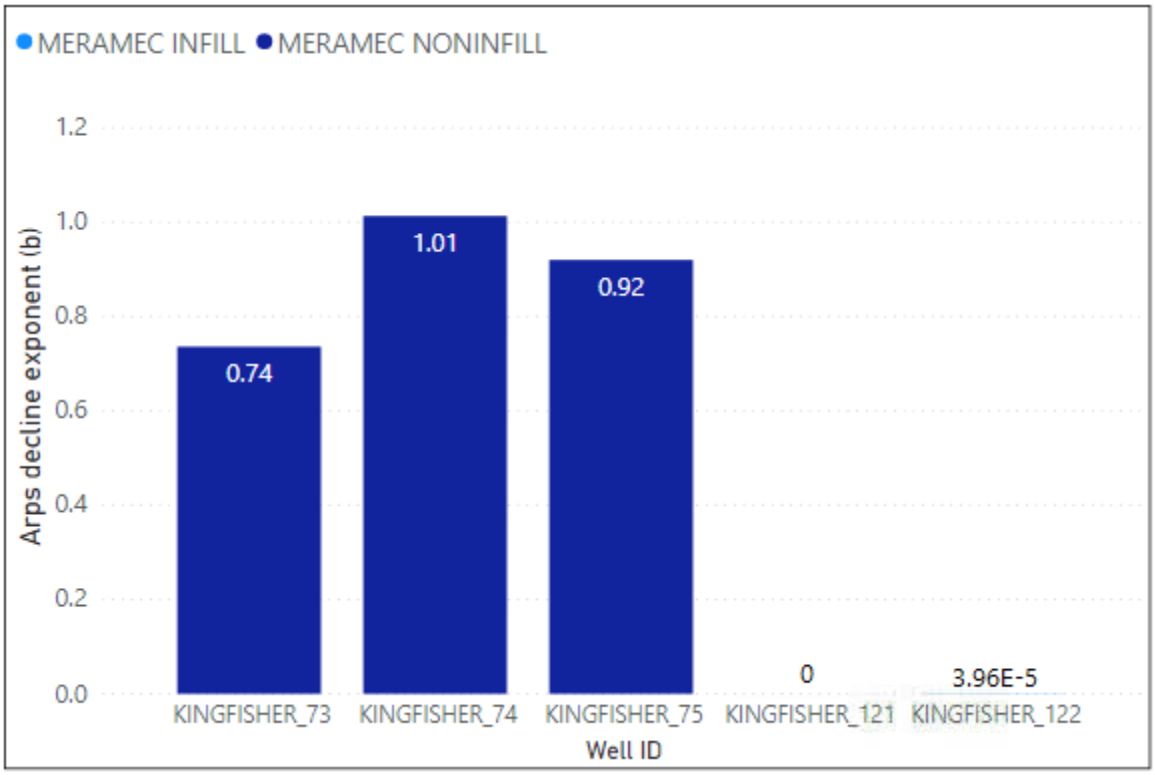

Figure 3.53 Arps decline exponent (b) for the five Kingfisher county oil wells.

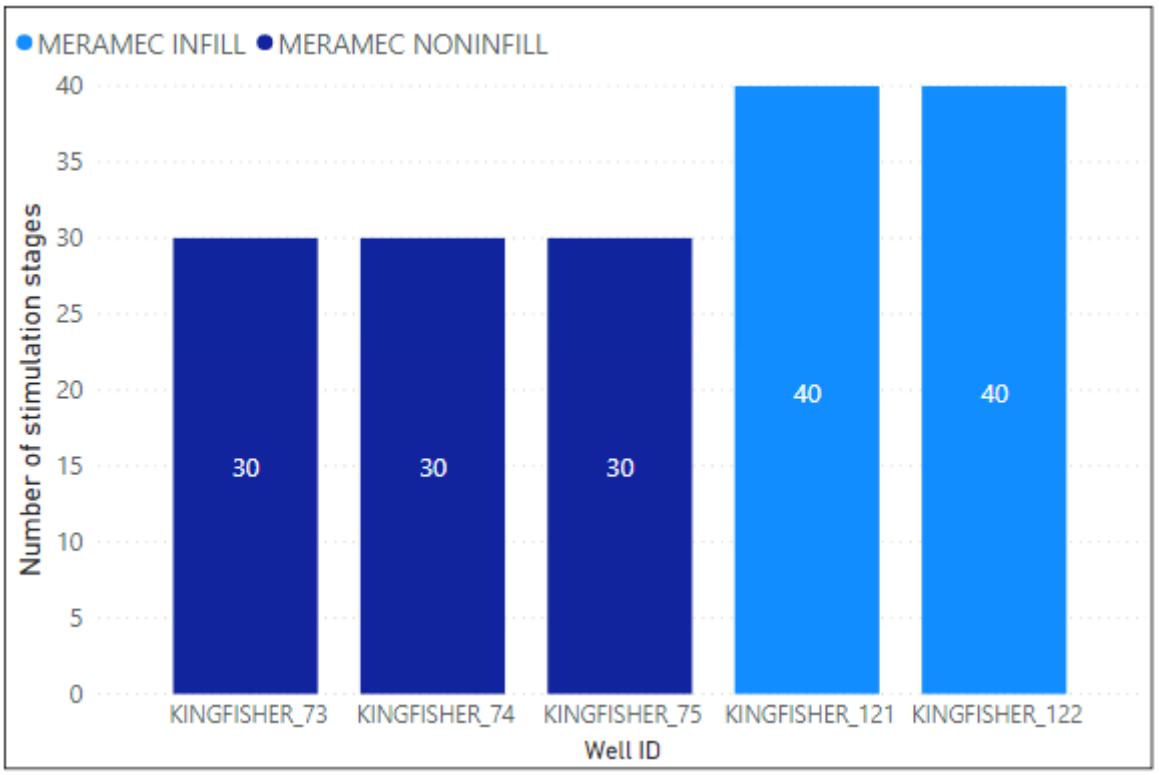

Figure 3.54 Number of stimulation stages for the five Kingfisher county oil wells.

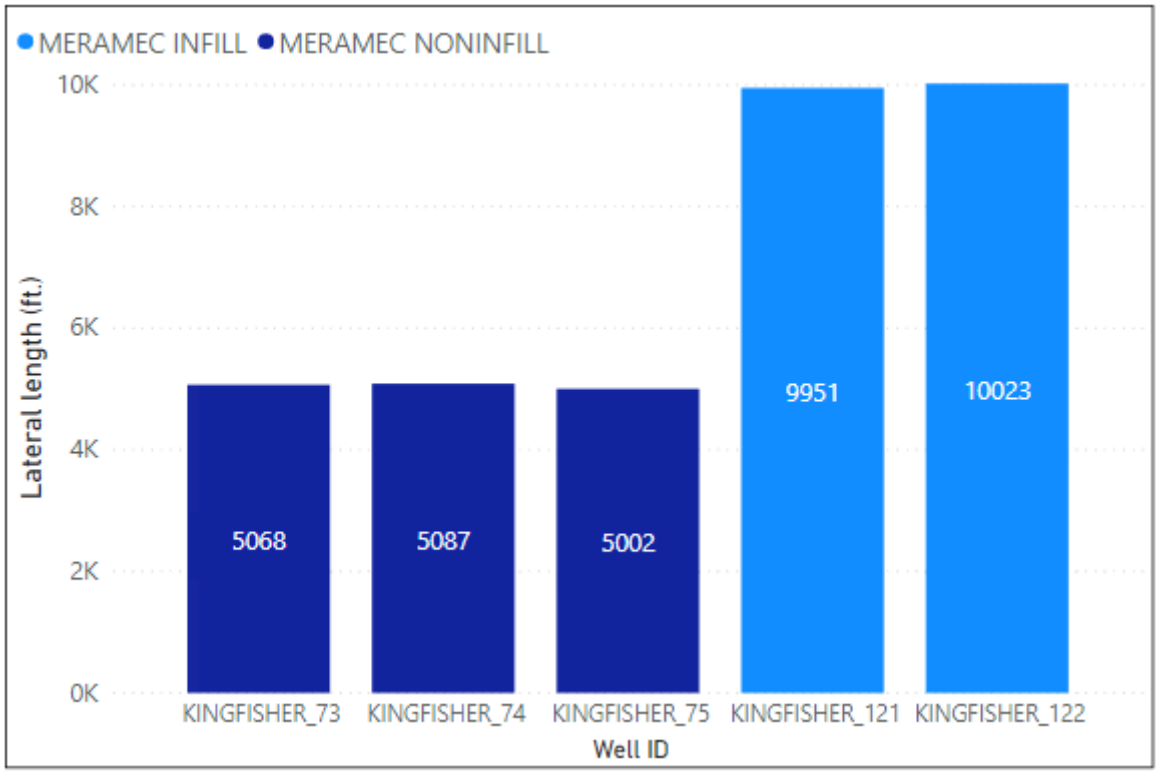

Figure 3.55 Lateral length (ft.) for the five Kingfisher county oil wells.

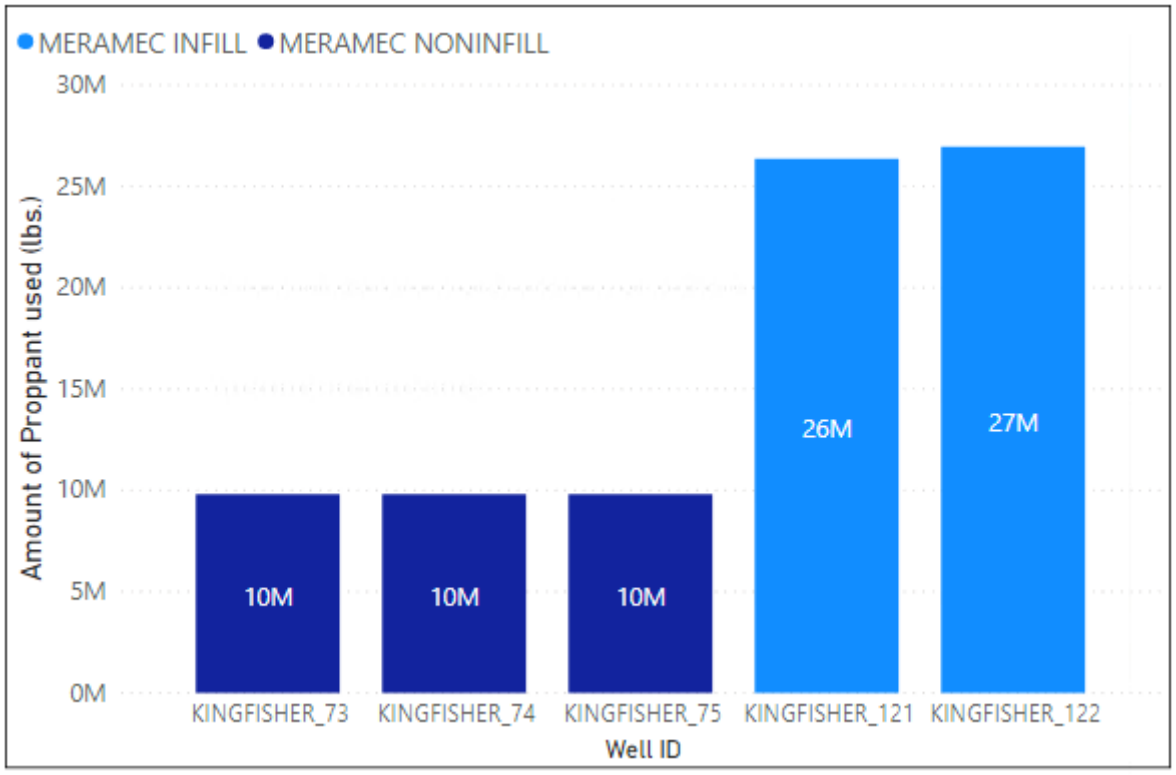

Figure 3.55 Amount of proppant used (lbs.) for the five Kingfisher county oil wells.

Chapter 4

Conclusions and Future Work

This chapter is divided into two sections. The first section presents conclusions based on the analysis conducted for this research. The second section presents the recommendations of this thesis.

4.1 Conclusions

In this study, I implemented a MCMC-based approach for decline curve parameter estimation and uncertainty estimation in production forecasting. I also compare the parameter values for infill and non-infill wells to provide insights into parent versus child well performance. The following conclusions can be drawn from this work:

1. MCMC using the Metropolis algorithm is a promising approach for uncertainty quantification in production forecasting.
2. The uncertainty in the production forecasts is seen to decrease with an increase in the availability of more production data.
3. Multi-segment forecasting is essential if offset activities and changing completion strategies impact the well production significantly. To avoid compromising production forecasts, it is always necessary to forecast ahead beginning with the termination of a significant disruptive event.
4. A comparison of infill (child) and non-infill (parent) well performance shows that child wells show reduced productivity despite being completed with longer laterals, and larger completion designs. This indicates that pressure depletion has already occurred in the region of the study.

4.2 Recommendations for future work

Given the potential of MCMC-based production forecasting and uncertainty estimation, it would be appropriate to extend this study to a larger set of wells and many other plays. This would allow for a more detailed description of infill versus non-infill well performance, identify sweet spots, diagnose anomalous declines, and make recommendations for completion practices.

Bibliography

Agarwal, A., Wei, Y., & Holditch, S. 2012. A Technical and Economic Study of Completion Techniques in Five Emerging US Gas Shales: A Woodford Shale Example. SPE Drilling & Completion. SPE-135396-PA. http://dx.doi.org/10.2118/135396-PA

Almasoodi M., Vaidya R., & Reza Z. 2019. Intra-Well interference in Tight Oil Reservoirs: What do We Need to Consider? Case Study from The Meramec . Paper URTEC-2019-83-MS presented at the Unconventional Resources Technology Conference, Denver, Co, 22-24 July. https://doi-org.ezproxy.lib.ou.edu/10.15530/urtec-2019-83

Anderson, D. M., Nobakht, M., Moghadam, S., & Mattar, L. (2010, January 1). Analysis of Production Data from Fractured Shale Gas Wells. Society of Petroleum Engineers. doi:10.2118/131787-MS

Andrieu, C., de Freitas, N., Doucet, A. et al. An Introduction to MCMC for Machine Learning. Machine Learning 50, 5–43 (2003). https://doi.org/10.1023/A:1020281327116

Arps, J. J. 1945. Analysis of Decline Curves. Transactions of the AIME. 1 December. SPE-945228-G. https://doi.org/10.2118/945228-G

Cheng, Y., Wang, Y., McVay, D., & Lee, W. J. 2010. Practical Application of a Probabilistic Approach to Estimate Reserves Using Production Decline Data. SPE Economics & Management, 1 April. SPE-95974-PA. https://doi.org/10.2118/95974-PA

Clark, A. J., Lake, L. W., & Patzek, T. W. 2011. Production Forecasting with Logistic Growth Models. SPE Annual Technical Conference and Exhibition, 30 October-2 November, Denver, Colorado, USA. SPE-144790-MS. https://doi.org/10.2118/144790-MS

Duong, A.N. 2011. Rate-Decline Analysis for Fracture-Dominated Shale Reservoirs. SPE Reservoir Evaluation & Engineering 14 (3): pp. 377-387. SPE-137748-PA. https://doi.org/10.2118/137748-PA

Fetkovich, M. J. 1980. Decline Curve Analysis Using Type Curves. Journal of Petroleum Technology. 1 June. SPE-4629-PA. https://doi.org/10.2118/4629-PA

Gong X., Gonzalez R., McVay D.A. et. al, 2011. Bayesian Probabilistic Decline-Curve Analysis Reliably Quantifies Uncertainty in Shale-Well-Production Forecasts. Presented at the Canadian Unconventional Resources Conference, Calgary, Alberta, Canada, 15-17 November. SPE-147588-MS. https://doi.org/10.2118/147588-MS

Gong, X., Gonzalez, R., McVay, D. A., & Hart, J. D. 2014. Bayesian Probabilistic Decline-Curve Analysis Reliably Quantifies Uncertainty in Shale-Well-Production Forecasts. SPE Journal. 19 (06): 1,047 - 1,057. SPE-147588-PA. http://dx.doi.org/10.2118/147588-PA

Guo B., Liu X., Tan X. 2017. Petroleum Production Engineering, Gulf Professional Publishing.

Gupta I., Rai C., Devegowda D., & Sondergeld C. 2020. A Data-Driven Approach to Detect and Quantify the Impact of Frac-Hits on Parent and Child Wells in Unconventional Formations. Paper URTEC-2020-2190-MS presented at the Unconventional Resources Technology Conference, Austin, TX, 20-22 July. https://doi-org.ezproxy.lib.ou.edu/10.15530/urtec-2020-2190

Ilk, D., Rushing, J. A., Perego, A. D., & Blasingame, T. A. 2008. Exponential vs. Hyperbolic Decline in Tight Gas Sands: Understanding the Origin and Implications for Reserve Estimates Using Arps Decline Curves. Presented at the SPE Annual Technical Conference and Exhibition, 21-24 September, Denver, Colorado, USA. SPE-116731-MS. http://dx.doi.org/10.2118/116731-MS

Jochen, V. A., & Spivey, J. P. 1996. Probabilistic Reserves Estimation Using Decline Curve Analysis with the Bootstrap Method. SPE Annual Technical Conference and Exhibition, 6-9 October, Denver, Colorado. SPE-36633-MS. https://doi.org/10.2118/36633-MS

Joshi, K. G., Awoleke, O. O., & Mohabbat, A. 2018. Uncertainty Quantification of Gas Production in the Barnett Shale Using Time Series Analysis. SPE Western Regional Meeting, 22-26 April, Garden Grove, California, USA. SPE-190124-MS. https://doi.org/10.2118/190124-MS

Lee, W.J. and Sidle, R. 2010. Gas-Reserves Estimation in Resource Plays. SPE Econ & Mgmt 2 (2): 86-91. SPE 130102-PA. http://dx.doi.org/10.2118/119197-MS

McVay, D. A., & Dossary, M. N. 2014. The Value of Assessing Uncertainty. SPE Economics & Management, 1 April. SPE-160189-PA. https://doi.org/10.2118/160189-PA

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machines. Journal of Chemical Physics, 21, 1087–1091. https://doi.org/10.1063/1.1699114

Metropolis, N., & Ulam, S. (1949). The Monte Carlo Method. Journal of the American Statistical Association, 44(247), 335-341. doi:10.2307/2280232

Odi U., Bacho S., & Daal J. 2019. Unconventional Reservoirs using a Variable Power Law Model: A Barnett Shale Example. Paper URTeC 2019-39-MS presented at the Unconventional Resources Technology Conference, Denver, CO, 22-24 July. https://doi-org.ezproxy.lib.ou.edu/10.15530/urtec-2019-39

Paryani, M., Awoleke, O. O, Ahmadi, M., Hanks, C., & Barry, R. 2017. Approximate Bayesian Computation for Probabilistic Decline-Curve Analysis in Unconventional Reservoirs. Society of Petroleum Engineers.SPE-183650-PA. https://doi.org/10.2118/183650-PA

Price, B., Haustveit, K., & Lamb, A. 2017. Influence of Stratigraphy on Barriers to Fracture Growth and Completion Optimization in the Meramec Stack Play, Anadarko Basin, Oklahoma. Paper URTeC 2697585 presented at the Unconventional Resources Technology Conference, Austin, TX, 14-16 July. https://doi.org/10.15530/URTEC-2017-2697585

S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-6, no. 6, pp. 721-741, Nov. 1984, doi: 10.1109/TPAMI.1984.4767596.

Salvatier J., Wiecki T.V., Fonnesbeck C. (2016) Probabilistic programming in Python using PyMC3. PeerJ Computer Science 2:e55 https://doi.org/10.7717/peerj-cs.55

Sun H. 2015. Advanced Production Decline Analysis and Application, Gulf Professional Publishing.

Valko, P. P., & Lee, W. J. 2010. A Better Way to Forecast Production from Unconventional Gas Wells. Society of Petroleum Engineers. SPE Annual Technical Conference and Exhibition, 19- 22 September, Florence, Italy. SPE- 134231-MS. https://doi.org/10.2118/134231-MS

W. K. Hastings, Monte Carlo sampling methods using Markov chains and their applications, Biometrika, Volume 57, Issue 1, April 1970, Pages 97–109, https://doi.org/10.1093/biomet/57.1.97

Xie, J., Efendiev, Y., & Datta-Gupta, A. (2011, January 1). Uncertainty Quantification in History Matching of Channelized Reservoirs using Markov Chain Level Set Approaches. Society of Petroleum Engineers. doi:10.2118/141811-MS