Article

Generalized Extreme Value Statistics, Physical Scaling and Forecasts of Oil Production in the Bakken Shale

Wardana Saputra[iD], **Wissem Kirati and Tadeusz Patzek** *[iD]

The Ali I. Al-Naimi Petroleum Engineering Research Center, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

*   Correspondence: tadeusz.patzek@kaust.edu.sa

Abstract: We aim to replace the current industry-standard empirical forecasts of oil production from hydrofractured horizontal wells in shales with a statistically and physically robust, accurate and precise method of matching historic well performance and predicting well production for up to two more decades. Our Bakken oil forecasting method extends the previous work on predicting fieldwide gas production in the Barnett shale[1] and merges it with our new scaling of oil production in the Bakken[2]. We first divide the existing 14,678 horizontal oil wells in the Bakken into 12 *static* samples in which depositional settings and completion technologies are similar. For each sample, we use a purely data-driven non-parametric approach to arrive at an appropriate generalized extreme value (GEV) distribution[3] of oil production from that sample's *dynamic* well cohorts with 1, 2, 3, . . . years on production. From these well cohorts, we stich together the P_{50}, P_{10}, and P_{90} statistical well prototypes for each sample. These statistical well prototypes are conditioned by well attrition, hydrofracture deterioration, pressure interference, well interference, progress in technology, etc. So far, there has been *no* physical scaling. Now we fit the parameters of our physical scaling model[4,5] to the statistical well prototypes, and obtain a *smooth* extrapolation of oil production into the future that is mechanistic, and not just a decline curve. At late times, we add a smattering of radial inflow from the outside[6]. By calculating the number of potential wells per square mile of each Bakken region (core and noncore), and scheduling future drilling programs, we stack up the extended well prototypes to obtain the plausible forecasts of oil production in the Bakken. We predict that Bakken will ultimately produce 5 billion barrels of oil from the existing wells, with the possible addition of 2 and 6 billion barrels from core and noncore areas, respectively.

Keywords: EUR; shale; infill wells; (re)fracturing; pressure depletion; average well; Middle Bakken; Three Forks

1. Introduction

Over the last decade, crude oil and gas production from hydrofractured shales in the United States has accounted for most of the net increase of global crude oil production. Therefore, it is important to have a reliable, quantitative method for delineating the possible futures oil and gas production in the data-rich US shales. The current industry-standard methods of forecasting production from shales are variants of the empirical decline curve analysis (DCA)[7], developed almost a century ago. Unlike other analytical and numerical methods that require numerous reservoir parameters and a lengthy calculation or simulation time, DCA only requires production data to predict future production by extrapolating oil or gas rate observed in a boundary-dominated flow regime. Due to its simplicity, most petroleum engineers adopt DCA for reserves assessment in shale (including Estimated Ultimate Recovery (EUR) predictions from USGS[8] and EIA[9]).

34 However, most shale wells do not reach the boundary-dominated flow regime for their entire
35 production lives because of the vanishing matrix permeability. Thus, the traditional DCA frequently
36 overestimates EURs of shale wells. To address this issue, many authors have suggested improved
37 DCA methods, specific to shale wells: the Power Law Exponential Decline[10], the Stretched
38 Exponential Decline[11], the Logistic Growth Model[12], the Extended Exponential DCA[13], and
39 the Extended Hyperbolic DCA[14]. To make things worse, the empirical DCA fits of particular wells
40 are ill-suited to forecasting production from a wide area of a given shale play in which reservoir
41 properties vary and uncertainties abound. Therefore, some authors have developed probabilistic
42 models to introduce a range of possible outcomes into their production forecasts [15–19]. The most
43 common assumption is that well productivities in shales are log-normally distributed.
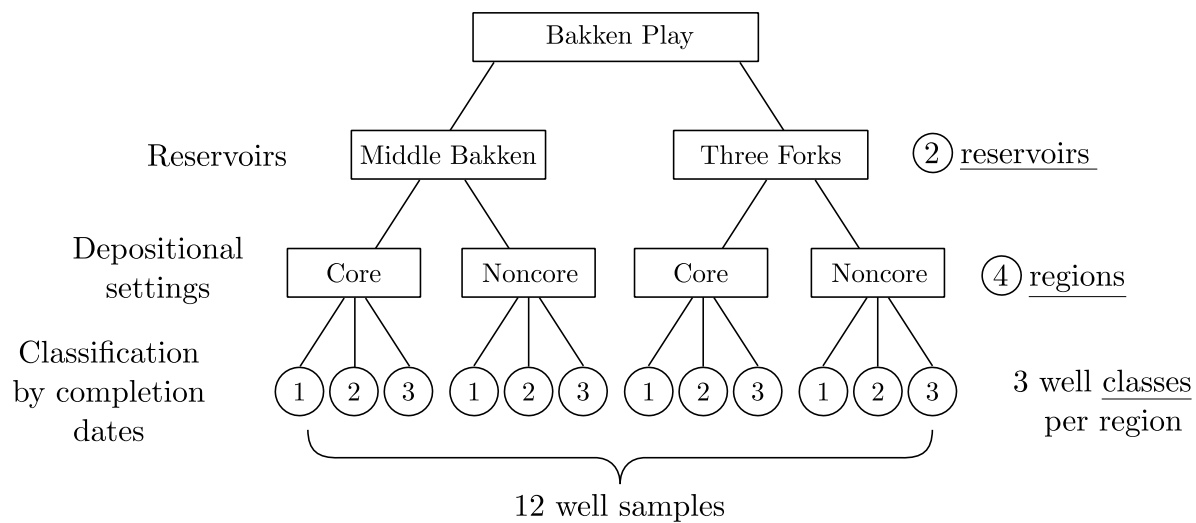

Figure 1. Division of the Bakken shale wells into 12 well samples.

44 In this paper, we adopt a hybrid data-driven *and* physics-based method of predicting oil or gas
45 production in shales that has been introduced in our previous work[1,2]. Here, we will deal only
46 with black oil production. First, we identify play **regions** in which depositional settings are similar,
47 see Figure 1. In each region, we identify well **classes** by different completion technologies. Finally,
48 a well class in a region constitutes a well **sample**. We ensure that oil production from all wells in
49 each sample is statistically uniform, i.e., has a unimodal distribution. For each well sample, we then
50 identify well **cohorts** with $1, 2, \ldots$ years on production. In general, different wells belong to a cohort
51 defined by each admissible elapsed time on production. It turns out that each cohort of wells is
52 superbly characterized by its unique Generalized Extreme Value (GEV) distribution of annualized
53 well rates or cumulative well production. Different cohorts in the same sample have different GEV
54 distributions, each with its unique expected value, median and mode. Here we choose the somewhat
55 better GEV fits of the production rate distributions. Each GEV distribution is statistically superior to
56 the corresponding log-normal distribution at the 95% confidence level. When we plot versus elapsed
57 time of production the expected values of the GEV distributions of all wells cohorts in a sample, we
58 obtain the average P_{50} statistical well prototype that is purely field data-driven.
59 Now we fit each statistical well prototype with a physical scaling curve to extend this prototype
60 to 30 years on production. The physical scaling curves are based on an analytical solution of
61 the pressure diffusion equation in the hydrofractured horizontal shale well geometry. In previous
62 papers, we comprehensively detailed the physical scaling solutions for shale gas wells[4,5] and shale
63 oil wells[2,20]. The late-time flow from the outer reservoir encompassing the stimulated reservoir
64 volume (SRV)[6] was also quantified. We have verified that our physical scaling is equivalent to
65 detailed numerical reservoir simulations[21]. This scaling is much simpler to set up and runs almost
66 infinitely faster than the corresponding reservoir simulations.

67   At this point, for each well sample in every play region, we have obtained a unique hybrid mean
68   well prototype with 30 years on production. In each play region, we know how many wells were
69   drilled and completed each month up until current time. We then multiply each well prototype by
70   the number of wells completed per month and stack them up to represent the total historical field
71   rate and future production decline. In this manner, we obtain a 'base' case forecast for all existing
72   wells. This base case forecast is a 'do nothing' scenario with no new wells drilled in the future. For all
73   other forecasts of future field production rate, we first determine the infill potential or the number of
74   wells that can possibly be drilled in the future. We cover each region with a fish-net grid that consists
75   of one-square-mile pixels. We then calculate the infill potential as the number of wells that can be
76   drilled so that the total number of wells in each pixel is less than the maximum allowable number of
77   wells without fracture hits. Next, based on the infill potentials for all regions, we create future drilling
78   programs to obtain the most plausible forecasts of oil or gas production. Based on current rig count
79   availability in the Bakken, we assume a constant overall drilling rate. Finally, we assign the correct
80   well prototype to every future well that will be drilled during each month of a postulated drilling
81   schedule and sum them up to obtain a forecast scenario.

82   In this paper, we select the Bakken shale, the current second-largest oil producer in the U.S. with
83   1.5 million bbl of oil per day, as an illustration. Being one of the oldest shale oil plays, the Bakken
84   shale has been a field laboratory to test drilling and completion technologies and well productivity.
85   Currently, Bakken has ∼15,000 active hydrofractured horizontal wells with a few wells that have 18
86   years of production data. In a previous paper[2], we scaled well-by-well all ∼15,000 wells in the
87   Bakken. We accounted for well re-fracturing and/or changes in downhole pressure. It turns out
88   that the 12 well prototypes obtained with our hybrid GEV – physical scaling method are as good in
89   duplicating the total field rate as the super-precise scaling of each individual well in our previous
90   work[2]. Since our results are not subject to bias, policy-makers should beware of assuming that the
91   production boom in the Bakken shale will last decades longer.

(a)

(b)
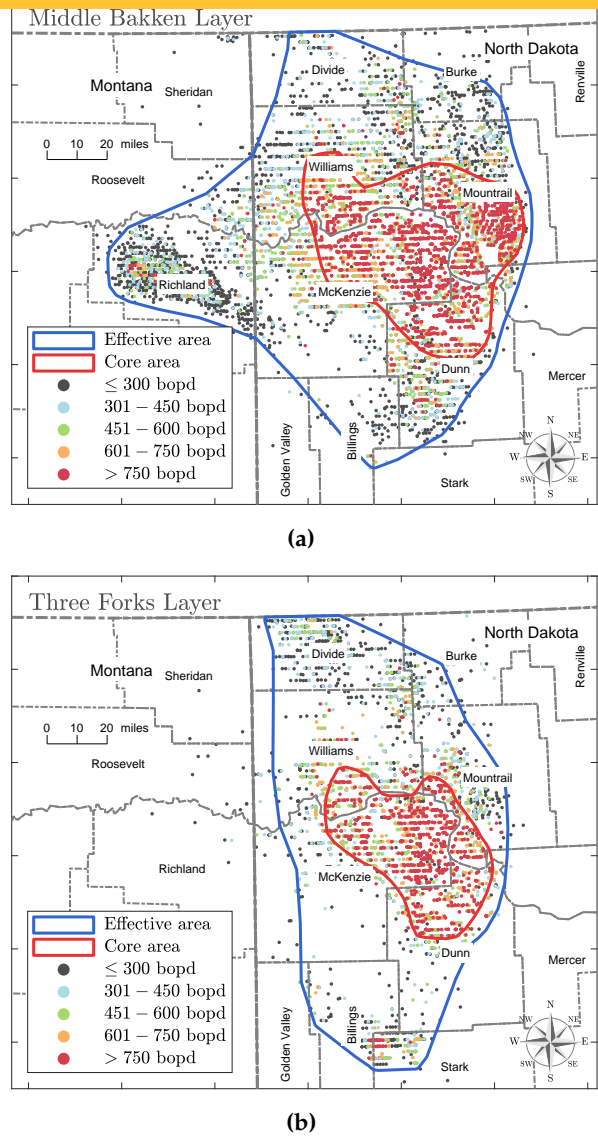
Figure 2. Maps of all 14,678 active horizontal wells completed in the Middle Bakken formation (**a**) and Three Forks formation (**b**) colored by maximum daily oil rate. The red lines define the *core* areas for each reservoir and delineate best producing wells with more than 750 barrels of oil per day. The blue lines define the *effective* areas for drilling neglecting a few poor producing wells outside. We define the *noncore* areas as the difference between the effective and core areas. In the Middle Bakken, there are currently 5,732 and 4,128 wells located in the core and noncore areas, respectively. The other 2,672 and 2,146 wells are located in the core and noncore areas of Three Forks, respectively.

2. Results

Our approach is as follows:

- Divide all 14,678 horizontal oil wells in the Bakken shale into 12 samples in which oil production is statistically uniform;
- Fit a generalized extreme value distribution to all wells in every sample and obtain 12 stable mean P_{50} well prototypes;
- Fit the physics-based scaling curves to every statistical well prototype and extend these prototypes smoothly to 30 years on production;
- Replace oil production from all existing wells with the 12 extended well prototypes and obtain a 'base' forecast;
- Calculate the infill potential for each of the 12 regions in the Bakken; and

103 • Create the most plausible infill drilling schedule and forecast total field oil production rate up
104   to the year 2050.

2.1. Design of Well Regions

106   In the previous work[1], we divided all 13,141 horizontal wells in the Barnett shale into 6 *static*
107 regions that corresponded to the geographical borders of 6 major gas-producing counties. We justified
108 this approach as follows: i) Productive Barnett shale is a single gas reservoir; ii) the Barnett play area is
109 relatively small, and the division by counties captures different depositional settings; iii) completion
110 technologies have not been revolutionized over time.

111   However, in the Bakken shale, we introduce a more complicated set of 12 static well samples
112 classified by reservoir, geography, and completion dates. In Bakken, there are two different producing
113 reservoirs, the Middle Bakken and Three Forks. Accordingly, we first split all Bakken wells into
114 the Middle Bakken and Three Forks groups. To account for different depositional settings, we then
115 divided each of these two well groups of into two macro regions: the core and noncore areas, see
116 Figure 2. Finally, we further refined the four Bakken macro regions by splitting them into three
117 completion date intervals, 2000–2012, 2013–2016 and 2017–2019, to reflect advancements in well
118 completion technology. In the end, we divided the existing 14,678 horizontal oil wells in the Bakken
119 into 12 unique samples listed in Table 1.
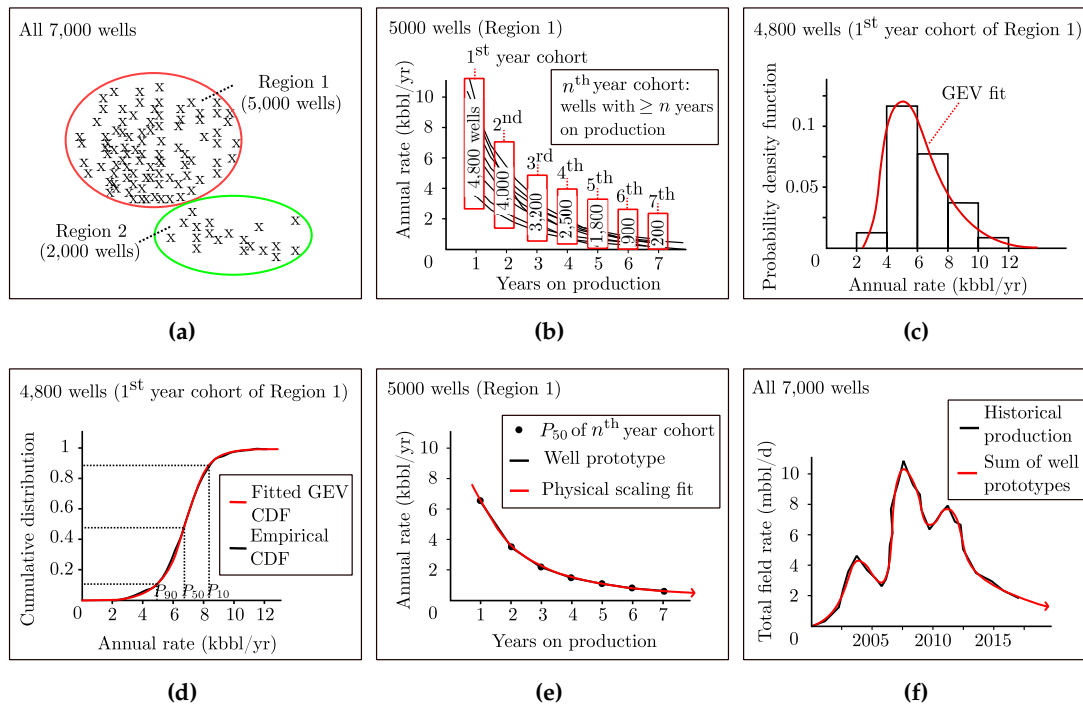
Figure 3. The procedure of arriving at GEV well prototypes in a given shale play. (**a**) Define *static* regions in which oil production is statistically uniform. (**b**) For each region, gather the dynamic cohorts of wells with $\geq i$ years on production, $i = 1, 2, 3, \ldots$. (**c**) Fit GEV probability density function (PDF) to each well cohort. (**d**) From the corresponding cumulative distribution function (CDF) pick the P_{10}, P_{50}, and P_{90} values for each cohort. (**e**) Construct time-lapse P_{50} well prototype for each region, by connecting all P_{50} values of all cohorts. (**f**) Time-shift and superpose the GEV well prototypes to match past production. Reproduced from our previous work[1].

2.2. Statistical Well Prototypes

121   For each of the 12 Bakken well samples defined in Section 2.1, we constructed a statistical well
122 prototype based on the Generalized Extreme Value (GEV) mean or the P_{50} values of annual rates from
123 all well cohorts that exist in that particular region. Figure 3 illustrates the procedure of arriving at the

GEV well prototypes. For every GEV distribution fit we acquired values of the location parameter (μ), scale parameter (σ), shape parameter (ξ), confidence interval (CI), mean (P_{50}), median, mode, P_{10} and P_{90}. In this paper, we show only two GEV fit examples (Figure 4 and 5). Please see Supporting Online Materials-1 to find all GEV fits for all 12 regions of the Bakken shale. The resulting statistical well prototypes are plotted vs the square root of time on production in Figure 6. In the initial year on production, for some well cohorts, time was shifted by subtracting the first 1–3 months on production (see the second column of Table 2) to discount all possible initial disturbances in well performance. After these time shifts, each mean or P_{50} well prototype starts from a straight line versus the square root of time on production. This is the expected in linear flow regime. Later, cumulative oil production bends down due to the inter-hydrofracture pressure interference and exponential decline of production rate[4]. For our GEV distribution fits, the mean or P_{50} prototype is always higher than the median, and the median is higher than the mode. The P_{10} and P_{90} prototypes diverge from the mean with time, indicating that the poorest wells are getting worse and best wells better as also observed in the Barnett shale[1].

**Table 1.** Static regions for the Bakken shale play.

Well sample	Reservoir	Area	Completion date	Sample size
1	Middle Bakken	Noncore	2000–2012	2,550
2	Middle Bakken	Noncore	2013–2016	1,355
3	Middle Bakken	Noncore	2017–2019	223
4	Three Forks	Noncore	2000–2012	735
5	Three Forks	Noncore	2013–2016	1,204
6	Three Forks	Noncore	2017–2019	207
7	Middle Bakken	Core	2000–2012	2,086
8	Middle Bakken	Core	2013–2016	2,534
9	Middle Bakken	Core	2017–2019	1,112
10	Three Forks	Core	2000–2012	428
11	Three Forks	Core	2013–2016	1,502
12	Three Forks	Core	2017–2019	742
	TOTAL			14,678

(a)   **(b)**   **(c)**
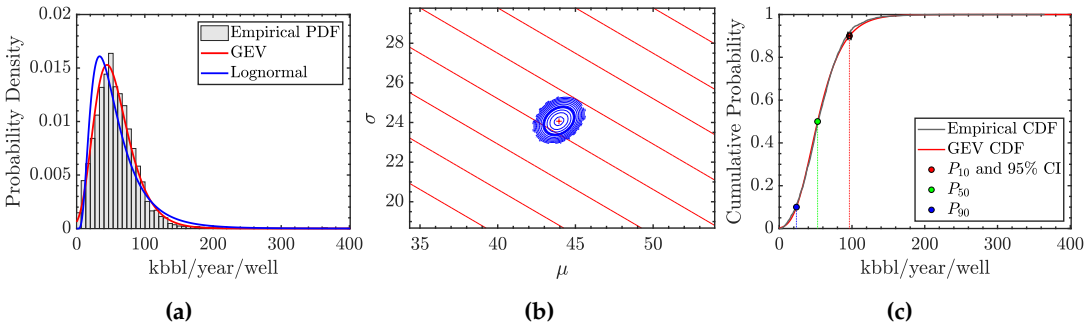
Figure 4. Distribution of oil production rates for 2540 horizontal wells with one year on production, completed between the year 2000 and 2012 in the Middle Bakken noncore area. (**a**) GEV PDF: $\xi = -0.0274$, $\mu = 43.9421$, $\sigma = 24.0635$. (**b**) Maximum Likelihood Estimate, 95% CI for μ and σ. (**c**) GEV CDF with the 90% confidence limits (CIs) on the residual for the P_{10} well.
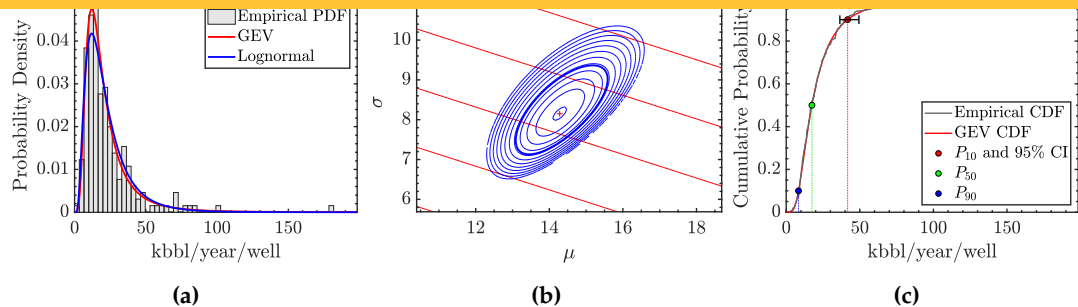
Figure 5. Distribution of oil production rates for 197 horizontal wells with seven years on production, completed between the year 2000 and 2012 in the Three Forks core area. (**a**) GEV PDF: $\xi = 0.3368$, $\mu = 14.2825$, $\sigma = 8.1499$. (**b**) Maximum Likelihood Estimate, 95% CI for μ and σ. (**c**) GEV CDF with the 90% confidence limits (CIs) on the residual for the P_{10} well.

2.3. Physical Scaling Fits

For each region, we fit the statistical P_{50} well prototype with a physical scaling curve, see Appendix B, to extend this prototype to 30 years on production. The fitting procedure is detailed in the Materials and Methods section. Briefly, we find optimum values of the scaling parameters τ and $\mathcal{M}$ so that the physics-based well prototypes match the statistical well prototypes. The results are shown in Figure 6. The red solid lines show the most pessimistic versions of each physics-based prototype, which assume that oil is produced only from the interior of the stimulated reservoir volume (SRV). The green solid lines show a more realistic forecast with an assumption that at late times there will be additional exterior flow of $a\sqrt{t}$ towards the SRV. Finally, one obtains the Estimated Ultimate Recovery (EUR) from each region by identifying the endpoint of cumulative production for the physics-scaling prototype. In Table 2, the scaling parameters, τ, $\mathcal{M}$ and a, and EUR values are listed for each region with and without exterior flow.

Table 2. Scaling parameters for an average well in the Bakken regions.

| Well sample	Shift years	Physical scaling			Physical Scaling with exterior flow			
		τ years	$\mathcal{M}$ kbbl/well	EUR kbbl/well	τ years	$\mathcal{M}$ kbbl/well	a	EUR kbbl/well
Middle Bakken noncore [2000–2012]	0.25	24.5	454.3	271.1	19.9	286.7	0.050	282.2
Middle Bakken noncore [2013–2016]	0.00	11.8	341.3	211.7	10.9	267.2	0.026	236.4
Middle Bakken noncore [2017–2019]	0.00	5.0	346.9	215.5	5.0	306.6	0.015	259.5
Three Forks noncore [2000–2012]	0.15	13.3	288.4	207.5	10.7	181.2	0.058	238.3
Three Forks noncore [2013–2016]	0.00	10.8	271.9	196.1	9.9	212.4	0.030	221.1
Three Forks noncore [2017–2019]	0.00	5.0	270.8	195.4	5.0	238.4	0.018	236.5
Middle Bakken core [2000–2012]	0.20	25.0	810.8	482.5	22.6	536.3	0.050	512.2
Middle Bakken core [2013–2016]	0.00	9.1	543.8	337.7	8.2	420.7	0.026	389.7
Middle Bakken core [2017–2019]	0.00	5.0	581.9	361.5	5.0	513.6	0.015	434.6
Three Forks core [2000–2012]	0.27	25.0	681.1	470.9	25.0	473.4	0.058	511.6
Three Forks core [2013–2016]	0.00	10.3	450.1	324.6	9.6	353.1	0.030	369.7
Three Forks core [2017–2019]	0.00	5.0	462.8	334.0	5.0	406.9	0.018	403.5
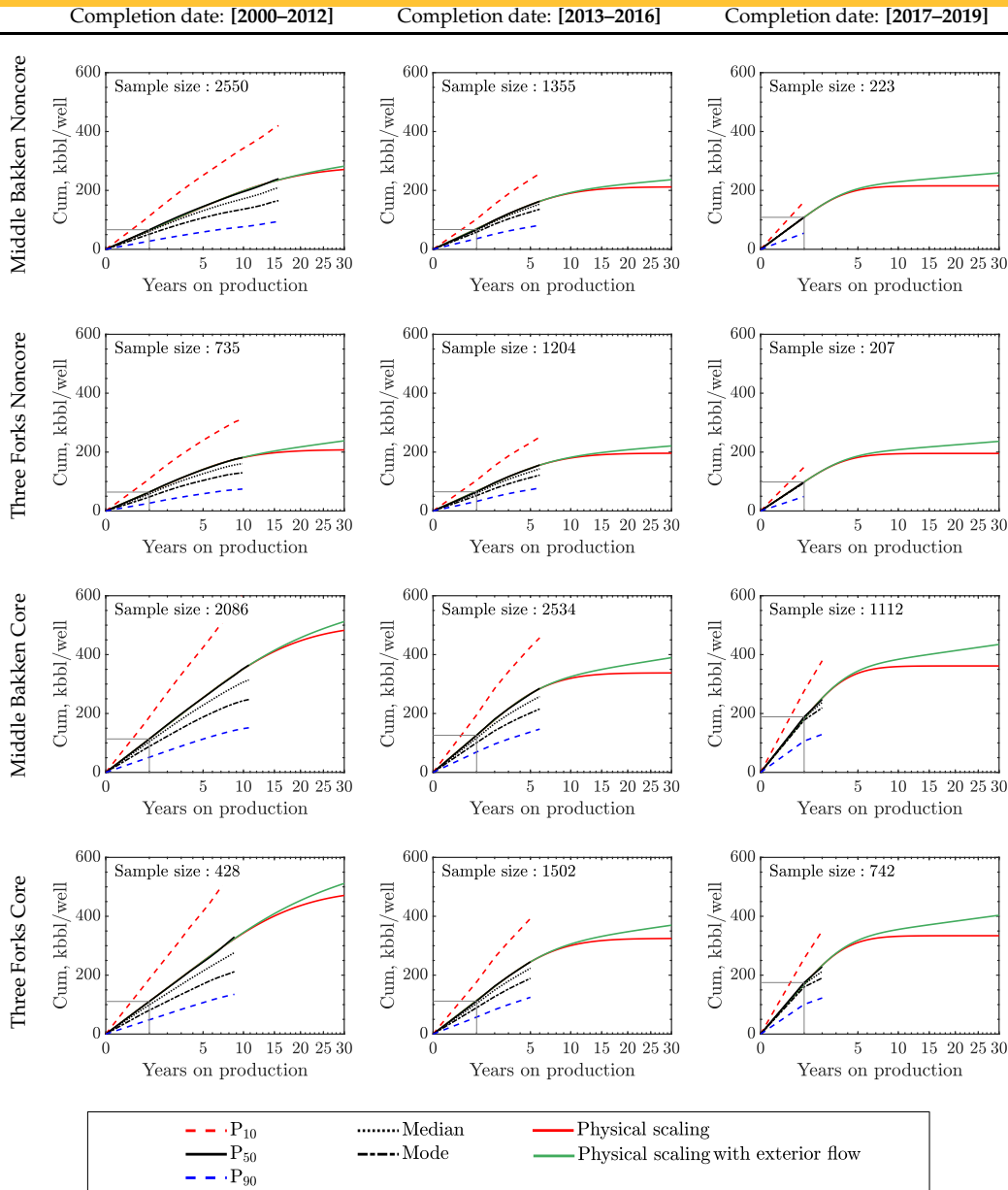
Figure 6. Average wells in 12 regions in the Middle Bakken and Three Forks. These wells are located in both the core and noncore areas and have three different completion periods, (2000–2012), (2013–2016) and (2017–2019). Each year in the past, every average well traces the expected values of the Generalized Extreme Value (GEV) distributions of all active horizontal wells in each well cohort, which have at least 1,2,...,15 years on production. The dashed lines labeled P_{10} and P_{90} denote wells whose cumulative production is exceeded by 10% and 90% of wells in each region. The red and green lines are the physics-based scaling curves that match each average well respectively with and without exterior flow during late time production. In general, ultimate oil recovery from the core areas is higher than that from the noncore ones. The Middle Bakken wells are slightly more productive than the Three Forks wells. The newly completed wells have much higher initial oil production, but they decline faster, resulting in more or less the same ultimate recovery as older wells.

150   *2.4. Base or 'Do Nothing' Forecasts*

151   We have replaced actual field production rate from all existing wells in each region with
152   its corresponding well prototypes multiplied by the numbers of actually completed wells and
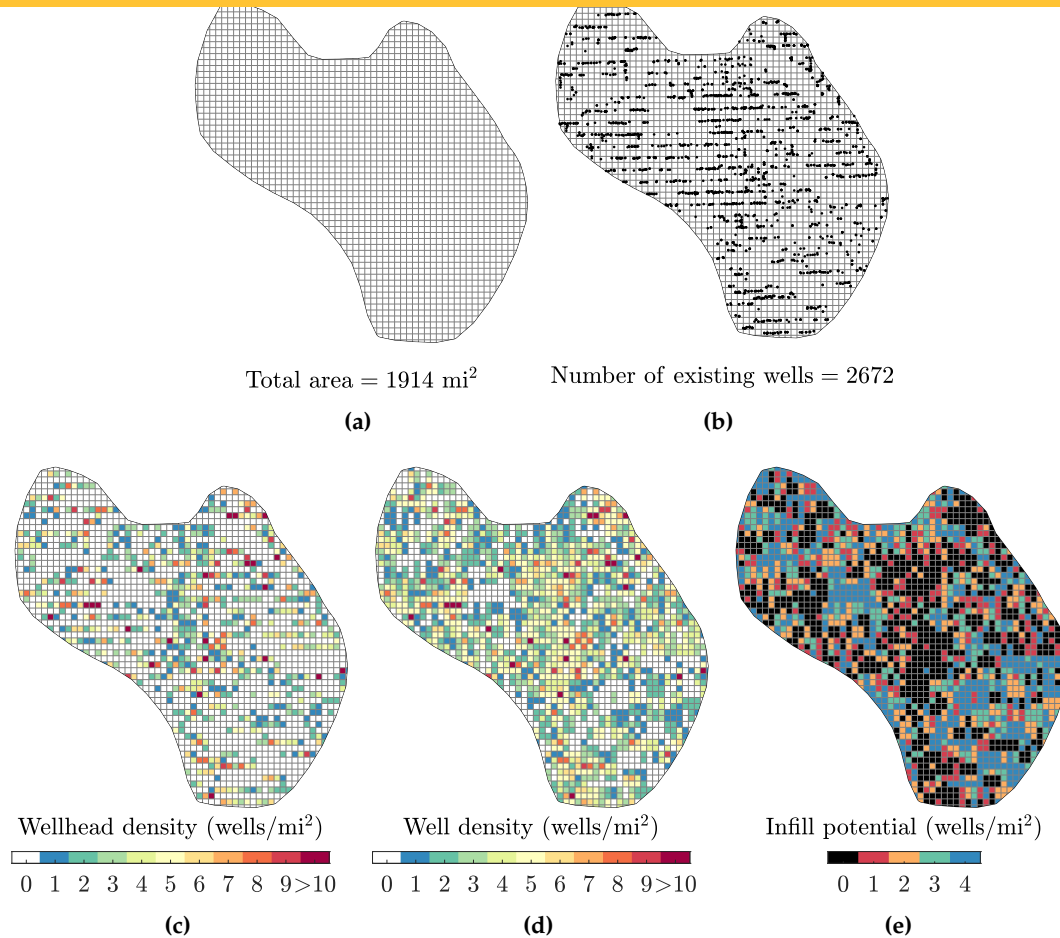
153 time-shifted. The superposed prototypes match the historical field rate. Figure 7 shows the sum of
154 well prototypes with and without exterior flows (green and red lines) versus the total field historical
155 oil rate and cumulative oil (black lines). The results are rather satisfactory. Our GEV prototypes
156 are robust. They capture all peaks of total historical oil rate and flawlessly match the cumulative
157 production. As stated before, the red line forecast is pessimistic. It assumes that there will be no
158 production contribution from the reservoir exterior to the SRVs of individual wells. This pessimistic
159 case may hold if there a shale play is already overdrilled. For the Bakken shale, we can infer that
160 the 14,678 existing wells will ultimately produce 4.5–5.3 billion of oil by 2050. This is a 'base' or 'do
161 nothing' case with no future drilling in the Bakken shale.
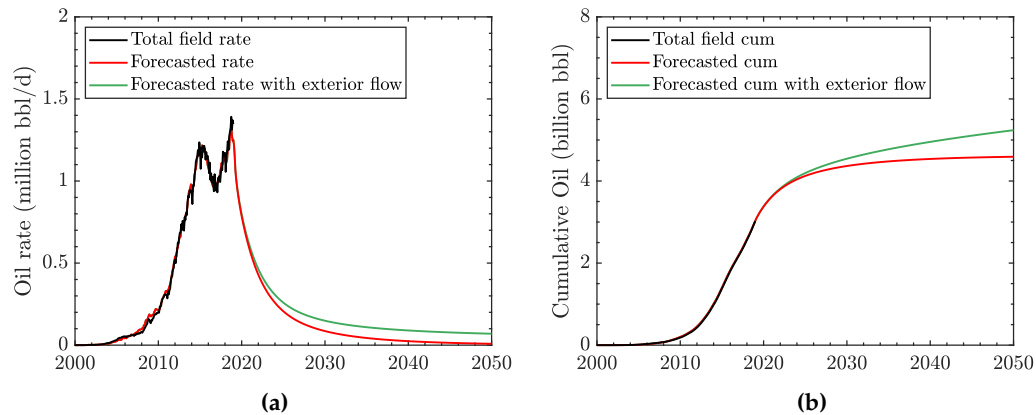
Figure 7. The actual and forecasted total field rate (**a**) and cumulative oil (**b**) in the Bakken shale.
Total field cumulative production curves were obtained by stacking calendar-shifted average wells.
Total production rates were obtained by differencing cumulative production. The red and green
curves are the physical scaling forecasts with and without exterior flow, while the black line shows the
historical production from the existing 14,678 wells. The red physical scaling gives the lower bound of
EUR estimate with about 4.5 billion bbl by 2050. Assuming reasonable exterior flow, EUR prediction
becomes slightly more than 5 billion bbl.

162 *2.5. Infill Potentials*

163   We have calculated the number of wells that can be drilled in the future in every one-square-mile
164 pixel of the grid that covers the entire Bakken play. In order to calculate infill potentials, one
165 should first determine well density. However, the publicly available data rarely provide information
166 about the bottomhole locations of the wells. Instead, only surface locations are reported as
167 latitude-longitude coordinates. Therefore, we have developed an algorithm that allows us to predict
168 the bottomhole well density from surface well locations. See Figure 8 and the Materials and Methods
169 section for more details. As mentioned before in Section 2.1, the Bakken shale has two reservoirs,
170 the Middle Bakken and Three Forks, and two significantly different depositional settings, the core
171 and noncore areas. Thus the infill potentials should be calculated for these four parts of the Bakken.
172 Supporting Online Materials-2 shows the gridding, well locations, wellhead densities, well densities,
173 and infill potentials for the four Bakken parts. Table 3 lists the numbers of wells that can potentially
174 be drilled in the future.

Total area = 1914 mi^2 Number of existing wells = 2672

(a)   **(b)**

Wellhead density (wells/mi^2)   Well density (wells/mi^2)   Infill potential (wells/mi^2)

0 1 2 3 4 5 6 7 8 9>10   0 1 2 3 4 5 6 7 8 9>10   0 1 2 3 4

**(c)**   **(d)**   **(e)**

Figure 8. A procedure to calculate infill well potential for each part of the Bakken play. This figure shows only the Three Forks core area. For other areas and reservoirs, please see the Supporting Online Materials. The procedure is as follows: (**a**) Create a one-square-mile fishnet inside the boundary of each area. In this case, there are 1914 grid squares that translate into the total area of 1914 mi^2. (**b**) Search all existing wells located inside the boundary. The black dots show the surface locations of 2672 existing wells in Three Forks core area. (**c**) Calculate wellhead density by counting the number of wells on each of the one-square-mile squares. This map is not the real well density map, because it only shows the wellhead density. (**d**) Calculate an approximate well density map from wellhead density map. The algorithm is as follows: 1) For each well, record its lateral length and calculate the number of squares, n, intercepted by the lateral. For example, a 5000 ft lateral will occupy one square and a 10000 ft lateral will occupy two squares because one mile is 5280 ft. 2) Search for the least occupied n squares in all possible directions (i.e. north, northeast, east, southeast, south, southwest, west, and northwest). 3) Increase the value of well density by 1 well/mi^2 for every least dense square found. 4) Repeat the process until all wells in the area of interest are exhausted. (**e**) Calculate an infill potential map by subtracting the calculated well density from the maximum number of wells, N_{max} (e.g., N_{max} = 4 wells/mi^2 to avoid frac hits). The summation of all values in the map is the infill potential for the one-square-mile grid. In this case, we obtain 3650 wells. As most wells have 10000 ft laterals and occupy two one-square-mile squares, we divide 3650 by 2 to obtain 1825 infill potential wells in the Three Forks core.

Table 3. Infill potential for Bakken region.

Reservoir	Region type	Total area sq miles	Existing wells	Infill potential
Middle Bakken	Noncore	8534.5	4128	14171
Three Forks	Noncore	6905.9	2146	12341
Middle Bakken	Core	3382.5	5732	2577
Three Forks	Core	1914.1	2672	1825
TOTAL		20737	14678	30914

### 2.6. Future Drilling Scenarios and Infill Forecasts

We have created several future drilling scenarios based on infill potentials listed in Table 3. We assumed that the rig count in the Bakken will not significantly change from the current value, and 120 wells will be drilled per month will be drilled between now and 2041 (Figure 9a). By looking at the current trend, many operators narrow their drilling choices to only the core areas to avoid spending money on less productive wells with high watercuts typical of the noncore areas[2]. Thus, it is reasonable to create a drilling schedule that exhausts all potential wells in the core areas before moving out to the less productive noncore areas. Figure 9b shows this scenario predicting that the Bakken core area will be completely drilled out by 2022. Later, the drilling in noncore areas can last up to 2041, according to our calculation. See Supporting Online Materials-2 for the more detailed schedules.
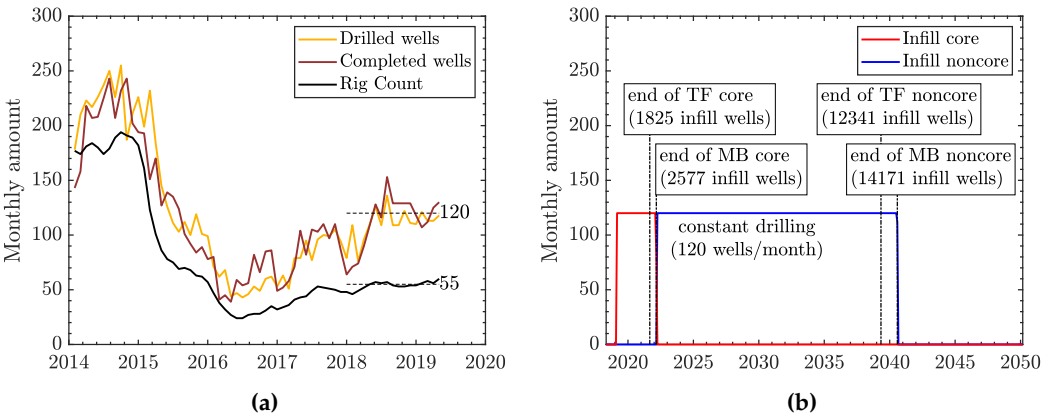
Figure 9. (**a**) The number of drilled wells, completed wells, and rig count per month for the Bakken play from the U.S. Energy Information Administration (EIA). The drilled and completed wells are almost the same for each month and are strongly correlated with the number of rigs available. These data reveal an increase in drilling efficiency. In 2015, one rig could drill about 1.2 wells per month, while in 2019 so far, one rig could drill two wells per month. The current drilling rate is constant at about 120 wells/month. (**b**) A future drilling scenario for the Bakken region up to 2041. The plan is to continue the current drilling rate at 120 wells/month for both core and noncore areas. The numbers of wells to be infilled in each region were previously calculated using the procedure detailed in Figure 8. The results show that the core areas will be fully drilled by mid 2022, leaving the less productive noncore areas for infilling until 2041.

Next, we assigned every single new well scheduled to be drilled in the future with its corresponding region prototype. The results are shown in Figure 10. The black solid lines are the historical total field oil rate and cumulative oil. The purple lines denote the 'base' forecast from all 14,678 existing wells displayed in Section 2.4. The red lines show the result of adding 4,402 new wells in the core area. Because all wells in the core area are high-grade wells, the total oil production rate climbs up its highest production peak at 1.6 million bbl/d in the year 2022. After 2022, there is no further drilling in the core area, and the production rate declines by a factor of two within one year. Ultimately, the total of 19,080 wells (existing + new core) will produce 7 billion bbl of oil by the year 2050. The blue lines show the most optimistic case with not only the core areas, but also the noncore

195 areas will be drilled by the operators. The drilling of less productive wells in the noncore areas will
196 maintain the plateau oil rate of 1 million bbl/d up to the year 2041. However, this plateau will require
197 the drilling of additional 26,500 wells (almost twice the current number of wells) which will have high
198 water cuts. Ultimately, the total of 45,580 wells (existing + core + noncore) will produce 13 billion bbl
199 of oil by 2050.
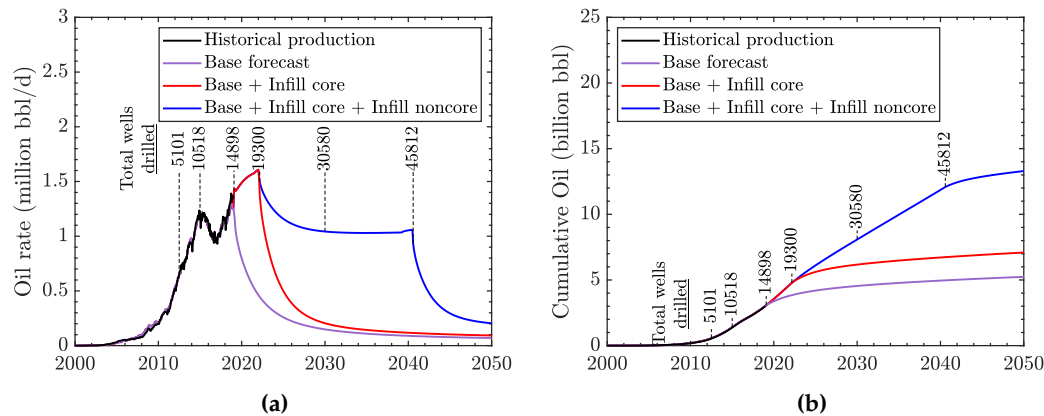
Figure 10. The forecasted total field rate (**a**) and cumulative oil (**b**) for Bakken based on the drilling
scenario in Figure 9b. Since we plan to infill the core areas first, the field oil rate will reach the all-time
production peak of about 1.6 million bbl/d by mid 2022, leaving no more 'sweet spots' in the core
areas. The infill of the less productive wells in the noncore areas will decline the production to a
plateau of 1 million bbl/d. After 2041, no more drilling locations will be left in the Bakken region,
and the oil rate will steeply decline to half a million bbl/d by 2042 and 0.2 million bbl/d by 2050.
Ultimately, the 14,898 existing wells will give an EUR of 5 billion bbl. By adding 4402 wells in core
areas, EUR will increase to 7 billion bbl. By adding another 26,512 wells in the noncore areas, EUR
will increase to 13 billion bbl.

3. Discussion

201 Our hybrid modeling approach seems to have done an outstanding job of matching current oil
202 production and predicting possible futures of the Bakken shale play. We suggest that the Securities
203 and Exchange Commission (SEC) ought to take notice and require operators to use our hybrid method
204 to book reserves and predict future production in the Bakken and elsewhere, most notably in the
205 Permian. We have successfully replaced the current industry-standard empirical forecasts of oil
206 production from hydrofractured horizontal wells in shales with a superior, statistically and physically
207 robust, accurate and precise method of matching historic well performance and forecasting future
208 well production for up to two decades. Our Bakken oil forecasting method extends the previous
209 work on predicting fieldwide gas production in the Barnett shale[1] and merges it with our new
210 scaling of oil production in the Bakken[2]. Our field data-driven statistical well prototypes are
211 conditioned by well attrition, hydrofracture deterioration, pressure interference, well interference,
212 progress in technology, etc. With *no* physical scaling, these prototypes follow the exact physics of
213 linear transient oil flow with pressure interference. Therefore these statistical well prototypes serve
214 as templates to calibrate the parameters of our physical scaling model (τ and $\mathcal{M}$)[4,5], and obtain
215 a *smooth* extrapolation of oil production into the future that is mechanistic, not merely an empirical
216 decline curve. At late times, we add to our extended prototypes some radial inflow from the outside
217 of well SRVs[6].

218 The extended P_{50} well prototypes in Figure 6 can be used to compare ultimate recovery in each
219 of the static regions we identified in the Bakken. The lower bounds are the extended P_{50} prototypes
220 without exterior flow (red lines). In most cases, wells completed in the upper Three Forks reservoir are
221 somewhat less productive than those in the Middle Bakken reservoir. The reasons for this difference
222 are: 1) higher water saturation and water cut in Three Forks, 2) faster decline rate (lower τ), and 3)
223 lower initial oil in place (lower $\mathcal{M}$)[2]. This difference is consistent with the stratigraphic column

of the Bakken total petroleum system, where Middle Bakken is sandwiched between two world-class source rocks with 10% TOC, the Upper Bakken Shale and the Lower Bakken Shale. On the other hand, the Three Forks formation is below the Lower Bakken Shale member, and is exposed to water-bearing formations beneath[8,22–24]. For the same reasons, the *core* and *effective* drilling areas in the Three Forks are smaller than those in the Middle Bakken.

In both reservoirs, production from the *core* areas is superior to that from the *noncore* areas. The core area located in the center of Williston Basin has been known as the most oil-prolific location in the Bakken region in North Dakota[8,22–24]. Since the 1950s, oil has been produced there in the thickest, naturally fractured Middle Bakken formation in the Nesson anticline. One cheap vertical well drilled in the core area in the 1950s has had ultimate recovery of 200 kbbl, the same as a $10 million hydrofractured horizontal well drilled in the Middle Bakken noncore area nowadays. The noncore areas are less productive because the three Bakken formations: Upper, Middle, and Lower Bakken are pinching out upward (are thinner and less mature) near the edges of the Williston basin. Consequently, the noncore areas are producing more water than oil, with watercut exceeding 50% on average.

The newly completed wells have much higher initial oil rates than the older ones, because they have: 1) longer lateral lengths, 2) bigger hydraulic fractures, and 3) more fracture stages[2,25–28]. However, the newly completed wells decline faster and have more less the same ultimate recovery as the older wells. The reasons for this behavior have been described elsewhere[29]. Interestingly, in most cases, older wells completed in 2000–2012 have higher ultimate recovery than newer ones, even though their initial production rates are lower. These older wells might have been drilled in the best locations ever in the Bakken region. In addition, shorter lateral lengths and fewer fracture stages may help in maintaining a stable pressure drawdown and prevent reservoir degassing that is unfavorable for future production. For comparison, the average lateral length in 2005 was 5,000 ft while the average lateral length in 2019 doubled to 10,000 ft. Historically, the number of hydrofracture stages in the Bakken has increased over time from 8 stages in 2007[30] to 18 stages in 2009[25], 35 stages in 2016[26], and to as many as 60 stages in 2019[27].

According to our records, more than 90% of the wells completed after 2017 are located in the core areas only. Operators have learned to drill only the best parts of the Williston Basin and avoid the less mature noncore areas. However, after calculating the infill potentials of all areas, we predict that by 2021 there will be no well locations left for future drilling in the core areas. Assuming a constant current drilling rate of 120 wells per month, the total field oil rate in the Bakken will reach record level of about 1.6 million bbl/d in 2021. Without further drilling, production will decline by one-half within a year. Later, operators will be forced to drill in the less productive, high watercut noncore areas along the edges of the Williston Basin. Our findings suggest that policy-makers should beware of assuming that the shale oil boom in the Bakken will last for several decades longer. We recommend that operators not focus only on increasing the initial oil rate. Maintaining reservoir pressure above the bubble point by preventing over-drilling is key to increasing ultimate oil recovery.

4. Materials and Methods

We mined public well data in the Bakken region from DrillingInfo database, the Montana Department of Natural Resources and Conservation – Board of Oil and Gas Conservation website, and the North Dakota Department of Mineral Resources – Oil and Gas Division website. From DrillingInfo, we selected only 14,678 horizontal oil wells that were completed in the Middle Bakken and Three Forks formations after January 2000. We designed an integrated MATLAB® software package to perform data cleanup, consolidation, and unit conversions. The clean production data for each well consist of a vector of elapsed times on production, a vector of oil production rates, and records of wellhead location (latitude and longitude), lateral length, completion date, reservoir description, and maximum oil rate. Our approach can be divided into six main steps:

I. Define play regions in which oil production is statistically uniform. In each region, follow the steps below:

   (a) Divide all wells in a play into $i = 1, 2, \ldots, n$ well groups, where n is the number of reservoirs[1]. $n = 2$ in the Bakken shale.

   (b) Further subdivide each of these well groups into j_i, $i = 1, 2, \ldots, n$, areas with different depositional settings[2]. In Bakken, $j_1 = 2$ and $j_2 = 2$, i.e., there are two depositional settings (core and noncore) for each of the reservoirs (Middle Bakken and Three Forks).

   (c) Subdivide wells in each of the four *regions* (two reservoirs × two depositional settings each) by time interval *classes* that encompass significant changes in well completion technology[3]. In Bakken, each of the 4 play regions has 3 completion classes, finally yielding 12 static well *samples*.

II. For each well sample, obtain a P_{50} well prototype by fitting a Generalized Extreme Value (GEV) distribution to all qualifying sample wells as follows:

   (a) From a given static well sample-k (a region further subdivided by completion dates), consider a dynamic cohort l_k that contains all wells that have minimum l_k years on production ($l_k = 1, 2, \ldots, t_{\max_k}$ and $k = 1, 2, \ldots, N_{\text{sample}}$)[4]. For Bakken, $N_{\text{sample}} = 12$.

   (b) Define a set X_{l_k} that takes values of oil production rate (kbbl/year) from all wells with l_k years on production in sample k.

   (c) For every X_{l_k}, fit a Generalized Extreme Value (GEV) distribution using Equation A1, and obtain the location parameter, μ, scale parameter, σ, and shape parameter, ξ. (All of these parameters are further subscripted by l_k, but we will skip this complication in the notation.)

   (d) Calculate $\bar{x}_{l_k}$ as the GEV mean of X_{l_k} for each year-l_k, $l_k = 1, 2, \ldots, t_{\max_k}$, using Equation A3.

---

1   In the Bakken play, for example, there are two main reservoirs. Thus, at this step, there are two groups of wells:

Well sample	Reservoir	Sample size
1	Middle Bakken	9,860
2	Three Forks	4,818

2   For example, in the Bakken play, the center of the basin has the thickest oil-prolific layer and hence the highest oil production[2,8,22–24]. Thus, we have delineated an area at the center of the basin (with maximum oil rate > 750 bopd) as *core* area, and the rest as *noncore* area, see Figure 2 for more details. At this step, we have four well groups that fall into four distinct, static play regions:

	Reservoir	Region	Sample size
1	Middle Bakken	noncore	4,128
2	Three Forks	noncore	2,146
3	Middle Bakken	core	5,732
4	Three Forks	core	2,672

3   In the Bakken play, for instance, the newly completed wells have longer lateral lengths, bigger hydraulic fractures, and more fracture stages[2,25–28]. Thus, we classify the wells in each of the four regions by three completion date intervals: [2000–2012], [2013–2016], and [2017–2019]. In the end, we have divided all 14,678 horizontal wells in the Bakken into 12 well groups (4 regions × 3 completion date classes) listed in Table 1.

4   For example: (i) There are 2,550 wells in *Middle Bakken noncore [2000–2012]* group. However, there are only 2,540 wells with at least one year on production (the other 10 wells have production records with less than 12 months). Thus, we retain these 2,540 wells as cohort-1 of this particular group, see Figure 4. (ii) There are 428 wells in *Three Forks core [2000–2012]* group. But, only 197 wells have production records of at least seven years. As such, these 197 wells are qualified as cohort-7 of this particular well group, see Figure 5. See Supporting Online Materials-1 for detailed GEV fits for all dynamic well cohorts and static well groups in the Bakken play.

(e) For each well sample k, obtain the P_{50_k} well prototype by connecting GEV mean values of all years $l_k = 1, 2, \ldots, t_{\max_k}$:

$$P_{50_k} = (\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_{t_{\max_k}}), k = 1, 2, \ldots, N_{\text{sample}}, \quad \text{kbbl/year.} \tag{1}$$

III. Extend P_{50_k} (we will subsequently skip the implied subscript k) well prototypes by fitting to them physical scaling curves as follows:

    (a) For a each well sample k, calculate the *observed* cumulative mass produced, $\mathbf{m}$ (ktons) from Equation A13. *Fix* $q_{o,\text{ST}} = P_{50_k}$ in kbbl/year and $\Delta t_i = 1$ year.

    (b) Adjust τ in Equation A22 or Equation A7 (with or without exterior flow), and $\mathcal{M}$ in Equation A11, so that the physical scaling curve matches the observed $\mathbf{m}$[6].

    (c) Get the extended cumulative mass produced, $\hat{\mathbf{m}}$ (ktons) by multiplying the fitted $\mathcal{M}$ and $RF(t/\tau)$ from the master curve where t is now calculated each month[7].

    (d) Obtain the extended P_{50} well prototypes for well sample k by differencing $\hat{\mathbf{m}}$ converted to volume:

$$\hat{P}_{50} = \frac{\Delta \hat{\mathbf{m}}}{\Delta t (\rho_{o,\text{ST}} + R_s \rho_{g,\text{ST}})}, \quad \text{kbbl/month} \tag{2}$$

IV. Obtain base forecast of total field oil production for existing wells as follows:

    (a) Create a calendar date series with monthly intervals e.g. (1/1/2000, 2/1/2000, ..., 12/1/2050) and assign N_{dates} as the length of this series.

    (b) Create an empty vector $\mathbf{s}$ with the length of N_{dates}: ($\mathbf{s} = [s_1\ s_2\ \cdots\ s_{N_{\text{dates}}}] = [0\ \cdots\ 0]$).

    (c) For each well, find index ii in the calendar date series that brackets well completion date.

    (d) For each well, add to vector $\mathbf{s}$ its corresponding $\hat{P}_{50}$ right-shifted by ii.

$$\mathbf{s} = \mathbf{s} + [\underbrace{0\quad \cdots \quad 0}_{ii\ \text{zeros}}\quad \hat{P}_{50}(1)\quad \cdots\quad \hat{P}_{50}(N_{\text{dates}} - ii + 1)] \tag{3}$$

    (e) Calculate total field oil rate, $\mathbf{q}$ and total field cumulative oil, $\mathbf{Q}$ for existing wells as follows:

$$\mathbf{q} = \mathbf{s}/(365.25/12) \times 10^{-3} \quad \text{million bbl of oil per day} \tag{4}$$

$$\mathbf{Q} = \left[s_1 \quad s_1 + s_2 \quad \cdots \quad \sum_{i=1}^{N_{\text{dates}}} s_i \right] \times 10^{-6} \quad \text{billion bbl} \tag{5}$$

V. Calculate infill potential as follows:

    (a) Create a one-square-mile fishnet inside the boundary of each area defined in step I(b).

    (b) Search all existing wells located inside the boundary.

    (c) Calculate wellhead density by counting the number of wells on each of the one-square-mile squares.

    (d) Calculate approximate well density following algorithm below:

        i. For each well, calculate the number of squares, n, intercepted by the lateral[8].

        ii. Search for the least-occupied n squares in all possible directions.

        iii. Increase the value of well density by 1 well/mi^2 for every least dense n square found.

        iv. Repeat points i–iii until all wells in the area of interest are exhausted.

---

[5] We can use the same procedure to obtain other statistic parameters for each well cohort. The P_{10_k} and P_{90_k} values can be calculated as the ninetieth and tenth percentiles of X_{l_k}, respectively. The median and mode can be calculated using Equation A5 and A6.

[6] Use $c_t/S_{oi}(p_i - p_f)$ and a values from the average Bakken reservoir properties listed in Table A1 and A2. The matching values of τ and $\mathcal{M}$ for two both scaling curves and for all well samples are detailed in Table 2

[7] The benefit of matching P_{50} with a physics-based scaling curve is that we can interpolate and extrapolate production data precisely. Thus, we can change time intervals from years to months and forecast production decades into the future. We recommend to use monthly intervals for precise forecasts i.e. $t = \frac{1}{12}, \frac{2}{12}, \ldots, 50$ years.

[8] For example, a 5,000 ft lateral occupies one square, a 9,000 ft lateral occupies two squares, and a 14,000 ft lateral occupies three squares because one mile is 5,280 ft.

318 (e) Calculate infill potential by subtracting the calculated well density from the maximum
319   allowable number of wells that still avoid frac hits[9].

320 VI. Obtain an infill forecast of total field oil production after adding future wells:

321 (a) Create monthly drilling schedules for every infill potential and assume a constant drilling
322   rate based on current rig availability, see Figure 9b, and SOM-2.
323 (b) Use the same calendar date series as in step IV(a) with N_{dates} as the length of date series.
324 (c) Create an empty vector $\mathbf{s}_f$ with the length of N_{dates}: ($\mathbf{s}_f = [s_{f,1} \; s_{f,2} \; \cdots \; s_{f,N}] = [0 \; \cdots \; 0]$).
325 (d) For every drilling schedule, find index ii in the calendar date series that brackets infill
326   schedule date. Let N_f be the number of wells to be drilled on that date.
     (e) For every drilling schedule, add to vector $\mathbf{s}_f$ its corresponding $\hat{P}_{50}$ after right-shifting it by
         ii and multiplying by N_f.

$$\mathbf{s}_f = \mathbf{s}_f + N_f \times [\underbrace{0 \quad \cdots \quad 0}_{ii \text{ zeros}} \quad \hat{P}_{50}(1) \quad \cdots \quad \hat{P}_{50}(N_{dates} - ii + 1)] \tag{6}$$

   (f) Calculate total field oil rate, $\mathbf{q}_f$ and total field cumulative oil, $\mathbf{Q}_f$ after infill drilling as
       follows:

$$\mathbf{q}_f = \mathbf{q} + \mathbf{s}_f / (365.25/12) \times 10^{-3} \text{ million bbl of oil per day} \tag{7}$$

$$\mathbf{Q}_f = \mathbf{Q} + \left[s_{f,1} \quad s_{f,1} + s_{f,2} \quad \cdots \quad \sum_{i=1}^{N_{dates}} s_{f,i} \right] \times 10^{-6} \text{ in billion bbl} \tag{8}$$

327   where $\mathbf{q}$ and $\mathbf{Q}$ are total field rate and field cumulative from existing wells calculated before
328   with Equations 4 and 5.

5. Conclusions

329
330 • We have provided a transparent hybrid method of forecasting oil production at shale basin
331   scale.
332 • Our statistical approach generates the non-parametric well prototype templates that are used to
333   calibrate our physics-based flow scaling with late-time radial inflow.
334 • In particular, our average P_{50} well prototypes follow the physics of linear transient flow and are
335   used to calibrate the physics-based scaling extensions to 30 years on production.
336 • A combination of GEV statistics with physical scaling matches historical production data almost
337   perfectly and gives a smooth, physics-based estimate of future production.
338 • Our prediction of the Bakken future is optimal in the least square sense[31–34]; in other words,
339   our prediction is as good as it gets given all data at hand.
340 • Regulators may want to consider our approach as a prerequisite to booking reserves in oil
341   shales.
342 • Newly completed wells have almost the same ultimate recovery as the older ones, despite their
343   much higher initial oil rates.
344 • Ultimately, we predict that the 14,678 existing wells in the Bakken will produce 5 billion bbl of
345   oil by 2050 (∼340 kbbl/well).
346 • After drilling additional 4,400 new wells at the rate of 120 wells/month, the core area of the
347   Bakken will be drilled out by 2021, and ultimate recovery will be 7 billion barrels of oil.
348 • With 26,500 more wells drilled in the noncore area until 2041, ultimate recovery in the Bakken
349   might be 13 billion barrels of oil, but drilling of such scale is unlikely to happen.
350 • Policy-makers should beware of assuming that oil boom in the Bakken shale will last decades
351   longer.

---

9   For example, in the Bakken play, the tip-to-tip hydraulic fracture length is roughly 1,200 ft. 5,280 ft/1,200 ft ≈ 4. Therefore,
    the maximum allowable number of wells to avoid frac hits is 4 wells per square mile. Suppose that at some location, the
    well density is 3 wells per square mile. Then, the infill potential is 4 − 3 = 1 well per square mile.

352 6. Supporting Online Materials

353   There are two files. The first one shows full details of infill potential maps for every sub-region in
354 the Bakken shale and a table listing drilling schedule program until year 2050. The second file shows
355 225 unique plots showing all the GEV distributions used to match all well cohorts in the Bakken shale.

366 **Abbreviations**

367 The following abbreviations are used in this manuscript:
368

CDF	Cumulative Distribution Function
EUR	Estimated Ultimate Recovery
GEV	Generalized Extreme Value
PDF	Probability Density Function
SRV	Stimulated Reservoir Volume

370 **Appendix A. The Generalized Extreme Value (GEV) Distributions**

371   The Generalized Extreme Value (GEV) distribution[3] combines three "Extreme Value"
372 distributions, Weibull[35], Gumbel[36], and Frechet[37] into a single functional form. The data decide
373 which of the three distributions is appropriate. In our case, Fréchet distribution almost always wins.

    The probability density function (PDF) of GEV distribution contains three parameters: location,
μ; scale, σ; and shape, ξ:

$$f(x) = \frac{1}{\sigma} t(x)^{\xi+1} e^{-t(x)}, \text{ where } t(x) = \begin{cases} \left(1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right)^{-1/\xi} & : \text{if } \xi \neq 0 \\ e^{-(x-\mu)/\sigma} & : \text{if } \xi = 0 \end{cases} \tag{A1}$$

374 If ξ is zero, Gumbel distribution results, followed by Weibull distribution, if ξ is negative, and by
375 Fréchet distribution if ξ is positive (Figure A1).

    Integrating the GEV PDF, one obtains the cumulative distribution function (CDF)
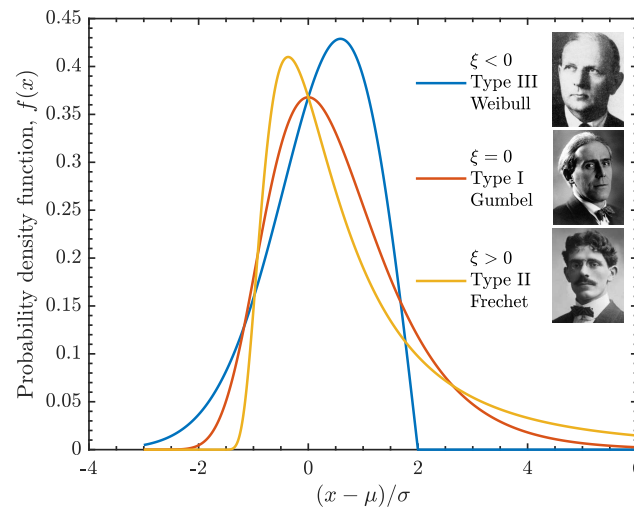
$$F(x) = \exp(-t(x)) \tag{A2}$$

The expected value (mean) of GEV distribution is defined as

$$E(X) = \begin{cases} \mu + \sigma(g_1 - 1)/\xi & : \text{if } \xi \neq 0 \text{ and } \xi < 1, \\ \mu + \sigma\,\gamma & : \text{if } \xi = 0, \\ \infty & : \text{if } \xi \geq 1 \end{cases} \tag{A3}$$

and the variance is

$$V(X) = \begin{cases} \sigma^2\,(g_2 - g_1^2)/\xi^2 & : \text{if } \xi \neq 0 \text{ and } \xi < \frac{1}{2}, \\ \sigma^2\,\frac{\pi^2}{6} & : \text{if } \xi = 0, \\ \infty & : \text{if } \xi \geq \frac{1}{2} \end{cases} \tag{A4}$$

376 where $g_k = \Gamma(1 - k\xi)$, and γ is Euler's constant. Notice that for $\xi \geq 1$, the expected value becomes
377 infinite and for $\xi \geq \frac{1}{2}$ the variance goes to infinity.

Figure A1. Examples of three GEV distributions with the location parameter, $\mu = 0$, the scale parameter, $\sigma = 1.$, and three values of the shape parameter i.e. $\xi < 0$ (Weibull), $\xi = 0$ (Gumbel), and $\xi > 0$ (Fréchet).

The median of GEV distribution is defined as

$$\text{Median}(X) = \begin{cases} \mu + \sigma \frac{(\ln 2)^{-\xi} - 1}{\xi} & : \text{if } \xi \neq 0, \\ \mu - \sigma \ln(\ln 2) & : \text{if } \xi = 0 \end{cases} \tag{A5}$$

and the mode is

$$\text{Mode}(X) = \begin{cases} \mu + \sigma \frac{(1+\xi)^{-\xi} - 1}{\xi} & : \text{if } \xi \neq 0, \\ \mu & : \text{if } \xi = 0 \end{cases} \tag{A6}$$

378 **Appendix B. Physical Scaling Approach to Forecasting Oil Production in Hydrofractured Shales**

379   The physical scaling approach used in this paper was first derived by Patzek et al. to predict
380 gas production from thousands of hydrofractured horizontal gas wells in the Barnett shale[4,5].
381 Later, Patzek et al. derived the scaling curve solution for oil production in the Eagle Ford shale[20].
382 Eftekhari et al. extended the model to include gas inflow from outside of SRVs for the Barnett shale
383 wells[6].

384 *Appendix B.1. Physical Scaling without Exterior Flow*

    Patzek et al.[20] derived the solution of oil production inside SRV based on a simple model illustrated in Figure A2a. This model assumes bi-linear flow towards hydraulic fracture planes inside SRV with the volume $H \times 2d \times 2L$. Briefly, by solving a one-dimensional nonlinear pressure diffusion equation analytically, we obtain a master curve equation for the *interior* oil production problem

$$\text{RF}_\text{I}(\tilde{t}) = \frac{c_t}{S_{oi}}(p_i - p_f)\left(1 - \frac{8}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}e^{(-(2n+1)^2\pi^2\tilde{t}/4)}\right) \tag{A7}$$

where $\tilde{t}$ is the dimensionless time defined as the ratio between elapsed time on production, t, and the characteristic time of pressure diffusion, τ.
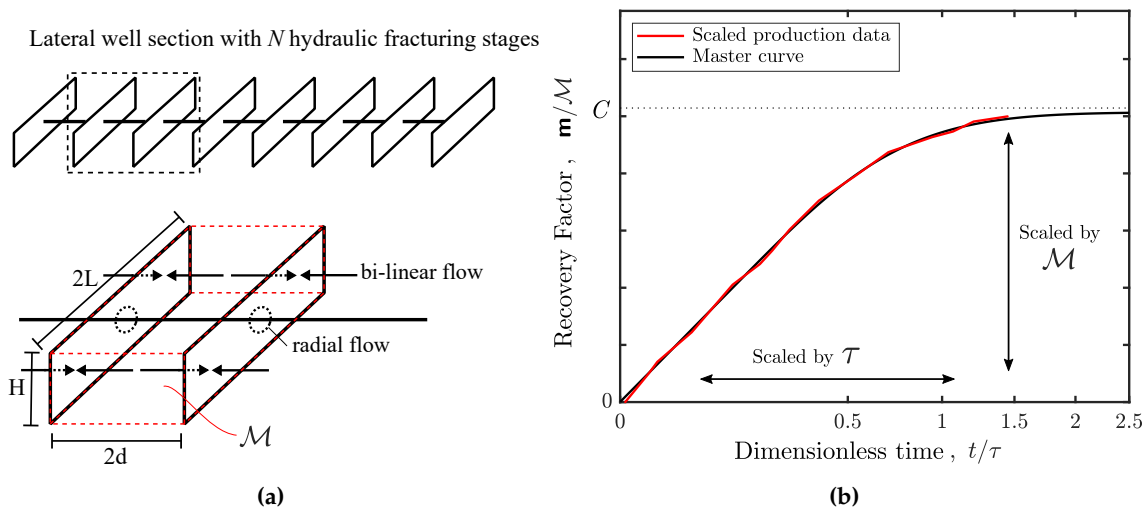
$$\tilde{t} = \frac{t}{\tau} \tag{A8}$$

τ is determined by the onset of pressure interference between two hydrofractures $2d$ apart

$$\tau = \frac{d^2}{\alpha_i} \tag{A9}$$

and α_i is the constant hydraulic diffusivity at initial reservoir conditions

$$\alpha_i = \frac{k}{\phi c_t \mu_{oi}} \frac{\rho_{oi}}{\rho_{\text{fluid},i}} \tag{A10}$$

Figure A2. (**a**) Schematic of bi-linear flow towards hydraulic fractures in a shale well inside the SRV with volume $H \times 2d \times 2L$ (*interior* problem). The early radial flow is neglected because the hydrofracture permeability is much higher than that of the matrix. Reproduced from Patzek et al.[20] (**b**) Illustration of physical scaling approach of the interior oil production problem. Reproduced from Saputra et al.[21]

The fractional oil recovery factor, RF, is the ratio between the cumulative oil mass production, **m** and the stimulated mass contained in the SRV, $\mathcal{M}$

$$\text{RF}(\tilde{t}) = \frac{\mathbf{m}}{\mathcal{M}} \tag{A11}$$

where

$$\mathcal{M} = \rho_{oi}V = \rho_{oi}A_f(2d)\phi S_{oi} \tag{A12}$$

and

$$\mathbf{m} = \left(\rho_{o,\text{ST}} + R_s\rho_{g,\text{ST}}\right)\left(\sum_{k=1}^{i} q_{o,\text{ST}}(t_k)\Delta t_k\right)_{i=1}^{i=t_{\max}} \tag{A13}$$

The physical scaling approach is as follows: First, plot the master curve from Equation A7; Second, scale the cumulative mass produced in Equation A13 to match the master curve by adjusting τ and $\mathcal{M}$. Figure A2b shows the master curve and scaled cumulative mass produced as the black and red lines.

389 *Appendix B.2. Physical Scaling with Exterior Flow*

Eftekhari et al.[6] numerically solved the *exterior* gas production problem in the Barnett shale. We later extended the same concept to oil shales and proposed an analytical solution of this problem. Figure A3a illustrates interior flow towards the hydraulic fractures inside SRV and exterior flow from the outer reservoir towards SRV. We write a new master curve equation for both interior and exterior flow as follows:

$$\mathrm{RF_T}(\tilde{t}) = \mathrm{RF_I}(\tilde{t}) + \mathrm{RF_E}(\tilde{t}) \tag{A14}$$

where $\mathrm{RF_I}(\tilde{t})$ is the master curve equation for interior oil production problem (Equation A7). Eftekhari et al.[6] defined the recovery factor of the exterior problem $\mathrm{RF_E}$ as a function of the new scaled time, $\tilde{t}' = t/\tau'$ where $\mathbf{m_E}(\tilde{t}')$ is the cumulative exterior production and $\mathcal{M}$ is the stimulated mass inside SRV defined in Equation A12:

$$\mathrm{RF_E}(\tilde{t}') = \frac{\mathbf{m_E}(\tilde{t}')}{\mathcal{M}} \tag{A15}$$

Here τ' is the characteristic time scale for exterior production, defined as the time it takes low pressure to diffuse from the wellbore over a distance equal to fracture half-length L:

$$\tau' = \frac{L^2}{\alpha_i} = \frac{d^2}{\alpha_i}\left(\frac{L}{d}\right)^2 = \tau\left(\frac{L}{d}\right)^2 \tag{A16}$$

The new scaled time for exterior production, $\tilde{t}'$ can be written as a function of the interior scaled time, $\tilde{t}$ in Equation A8:

$$\tilde{t}' = \frac{t}{\tau'} = \frac{t}{\tau}\left(\frac{d}{L}\right)^2 = \tilde{t}\left(\frac{d}{L}\right)^2 \tag{A17}$$

Because oil flow from the exterior reservoir is another one-dimensional pressure diffusion problem, one can assume the square root of time solution[6] with a constant slope c_1:

$$\mathrm{RF_E}(\tilde{t}') \approx c_1\sqrt{\tilde{t}'} \tag{A18}$$

Because both the interior and exterior reservoir likely contain similar shale matrix with similar reservoir properties, the constant c_1 can be assumed to be the slope of the interior master curve vs square-root of time (Equation A7 and Figure A2b), which is constant at early times, say, $0 < \tilde{t}^* < 0.2$.

$$c_1 = \frac{4c_t\sqrt{\tilde{t}^*}}{S_{oi}}(p_i - p_f)\sum_{n=0}^{\infty}\mathrm{e}^{(-(2n+1)^2\pi^2\tilde{t}^*/4)} = 1.1284\frac{c_t}{S_{oi}}(p_i - p_f) \tag{A19}$$

$\mathrm{RF_E}$ can be written as a function of the scaled time for the interior production by substituting Equation A17 to A18:

$$\mathrm{RF_E}(\tilde{t}) = c_1\frac{d}{L}\sqrt{\tilde{t}} = a\sqrt{\tilde{t}} \tag{A20}$$
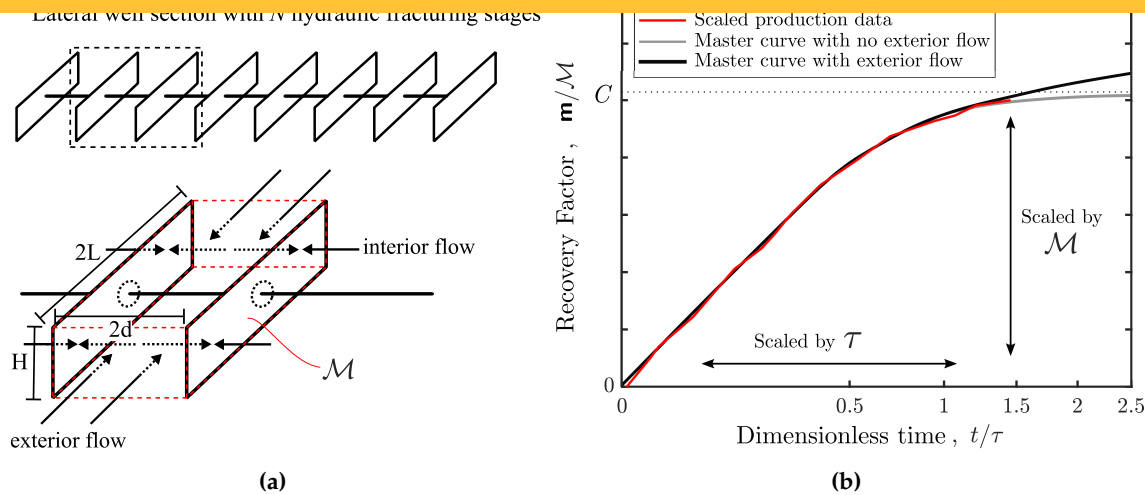
where a is defined as:

$$a = 1.1284\left(\frac{d}{L}\right)\frac{c_t}{S_{oi}}(p_i - p_f) \tag{A21}$$

Recalling Equation A14, we obtain a new master curve equation that embodies interior and exterior oil flow as follows:

$$\mathrm{RFT}(\tilde{t}) = \frac{c_t}{S_{oi}}(p_i - p_f)\left(1 - \frac{8}{\pi^2}\sum_{n=0}^{\infty}\frac{1}{(2n+1)^2}\mathrm{e}^{(-(2n+1)^2\pi^2\tilde{t}/4)}\right) + a\sqrt{\tilde{t}} \tag{A22}$$

(a)   **(b)**

Figure A3. (**a**) Illustration of *interior* flow towards the hydraulic fractures of a shale well inside SRV and *exterior* flow from the reservoir beyond SRV. Adapted from Patzek et al.[20] and Eftekhari et al. [6] (**b**) Illustration of physical scaling approach of the interior and exterior oil production problem. Adapted from Saputra et al.[21]

The physical scaling approach for this problem is as follows: First, plot the master curve from Equation A22; Second, scale the cumulative mass produced in Equation A13 to match the master curve by adjusting τ and $\mathcal{M}$. Figure A3b shows the scaled cumulative mass as the red line and the new master curve as the black line that deviates from the interior master curve in Equation A7 (shown as the gray line) at late times on production.

Appendix C. Reservoir Properties of the Bakken Shale

Table A1. Reservoir properties used in scaling oil production in the Middle Bakken and Three Forks (part-1).

Parameter	Middle Bakken		Three Forks		Data source
	SI units	Field units	SI units	Field units	
Initial pressure, p_i	36.8 Mpa	5340 psia	37.1 Mpa	5380 psia	[2]
Fracture pressure, p_f	3.4 Mpa	500 psia	3.4 Mpa	500 psia	[2]
Connate water saturation, S_{wc}	0.57	0.57	0.65	0.65	[2]
Initial oil saturation, S_{oi}	0.43	0.43	0.35	0.35	[2]
Rock porosity, ϕ	0.046	0.046	0.058	0.058	[2]
Rock permeability, k	4.4×10^{-17} m^2	0.045 md	4.6×10^{-17} m^2	0.047 md	[2]
Rock compressibility, c_ϕ	4.3×10^{-10} Pa^{-1}	3.0×10^{-6} psi^{-1}	4.3×10^{-10} Pa^{-1}	3.0×10^{-6} psi^{-1}	[38]
Water compressibility, c_w	4.3×10^{-10} Pa^{-1}	3.0×10^{-6} psi^{-1}	4.3×10^{-10} Pa^{-1}	3.0×10^{-6} psi^{-1}	[38]
Oil compressibility, c_o	1.4×10^{-9} Pa^{-1}	1.0×10^{-5} psi^{-1}	1.4×10^{-9} Pa^{-1}	1.0×10^{-5} psi^{-1}	[38]
Total compressibility, c_t	6.3×10^{-10} Pa^{-1}	9.0×10^{-6} psi^{-1}	5.9×10^{-10} Pa^{-1}	8.5×10^{-6} psi^{-1}	
Oil viscosity, $\mu_{o,i}$	3.9×10^{-5} Pa s	0.392 cp	2.8×10^{-5} Pa s	0.276 cp	[39]
API gravity	42.2° API	42.2° API	38.7° API	38.7° API	[39]
$(c_t / S_{oi})(p_i - p_f)$	0.1014	0.1014	0.1178	0.1178	

Table A2. Reservoir properties used in scaling oil production in the Middle Bakken and Three Forks (part-2).

Completion date	Number of frac stages, N	$2d$ (ft)	$\frac{d}{L}$	a Middle Bakken	a Three Forks
2000–2012	18[25]	528	0.44	0.050	0.058
2013–2016	35[26]	271	0.23	0.026	0.030
2017–2019	60[27]	158	0.13	0.015	0.018

References

1. Patzek, T.W.; Saputra, W.; Kirati, W.; Marder, M. Generalized Extreme Value Statistics, Physical Scaling and Forecasts of Gas Production in the Barnett Shale. *Energy & Fuels* **2019**, *submitted*.

2. Saputra, W.; Kirati, W.; Patzek, T.W. A Physical Scaling Approach to Predicting Ultimate Recovery From Fifteen Thousand Oil Wells in the Bakken Shale. *Journal of Petroleum Science and Engineering* **2019**, *In review*.

3. Gumbel, E.J. *Statistics of Extremes*; Columbia University Press: New York, 1958.

4. Patzek, T.W.; Male, F.; Marder, M. Gas production in the Barnett Shale obeys a simple scaling theory. *Proceedings of the National Academy of Sciences* **2013**, *110*, 19731 – 19736. doi:10.1073/pnas.1313380110.

5. Patzek, T.; Male, F.; Marder, M. A simple model of gas production from hydrofractured horizontal wells in shales. *AAPG Bulletin* **2014**, *98*, 2507 – 2529.

6. Eftekhari, B.; Marder, M.; Patzek, T.W. Field data provide estimates of effective permeability, fracture spacing, well drainage area and incremental production in gas shales. *Journal of Natural Gas Science and Engineering* **2018**, *56*, 141 – 151. doi:10.1016/j.jngse.2018.05.027.

7. Arps, J.J.; others. Analysis of decline curves. *Transactions of the AIME* **1945**, *160*, 228–247.

8. Gaswirth, S.B.; Marra, K.R.; Cook, T.A.; Charpentier, R.R.; Gautier, D.L.; Higley, D.K.; Klett, T.R.; Lewan, M.D.; Lillis, P.G.; Schenk, C.J. Assessment of undiscovered oil resources in the bakken and three forks formations, Williston Basin Province, Montana, North Dakota, and South Dakota, 2013. *US Geological Survey Fact Sheet* **2013**, *3013*.

9. EIA. Assumptions to the Annual Energy Outlook 2019: Oil and Gas Supply Module. *US Energy Information Administration: Annual Energy Outlook* **2019**, p. 7.

10. Ilk, D.; Anderson, D.M.; Stotts, G.W.; Mattar, L.; Blasingame, T. Production Data Analysis–Challenges, Pitfalls, Diagnostics. *SPE Reservoir Evaluation & Engineering* **2010**, *13*, 538–552.

11. Valko, P.; Lee, W. A Better Way To Forecast Production From Unconventional Gas Wells. Paper SPE 134231 presented at the SPE Annual Technical Conference and Exhibition, Florence, Italy, 19–22 September, 2010.

12. Clark, A.J.; Lake, L.W.; Patzek, T.W. Production forecasting with logistic growth models. SPE Annual Technical Conference & Exhibition. Society of Petroleum Engineers, 2011.

13. Zhang, H.; Rietz, D.; Cagle, A.; Cocco, M.; Lee, J. Extended exponential decline curve analysis. *Journal of Natural Gas Science and Engineering* **2016**, *36*, 402–413.

14. Du, D.; Zhao, Y.; Guo, Q.; Guan, X.; Hu, P.; Fan, D. Extended Hyperbolic Decline Curve Analysis in Shale Gas Reservoirs. Proceedings of the International Field Exploration and Development Conference 2017. Springer, 2019, pp. 1491–1504.

15. Gong, X.; Gonzalez, R.; McVay, D.A.; Hart, J.D. Bayesian probabilistic decline-curve analysis reliably quantifies uncertainty in shale-well-production forecasts. *SPE Journal* **2014**, *19*, 1–47.

16. Paryani, M.; Awoleke, O.O.; Ahmadi, M.; Hanks, C.; Barry, R. Approximate Bayesian Computation for Probabilistic Decline-Curve Analysis in Unconventional Reservoirs. *SPE Reservoir Evaluation & Engineering* **2017**, *20*, 478–485.

17. Fulford, D.S.; Bowie, B.; Berry, M.E.; Bowen, B.; Turk, D.W.; others. Machine learning as a reliable technology for evaluating time/rate performance of unconventional wells. *SPE Economics & Management* **2016**, *8*, 23–39.

18. de Holanda, R.W.; Gildin, E.; Valkó, P.P. Mapping the Barnett Shale Gas with Probabilistic Physics-Based Decline Curve Models and the Development of a Localized Prior Distribution. Unconventional Resources Technology Conference, Houston, Texas, 23-25 July 2018. Society of Exploration Geophysicists, American Association of Petroleum . . . , 2018, pp. 217–232.

19. Holanda, R.W.d.; Gildin, E.; Valko, P.P.; others. Combining physics, statistics, and heuristics in the decline-curve analysis of large data sets in unconventional reservoirs. *SPE Reservoir Evaluation & Engineering* **2018**, *21*, 683–702.

20. Patzek, T.; Saputra, W.; Kirati, W. A Simple Physics-Based Model Predicts Oil Production from Thousands of Horizontal Wells in Shales. SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers, 2017.

446 21.   Saputra, W.; Albinali, A.A. Validation of the Generalized Scaling Curve Method for EUR Prediction in
447   Fractured Shale Oil Wells. SPE Kingdom of Saudi Arabia Annual Technical Symposium and Exhibition.
448   Society of Petroleum Engineers, 2018.
449 22.   Sonnenberg, S.A.; Pramudito, A. Petroleum geology of the giant Elm Coulee field, Williston Basin. *AAPG*
450   *bulletin* **2009**, *93*, 1127–1153.
451 23.   Mason, J. Oil production potential of the North Dakota Bakken. *Oil & Gas Journal* **2012**, *110*.
452 24.   Meissner, F.F. Petroleum geology of the Bakken Formation Williston Basin, North Dakota and Montana
453   **1991**.
454 25.   Buffington, N.; Kellner, J.; King, J.G.; David, B.L.; Demarchos, A.S.; Shepard, L.R. New technology in
455   the Bakken Play increases the number of stages in packer/sleeve completions. SPE Western Regional
456   Meeting. Society of Petroleum Engineers, 2010.
457 26.   Weddle, P.; Griffin, L.; Pearson, C.M. Mining the Bakken: Driving Cluster Efficiency Higher Using
458   Particulate Diverters. SPE Hydraulic Fracturing Technology Conference and Exhibition. Society of
459   Petroleum Engineers, 2017.
460 27.   Martin, E. *United States Tight Oil Production 2018*, 2009 (accessed March 25, 2019).
461   www.allaboutshale.com/united-states-tight-oil-production-2018-emanuel-omar-martin/.
462 28.   Weijermars, R.; van Harmelen, A. Shale Reservoir Drainage Visualized for a Wolfcamp Well (Midland
463   Basin, West Texas, USA). *Energies* **2018**, *11*, 1665.
464 29.   Haider, S.; Patzek, T.W. The key physical factors that yield a good horizontal hydrofractured gas well in
465   mudrock. *Journal of Natural Gas Science and Engineering* **2019**, *Submitted*. doi:10.1016/j.jngse.2018.05.027.
466 30.   Paneitz, J. Evolution of the Bakken Completions in Sanish Field, Williston Basin, North Dakota. *SPE*
467   *ATW: Maximizing Tight Oil in the Bakken Keystone, Colorado. PowerPoint Presentation. August* **2010**, *6*.
468 31.   Chorin, A.J.; Kast, A.; Kupferman, R. Optimal prediction of underresolved dynamics **1999**. *95*, 4094–98.
469 32.   Chorin, A.J.; Kast, A.; Kupferman, R. Unresolved computation and optimal prediction. *Comm. Pure Appl.*
470   *Math.* **1999**, *52*, 1231–1254.
471 33.   Chorin, A.J.; Kast, A.; Kupferman, R. On the prediction of large-scale dynamics using unresolved
472   computations. *Contemp. Math.* **1999**, *238*, 53–75.
473 34.   Chorin, A.J.; Hald, O.; Kupferman, R. Optimal prediction and the Mori-Zwanzig representation of
474   irreversible processes **2000**. *97*, 2968–73.
475 35.   Weibull, W.; others. A statistical distribution function of wide applicability. *Journal of applied mechanics*
476   **1951**, *18*, 293–297.
477 36.   Gumbel, E.J. Statistical theory of extreme values and some practical applications. *NBS Applied Mathematics*
478   *Series* **1954**, *33*.
479 37.   Fréchet, M. Sur la loi de probabilité de l'écart maximum. *Ann. Soc. Math. Polon.* **1927**, *6*, 93–116.
480 38.   Tran, T.; Sinurat, P.D.; Wattenbarger, B.A. Production characteristics of the Bakken shale oil. SPE Annual
481   Technical Conference and Exhibition. Society of Petroleum Engineers, 2011.
482 39.   Kurtoglu, B. *Integrated reservoir characterization and modeling in support of enhanced oil recovery for Bakken*;
483   Colorado School of Mines, 2013.