CrossMark

Original Paper

Statistical Detection of Flow Regime Changes in Horizontal Hydraulically Fractured Bakken Oil Wells

E. D. Attanasi,[1,3] T. C. Coburn,[2] and B. Ran-McDonald[2]

The application of horizontal and hydraulically fractured wells for producing oil from low permeability formations has changed the face of the North American oil industry. One feature of the production profile of many such wells is a transition from transient linear oil flow to boundary-dominated flow. The identification of the time of this transition is important for the calibration of models that forecast the well's future production and the expected ultimate recovery. It is preferable that such models generally use data from the boundary-dominated flow regime for parameter calibration. Accurate well production forecasts are needed for operational decisions, long-term planning, commercial transactions, regulatory proceedings, and asset valuation. Petroleum engineers frequently make the call on the transition point based on subjective visual interpretations of log–log plots for individual wells. This is time-consuming and is generally not repeatable by other analysts. This note evaluates statistical approaches that can serve as alternatives to the subjective visual interpretations. Specifically, the predictive performance of production models calibrated with boundary-dominated data based on transition dates calculated with constrained non-linear least squares and Bayesian regressions was very close to that obtained using the visual method, suggesting that statistical approaches may indeed be constructed to replace less objective visual approaches without loss of accuracy.

KEY WORDS: Linear transient flow, Boundary-dominated flow, Flow regime transition, Bayesian regression, Constrained least squares.

INTRODUCTION

The application of horizontal and hydraulically fractured wells to produce oil from low permeability formations has changed the face of the North American oil industry. For horizontal hydraulically fractured Bakken wells, production commonly follows two general flow regimes (Kabir et al. 2011). Early in the life of the well, production is dominated by linear flow as the reservoir fluids flow through the induced fractures. Transient linear flow is assumed to characterize infinite fracture conductivity with no "skin" effect, i.e., formation damage that impedes flow to the wellbore. After a period of transient flow, a new flow regime, boundary-dominated flow, begins as the flow reaches the boundaries (extent of the well's drainage area) where a constant reservoir pressure can no longer be sustained for the fluids driven through the fractures. The transient production phase may last up to 3 years, but the boundary-dominated flow regime characterizes the well's production behavior for the remainder of the well's life.

Seidle and O'Connor (2016) recently presented case studies where three production decline models

---
[1]U.S. Geological Survey, 12201 Sunrise Valley Dr., MS 956, Reston, VA 20192, USA.
[2]University of Tulsa, Tulsa, OK 74104, USA.
[3]To whom correspondence should be addressed; e-mail: tim-coburn@utulsa.edu

that are commonly applied to forecasting fractured horizontal well production were calibrated with transient period data to predict future production. In previous studies, these three techniques—the modified Arps model (Long and Davis 1988), the Duong model (Duong 2011), and the stretched exponential model (Yu et al. 2013), have been found to perform reasonably well when data from the boundary-dominated production phase are used for calibration. In their conclusions, Seidle and O'Connor (2016) advise caution in the use of transient phase data for calibration of these models because of the inaccurate forecasts and estimated ultimate recovery (EUR) predictions. This leads to the question of how to best detect the change in the well's flow regimes.

Anderson and Mattar (2003) demonstrate a graphical method using a log–log plot of production versus time to identify well flow regimes. Figure 1 shows a variation of the log–log plot that captures the flow regime change where only well production data are available. The graph shows the inverse of daily production (ordinate) versus time in number of production days (abscissa). When plotted on log–log paper, during the transient flow period the graph follows a linear trend with a slope of ½. At the transition from linear to boundary-dominated flow, the slope of the trend line increases to 1.0.

In applied situations the analyst may identify the transition point visually by examining such plots for each individual well in the suite of wells for which production predictions or EURs are required. Once the transition date is identified, analysts use the production data from the boundary-dominated flow phase to calibrate a decline function in order to predict production or well recovery. This approach assumes the transition from transient to boundary-dominated (pseudo-steady state) flow is characterized by an abrupt change. However, such an assumption oversimplifies the physical situation. For a perfectly circular reservoir, the transition between transient and pseudo-steady state may, indeed, be a near-discrete step change; but for reservoirs having other shapes, the transition between flow regimes may take longer (because more time is necessary to reach all boundaries before boundary-dominated flow takes over) resulting in a more gradual, nearly continuous conversion. Because of the difficulty in modeling such a continuous change, software used by engineers to pick the change point (e.g., Harmony (IHSMarkit, n.d.)) typically assumes the "abrupt change" formulation (Kelkar 2018).

Visual identification of the transition point (henceforth referred to as the visual method) is subjective and may be different from one analyst to the next, a situation which can contribute unnecessarily to the variability among well recovery values for a suite of wells. Hence, a more rigorous approach is needed that alleviates the subjectivity and is verifiable over time. This note reports on the performance of several statistical procedures applied to the detection of flow regime changes that may lead to more consistent, reliable, and repeatable results.

STATISTICAL METHODS

The problem of detecting a change (or break-point) in an underlying data generating process is common to many business and operational settings, including the identification of changes in financial market trends, shifts in product manufacturing quality, and monitoring of hospital emergency admissions for early identification of epidemics. The change point detection literature is vast. However, the present setting has a physically based conceptual structure that is absent in other applications. The graphical characterization of the problem is shown in Figure 1, which provides a framework for the analysis. Embedded in that framework is the assumption (previously noted) that in log space the inverse of daily production versus the number of days during the transient period follows the linear path with a slope of 1/2 and that immediately following the change to boundary-dominated flow the inverse of production follows a linear path (in log space) with a slope of 1.0 (Kabir et al. 2011).

In this paper three different change point regression approaches are proposed for purposes of detecting the change in flow regimes. The presentation of these techniques reflects an increasing degree of structure in the methodology. The first approach is an application of piecewise linear regression (Toms and Lesperance 2003). A useful tutorial on piecewise linear regression, also called segmented or 'broken stick' regression, is given by Ryan and Porth (2007). The model is specified as

$$y_i = \begin{cases} \beta_o + \beta_1 x_i + e_i & x_i \le t^* \\ \beta_o + \beta_1 x_i + \beta_2(x_i - t^*) + e_i & x_i > t^* \end{cases} \quad (1)$$

where y_i is the $\log_{10}$ of the inverse of the daily flow, x_i is the $\log_{10}$ of the production day number (see Fig. 1) and t^* is the $\log_{10}$ of first day of boundary-
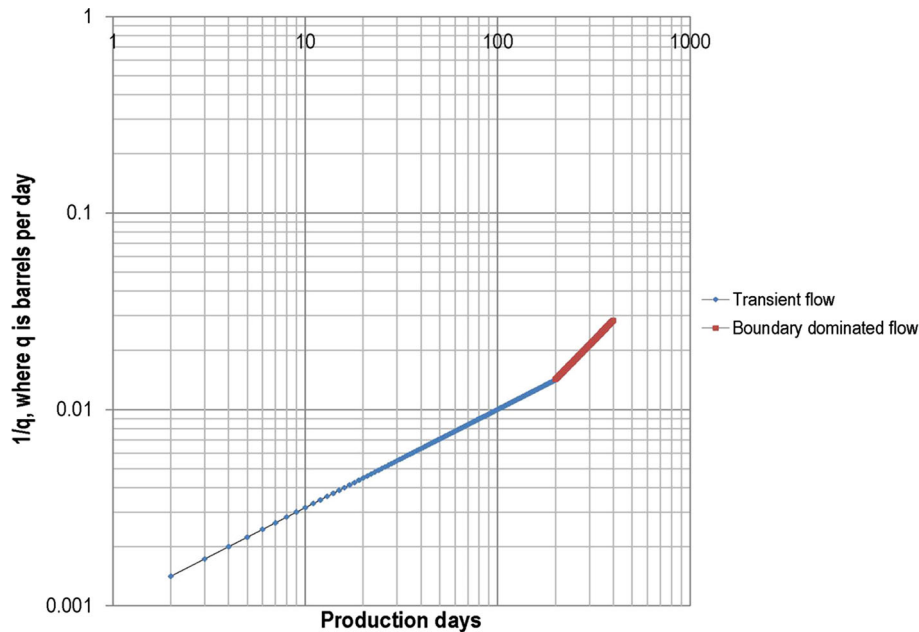
Figure 1. Schematic of the change in oil well flow regimes for daily production (q) from transient followed by boundary-dominated flow at 200 production days. The graph is in log–log scale, and the label that identifies the particular day of production is on top of the graph..

dominated flow. The calculations involve estimation of the spline function of degree 1 and 1 knot (Toms and Lesperance 2003). No constraints are placed on the parameters. The relationship between the inverse of daily production and days is linear (Fig. 1) in log space, so values of the regression variables are stated in logarithms.

A second approach is to estimate t^* as in the first formulation but to impose restrictions on the feasible values of the parameters for β_1 and β_2 that constrain their values to roughly conform to the production regimes presented in Figure 1. One possible set of these restrictions could be to constrain the values of β_1 and β_2 within some interval of 0.45–0.55. The imposition of such constraints requires that a nonlinear algorithm be used to estimate the parameter values, including the estimate of t^*. The solution to the resulting nonlinear regression depends on the autocorrelation structure of the production series so that the variance–covariance matrix is invertible.

The third technique is based on the approach developed by Bacon and Watts (1971) in which Bayesian regression is used to estimate the statistical distribution of the uncertain change point. Carlin and others (1992) facilitated implementation of the Bayesian approach by using a Gibbs sampler to compute parameter posterior density functions, thus

avoiding any need for the higher dimensional integration procedures. The Gibbs sampler is a Markov chain Monte Carlo (MCMC) method for calculating a sequence of observations which are approximated from a specified multivariate distribution and can be used to approximate a joint distribution, a marginal distribution of one of the variables or the distribution of a subset of variables (Casella and George 1992). The change point problem in this case is specified in (2) where the change point is t^*. This formulation follows Spiegelhalter et al. (1996):

$$y_i \sim \begin{cases} \text{Normal}\big(\theta_0 + \theta_1(x_i-t^*), \sigma^2\big) & \text{if } x_i < t^* \\ \text{Normal}\big(\theta_0 + \theta_2(x_i-t^*), \sigma^2\big) & \text{if } x_i \geq t^* \end{cases} \quad (2)$$

For the present investigation the following priors are used for illustrative purposes:

$$t^* \sim \text{Uniform}(a, b);$$
$$\theta_1 \sim \text{Uniform}(0.45, 0.55);$$
$$\theta_2 \sim \text{Uniform}(0.90, 1.10);$$
$$\theta_0 \sim \text{Normal}(0, v = 100);$$
$$\sigma^2 \sim \text{Uniform}(0, 1000);$$

The variables y_i and x_i and the parameter t^* are defined as in (1), and distributions for the parameters t^* and θ_i are defined in (2). The prior density

function on the change point (t^*) assumes a uniform distribution where the endpoints, a and b, are expressed in $\log_{10}$ of days. For the current illustration the original values of a and b are taken to be 100 and 1000 days in real space, respectively. In theory, if one has the point estimates of the reservoir parameters plus information about well spacing, then one can estimate parameters for a prior distribution for t^* (Kabir et al. 2011; Ran 2016). The implementation of the Bacon–Watts approach presented here, along with the associated computations, follows the approach of the SAS Institute (2016).

In summary, the only restriction on the piecewise linear (spline-type) function (Method 1) is linearity and that the linear segments meet at a single point. For Method 2, the nonlinear least squares places constraints on the slope coefficients; and for the Bacon–Watts framework (Method 3), prior distributions are placed on the slope and intercept coefficients, as well as on t^*, the uncertain change point.

SIMULATION STUDIES OF CHANGE POINT DETECTION METHODS

The synthetic data displayed in Figure 1 represent a 300-day daily production stream from a fictitious well that exhibits a transient production phase during the first 200 days. At that point (the change point) production switches to a boundary-dominated production phase. Using data from Figure 1 for the initial 300 days of production, all three methods described above return the exact 200-day change point. To examine the robustness of the methods, random variations in daily production were simulated and added to the existing synthetic data stream, as described below, to represent a more realistic production setting. The three methods were then subsequently applied to this "noisy" data.

Three sets of 100 production stream simulations were prepared, each representing an increased weight of the random component. Suppose p_i is daily production and is composed of deterministic and random components as indicated in (3) below; i.e., p_{id} is the deterministic component of daily production, p_{icr} is a random variable (rv) increment (either positive or negative) of daily production, and fac represents the increasing noise levels at 5, 10, and 20% of the deterministic daily production volume. The noisy production streams were generated as follows:

$$p_i = p_{id} + rv(p_{icr}) \text{ with } p_{icr} \sim \text{Normal}(0, \text{fac} * p_{id})$$

with a different production stream for each noise level (i.e., value of fac). Equation 3 shows that the noise for each production day was generated with a Normal distribution of mean zero and standard deviation equivalent to the product fac $* p_{id}$. Figure 2 shows the original production curve, plus one realization of the daily production streams for the 10 and 20% daily production variation levels (in log space). Each of the three change point detection algorithms was then applied to sets of 100 generated production streams, resulting in an empirical probability distribution for the day at which the production flow regime changed. The mean, 2.5 percentile and 97.5 percentile values of these distributions are presented in Table 1.

The constrained nonlinear least squares algorithm (Method 2) and the Bayes algorithm (Method 3) generally returned mean values close to the day of the flow change (see Table 1). Moreover, Table 1 indicates that the 95% confidence intervals are very narrow. The performance of the piecewise linear model (Method 1), however, deteriorated, compared to the other algorithms, as the random noise became more intense. Without constraints on the slope coefficients, the transition days identified by the piecewise linear model (Method 1) varied wildly. As might be expected, increases in *fac* increase the spread of the estimated change day distributions, as evidenced by the widening gap between the 2.5 percentile and the 97.5 percentile values of the change point.

Under the case of 20% daily production variation, the data series after the transition (at 200 days) was truncated and each of the change point estimation algorithms were applied. Table 1 shows that the performance of the linear piecewise algorithm continues to deteriorate. The calculated mean value of the constrained nonlinear least squares method is about 5% in error with a series length of 225 data points. The estimated Bayes algorithm mean transition days remained very close to the actual value. Because the Bayes formulation imposes a prior distribution on t^*, it seems to be more adaptable than the other methods to accommodating the shorter data series. With respect to the Bayes method, if the prior ranges for the slope coefficients are increased to about 0.35–0.65 and 0.7–1.3, then the confidence interval about t^* shows about a 10% increase but the change date estimate does not change significantly.
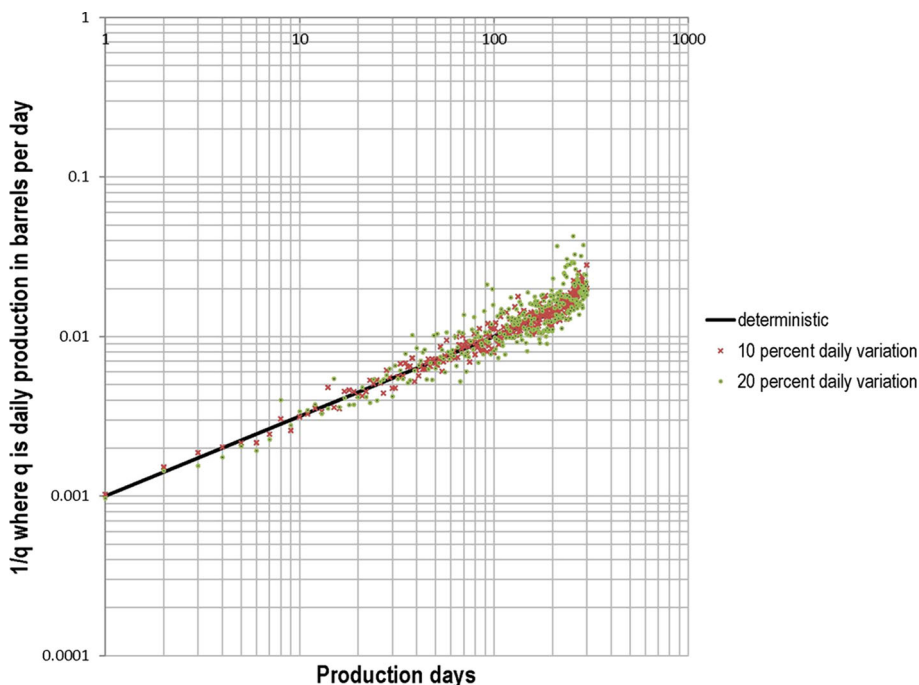
Figure 2. The representative well production data from Figure 1 (designated deterministic) and the well with one realization of the introduction of random noise with a standard deviation of 10 and 20% of daily production (q)..

EMPIRICAL APPLICATION OF CHANGE POINT DETECTION METHODS TO BAKKEN PRODUCTION WELLS

Data Considerations

The simulation experiments described above evaluated the predictive power of the three approaches when increasing levels of noise were introduced into the behavior of the synthetic daily well production stream. Unfortunately, daily production records are usually only available to the well operator and not to others. In most cases the only production data available are monthly totals provided by state agencies or commercial entities, and often such values are reported on a lease by-lease basis rather than for a specific well. Some additional pre-analysis processing of the actual well production data is required before any of change point detection approaches can be applied (this is also true for the traditional visual approach). The following describes the pre-analysis processing of actual production data from the Bakken wells in North Dakota.

For the part of the Bakken play located in North Dakota, the North Dakota Industrial Com-

mission, Oil and Gas Division (NDIC/OGD) publishes well level monthly production data. These data consist of the number of days the well was in production and the total monthly production for oil, gas, and water. Using data obtained from NDIC/OGD, a daily production series was constructed for each well by dividing the well's monthly production by the number of days the well operated and assigning that quotient to the days the well was in production for the month. This necessary procedure obviously results in a distortion of the actual daily production but it may convey sufficient information to detect the change in flow regimes. Figure 3 shows such monthly data plotted in log–log format. The log–log graph derived from monthly data appears as a series of step functions rather than a piecewise linear function as shown in Figure 1 (Coburn and Ran-McDonald 2017).

Estimating the Flow Regime Change Point in Bakken Wells

Following generally accepted engineering practice, the visual method was first employed to estimate the change point in flow regimes for each of

Table 1. Performance of the three statistical approaches in simulation experiments based on a 300-day series of synthetic production data with a change in flow regime set at 200 days and where the increasing fac level represents an increasing variation from the mean of daily production

fac level	Mean	2.50%	97.50%	Length of data series
Distribution: production day of transition				
Piecewise linear (No constraints)				
5%	200	193	211	300
10%	186	155	219	300
20%	93	2	246	300
20%	45	2	227	275
20%	20	1	219	250
20%	17	1	155	225
Nonlinear constraint				
5%	200	195	206	300
10%	200	190	210	300
20%	200	176	226	300
20%	201	176	230	275
20%	199	173	232	250
20%	191	171	212	225
Bayes (with priors)				
5%	200	195	205	300
10%	200	190	210	300
20%	203	178	249	300
20%	203	179	234	275
20%	203	173	228	250
20%	197	176	210	225

Assume p_i is daily production and is composed of deterministic and random components; i.e., p_{id} is the deterministic component of daily production and p_{icr} is a random increment (either positive or negative) of daily production. "Noisy production" was generated using the following formulation: $p_i = p_{id} + rv(p_{icr})$ with $p_{icr} \sim N(0, \text{fac} * p_{id})$ where N is the Normal distribution of zero mean and standard deviation equivalent to the product fac $* p_{id}$

the 37 wells. A petroleum engineer was tasked with visually fitting the linear trends (as in Fig. 3) to the constructed daily production data for each well (Coburn and Ran-McDonald 2017). Prior to fitting the trends, the engineer employed the common approach of censoring the data by ignoring a number of consecutive production days at the start of each well's production stream for which the values were deemed to be erratic and uncharacteristic of normal well operations. As is commonly observed, the number of days ignored were different for each well. Of the three statistical-change point estimation procedures originally proposed, Methods 2 (constrained nonlinear least squares) and Method 3 (Bayesian) were then applied to these same censored daily data streams. Method 1 was eliminated from further consideration based on results obtained on the synthetic data.

Table 2 presents the number of production days ignored for each well along with estimates of the production day at which the flow regime changed (change point) using the visual method, Method 2, and Method 3. The values reported for Methods 2 and 3 are the estimated expected values based on the statistical assumptions underlying the change point model, whereas the values reported for the visual method represent single-point estimates subjectively identified by the engineer. Note that the visual estimates appear as approximations, being expressed as values rounded to hundreds or fifties of days, whereas the estimates determined with Methods
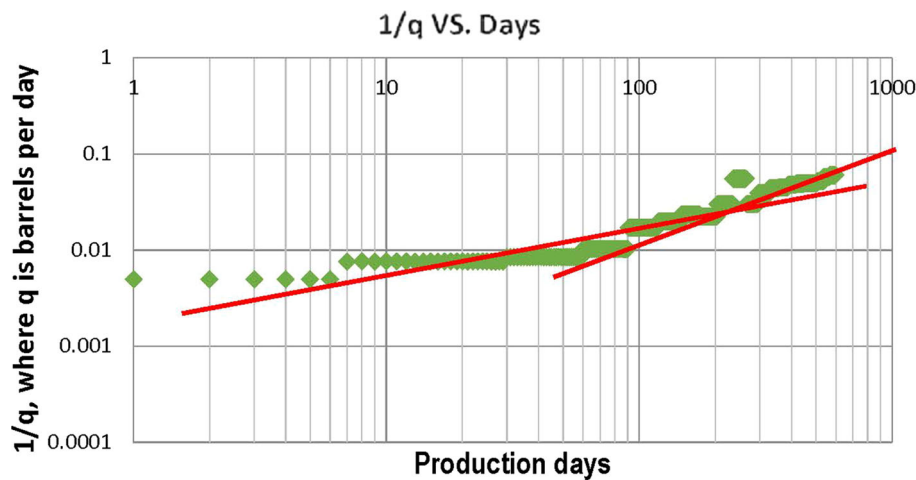
Figure 3. Identification of the flow regime change for a Bakken horizontal hydraulically fractured oil well where daily production (q) is computed from monthly production divided by the number of days the well was productive during the month..

Table 2. Bakken production well case studies comparing the informed visual estimates for the day of flow regime change with the estimates and the 95% confidence intervals of constrained nonlinear least squares and the Bayesian regressions

N. Dakota Permit No.	Well ID	Days ignored	Visual estimate	Nonlinear constrained least sq.			Bayesian regression		
				Mean	2.50%	97.50%	Mean	2.50%	97.50%
Production day of transition from transient to boundary-dominated flow									
20802	2	30	400	299	276	324	301	279	324
22396	4	0	350	393	331	467	392	368	422
22795	5	10	350	87	52	146	428	385	482
22929	6	20	300	277	245	313	282	269	296
21213	7	60	400	309	282	340	340	326	356
22573	8	60	200	113	57	225	114	100	133
22158	9	30	300	306	255	368	330	311	351
21006	10	30	400	379	324	444	388	364	414
21045	11	30	500	536	481	597	544	528	559
21654	12	60	400	115	86	155	285	264	306
21034	13	20	150	170	139	208	174	157	191
22483	14	30	300	352	326	380	326	307	343
22572	15	60	200	120	99	146	123	113	138
19265	16	30	250	207	164	261	348	310	395
22990	17	30	400	461	440	483	459	442	472
22495	18	40	300	326	277	384	332	313	354
21106	19	20	300	295	254	343	299	283	316
22891	20	30	300	287	251	329	294	278	314
21860	23	20	600	563	529	599	564	550	576
22361	24	20	250	375	332	423	374	364	385
23467	25	0	300	176	157	197	177	169	185
23420	26	30	300	230	207	254	336	323	350
20394	28	50	600	374	305	459	394	354	446
19698	30	0	400	477	375	606	496	446	554
20831	31	80	500	119	90	157	123	109	142
23504	32	30	350	352	316	393	361	345	376
22285	33	7	300	228	191	271	240	215	276
22166	34	30	200	167	132	212	323	295	351
21728	36	0	350	379	348	412	364	339	388
22258	38	20	300	79	62	101	103	100	110
21402	39	80	300	168	149	189	187	176	201
22152	40	40	300	244	223	267	243	234	251
20917	41	40	300	417	371	469	443	402	483
22280	43	20	350	496	398	618	530	475	638
20476	44	40	400	243	220	268	431	398	464
21937	45	100	200	100	91	110	643	639	646
22769	46	0	250	233	214	254	236	226	248

2 and 3 are stated more strictly as exact numbers of days.

On whole, the change point estimates obtained using Methods 2 and 3 are generally closely aligned, as shown in Figure 4, though not uniformly so. It is likely that the nonlinear least squares algorithm (Method 2) obtained a local rather than global optimum value in cases where the corresponding Method 2 and Method 3 estimates are considerably different. Table 2 shows 37 cases, there were 8 other cases where the constrained nonlinear least squares

algorithm failed because of singularity of the Hessian matrix. For 29 of the 37 wells, the expected values obtained with these two methods are within 10% of each other, with the expected value of the change point obtained using Method 3 being the higher value for 31 of the 37 wells. This close alignment can also be observed by considering the 95% confidence intervals associated with the expected values of the change points obtained using the two methods, also shown in Table 2. For 25 of the 37 wells, the 95% confidence interval on the
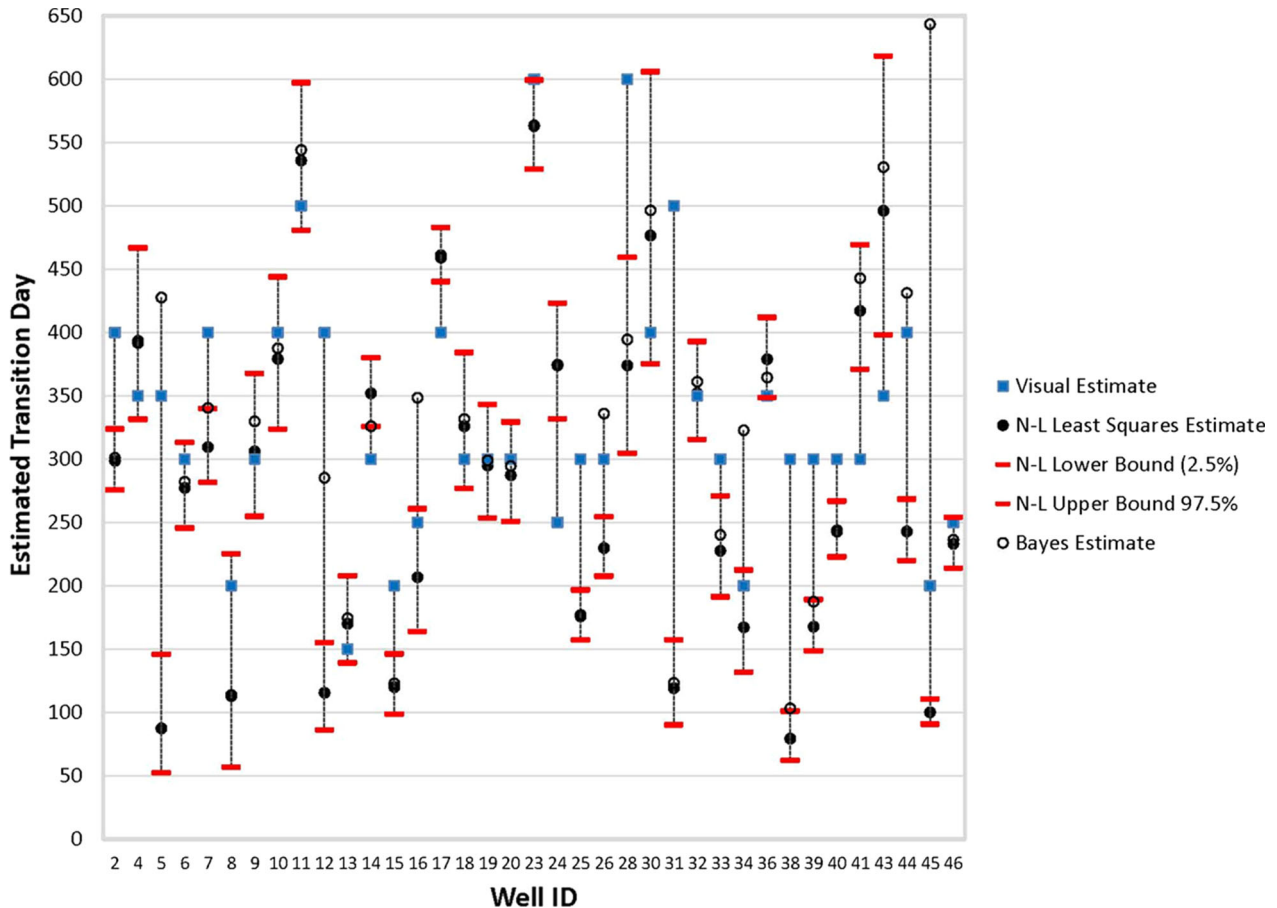
Figure 4. Display of the estimated transition dates from transient to boundary-dominated oil well flow for 37 hydraulically fractured Bakken oil wells based on the visual method, constrained nonlinear least squares regression method and the Bayes regression method. The 95% confidence bounds (2.5 and 97.5%) for the constrained nonlinear least squares estimate are shown. For 30 for 37 cases the Bayes regression estimate of the transition day fall within that 95% confidence interval..

expected value of the change point obtained using Method 3 encompasses the corresponding expected value obtained using Method 2; and similarly, for 30 of the 37 wells, the 95% confidence interval on the expected value change point obtained using Method 2 includes the expected value change point obtained using Method 3 as shown in Figure 4. Hence, for at least 2/3 of these wells, Methods 2 and 3 do not lead to mean change point values that are significantly different from a statistical standpoint. Still, the confidence intervals with Method 2 are wider on, average, than those associated with Method 3 suggesting Method 2 (constrained nonlinear least squares) may lead to somewhat more uncertainty than Method 3 (Bayesian). No confidence intervals are provided for estimates of the change point day produced using the visual method because these are single-point values subjectively produced by the engineer that do not encompass uncertainty.

As previously described, the visual method and Methods 2 (nonlinear least squares) and 3 (Bayesian) are quite different in nature, with the visual approach being deterministic and Methods 2 and 3 being stochastic. Still, comparisons among the results associated with all three approaches are instructive. For some wells, the estimates obtained using the visual method are quite close to the corresponding estimates obtained using Methods 2 and 3, while for other wells this is not the case. In particular, for nearly half of the wells (16 out of 37) the visual estimate was within the 95% confidence interval around the mean change point estimated using Method 2, whereas this was true for a smaller percentage of the wells (9 out of 37) relative to

Method 3 (an unsurprising result because, as noted above, the confidence intervals associated with Method 2 are, on average, wider than the corresponding ones associated with Method 3).

Implications of Alternative Estimates of the Transition Date for Predictions of Production

It is instructive to consider the short-term implications of the change point estimates obtained for the Bakken wells using the methods described above. To do this, the daily production on the initial (estimated) day of boundary-dominated flow plus production for the next 179 production days (total 180 days, hereafter called the in-sample period) was used to calibrate a hyperbolic decline function with the ''Arps Decline Curve Analysis in R'' package (Turk 2017). The hyperbolic decline was applied as a uniform method for projecting ahead relatively short time periods. The wells represented in Table 2 had relatively short production histories, and it was beyond scope of this study to fully address best modeling practices for late well life production forecasting. The decline model parameters were chosen to minimize the sum of squared errors. This model was then used to predict production for the subsequent 180-day period (designated the out-of-sample period). The criteria used for comparing the out-of-sample performance was to find the differences in the sums of the actual and predicted well production during the period and to divide by the actual cumulative well production. The result of this calculation is called the percent prediction error of out-of-sample cumulative production.

Table 3 shows the root mean square errors (RMS) of predicted production for the in-sample period and for the out-of-sample period. RMS is calculated by taking the square root of the sum of the squared prediction errors divided by the number of days of production. The RMS predictions may not be strictly comparable because the estimated change point is different for each well and, as a result, a different part of each well's production history is being addressed (i.e., the production decline models used for calibration during the in-sample period start at different production days which represent when boundary-dominated flow is estimated to begin). However, the numbers in Table 3 do provide an idea of the scale of the errors in daily production rates.

As expected, the RMS prediction errors are generally larger for the out-of-sample period than for the in-sample period. In almost two-thirds of the cases the out-of-sample RMS prediction errors for one of the proposed methods, as well as for the visual method, were within 10 barrels per day of each other. However, for this set of wells the visual method had the absolute lowest RMS prediction errors for about two-thirds of the cases.

Table 4 shows the percent of prediction error of cumulative production during the in-sample and out-of-sample periods. The calibration method for the hyperbolic function assured a near match during the in-sample period of the predicted and actual volume produced. The 180-day out-of-sample percent errors are often much larger, but no method consistently has the lowest percent error. For many of the cases where the out-of-sample percent error was large, all the methods had relatively large errors. For about 70% of the cases the difference between percent error for the visual method and one of the statistical methods is no more than 10%. Notwithstanding these small differences, the visual method had the smallest out-of-sample percent error in about 56% of the cases.

The predictive performance of the models based on a 90-day out-of-sample period was compared to the performance for the 180-day out-of-sample period. With only a few exceptions the RMS for the 90-day period was lower than that of the 180-day period. In about 60% of cases the cumulative percent prediction errors were smaller for 90-day period. For the 90-day out-of-sample period there was a slightly better predictive performance for Method 2 and Method 3 relative to the visual method than observed in Tables 3 and 4.

DISCUSSION

Establishing the time or point at which production from a hydraulically fractured horizontal well transitions from transient flow to boundary-dominated flow has significant implications for forecasting the ultimate cumulative production of the well. Engineers implicitly rely on the boundary-dominated flow regime to establish production decline, so the ability to identify this particular point in time with some degree of precision is extremely important from an economic perspective. In fact, as noted earlier, Seidle and O'Conner (2016) caution

Table 3. Bakken production well case studies comparing the root mean square error associated with different transition dates to boundary-dominated flow (see Table 2): 180-day in-sample period for estimation of hyperbolic decline and 180-day out-of-sample forecasts*

North Dakota Permit no	Well ID No	Root mean square error, in-sample			Root mean square error out-of-sample		
		Visual (boe/d)	Nonlinear constrained least squares (boe/d)	Bayesian regression (boe/d)	Visual (boe/d)	Nonlinear constrained least squares (boe/d)	Bayesian regression (boe/d)
20802	2	13	12	12	16	25	25
22396	4	6	6	6	19	25	25
22795	5	14	51	16	68	164	109
22929	6	12	13	13	18	17	20
21213	7	3	4	4	8	4	3
22573	8	12	17	17	23	48	48
22158	9	73	73	93	160	136	67
21006	10	38	32	35	43	73	58
21045	11	13	12	12	27	40	41
21654	12	4	16	10	11	17	11
21034	13	77	42	39	41	57	62
22483	14	3	2	3	7	8	8
22572	15	8	13	10	15	6	6
19265	16	11	11	6	8	9	8
22990	17	3	3	3	11	13	13
22495	18	25	28	27	148	124	129
21106	19	8	8	8	21	22	21
22891	20	51	50	51	68	59	65
21860	23	9	4	4	8	25	25
22361	24	122	27	28	119	137	138
23467	25	7	12	12	6	13	13
23420	26	26	32	10	13	25	20
20394	28	10	5	4	6	9	9
19698	30	4	5	5	5	3	3
20831	31	3	7	7	8	10	10
23504	32	8	8	8	15	15	16
22285	33	4	9	7	26	9	9
22166	34	3	3	3	5	6	6
21728	36	3	3	3	10	27	10
22258	38	0	7	6	2	3	7
21402	39	12	10	8	20	36	51
22152	40	12	20	20	26	20	21
20917	41	10	5	4	7	22	35
22280	43	25	7	5	28	19	13
20476	44	7	21	8	26	9	21
21937	45	3	7	2	4	5	6
22769	46	8	9	8	22	16	21

*The 180-day period following the transient flow period (in-sample data) was used to calibrate a hyperbolic decline curve, and out-of-sample root mean squares (RMS) show prediction errors for the 180-day period following the data series used for calibration

against using the transient flow data to calibrate common decline function models.

This situation necessitates the establishment of a protocol to determine the point at which the flow regime actually changes. Historically, engineers have visually examined each individual well's production stream to identify the change point, a process that can be tedious, time-consuming, and highly subjective, particularly when multiple wells are involved, even with the help of contemporary software packages. The difficulty with this approach is that it is not necessarily repeatable or reproducible; i.e., the same engineer looking at the same data at some time in the future might select a different transition point, and a different engineer looking at the same data certainly might pick a different point. Unfortunately, small variations in the selection can have substantial impacts on the forecast of cumulative production because of the assumed nonlinear decline after boundary-dominated flow is established.

Table 4. Bakken production well case studies comparing the prediction of cumulative well production associated with different transition dates to boundary-dominated flow (see Table 2): 180-day in-sample period to calibrate a hyperbolic decline and 180-day out-of-sample forecasts*

North Dakota Permit no	Well ID No	Percent prediction error cumulative production					
		180-day in sample			180-day out-of-sample		
		Visual	Nonlinear constrained least squares	Bayesian regression	Visual	Nonlinear constrained least squares	Bayesian regression
20802	2	0.07	0.00	0.00	− 20.09	31.27	31.74
22396	4	0.01	0.00	0.00	− 28.80	− 39.35	− 39.40
22795	5	0.00	− 0.39	0.17	− 11.71	− 75.86	− 63.10
22929	6	0.02	0.00	0.00	− 15.14	− 13.33	− 17.05
21213	7	− 0.01	0.02	0.00	33.28	− 9.32	− 0.55
22573	8	0.16	0.03	0.03	29.72	89.07	90.36
22158	9	− 0.01	0.00	0.11	58.75	48.14	− 4.75
21006	10	− 0.04	− 0.02	− 0.03	− 18.92	− 39.25	− 29.57
21045	11	0.00	0.03	0.02	− 14.43	− 26.09	− 27.22
21654	12	0.00	− 0.02	− 0.04	− 5.69	10.16	4.59
21034	13	− 0.04	− 0.10	− 0.09	− 0.71	22.67	27.53
22483	14	0.00	0.00	0.00	− 0.72	0.98	− 1.61
22572	15	− 0.01	0.08	0.12	− 22.77	− 0.27	2.82
19265	16	0.01	0.01	0.00	16.61	18.01	− 20.14
22990	17	0.00	0.00	0.00	− 3.63	− 10.94	− 10.93
22495	18	− 0.01	0.00	0.00	− 40.54	− 32.07	− 34.50
21106	19	0.00	0.00	0.00	− 1.51	− 4.80	− 2.32
22891	20	0.14	0.11	0.13	− 30.29	− 23.67	− 27.61
21860	23	0.02	0.00	0.00	4.86	31.61	32.11
22361	24	0.00	− 0.01	− 0.02	9.63	− 21.81	− 22.20
23467	25	− 0.01	0.01	0.01	0.79	8.49	8.30
23420	26	0.00	0.00	− 0.08	− 3.77	− 9.25	7.32
20394	28	1.19	0.00	0.00	− 1.08	− 7.93	− 1.82
19698	30	0.00	0.00	0.01	4.50	1.31	− 1.41
20831	31	− 0.01	− 0.02	− 0.02	− 16.47	− 4.02	− 4.22
23504	32	0.01	0.00	0.00	− 3.85	− 4.20	− 0.12
22285	33	0.00	0.01	0.01	29.18	− 0.58	1.65
22166	34	0.00	− 0.06	− 0.01	1.46	21.14	− 14.33
21728	36	0.00	0.00	0.00	− 7.46	− 10.72	− 8.11
22258	38	0.00	0.13	0.04	7.86	− 0.24	− 30.63
21402	39	0.00	0.00	0.00	− 6.23	18.27	30.50
22152	40	− 0.02	− 0.05	− 0.06	17.90	− 13.05	− 14.10
20917	41	0.00	0.01	0.00	4.05	7.85	− 7.14
22280	43	− 0.02	− 0.02	− 0.03	27.69	− 17.39	− 10.92
20476	44	− 0.01	− 0.01	0.00	12.78	− 2.00	11.16
21937	45	0.00	− 0.01	0.00	2.17	− 3.90	− 10.80
22769	46	0.00	0.00	0.00	− 11.57	− 5.96	− 11.81

*The 180-day period following the transient flow period (in-sample data) was used to calibrate a hyperbolic decline curve and out-of-sample percent prediction errors for the cumulative production for the 180-day period following the date series used for calibration

Consequently, there is a definitive need to develop an alternative process—one that is objective, consistent, reliable, repeatable, and verifiable—and that might possibly be automated to promote overall operational efficiency. The results reported herein represent an attempt to achieve these goals.

It must be recognized that the change from transient to boundary-dominated flow is not necessarily an abrupt one. In fact, for many reservoirs, the transition would be expected to occur more gradually over time. Unfortunately, a gradual transition is often difficult to detect because the well production data, often reported as monthly totals, are not sufficiently granular to capture such small changes in the flow regime. This is the situation faced by most entities that are seeking to gain knowledge about a producing area they do not already own because, at best, the only data available is monthly data that can

be obtained from state agencies or vendors. As demonstrated in the present study, monthly data totals must be disaggregated and allocated to individual days if there is to be any hope for improving the estimate of the change point from one flow regime to the other. Still, disaggregation procedures are not without limitations. The present study relies on the use of one such procedure, and its limitations and their potential impacts are described.

Three statistical approaches were proposed that directly estimate the point (time) of change between flow regimes: piecewise linear regression (Method 1), nonlinear least squares with parameter constraints (Method 2), and Bayesian regression (Method 3). These methods were first evaluated using synthetic production data (with and without random variation added) and then applied to factual production data from hydraulically fractured horizontal Bakken oil wells in North Dakota.

Simulation experiments were designed to initially evaluate the performance of the three proposed methods. A synthetic 300-day production stream containing a pre-established change point at the 200th day was constructed for use in these experiments. When applied to this data, all three methods exactly reproduced the established change point, suggesting all three as possible change point identification contenders.

Next, various amounts of random variation (up to 20%) were added to the synthetic daily data (to reasonably approximate factual, but noisy, daily production data), and the three methods were reapplied in a second round of simulation experiments. The constrained nonlinear least squares approach (Method 2) and Bayesian regression (Method 3), which permits inclusion of prior engineering information about the parameters, again nearly exactly reproduced the pre-established change point—another encouraging finding. However, linear piecewise regression (Method 1) was unable to do so when the amount of added random noise exceeded 5%. The simulation experiments were again repeated using the synthetic daily data with 20% random variation added, but with the series length truncated to 275, 250, and 225 production days, respectively. The results for Methods 2 and 3 were similar but indicated a slight reduction in performance as the data series was shortened. The results for Method 1 were substantially worse.

From these experiments it can be concluded that constrained nonlinear least squares (Method 2) and Bayesian regression (Method 3) are likely viable

candidate techniques with which to identify the transition from transient to boundary-dominated flow in daily production data, but that linear piecewise regression (Method 1) is not. Moreover, Methods 2 and 3 seem to detect the transitions within a reasonable time period of the occurrence of the transition. These protocols, findings and results were further evaluated using "live" production data from 37 hydraulically fractured horizontal Bakken wells. Monthly data obtained from NDIC/OGD were converted to daily data using a simple disaggregation procedure. After eliminating Method 1 from contention, Methods 2 and 3 were applied to the data to estimate the expected value of the change point, as well as the 95% confidence interval about that estimate. Visual "picks" were determined by an engineer.

For 28 of the 37 wells, the expected values of the change points of Methods 2 and 3 were within 10% of each other, with the expected value of the change point using Method 3 being higher than for Method 2 for 31 of the 37 wells. For at least 2/3 of these wells, Methods 2 and 3 yielded expected change point values that were not significantly different, even though the estimates obtained with Method 2 were observed to be somewhat more variable than those obtained with Method 3. The two different methods lead to similar results which further underscores their viability as protocols for determining the transition between transient and boundary-dominated flow, thereby promoting the possibility of achieving this goal using an objective, quantitative, and at least semi-automated, approach.

As opposed to Methods 2 and 3 which are stochastic in nature, the visual method of selecting the change point is quite different. Being deterministic (and somewhat subjective), it yields a single-point estimate for which there is no measure of associated uncertainty (i.e., there is no way to estimate what the result might be if the process were to be repeated). Nonetheless, comparing results obtained with the visual method to those obtained with Methods 2 and 3 can be instructive. For at least half of the wells, the change points determined with the visual method occurred later in the data series than those determined with Methods 2 and 3. On the other hand, for nearly half of the 37 wells investigated, the visual estimate was encompassed by the 95% confidence interval around the expected value of the change point obtained with Method 2, whereas this was true for a smaller percentage of the wells relative to Method 3. Even though the visual

method is not positioned here as the ''gold standard'' (because of its deterministic, subjective nature), and acknowledging that the relationship is not yet fully understood, these findings do suggest there is some degree of alignment of the visual method with Methods 2 and 3.

The performance of all three methods was further investigated in terms of what happens to production beyond the estimated point of transition to boundary-dominated flow. For each well, a hyperbolic decline curve was initially calibrated to the first 180 days of production data (estimated day of transition obtained with each of the three methods, plus the next 179 days). Once calibrated, the well decline curves were then used to further project production for the subsequent 180 days.

On average, all three methods almost exactly reproduced cumulative production in the calibration phase (first 180 days or in sample) of each well's data; but for at least two-thirds of the wells, all three under-predicted cumulative production beyond that point (the next 180 days or out-of-sample period). For these 37 wells collectively, average percentages underestimation are about 0.1% for the visual method, about 1.8% for Method 2, and about 4.5% for Method 3. As often observed in forecasting scenarios, the average root mean square (RMS) errors for the out-of-sample period were higher than for the in-sample period, on a well-by-well basis, for all three methods. With regard to all three methods the average out-of-sample RMS improved when the out-of-sample period was shortened to 90 days.

CONCLUSIONS

These results imply that (1) all three methods are reasonable approaches for identifying transition from transient to boundary-dominated flow, (2) Methods 2 and 3, being purely statistical techniques, are both reasonable alternatives to the more subjective visual approach, and (3) the longer the production forecast period, the less accurate all three methods become (which is to be expected when trying to forecast into the future). While further improvements might be realized if the decline curve calibration function could be optimized, or if actual daily production data were available, the above results do suggest that statistical approaches may, indeed, be constructed to replace less objective and repeatable visual approaches for determining the flow regime transition point.

With regard to the disaggregation issue previously discussed, there are various ways to handle the situation (e.g., methods suggested in the financial or hydrologic literature), but because production days are not necessarily tied to calendar days, a disconnection still exists. Further, the performance of all the change point detection methods described here (including the visual method) depends on the subjective choice of the number of days at the beginning of production deemed to be unrepresentative. This is both an engineering and data handling challenge that is unlikely to be easily resolved with purely statistical methods, but which may be treatable with procedures that are more data-driven such as neural networks.

The problem of detecting the change in production flow regimes has an important bearing on the estimated ultimate recovery of a well. The work summarized here suggests that statistical approaches can be developed that are consistent and repeatable, and which possibly might be automated, to alleviate the inherent variability and bias that accompanies the more traditional and subjective approach, and which may account, in part, for unexplained spatial variation in the performance of horizontal, hydraulically fractured wells.

APPENDIX

SI metric conversion factors			
ft	x	3.048	m
stb	x	0.1589	m^3
cf	x	0.0283	m^3

ACKNOWLEDGMENTS

OPEN ACCESS

REFERENCES

Anderson, D. M., & Mattar, L. (2003). Material-balance-time during linear and radial flow. In *Canadian international petroleum conference, 10–12 June, Calgary, Alberta*. https://doi.org/10.2118/2003-201.

Bacon, D. W., & Watts, D. G. (1971). Estimating the transition between two intersecting straight lines. *Biometrika, 58*(3), 525–534.

Carlin, B. P., Gelfand, A. E., & Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Journal of the Royal Statistical Society, Series C, 41*(2), 389–405.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. *The American Statistician, 46*(3), 167–174.

Coburn, T. C., & Ran-McDonald, B. (2017). Investigation of statistical approaches for estimating/predicting ultimate recovery in hydraulically fractured horizontal Bakken oil wells to facilitate economic assessment of unconventional resources, project report: USGS Grant G16AS001.

Duong, A. N. (2011). Rate-decline analysis for fracture-dominated shale reservoirs. *SPE Reservoir Evaluation & Engineering, 14*(3), 13. https://doi.org/10.2118/137748-PA.

IHSMarkit, n.d., Harmony & Harmony Enterprise. https://ihsmarkit.com/product/harmony-oil-well-performance-software.html. Accessed 14 April 2018.

Kabir, C. S., Rasdi, F., & Igboalisi, B. (2011). Analyzing production data from oil tight oil wells, SPE Paper 147414. *Journal of Canadian Petroleum Technology, 50*(5), 48–58.

Kelkar, M. (2018). *Personal communication*. Tulsa: Department of Petroleum Engineering, University of Tulsa.

Long, D. R., & Davis, M. J. (1988). A new approach to the hyperbolic curve. *Journal of Petroleum Technology, 70*(7), 909–912. https://doi.org/10.2118/16237-PA.

Ran, B. (2016). Well completion optimization of horizontal wells in Bakken shale. Unpublished, Masters Thesis, University of Tulsa, Tulsa, Oklahoma.

Ryan, S. E., & Porth, L. S. (2007). A tutorial on the piecewise regression approach applied to bedload transport data. General Technical Report RMRS-GTR-189. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

SAS Institute. (2016). SAS 9.4 Statistical Software. Cary, NC: SAS Institute.

Seidle, J. P., & O'Conner, L. S. (2016). Estimation of unconventional well recoveries and economics from transient flow data. In *SPE/IAEE hydrocarbon economics and evaluation symposium, 17–18 May, Houston*. https://doi.org/10.2118/179983-MS.

Spiegelhalter, D. J., Thomas, A., Best, N. G., & Gilks, W. R. (1996). ''BUGS Examples, Volume 2.'' Version 0.5 (version ii).

Toms, J. D., & Lesperance, M. L. (2003). Piecewise regression: A tool for identifying ecological thresholds. *Ecology, 84,* 2034–2041.

Turk, D. (2017). Arps decline curve analysis in R. available at https://github.com/derrickturk/aRpsDCA. Accessed Sept 2017.

Yu, S., Lee, W. J., Miocevic, D. J., Li, D., & Harris, S. (2013). Estimating proved reserves in tight/shale wells using the modified SEPD Method. In *Paper presented SPE annual technical conference and exhibition, 30 September–2 October, New Orleans*. https://doi.org/10.2118/166198-MS.