Probabilistic Decline Curve Analysis in the Permian Basin using Bayesian and Approximate Bayesian Inference

Anand Korde, Scott D. Goddard, and **Obadare O. Awoleke,** University of Alaska

Summary

In this work, a probabilistic methodology for decline curve analysis (DCA) in unconventional reservoirs is presented using several Bayesian model-fitting algorithms and deterministic models. The deterministic models considered are the power law exponential (PLE) model, the stretched exponential production decline (SEPD) model, Duong's model, and the logistic growth analysis (LGA) model. Accurate production forecasting and uncertainty quantification were the primary objectives of this study.

The Bayesian inferencing techniques described in this work utilize three sampling vehicles, namely the Gibbs sampling (implemented in OpenBUGS, an open-source software), the Metropolis-Hastings (MH) algorithm, and approximate Bayesian computation (ABC) to sample parameter values from their posterior distributions. These different sampling algorithms are applied in conjunction with DCA models to estimate DCA parameter prediction intervals. Using these prediction levels, production is forecasted, and uncertainty bounds are established.

To examine its reliability, the methodology was tested on over 74 oil and gas wells located in the three main subplays of the Permian Basin, namely, the Delaware play, the Central Basin Platform, and the Midland play. Results show that the examined DCA-Bayesian models are well calibrated, result in low production errors, and narrow uncertainty bounds for the production history data sets. Also, the LGA model is best in terms of prediction errors for all algorithms except MH. The Gibbs algorithm is nearly the best algorithm in terms of prediction error for all DCA models except the Arps model. Prediction errors are the highest in the Central Basin Platform. The methodology was also successfully applied to unconventional reservoirs with as low as six months of available production history. Depending on the amount of production history data available, the probabilistic model that provides the best fit can vary. It is therefore recommended that all possible combinations of the deterministic and Bayesian model-fitting algorithms be applied to the available production history. This is to obtain more confidence in the conclusions related to the production forecasts, reserves estimate, and uncertainty bounds.

The novelty of this methodology relies on using multiple combinations of DCA-Bayesian models to achieve accurate reserves estimates and narrow uncertainty bounds. This paper can help assess shale plays because some of the shale plays are in the early stages of development when productivity estimations are carried out.

Introduction

To facilitate the prediction of expected productivity and ultimate recovery in shale plays, researchers have suggested numerous deterministic models in the past one to two decades. The models developed to address the drawbacks of the traditional Arps equation include: the PLE model (Ilk et al. 2008), the SEPD model (Valko and Lee 2010), Duong's model (Duong 2011), and the LGA model (Clark et al. 2011). These models can be combined with various statistical algorithms to provide uncertainty ranges for production forecasting and reserves estimation purposes.

There are two main paradigms of incorporating uncertainty into the workflow while using the abovementioned methods, namely the frequentist and the Bayesian approach. Frequentist methods that have been used in the technical literature include the bootstrap (Jochen and Spivey 1996), the modified bootstrap (Cheng et al. 2010), and time-series analysis methods (Joshi et al. 2018). Bayesian methods that have been used in the past include the Markov chain Monte Carlo (MCMC) (Gong et al. 2011) and ABC methods (Paryani 2015; Paryani et al. 2017). We intend to focus the present work on using the Bayesian approach to incorporate uncertainty into productivity predictions from the aforementioned decline curve models. Bayesian methods may have an advantage over the frequentist approach in this setting because they add model structure and consider the prior information (background knowledge) concerning the parameters. Considering good prior information can lead to improved predictions. A more detailed summary of the preceding papers is found in the following paragraphs.

Jochen and Spivey (1996) introduced the bootstrap method, which is a statistical approach to DCA using Monte Carlo analysis and nonlinear regression. The process involves generating many synthetic data sets from the original data set by random sampling with replacement. Each synthetic data set determines a set of decline curve parameters. Every set of decline curve parameters determines the reserves estimates. A set of all such reserves estimates form a distribution of the reserves estimates and is used to quantify uncertainty. The advantage of this method apart from providing a means to quantify uncertainty is that it does not require knowledge of the prior distribution of the decline curve parameters. However, this method assumes that the original data are independent and identically distributed. This assumption is usually not accurate.

Cheng et al. (2010) addressed the independent and identically distributed issue by introducing a more rigorous bootstrap technique called the modified bootstrap method (MBM). The MBM involves generating synthetic data sets using block resampling of the residuals obtained by the application of decline models (in this case, the Arps model) to the actual data set. The MBM technique was tested on 100 mature conventional oil and gas wells having sufficient production history by Cheng et al. (2010). Their results show that MBM has better coverage rate (CR; 83%) as compared to the bootstrap method (34%). Gong et al. (2011) further demonstrated that the MBM method was well calibrated for conventional and unconventional reservoirs.

Gong et al. (2011) were the first researchers to apply Bayesian statistics to DCA models with the objective of reserves estimation and uncertainty quantification in unconventional reservoirs. Their methodology involved a combination of the Arps DCA model and an MCMC sampling technique. The Metropolis algorithm was the sampling strategy used to sample from the target or posterior

Original SPE manuscript received for review 31 August 2020. Revised manuscript received for review 27 October 2020. Paper (SPE 204477) peer approved 2 November 2020.

distribution. The main consideration in this technique is how the prior distribution, the likelihood function, and the posterior distribution are defined and computed. They considered the Arps model parameters as uniformly distributed random variables. The likelihood function was defined as the conditional probability of the available historical production data given the DCA parameters. The posterior distribution was defined as the distribution of the DCA parameters after the available historical production data is considered. Posterior distributions for each of the random variables are used to determine the production forecast and the uncertainty estimates. Gong et al. (2011), in their case study of 167 wells, showed that their Bayesian method required far less computational time (25 minutes) compared to the MBM technique (6.5 hours). They also indicated that the Bayesian method had a low relative error, low absolute errors, and low error in true reserves compared to the MBM.

Paryani et al. (2017) used the ABC methodology to quantify the uncertainty associated with DCA models. The primary purpose of developing this methodology was to simplify the Bayesian inferencing procedure by approximating the complicated likelihood function. For a set of parameters drawn from their priors, a data set is simulated from the DCA model parameters. The likelihood is replaced with a distance measure that quantifies how different the simulated data are from the observed data; many distance measures are possible for this quantification. If the misfit in the summary statistic (observed vs. simulated) is greater than some set threshold, the drawn parameters are discarded; otherwise, they are accepted. Using a data set consisting of 121 wells (100 gas wells and 21 oil wells), Paryani et al. (2017) concluded that the ABC methodology is computationally faster than the likelihood-based numerical approximation because the likelihood is not evaluated directly. They also indicated that the LGA model provides better CRs than other known deterministic models when employed in the ABC framework.

Based on the information above, in the present work, we will extend Gong et al.'s contribution (Gong et al. 2011) and consider how Bayesian and ABC methodologies can be used to incorporate uncertainty using deterministic models other than the Arps equation. We will also determine how these probabilistic models perform if limited production history is available. Accordingly, the objectives of our paper are as follows: to identify and discuss various MCMC-based sampling techniques that can be used to perform Bayesian data analysis; to identify the best probabilistic model for different sectors of the Permian Basin using our dataset; and to test the performance of these probabilistic models in scenarios in which there is limited production history. All the above analysis is based on an analysis of 74 wells from the Permian Basin.

Overview of the Permian Basin

The Permian Basin is an old and widely acknowledged hydrocarbon-producing region in North America. It extends from the southeastern part of New Mexico to the western part of Texas, covering around 86,000 square miles over 52 counties. Mass deposition, continental collision, and subbasin sediment filling are the three main attributes leading to the evolution of the basin. The basin is divided into three main subbasins, namely, the Midland Basin, the Central Basin Platform, and the Delaware Basin, as shown in **Fig. 1.**

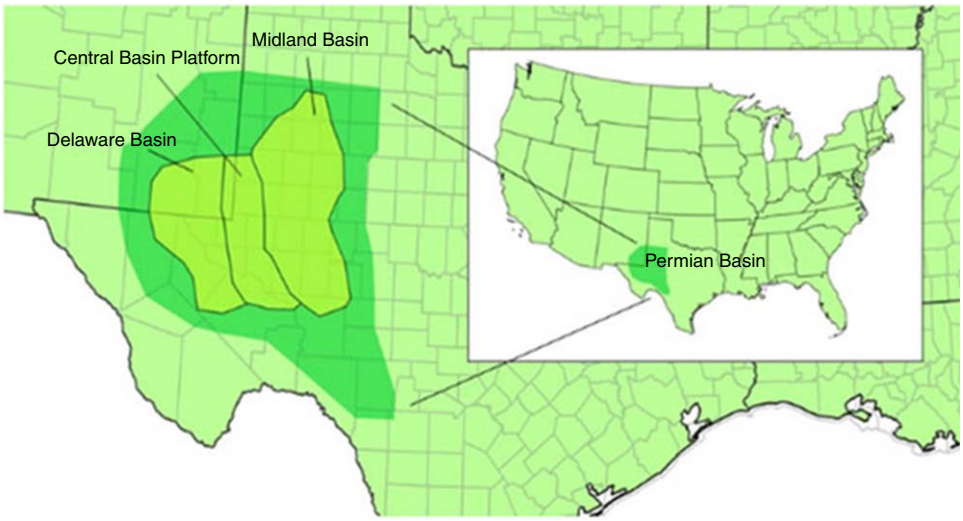

Fig. 1—Location map of the Permian Basin (Tarka Resources 2019).

The Midland Basin is located toward the eastern region of the Permian Basin. The two stratigraphic sections that make up the Leonardian and Wolfcampian strada are the Spraberry (along with the Dean) and the Wolfcamp formations (see **Fig. 2**). The basin has a multilayered stratigraphy with diverse lithologic zones. Oil and gas companies initially focused on drilling vertical wells through the Spraberry formation, extending into the Wolfcamp formation. However, in recent time, multistage hydraulic fracturing techniques have facilitated the exploration of deeper intervals of the Wolfcamp and Spraberry shales. Wolfcamp and Spraberry plays are together known as the Wolfberry play.

The Central Basin Platform lies in the central part of the Permian Basin (Fig. 1) and has a higher elevation when compared to the other two basins (Midland and Delaware), resulting in a dissimilar depositional environment. The Permian Basin's stratigraphic interval rapidly thins to the Central Basin Platform, in comparison to the gradual thickness decline toward western Delaware and eastern Midland Basins. The Central Basin is divided into several formations such as Wolfcamp, Abo, San Andreas, Seven Rivers, and Yates. The sequence mainly comprises carbonate reef deposits and shallow marine clastic sediments (Ward et al. 1986).

The Delaware Basin is located toward the western region of the Permian Basin. It is around 2,000 feet deeper than the Midland Basin, and thus, sediments experience significantly higher pressure during burial. The Delaware Basin has different stratigraphy compared to the Midland Basin in the Leonardian strata (Fig. 2). However, both basins share an analogous Wolfcamp formation, which is also an ideal heterogenetic resource of hydrocarbons. Bone Spring and Avalon are the other main formations in the Delaware Basin. Wolfcamp and Bone Spring formations are together known as the Wolfbone formation. The deeper part of the Delaware Basin is multistacked like the Midland Basin.

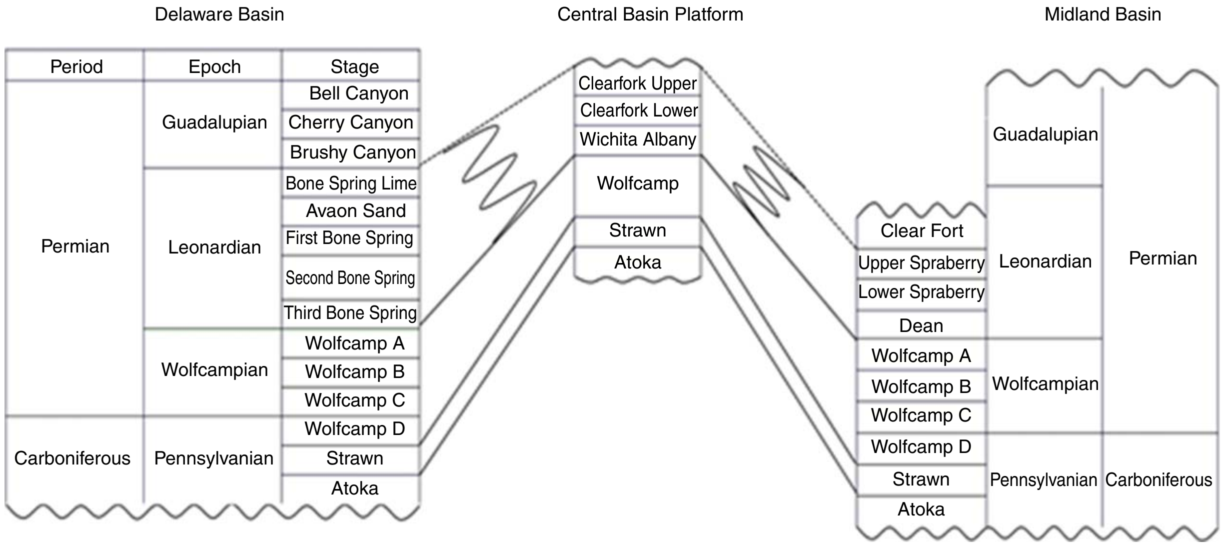

Fig. 2—General stratigraphic schema of Upper Carboniferous to Upper Permian intervals for the Permian region (EIA 2018).

In this study, we focus on reserves estimation and uncertainty quantification of hydrocarbons produced from the Midland (Midland, Terry, and Upton Counties), Central Platform (Ward County) and Delaware (Ward County) subplays.

Methodology

This study assesses the performance of five DCA models in conjunction with each of three Bayesian samplers. The DCA models chosen for inclusion are Arps hyperbolic, SEPD, Duong, PLE, and LGA. The Bayesian samplers are known as the Metropolis algorithm, the Gibbs samplers, and ABC. The process starts by dividing the historical production data for a given well into two parts. The earlier portion is used for fitting the DCA-Bayesian models and is known as the *hindcast*. The fitted models are then used to make predictions for the later portion of the data, which is used for validating the forecast predictions. By comparing these predictions to the validation data, the strength of the various DCA-Bayesian models can be evaluated.

As compared to the traditional frequentist analyses that engineers have customarily relied upon in DCA uncertainty analysis, the Bayesian approaches employed here offer a tool to update individual beliefs in the presence of new data. The representation of uncertainty in parameters' values before looking at any data is accomplished by defining prior probability distributions for the parameters. Sampled data are then used to update these distributions to result in posterior distributions. This updating process is done through a simple application of Bayes' theorem:

$$P(\theta|Y) = \frac{f(Y|\theta) * P(\theta)}{P(Y)}. \quad (1)$$

Here, Y refers to the set of observations in the data, and θ is a set of the parameters in the model. $P(\theta|Y)$ represents the posterior distribution of the parameters, $f(Y|\theta)$ is the likelihood function, $P(\theta)$ is the prior distribution of the parameters, and $P(Y)$ is the marginal likelihood function (i.e., a normalizing constant). We will discuss each of the components of Eq. 1 in the following paragraphs.

The prior is a probability distribution that incorporates background knowledge concerning each of the parameters of the DCA model in question. Background knowledge may include expert opinions, previous analyses, and other dependable sources. Priors may also incorporate little or no background information when such information is lacking or of contested validity. These priors may be termed noninformative; they give the data a large amount of latitude in influencing the posteriors. We use noninformative priors in this study to represent our substantial lack of foreknowledge regarding parameters' values in each DCA model. Specifically, these noninformative priors are set as uniform distributions with varying lower and upper limits for the various parameters included. For example, in the case of the Arps model, which uses q_i, b, and D_i as parameters, the prior for parameter b is a uniform distribution between 0 and 2. **Table 1** provides a list of parameters of the different DCA models and the prior distributions applied to these parameters in this study. The distribution boundaries are set large enough to include all physically plausible values of the parameters in the analysis.

The likelihood function represents the presumed distribution of the data given specific parameter values. The likelihood function contains all the information about parameter values that is obtained from the data itself and is the way through which the data affect the posterior distribution. A Bayesian model can be defined entirely only when both the prior distribution and the likelihood function are fully specified. For our study, we use the same likelihood function as described by Gong et al. (2011), as shown in Eq. 2.

$$f(Y|\theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{\mathrm{r}(\theta)^2}{\sigma^2}\right). \quad (2)$$

Here, Y is the observed data, θ is the set of DCA parameters, σ is the standard deviation of the statistical errors, and $\mathrm{r}(\theta)$ is the sample standard deviation of residuals between the observed data and the decline curve.

The posterior distribution is a compromise of the prior distributions and the observed data. It reflects the updated knowledge about the parameters' values and involves balancing prior knowledge with observed evidence. It expresses remaining uncertainty about the set of parameters in question, conditional on observed data.

The posterior predictive distribution represents the distribution of future, additional observations drawn from the population after accounting for the uncertainty in parameters' values that is expressed in the posterior distribution. It can be estimated through compositional sampling, wherein parameter draws are made from the posterior, and these are put into the likelihood function so that new observations can be drawn from it.

DCA Model	Parameter	Units	Lower Limit	Upper Limit
Arps	q_i	Mcf/D	1	100,000
	b	–	0	2
	D_i	1/year	0.1	50
PLE	q_i	Mcf/D	1	100,000
	$\hat{D}$	–	0.001	10
	D_∞	–	0.0001	1
	n	–	0.001	2
SEPD	q_i	Mcf/D	1	100,000
	η	–	0.01	5
	τ	–	0.15	10
Duong	q_i	Mcf/D	0.01	100,000
	a	–	0.5	5
	m	–	0.5	2
LGA	k	Mcf	1,000	5,000,000
	n	–	0.01	1.2
	a	days	1	900

Table 1—Prior distributions for the parameters of the various deterministic models used in this study.

Sampling Algorithms

Constructing production forecasts and prediction intervals requires finding the DCA model parameters' joint posterior distribution. In principle, this would entail the computation of the marginal distribution $P(Y)$, which involves solving intricate integrals. A more tractable approach is to use an algorithm that can estimate the joint posterior distribution by sampling from it. Various algorithms exist to sample from the joint posterior; the likelihood-based MCMC techniques known as the Metropolis algorithm and the Gibbs sampler algorithm, as well as the nonlikelihood-based ABC algorithm form the underlying strategies used in this work.

Metropolis Algorithm. The Metropolis algorithm (and its more famous generalization, the MH algorithm) generates a Markov chain of values drawn from a target distribution. This is done via a proposal distribution, which proposes draws that are either accepted or rejected. The proposal distribution is given by $\theta_{\text{proposal}} \sim N(\theta^{(t-1)}, \sigma_i^2)$. Here, the initial standard deviation, σ_i, for different DCA models is chosen to acquire good mixing of the algorithm as discussed by Gonzalez et al. (2012). The standard deviation values we have used are shown in **Table 2.**

Method	Parameter	σ_i
Arps	$\ln q_i$	0.2
	b	0.2
	$\ln D_i$	0.4
PLE	$\ln q_i$	0.2
	$\ln \hat{D}$	0.4
	$\ln D_\infty$	0.2
	n	0.4
SEPD	$\ln q_i$	0.2
	$\ln \eta$	0.4
	$\ln \tau$	0.2
Duong	$\ln q_i$	0.2
	a	0.2
	m	0.2
LGA	$\ln k$	0.4
	n	0.2
	a	0.4

Table 2—Values of standard deviation for MH proposal distribution in DCA models.

The algorithm begins with presetting an initial parameter value $\theta^{(0)}$ consistent with the prior distribution. It then continues as follows:

1. A new candidate θ_{proposal} is drawn from the proposal distribution, which is centered on the previously accepted value $\theta^{(0)}$.
2. An acceptance ratio is computed that divides the posterior density evaluated at the proposed parameter θ_{proposal} by the posterior density evaluated at the previously-accepted value $\theta^{(0)}$.
3. θ_{proposal} is accepted with probability equal to the acceptance ratio or 1 (whichever is smaller). A uniformly distributed random-number generator is typically used for determining whether to accept θ_{proposal}.
4. If accepted, θ_{proposal} replaces the current candidate value $\theta^{(0)}$ and becomes $\theta^{(1)}$. If θ_{proposal} is rejected, the current value $\theta^{(0)}$ becomes $\theta^{(1)}$. Therefore, if the proposed value is accepted, the algorithm moves to a new position; otherwise, it stays in the same position.
5. At this point, the algorithm repeats the previous four steps a large number of iterations in which, on iteration t, $\theta^{(0)}$ would be notated as $\theta^{(t-1)}$ and $\theta^{(1)}$ would be notated as $\theta^{(t)}$.

If the Markov chain formed is long enough and certain mild regularity conditions hold, the Metropolis algorithm is known to converge to the desired posterior distribution (Gong et al. 2011) (i.e., the distribution of many such sampled values approximates the target posterior distribution). In our implementation of the Metropolis algorithm, we defined the prior distribution and the initial proposal distribution for each DCA model according to Tables 1 and 2, respectively. The likelihood function is given in Eq. 2. Together these priors, these proposals, and this likelihood allow for the estimation of the parameters' posterior distributions.

Successive posterior draws were then plugged into the relevant DCA model to generate forecasts of the validation data, resulting in posterior predictive distributions of the forecast at each timestep. The 10th, 50th, and 90th percentiles of these posterior predictive distributions constitute the P90, P50, and P10 prediction bounds, respectively. We measured the accuracy of the predictions by computing scaled prediction errors between actual and predicted values. We also measured the P10 to P90 interval widths. The aim for each well forecast is: to have at least 80% of the actual production values bounded within the predicted P10 to P90 range, to have a minimal prediction error, and to have a narrow P10 to P90 interval width.

Although the number of iterations required for the convergence of each well data can vary, we discarded the first 2,000 iterations of the algorithm as a burn-in period wherein the algorithm is working toward convergence. Samples are collected after the burn-in period with the expectation that the Markov chain formed has converged during the burn-in period. As a standard, 20,000 iterations are simulated for each well and model after the burn-in period. Computation for this methodology was carried out with the help of the R programming language (R Core Team 2018).

Gibbs Sampling Using OpenBUGS. The Gibbs sampler is another MCMC algorithm because it also produces a Markov chain of random draws from a distribution that approximates the posterior. Gibbs sampling forms the basic architecture of common MCMC software packages like BUGS and JAGS.

This algorithm repeatedly samples from the conditional posterior distribution of each variable, given all the other variables present in the model. These conditional posterior distributions, if not available analytically, can be sampled by using some other class of MCMC algorithm within the Gibbs sampling framework. The algorithm cycles through the parameters of the DCA model, drawing each from its conditional posterior distribution, given each of the other parameters. One cycle through all parameters constitutes a single iteration of the sampler. As with the Metropolis algorithm, the joint distribution of the parameters over many iterations of the Gibbs sampler has been shown to converge to the target posterior distribution under mild regularity conditions.

We defined the prior distribution and the likelihood function in the same manner as in the case of the Metropolis algorithm and used the open-source software OpenBUGS (Lunn et al. 2009) to carry out the sampling process mentioned above. The number of iterations and burn-in value limits were kept the same as in the Metropolis sampler to induce parity in the methods' evaluation. Statistical theory holds that the Gibbs and the Metropolis algorithm converge to the same posterior distribution. In including both samplers in this study, we do not mean to imply that we expect to get distinct predictions from them under ideal circumstances, but rather, we have opted to try both in order to see which one performs better under real-world conditions. Practical constraints such as the need to choose proposal distributions may translate into differential mixing rates while necessarily finite numbers of sampler iterations suggest that the two algorithms may, in practice, perform differently in material ways.

Approximate Bayesian Computation. Another method used to estimate the posterior distribution of model parameters rooted in Bayesian methodology is the ABC technique. Although MCMC samplers are theoretically well founded and often straightforward to implement in existing software, they depend on the specification of a likelihood function. This is because the likelihood function is used to portray the probability of the observed data and helps to quantify the support lent to particular values of the parameters by the data. The formulation of a likelihood function for simple models can be relatively simple, but for complex models it can be difficult. The ABC technique is used to bypass the requirement of a likelihood function while approximating the posterior distribution. It is potentially quicker to run than MCMC samplers because the likelihood is not evaluated directly but replaced with an approximation that may be easier to calculate (Paryani 2015). In this paper, the method involves three essential pieces: prior distributions, a distance measure, and the DCA model. This method works on the principle of rejection sampling; during each iteration, a set of parameters is drawn from the prior distribution so that production rate at the given timestep can be predicted using a given DCA model. This predicted production is then compared to the actual production using the predetermined distance measure. If the measure is beyond a predetermined threshold, the sampled set of parameters gets rejected; otherwise, it is accepted.

Once the algorithm terminates, the set of all the accepted parameters form a Monte Carlo sample drawn from a distribution may closely resemble the posterior distribution that would result if a likelihood function were specified. As with MCMC methods, posterior predictive distributions can then be obtained through compositional sampling, and these produce the prediction bounds P10, P50, and P90. A critical difference between the likelihood-free ABC method and the likelihood-based MCMC methods is that the parameters drawn in the ABC method at successive iterations are not dependent on each other.

As with the likelihood-based methods, the ABC method was evaluated by partitioning the data into a hindcast and validation data. We define the prior distributions in the same manner as the likelihood-based methods (given by Table 1). For this study, we sampled 100,000 simulated data sets to estimate parameter posterior distributions. The distance measures or summary statistics we selected are a mean of cumulative production, median absolute deviation of cumulative production, and standard deviation of cumulative production. Simulated data sets that are not within the set threshold limits are rejected. The maximum threshold value used is 0.01, which means only a 1% difference in the summary statistics is considered acceptable. All the accepted simulated parameters then form a distribution

and are used for uncertainty analysis and reserves forecasting. We employed the R statistical software "abc" package (Csillery et al. 2012) to perform these computations.

Note that in this study, the ABC method is not applied to the PLE model because the results obtained were not satisfactory. The main reason for this inadequacy might be related to the larger number of parameters (four) that are involved in the PLE model as compared to the other DCA models (three). It may also be related to our choice of distance measures.

Initial Validation of the Methodology

For reference and validation, we compare our production forecast and uncertainty estimation results to those generated in Gong et al. (2014) (see **Fig. 3**). Using the same data set, we applied the Arps model in conjunction with Bayesian probabilistic DCA for the likelihood-based Metropolis sampler (see **Fig. 4**), the Gibbs sampler (see **Fig. 5**), and the likelihood-free ABC method (see **Fig. 6**). The length of the hindcast in all the foregoing methods was 45 months. The burn-in period was 2,000 iterations, and the total number of iterations was 20,000. Trace plots and posterior distributions for the parameters of the Arps model for the three different sampling regimes are shown in **Figs. 7 through 9.** We observe that this data set is not monotonically decreasing and could benefit from filtering. We did not do this, however, for two reasons: 1) We wanted to compare our results directly with the work of Gong et al. (2014) using the same data set they used, and 2) our preference is not to modify data sets unless the production data trends change drastically due to operational changes like shut ins/restarts and refracturing. In the case of this data set, there was no suggestion in Gong et al.'s paper that this well was shut in at any time during the production period under investigation. In the special case of refracturing, we recommend that the engineer analyzes the pre- and post-fracturing production data separately.

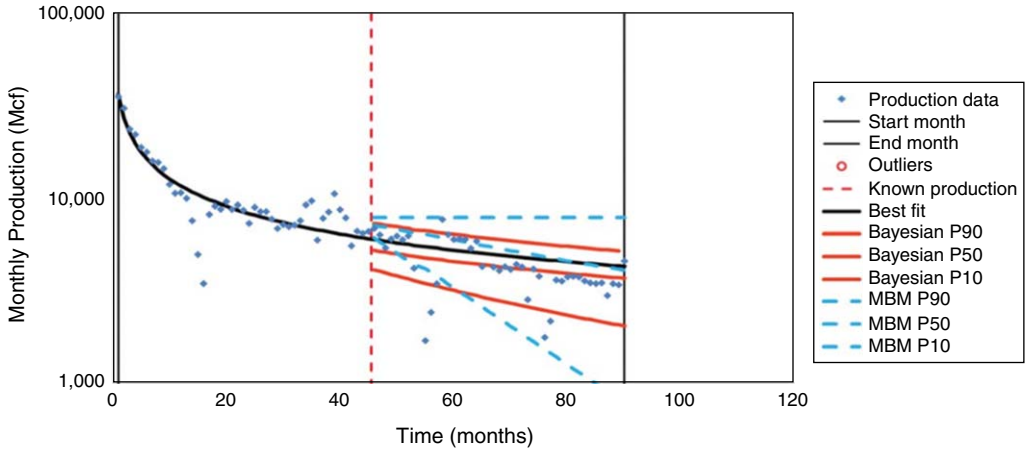

Fig. 3—Probabilistic forecast (Gong et al. 2014).

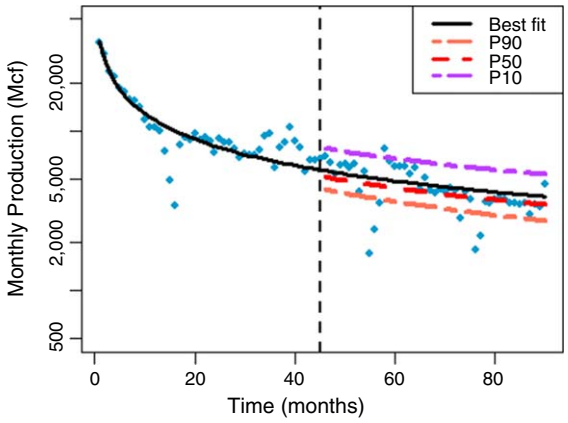

Fig. 4—Probabilistic forecast using the Metropolis algorithm.

Convergence Diagnostics

In Bayesian inference using Monte Carlo methods, convergence is said to occur when the Markov chain produced sufficiently approximates the posterior distribution of interest. Although there does not exist an infallible method for verifying the convergence of a Markov chain, there are methods that can provide evidence for or against convergence.

Trace Plots. A trace plot is a time-series plot of a parameter's values generated from the Markov chain across iterations. If the plot is centered on a constant mean, it provides evidence of convergence. Likewise, if the plot forms a clear pattern (for example, a constant increase/decline), it is indicative of nonconvergence, as is the failure of parameter values to adequately traverse the parameter space. Similarly, multiple independent chains of the same parameters of a given model can be constructed and overlaid on the same graph to assess the quality of mixing. Poor quality of mixing is evidence of the chains being nonconvergent.

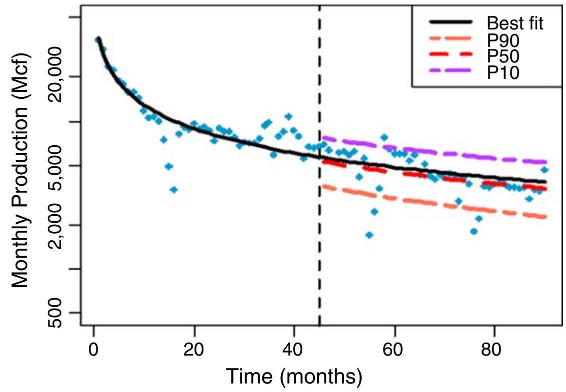

Fig. 5—Probabilistic forecast using the Gibbs sampler.

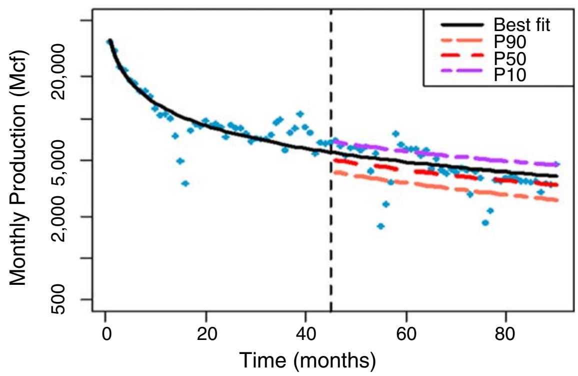

Fig. 6—Probabilistic forecast using the ABC method.

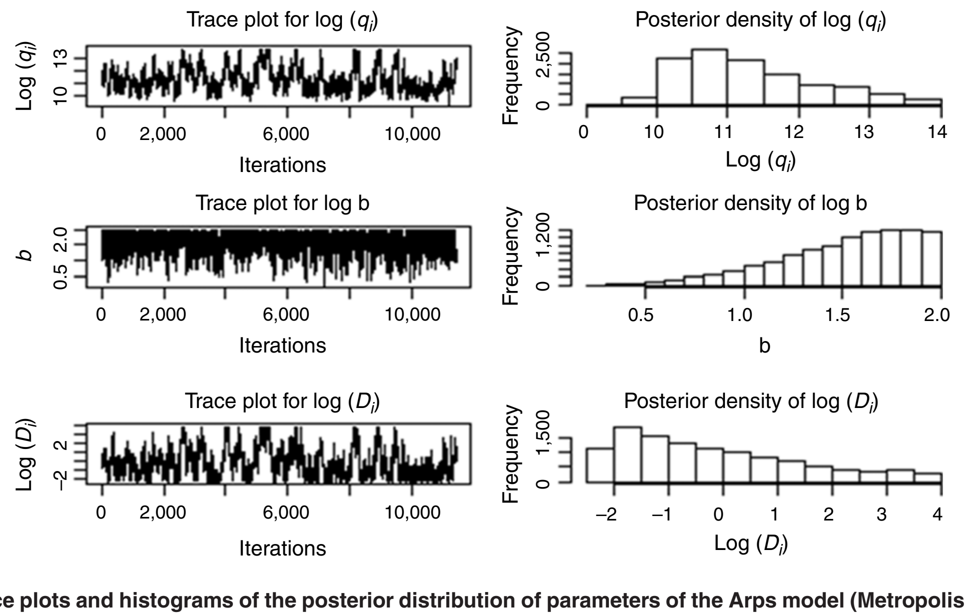

Fig. 7—Trace plots and histograms of the posterior distribution of parameters of the Arps model (Metropolis algorithm).

As an example, a trace plot containing two posterior chains for the parameter q_i from the Arps model is shown in **Fig. 10.** Although the two chains show no apparent pattern, their divergence is suggestive of poor mixing. In contrast, in **Fig. 11** the chains traverse the sample space better and do not show any obvious incongruence, demonstrating good mixing. Convergence is only validated in Fig. 11.

Gelman and Rubin Convergence Test. The Gelman and Rubin convergence test (Gelman and Rubin 1992) is another method that can be used to provide evidence against nonconvergence of an MCMC chain. The method requires the construction of two or more chains of sampled values of the parameters and uses only the second half of the total iterations. It first calculates the potential scale reduction factor (PSRF) as defined in Eq. 3.

$$\mathrm{PSRF} = \sqrt{\frac{n-1}{n} + \frac{(m+1)B/n}{mW}}. \quad (3)$$

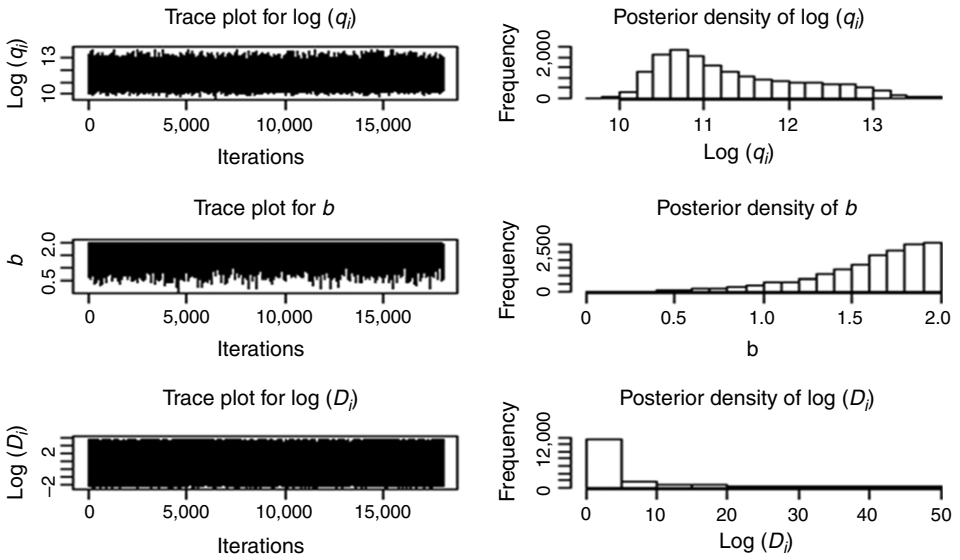

Fig. 8—Trace plots and histograms of the posterior distribution of parameters of the Arps model (Gibbs sampler).

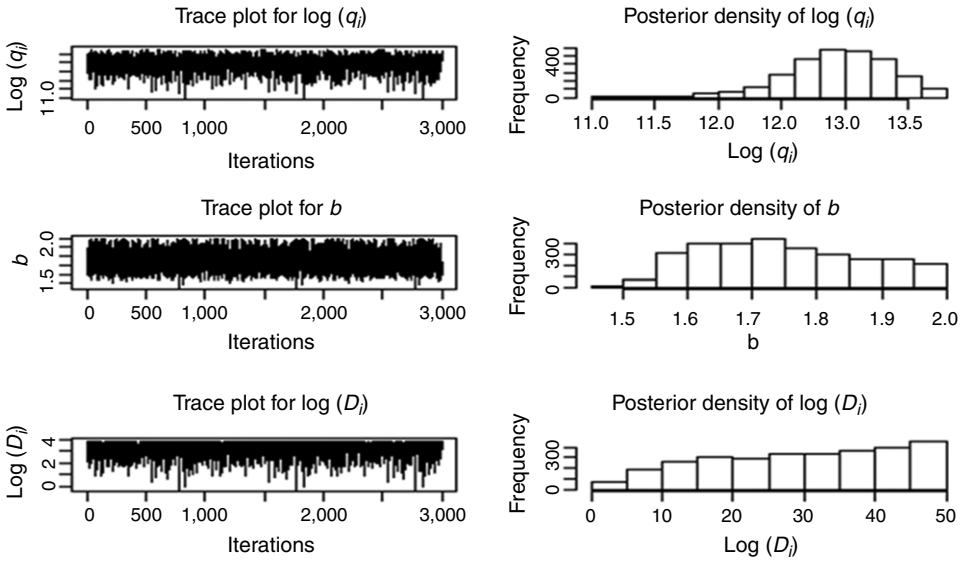

Fig. 9—Trace plots and histograms of the posterior distribution of parameters of the Arps model (ABC method).

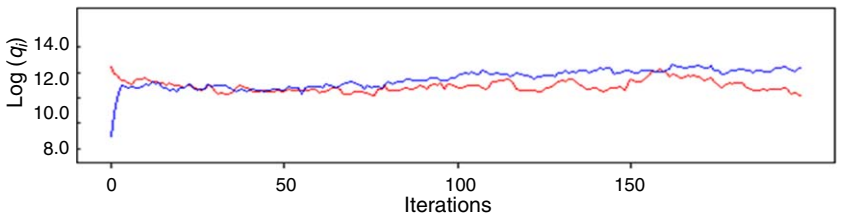

Fig. 10—Trace plot of parameter q_i with 200 iterations.

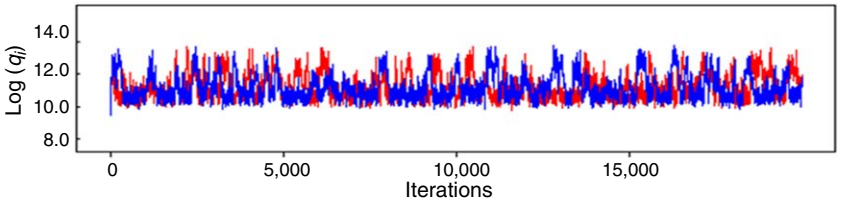

Fig. 11—Trace plot of parameter q_i with 20,000 iterations.

Here, B/n is the variance between the chains, n is the number of iterations used, m is the number of chains, and W is the mean of variance within the chains. The variance between the chains (B/n) should continue to grow smaller relative to the mean of variance within the chains as the chains converge to a common target distribution. Thus, PSRF should grow close to 1 as iterations increase; a large PSRF would indicate that the variance in between the chains is greater than the variance within the chains. To assess the convergence of a set of parameters simultaneously, the multivariate potential scale reduction factor (MPSRF) is used, based on Eq. 4.

$$\mathrm{MPSRF} \geq \max\{\mathrm{PSRF}_i\}. \quad (4)$$

Here, i indexes the parameter in use. A 97.5% quantile of the MPSRF is computed for each parameter; if the computed quantile has a value greater than 1.2, it represents nonconvergence. The Gelman and Rubin test results can be easily computed using a function from the package "coda" (Plummer et al. 2006) in the R computational software. **Fig. 12** shows the changes in the median PSRF through the iterations for a posterior distribution of parameters q_i, b, and D of the Arps model. The distribution constructed using two chains per parameter and 20,000 iterations generates a 97.5% quantile value of 1 for parameters b and D. The 97.5% quantile value generated for parameter q is 1.01. The MPSRF is calculated to be 1. Because the 97.5%, quantile values of each parameter are less than 1.2 and because the MPSRF is 1 (less than 1.2), we can confirm that the Gelman and Rubin test provides evidence of convergence.

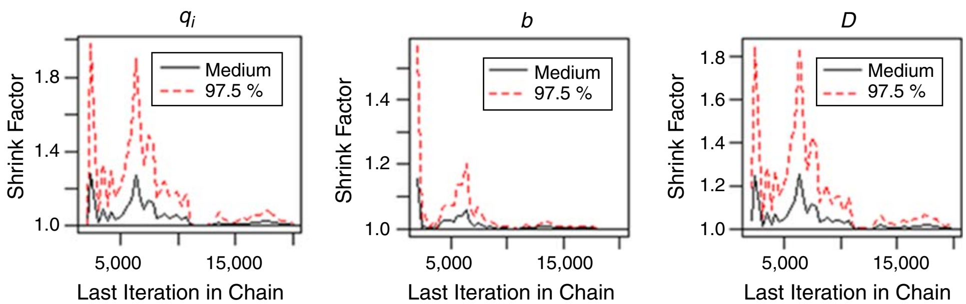

Fig. 12—Gelman and Rubin diagnostic plots for Arps model parameters.

Results and Discussion

Data Collection and Set of Criteria for Deterministic-Probabilistic Model Selection. The production history data used for the oil and gas wells in this study were obtained from the Drillinginfo website (Drillinginfo 1998). Various operators producing oil and gas reported their monthly production data to state agencies over time. These data were then checked for gross abnormalities, and corrections were made as needed, whereupon it was published. Our criteria for the inclusion of Permian Basin wells consisted of wells drilled horizontally with single stage or multistage fracturing. The minimum required production time for each well was set to six and a half years, or 78 months. Wells were chosen after the year 2003 and until the year 2018. Among the selected wells, some of the wells had been restimulated. For such wells, the portion of the production history after the restimulation was considered for the analysis, provided it satisfied the minimum production time criteria (at least 78 months of production).

Test Scenario. To identify the best model, all the probabilistic models are assessed based on a prediction error comparison, CR comparison, and an interval width comparison (bound interval) for each well production history data set using the same set of priors (Table 1). The prediction error is calculated using Eq. 5 below and represents the scaled difference between the actual cumulative production at the end of the production history and the prediction at the same time.

$$\text{Prediction error} = \frac{|\text{P50} - \text{Actual}|}{\text{Actual}}. \quad (5)$$

The CR represents the percentage of wells analyzed that have their actual production within the predicted P10 to P90 intervals. The interval width comparison represents the width of the predicted P10 to P90 interval. Ideally, it is not only desirable to have the actual production within the predicted P10 to P90 interval in approximately 80% of the cases but also to have a narrowest possible P10 to P90 interval that can be supported based on the data, DCA model, and the Bayesian model-fitting algorithms.

Table 3 shows the well statistics for oil and gas wells chosen from different counties of the Permian Basin. The location of these wells is also shown on the map in **Fig. 13.** Results for varying lengths of production history (6, 9, 12, 15, 30, 45, and 60 months) for the Delaware Basin, Midland Basin, and Central Basin Platform are represented in **Figs. 14 through 18.**

Primary Basin	Subbasin	County	No. Wells	Production Type
Permian	Delaware	Ward	22	Gas
	Midland	Midland	19	Gas
		Terry	14	Oil
		Upton	9	Oil
	Central Platform	Ward	10	Gas

Table 3—Well statistics.

Model Performance. Fig. 14 represents the results for the prediction error comparison (top row), CR comparison (middle row), and interval width comparison (bottom row) for Gibbs, Metropolis, and ABC sampling, respectively, in the Delaware Basin.

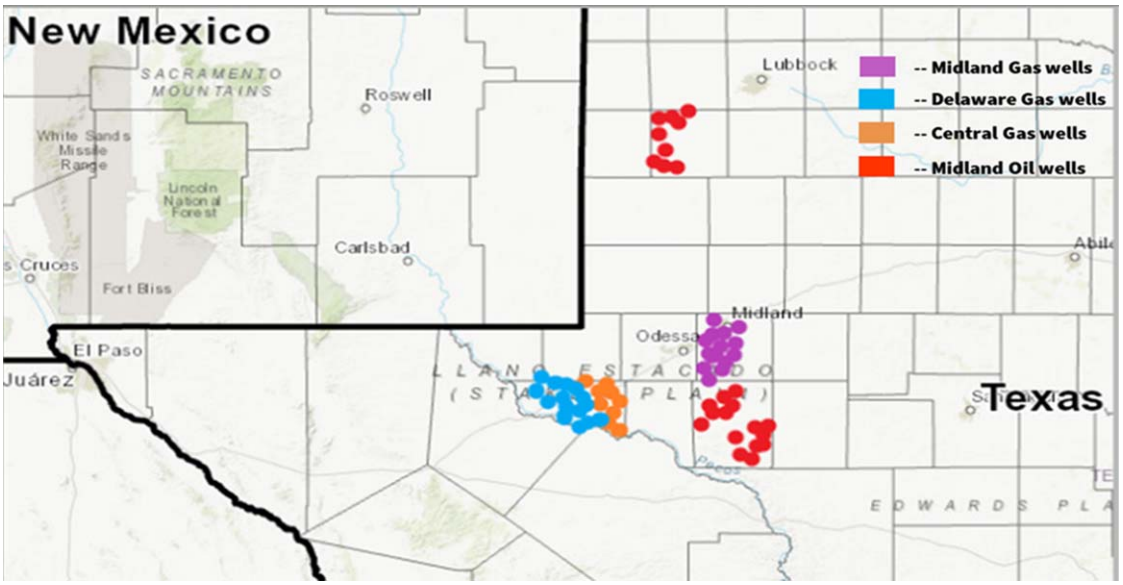

Fig. 13—Location of wells in the Permian Basin.

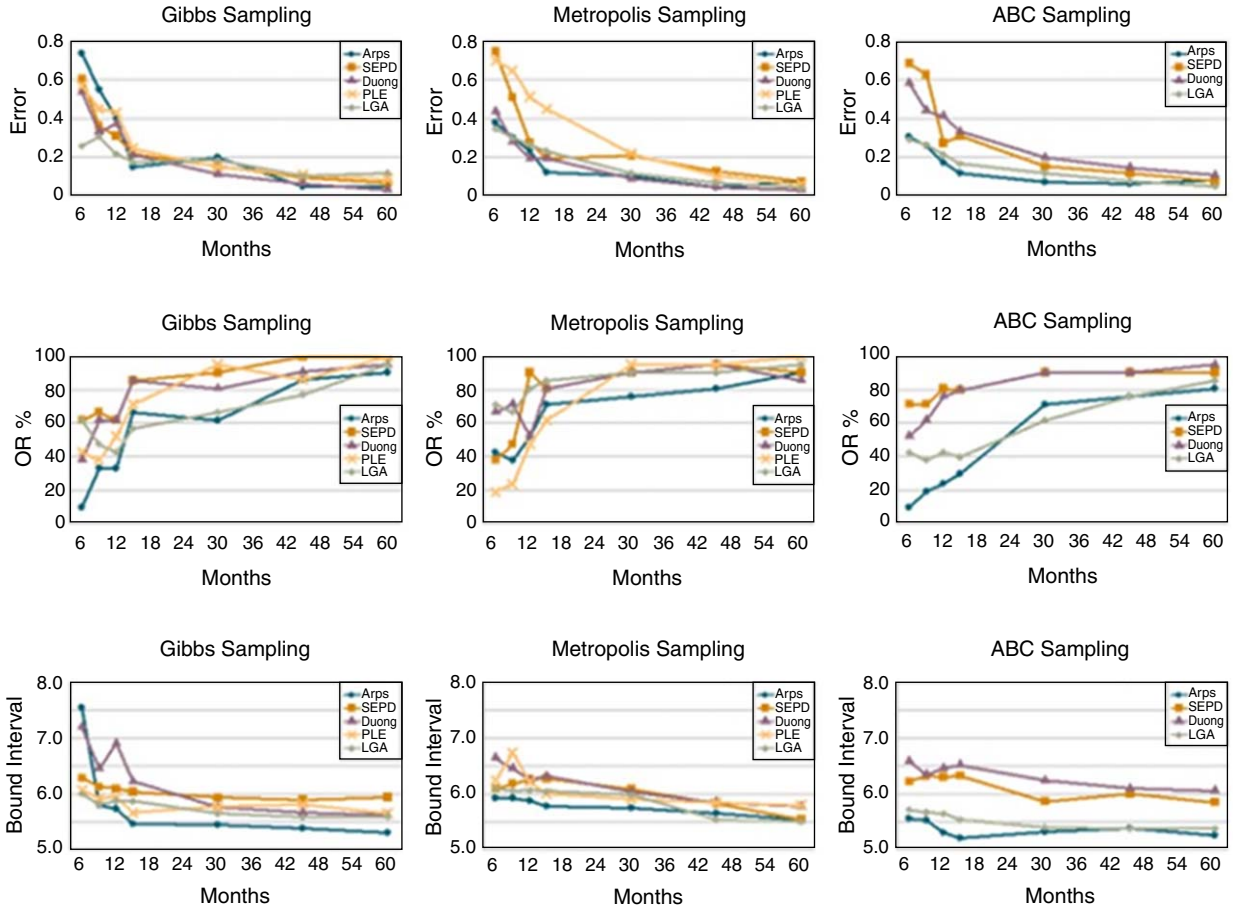

Fig. 14—Errors plots (top row), CR plots (middle row), and bound interval plots (bottom row) as a function of production history using well data from the Delaware Basin.

Fig. 15 represents the results for the prediction error comparison (top row), CR comparison (middle row), and interval width comparison (bottom row) for Gibbs, Metropolis, and ABC sampling, respectively, in the Midland Basin (gas wells).

Fig. 16 represents the results for the prediction error comparison (top row), CR comparison (middle row), and interval width comparison (bottom row) for Gibbs, Metropolis, and ABC sampling, respectively, in the Midland Basin (oil wells).

Fig. 17 represents the results for the prediction error comparison (top row), CR comparison (middle row), and interval width comparison (bottom row) for Gibbs, Metropolis, and ABC sampling, respectively, in the Central Basin platform.

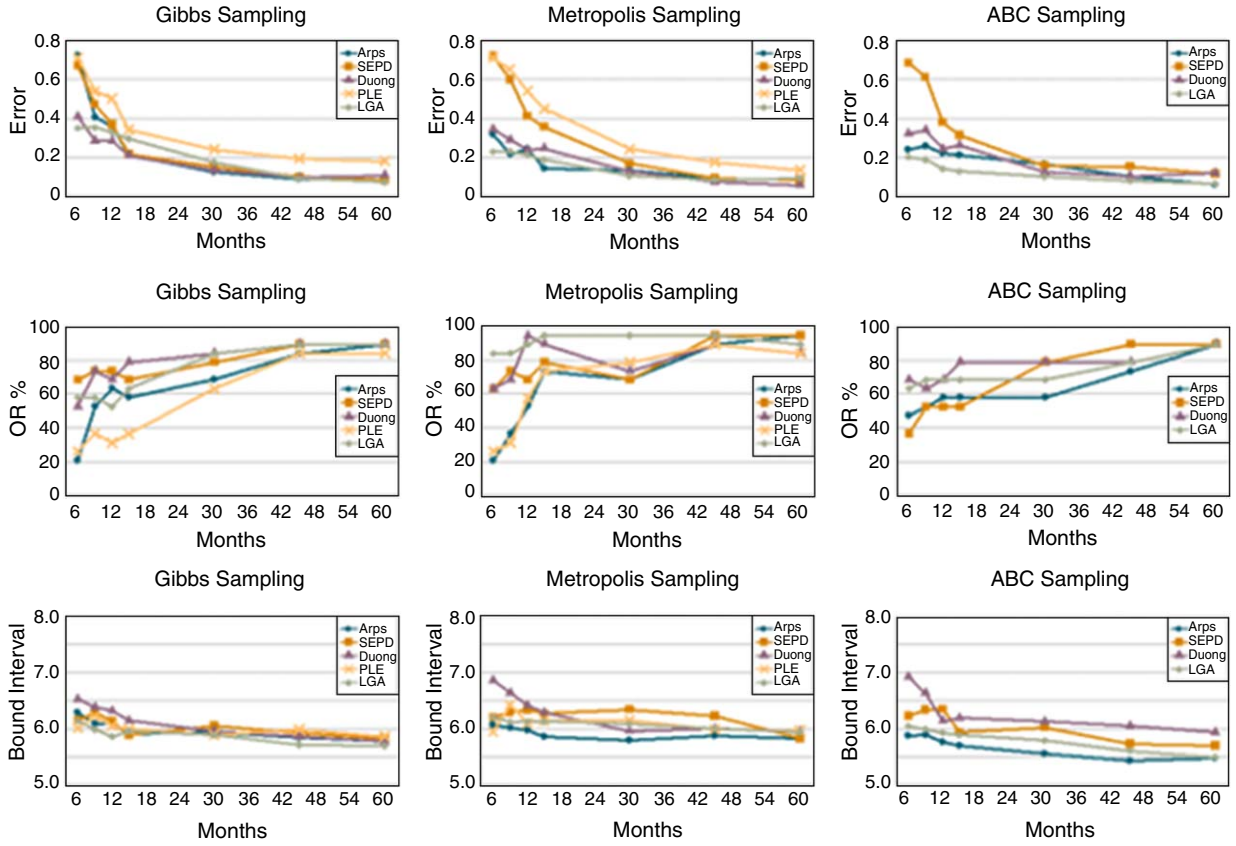

Fig. 15—Errors plots (top row), CR plots (middle row), and bound interval plots (bottom row) as a function of production history using well data from the Midland Basin (gas wells).

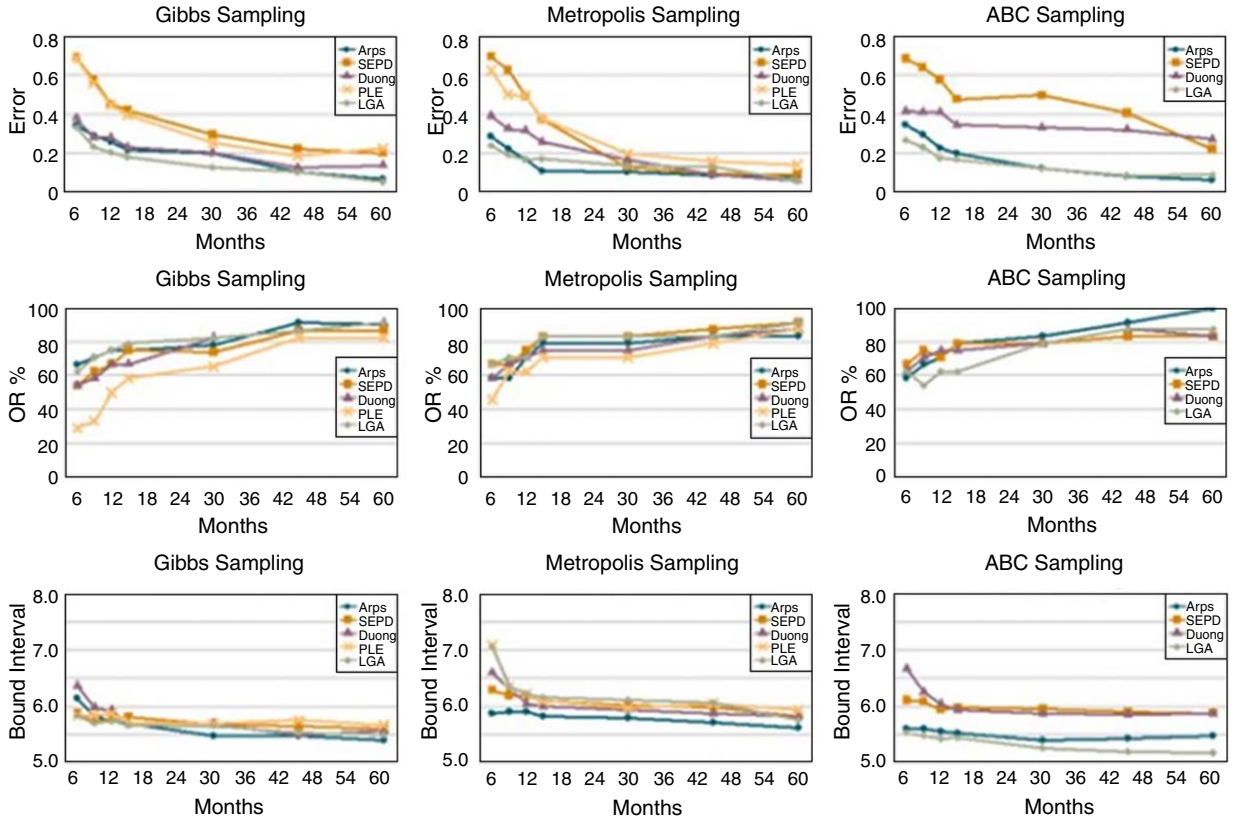

Fig. 16—Errors plots (top row), CR plots (middle row), and bound interval plots (bottom row) as a function of production history using well data from the Midland Basin (oil wells).

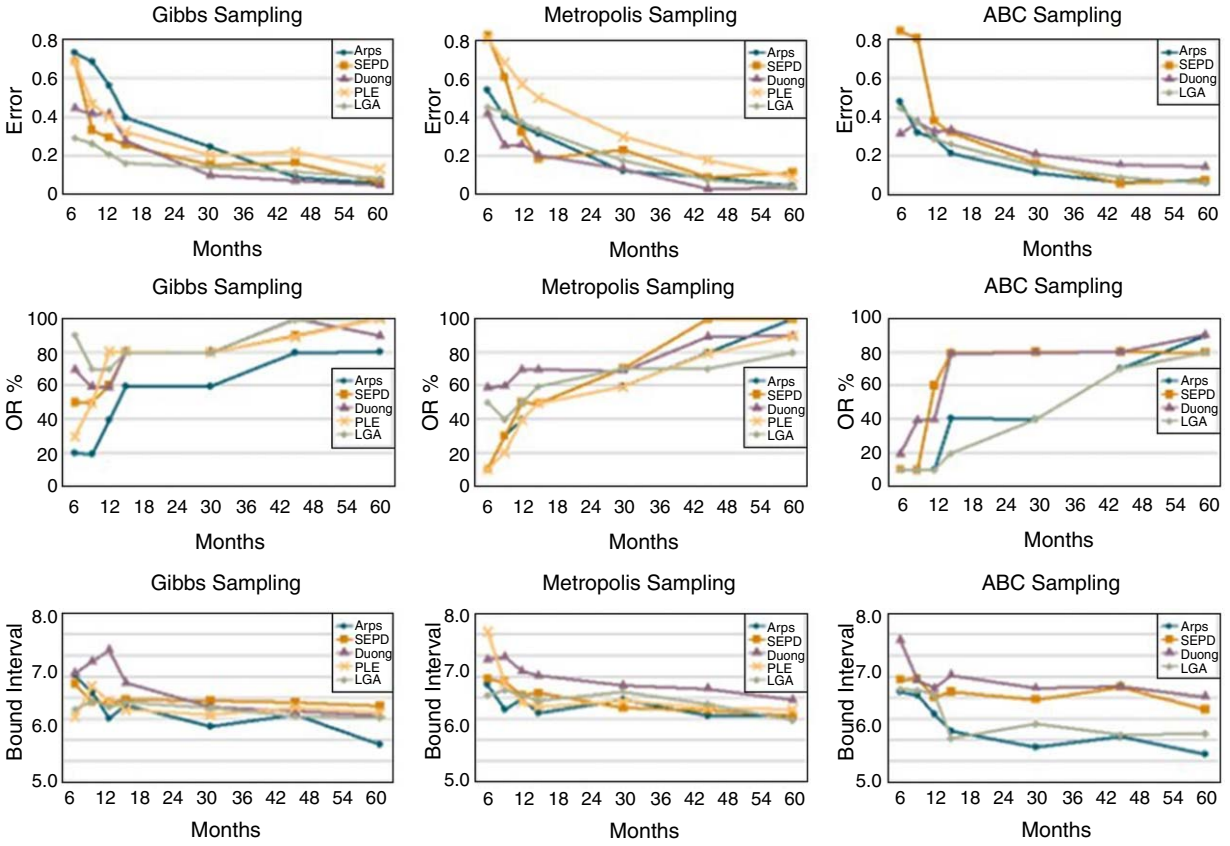

Fig. 17—Errors plots (top row), CR plots (middle row), and bound interval plots (bottom row) as a function of production history using well data from the Central Basin Platform.

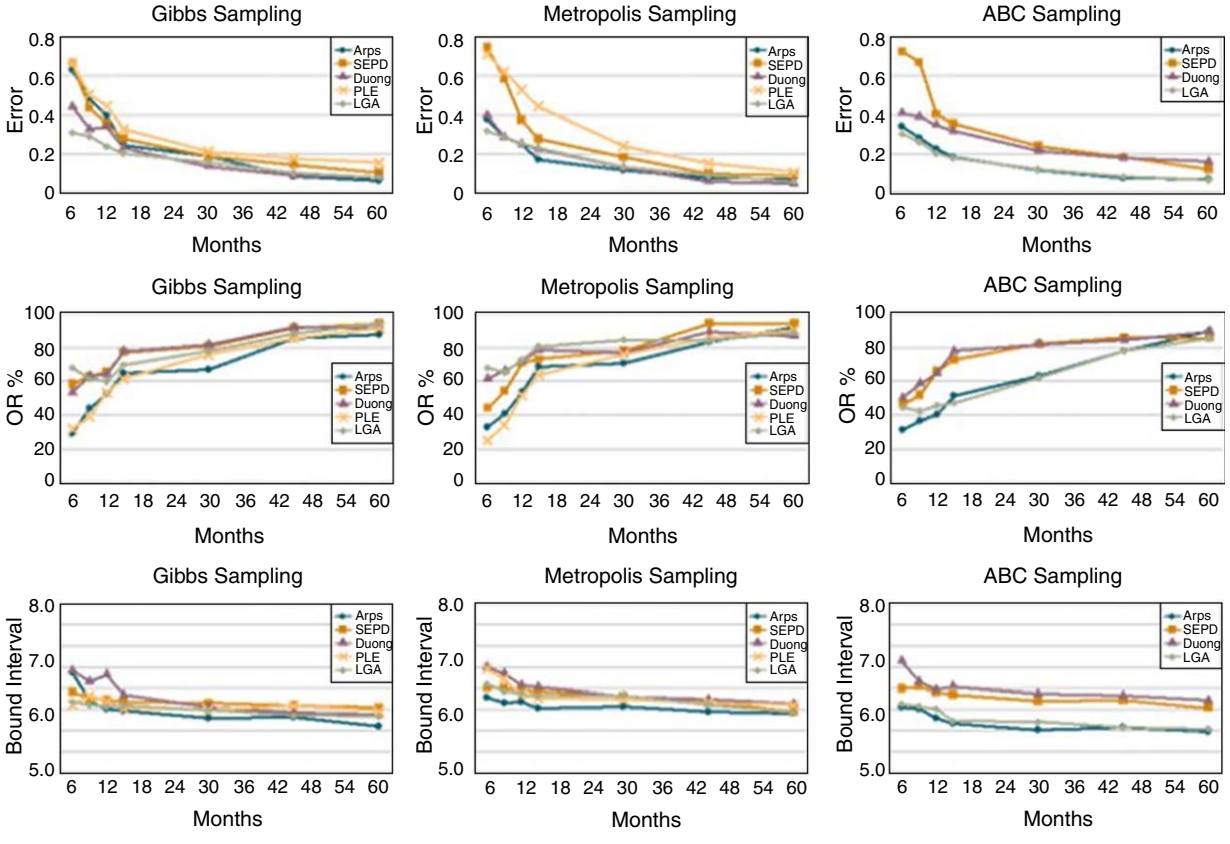

Fig. 18—Errors plots (top row), CR plots (middle row), and bound interval plots (bottom row) as a function of production history using well data from the overall Permian Basin.

Fig. 18 represents the results for the prediction error comparison (top row), CR comparison (middle row), and interval width comparison (bottom row) for Gibbs, Metropolis, and ABC sampling, respectively, in the Permian Basin overall, averaged over the subbasins plotted in Figs. 14 through 17.

The findings shown in Figs. 14 through 18 reflect results obtained from models fit to the current data studied. Their reliability and generalizability could only be ascertained by repeating a similar analysis on additional data. For ease of use, the findings of these figures have been summarized by three linear regression models that fit a response (prediction error, CR, and bound interval width) to the factor variables model, algorithm, and Basin as well as the continuous covariate time (meaning length of time in the hindcast). **Fig. 19** displays the effect plots as produced by the R "effects" package (Fox et al. 2020) for these predictors and using the prediction error response. The linear model was fit with an interaction between model and algorithm due to the interaction's strong significance. *P*-values of the analysis of variance test of each predictor's significance are shown above each plot. For the two interacting factors, the relevant factor's significance was tested at each level of the other factor, resulting in multiple *p*-values. We caution that these *p*-values are meant to illustrate the relatively dramatic nature of (most of) the factors' significance, which is supported by the trends seen in the earlier figures. The *p*-values should not be taken to perfectly accurately quantify the uncertainty with which decisions regarding these factors' significance can be made because they are calculated on response data (e.g., prediction errors), which are themselves outputted from fitted models. In particular, the same data sets were used multiple times in the earlier models, likely leaving correlation between the errors in the later models, which has not been accounted for.

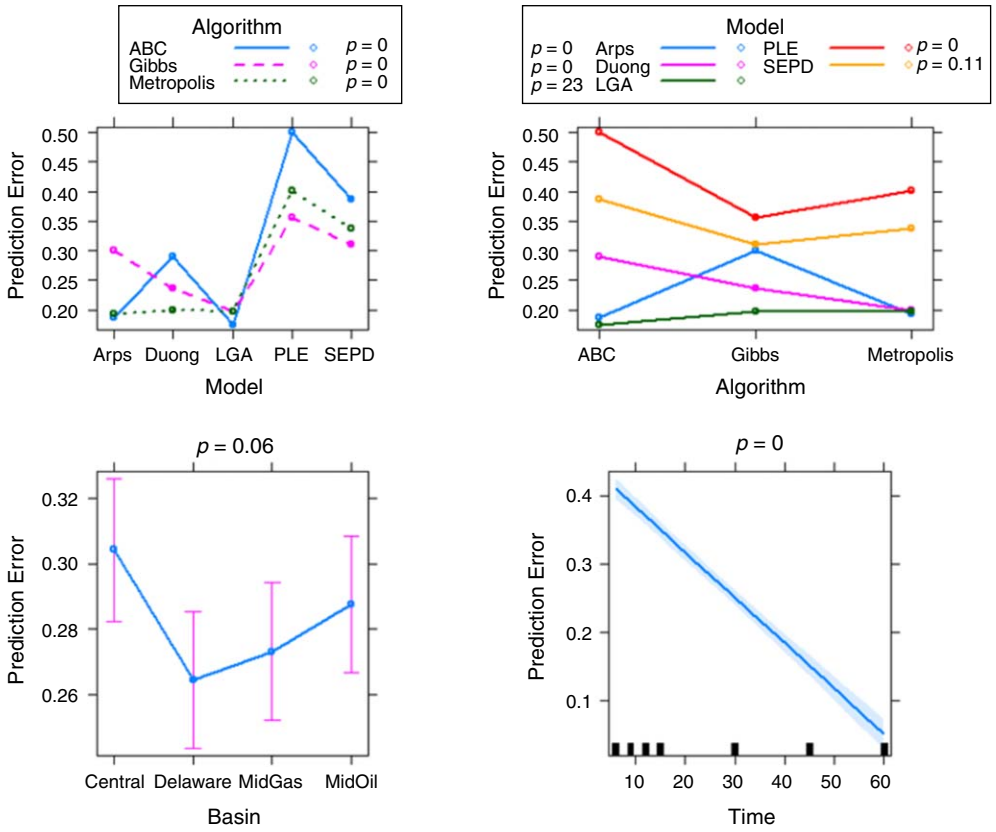

Fig. 19—Effect plots from a linear model fit using the prediction errors from Figs. 14–18 as the response and predictors model, algorithm, basin, and time. *P*-values listed above each plot test the hypothesis that the relevant term belongs in the model. In the upper two panes, the relevant factors were tested at each level of the other factor, owing to a significant interaction between them.

The takeaway message from Fig. 19 is that an engineer who desires to minimize prediction error would need to base the selection of a Bayesian model-fitting algorithm on the DCA model selected (or vice-versa). Despite this, the LGA model performs well for all algorithms. The same can be said of the Gibbs algorithm for nearly all models. Further, the basin to which well data belong appears to somewhat influence the prediction error (prediction errors are highest in the Central Basin), as certainly the hindcast data length does. These same conclusions apply to the effect plots (not shown here to reduce paper length) that result from the linear models fit to response variables CR and bound interval width. All the above conclusions relate specifically to our data set and will need to be tested in further studies with additional data.

The convergence of the samplers for each of the underlying model parameters can tested using the diagnostic tools discussed previously. A Gelman and Rubin plot and a trace plot for a sample well from the Delaware Basin (Arps-ABC model) are shown in **Figs. 20 and 21,** respectively. For Fig. 20, the MPSRF value generated for the Gelman and Rubin diagram is calculated as 1.01. The trace plot in Fig. 21 consists of two chains running in parallel for each of the model parameters of the Arps-ABC model. These chains show good parameter space exploration, good mixing, and a mean around their individual centers. Thus, both diagnostic tools confirm evidence against nonconvergence.

Table 4 provides the probabilistic model that performs best under each assessment criterion, broken down by basin. Comparing the satisfactory (12 months) and excellent (45 months) results, the probabilistic model producing the best-fit model predictions varies for every production history data set. Likewise, the best-fit probabilistic model also varies when additional historical production data are available. Thus, it appears prudent to initially apply and assess all probabilistic models during reserves estimation and uncertainty

quantification of a given production history data set. As time progresses and forecasts are compared against experience, future forecasts may increasingly rely upon whichever probabilistic model has been ascertained to perform best in the case. Using all probabilistic models at the outset can also provide assurance of the results when they agree, thereby increasing users' confidence in the reliability of the method.

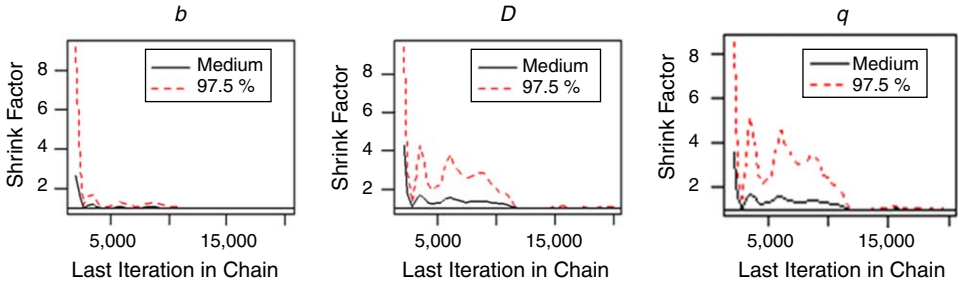

Fig. 20—Sample Gelman and Rubin plot for a Delaware Basin well with MPSRF value 1.01.

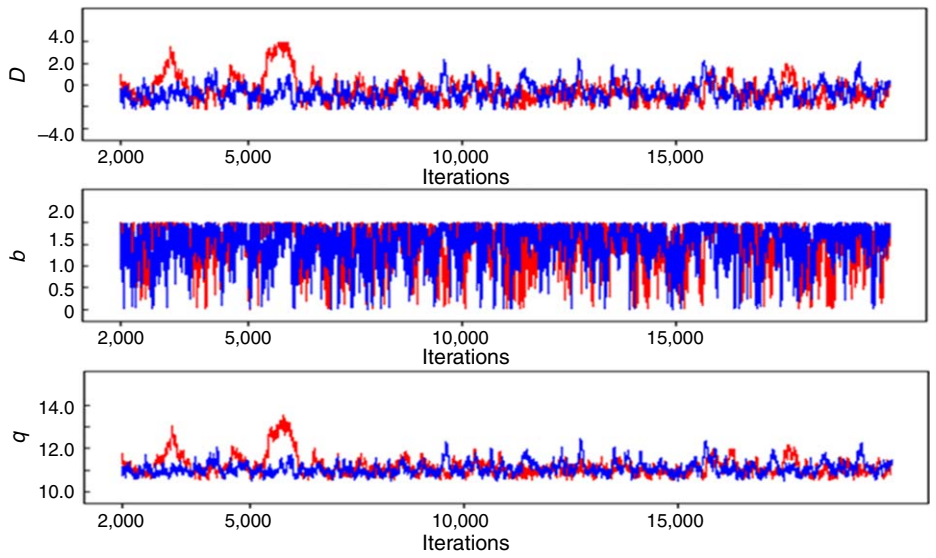

Fig. 21—Sample trace plot for a Delaware Basin well.

		12-Month Production History			45-Month Production History		
Basin	Figures	Prediction Error	Coverage Rate	Interval Width	Prediction Error	Coverage Rate	Interval Width
Delaware gas wells	14	17.4% (Arps-ABC)	80.95% (LGA-MH; SEPD-ABC)	5.65 (LGA-ABC)	4.2% (Arps-MH)	80.95% (Arps-MH)	5.38 (Arps-Gibbs)
Midland gas wells	15	14.34% (SEPD-Gibbs)	73.68% (LGA-MH)	5.65 (Arps-ABC)	7.5% (Duong-MH)	78.94% (Duong-ABC; LGA-ABC)	5.44 (Arps-ABC)
Midland oil wells	16	16.89% (Arps-MH)	83.33% (PLE-ABC)	4.42 (LGA-ABC)	8% (Arps-ABC)	79.17% (PLE-MH)	4.20 (LGA-ABC)
Central gas wells	17	69.89% (LGA-Gibbs)	80.50% (PLE-Gibbs)	5.51 (Arps-Gibbs)	3.6% (Duong-MH)	79.93% (Arps-Gibbs)	5.09 (Arps-ABC)
Overall Permian	18	20.58% (LGA-ABC)	80.50% (PLE-Gibbs)	5.31 (Arps-ABC)	7.6% (Arps-MH)	79.93% (Arps-Gibbs)	5.08 (Arps-ABC)

Table 4—Summary of best model performance statistics by basin. Models that appear repeatedly have been consistently colored for effect: Arps-ABC, LGA-ABC, Duong-MH, Arps-Gibbs, Arps-MH, LGA-MH, PLE-Gibbs.

It is also important to realize that each of the probabilistic models perform better (i.e., lowered prediction error, better-calibrated CR, and lowered the interval width) as more production history data become available. Table 4 demonstrates that it is possible, and even perhaps likely, that whichever probabilistic model performs best for a short hindcast length may be outperformed by another for a longer hindcast horizon. This, again, suggests that using all possible probabilistic models from the beginning, and perhaps even later in a reservoir's lifespan, may be judicious.

The CRs that exceed the prediction intervals' nominal coverage rate of 80% reflect the noninformative nature of the priors that we have selected. These noninformative priors have a diffusive effect on the posterior distributions, effectively widening their posterior predictive intervals. As a result, CRs are high and interval widths are high. If the analyses were repeated with more informative priors, the interval widths would naturally shrink. CRs would also shrink toward 80%, as long as the informative priors were representative of the distributions of the parameters over repeated wells. If they were not, they would introduce bias into the posteriors, and the CRs would drop below 80%. The takeaway message from this is that we have, in this study, achieved strong CRs at the expense of not having narrow prediction intervals, and this tradeoff is modulated through the choice of priors.

Computation time required for each of the probabilistic models to generate the well prediction varies depending on the production history data under analysis. Average time per run required for each model is between 5 and 25 seconds.

Other Considerations/Limitations of Work

The chief limitation of this work is that the integrated probabilistic models do not consider the physics of the process. Physics-based models are complex, computationally expensive, and require data that are not easily available. However, if physics-based models are available and are formed based on certain important parameters, they can be incorporated in the above methodology. The PLE-ABC probabilistic model does not perform well with the distance measures we have selected here for ABC. The choice of distance measure is somewhat arbitrary because there is no principled way to select it. Plausible solutions for achieving better PLE-ABC model performance might include experimenting with other distance measures or decreasing the set threshold percentage for the accepted sample sets.

Future work may include the development of a software tool for accessing the probabilistic models described in this study and for identifying the best probabilistic model, depending on the production history data available. An exploration into the use of more informative priors might also increase the accuracy of the methods' uncertainty quantification.

Conclusions

We summarize our conclusions as follows:

- The three MCMC sampling techniques can be integrated with modern deterministic models of conventional and unconventional oil and gas wells to yield reliable cumulative production forecasts and uncertainty quantifications with low available historic production data.
- Notably, for the overall Permian Basin, the methodology achieved a CR of between 80 and 100% when 45 and 60 months of production history were available. Likewise, it achieved between 5 and 18% prediction errors when 45 and 60 months of production data were available.
- Based on the data analyzed in this study, LGA is best in terms of prediction errors for all algorithms except MH. The Gibbs algorithm is nearly the best algorithm in terms of prediction error for all DCA models except Arps. Prediction errors are the highest in the Central Basin platform.
- The hindcast length is highly significant and steeply influential in determining the prediction error, CR, and interval width of a model's predictions.
- The practice of fitting multiple probabilistic models is recommended because it allows users to have greater confidence in the model predictions if the various models' predictions concord.
- Each probabilistic model performs better (by reducing prediction error, better calibrating CR, and reducing the interval width) as more historic production data become available.
- Each of the probabilistic models require a low computation time of around 5 to 25 seconds, depending on the available hindcast data.

Nomenclature

Actual = true cumulative production
B/n = variance between chains
D_∞ = power law decline at infinite time in PLE
k = carrying capacity (STB or MCF)
m = dimensionless slope parameter in Duong's model
n = dimensionless time exponent
$P(\theta)$ = prior distribution
$P(Y|\theta)$ = likelihood function
$P(\theta|Y)$ = posterior distribution
q_i = initial production
q_t = production at time t
t = time
W = mean variance within chains
Y = observed data (monthly production)
θ_{proposal} = parameters drawn from proposal distribution

List of Symbols

$\hat{\text{D}}$ = power law decline constant in PLE
f = density function
α = intercept of Duong's model
ε = difference between simulated results and observed data
η = dimensionless exponent parameter in SEPD

θ = model parameters
σ^2 = variance
τ = characteristic time parameter in SEPD

References

Cheng, Y., Wang, Y., McVay, D. et al. 2010. Practical Application of a Probabilistic Approach to Estimate Reserves using Production Decline Data. *SPE Econ & Mgmt* **2** (1): 19–31. SPE-95974-PA. https://doi.org/10.2118/95974-PA.

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

Csillery, K., Francois, O., and Blum, M. G. 2012. abc: An R Package for Approximate Bayesian Computation (ABC). *Meth Ecol Evol* **3** (3): 475–479. https://doi.org/10.1111/j.2041-210X.2011.00179.x.

Drillinginfo. 1998. Drillinginfo https://app.drillinginfo.com/ (accessed 1 August 2017).

Duong, A. N. 2011. Rate-Decline Analysis for Fracture-Dominated Shale Reservoirs. *SPE Res Eval & Eng* **14** (3): 377–387. SPE-137748-PA. https://doi.org/10.2118/137748-PA.

EIA (U.S. Energy Information Administration). 2018. Permian Basin Wolfcamp Shale Play Geology Review, https://www.eia.gov/maps/pdf/PermianBasin_Wolfcamp_EIAReport_Oct2018.pdf (accessed 22 March 2019).

Fox, J., Weisberg, S., Price, B. et al. 2020. Package "Effects." https://cran.r-project.org/web/packages/effects/effects.pdf.

Gelman, A. and Rubin, D. B. 1992. Inference from Iterative Simulation Using Multiple Sequences. *Stat Sci* **7** (4): 457–472. https://doi.org/10.1214/ss/1177011136.

Gong, X., Gonzalez, R., McVay, D. A. et al. 2011. Bayesian Probabilistic Decline-Curve Analysis Reliably Quantifies Uncertainty in Shale-Well-Production Forecasts. Paper 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. et al. 2014. Bayesian Probabilistic Decline-Curve Analysis Reliably Quantifies Uncertainty in Shale-Well-Production Forecasts. *SPE J*. **19** (6): 1047–1057. SPE-147588-PA. https://doi.org/10.2118/147588-PA.

Gonzalez, R. A., Gong, X., and McVay, D. A. 2012. Probabilistic Decline Curve Analysis Reliably Quantifies Uncertainty in Shale Gas Reserves Regardless of Stage of Depletion. Paper presented at the SPE Eastern Regional Meeting, Lexington, Kentucky, USA, 3–5 October. SPE-161300-MS. https://doi.org/10.2118/161300-MS.

Ilk, D., Rushing, J. A., Perego, A. D. et al. 2008. Exponential vs. Hyperbolic Decline in Tight Gas Sands: Understanding the Origin and Implications for Reserves Estimates using Arps Decline Curves. Paper presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, USA, 21–24 September. SPE-116731-MS. https://doi.org/10.2118/116731-MS.

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

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

Lunn, D., Spiegelhalter, D., Thomas, A. et al. 2009. The BUGS Project: Evolution, Critique, and Future Directions (with Discussion). *Stat Med* **28** (25): 3049–3082. https://doi.org/10.1002/sim.3680.

Paryani, M. 2015. *Approximate Bayesian Computation for Probabilistic Decline Curve Analysis in Unconventional Reservoirs*. MS thesis, University of Alaska, Fairbanks, Alaska, USA (December 2015).

Paryani, M., Awoleke, O. O., Ahmadi, M. et al. 2017. Approximate Bayesian Computation for Probabilistic Decline-Curve Analysis in Unconventional Reservoirs. *SPE Res Eval & Eng* **20** (2): 478–485. SPE-183650-PA. https://doi.org/10.2118/183650-PA.

Plummer, M., Best, N., Vines, K. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* **6**: 7–11. https://www.r-project.org/doc/Rnews/Rnews_2006-1.pdf.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Tarka Resources, Inc. 2019. Permian Basin Oil Exploration and Operations. https://tarka.com/permian-basin/ (accessed 1 March 2019).

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

Ward, R. F., Kendall, C. G., and Harris, P. M. 1986. Upper Permian (Guadalupian) Facies and Their Association with Hydrocarbons—Permian Basin, West Texas, and New Mexico. *AAPG Bull* **70**: 239–262. https://doi.org/10.1306/9488566F-1704-11D7-8645000102C1865D.