

Big-Data Analytics for Production-Data Classification Using Feature Detection: Application to Restimulation-Candidate Selection

Egbadon Udegbe, Eugene Morgan, and Sanjay Srinivasan, Pennsylvania State University

Summary

In recent years, there has been a proliferation of massive subsurface data sets from sources such as instrumented wells. This places significant challenges on traditional production-data-analysis methods for extracting useful information in support of reservoir management and decision making. In addition, with increased exploration interest in unconventional-shale-gas reservoirs, there is a heightened need for improved techniques and technologies to enhance the understanding of induced- and natural-fracture characteristics in the subsurface, as well as their associated effects on fluid flow and well performance.

These challenges have the potential to be addressed by developing big-data-analytics tools that focus on uncovering masked trends related to fracture properties from large volumes of subsurface data through the application of pattern-recognition techniques. We present a new framework for fast and robust production-data classification, which is adapted from a real-time face-detection algorithm. This is achieved by generalizing production data as vectorized 1D images with pixel values proportional to rate magnitudes. Using simulated shale-gas-production data, we train a cascade of boosted binary classification models that are capable of providing probabilistic predictions. We demonstrate the viability of this approach for identifying hydraulically fractured wells that have the potential to benefit from restimulation treatment. The results show significant improvements over existing type-curve-based approaches for recognizing favorable-candidate wells, using only gas-rate profiles.

Introduction

In the past decade, there has been a significant increase in the application of streaming-monitoring technologies, such as permanent downhole sensors and distributed-acoustic-sensing tools in instrumented wells. For example, we now have shale fields that contain several wells that generate long-duration-production data (such as fluid-flow rate, density, pressure, and temperature), which are recorded in intervals of minutes to seconds. In the context of flow-rate data, big data takes the form of abundant fluid-volume time-series measurements from multiple producing wells within a field, which must be analyzed and inferred for reservoir-related information. However, there has been a lag in the development of efficient techniques to harness the vast amount of reservoir-related information embedded in these well data. Recently, there has been increased research interest in big-data analytics for upstream applications, with specific focus on new tools and techniques for addressing problems related to infrastructure, data organization and management, analysis and discovery, and decision support (Febowitz 2013). We define the scope of this paper within the context of analysis and knowledge discovery from big data, more specifically for applications involving well-production data. A number of recent publications within this focus area have applied data-mining techniques in an attempt to extract useful information from voluminous subsurface time-series data (Seemann et al. 2013; Aulia et al. 2014; Patri et al. 2014). Although these studies attest to the value of developing new analysis tools for extracting information from large volumes of data, there is currently insufficient literature on harnessing subsurface big data to improve the characterization of hydraulically fractured wells within naturally fractured shale reservoirs.

In the past decade, technological advances in directional drilling and hydraulic fracturing have led to more economically feasible exploitation of unconventional reservoirs such as shale. To infer important reservoir properties and performance-related parameters, historical flow rate and pressure data from producing wells are typically analyzed using techniques such as straight-line (regime-flow) analysis, analytical models, and numerical-simulation approaches (Clarkson 2013). In addition, techniques such as empirical decline-curve models are commonly used to fit trends to well-production data. However, most of these traditional methods (specifically type-curve analysis and regime-flow methods) are formulated for vertical wells in conventional reservoirs, which means that they are generally predisposed to boundary-dominated flow and only provide average information about bulk reservoir properties (Anderson et al. 2010). In addition, they do not consider the complex physics of fluid storage and nanoscale fluid flow associated with shale. A number of dual-porosity analytical models have therefore been proposed to characterize production from multistage hydraulically fractured shale wells (Ozkan et al. 2009; Bello and Wattenbarger 2010; Brown et al. 2011). However, these analytical models are often developed in Laplace space corresponding to specific boundary conditions, which means that numerical-inversion schemes must be used to express the solutions in the time domain. In addition, they typically require several matrix, fracture, and well input parameters, which are often expensive or impossible to infer accurately.

Restimulation treatment has the potential to improve the economic performance of producing shale wells by increasing the conductivity of existing fractures and/or enhancing their contact with the formation (Moore and Ramakrishnan 2006). However, the influences of matrix and fracture characteristics on the success of restimulation are not completely understood, which has led to uncertainty in determining favorable-candidate wells and the optimal timing of restimulation treatment. A number of restimulation-candidate-selection methods have been proposed in the literature. For example, favorable candidates have been identified by predicting untapped production capacity (as characterized by real-valued targets, such as 5-year cumulative production) using artificial-neural-network (ANN) models and detailed data such as geographical, reservoir, production, and well-completion information (Shelley 1999; Mohaghegh et al. 2000). However, under these approaches, there is the potential to impose our understanding of the inputs that

influence restimulation success on the model. In other words, there is the potential that some essential unmeasured properties are not captured as model inputs. In addition, this approach might potentially lead to nonintuitive results (e.g., the influence of a parameter such as surface elevation might be assigned nonzero weight). Also, some of the inputs used in training might be unavailable for new wells, which limits the generality of these models. In another study, dimensionless groups related to the fracture and other physical parameters have been defined and correlated with restimulation success (Roussel and Sharma 2011). Numerical-simulation studies have also been used to study field-production data to define a set of heuristics for candidate selection and restimulation timing (Tavassoli et al. 2013). Overall, many of these existing methods tend to require detailed input data and/or lack generality to other reservoir settings, in addition to being time consuming to implement, which limits their applicability to problems involving massive data sets.

In this paper, we aim to address the previously discussed challenges by developing a new methodology for fast and robust analysis of production data from hydraulically fractured wells. In the context of this work, robustness refers to consistent model performance in support of flexible development strategies (risk-averse or aggressive) under a wide array of scenarios. We use a dual-permeability forward-flow-modeling approach to generate multiple realizations of rate/production profiles by modifying fracture, reservoir, and operational parameters. Using these data, we apply pattern-recognition tools to help uncover trends associated with favorable and unfavorable restimulation candidates. This is achieved using a binary classification framework adapted from real-time face detection, which uses simple numerical criteria computed directly from raw flow-rate data, thus eliminating the need for detailed information and promoting computational efficiency. We show that these features carry information about flow regimes and curvature attributes within well data at various time scales and time periods during the life of the well. The algorithm also provides probabilistic predictions, which serve as a means to rank candidate wells. Although the process of training the classifier can potentially be computationally intensive, the application of the trained classifier on the observed data is extremely fast, thereby rendering the method amenable to real-time classification on the basis of well performance. This in turn has the potential to affect decision-making work flows involving massive data sets collected from producing wells within a field, which helps to mitigate uncertainty in reservoir management and decision making through continuous analysis.

Approach

In this paper, we propose a probabilistic classification framework that is aimed at addressing analysis challenges involving the massive well-production data sets. This approach is demonstrated for a use case in restimulation-candidate selection. To achieve this, we have implemented a supervised-learning framework. In the context of restimulation-candidate selection, this means that we predefine two categories of example wells depending on their observed rate response to restimulation (i.e., “favorable” and “unfavorable” restimulation candidates). Using this information, we train a statistical model to distinguish between both sets of examples and to estimate the probability of restimulation success. The set of quantitative criteria used in defining favorable and unfavorable candidates will be addressed in the upcoming sections.

As an alternative to data-driven techniques that use a limited set of reservoir and completion data to forecast real-valued fluid rates after restimulation, we propose an approach that makes categorical predictions on post-refracturing well performance (i.e., successful or unsuccessful restimulation treatment) by taking into account subtle decline characteristics embedded within the production/rate data. On this basis, the statistical approach in this study is adapted from real-time face-detection techniques that use texture-based features in characterizing input data, with the objective of fast and robust classification performance.

In the abundance of available field-production records where the outcome of the restimulation treatment is known, these well profiles can be directly used in training the classification model under the proposed framework. The trained model can then be used to predict if other candidate wells are likely to benefit from restimulation treatment. On the other hand, in the absence of field data sets, a flow-simulation approach can be used to generate an ensemble of well-production profiles to serve as training data. However, one important advantage of the proposed classification framework is that instead of explicitly forecasting a real-valued flow rate after restimulation, the classification model generates a binary categorical response (“favorable” or “unfavorable” restimulation candidate) along with an estimated probability of refracturing success. This enables us to generate training data using less-detailed flow-simulation models, which need only capture the broad reservoir and fracture characteristics in the productive region surrounding the wellbore.

Fig. 1 summarizes the overall work flow we have implemented in the development of the proposed production-data-classification algorithm.

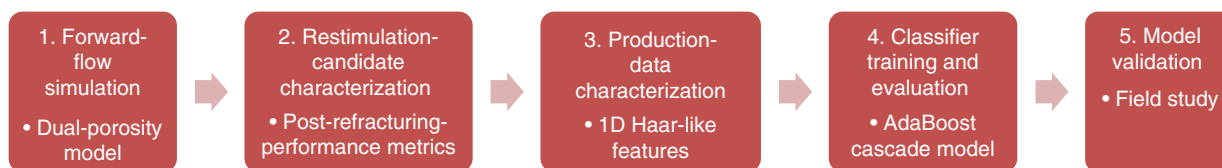


Fig. 1—Summary of overall work flow for developing the proposed production-data-classification model.

In the next section, we discuss the details of the forward-flow-simulation approach used in generating training examples and criteria for specifying favorable and unfavorable restimulation candidates. We then present the proposed statistical framework for production-data classification. Finally, we discuss the model training and evaluation before considering a field study.

Input Production Data

In line with the supervised learning approach proposed in this paper, the first step in implementing a classification scheme for identifying restimulation candidates is to define examples of wells that have undergone both successful and unsuccessful restimulation. As previously mentioned, in the abundance of field-production data from restimulated wells with accompanying completion information, the proposed classification approach can be applied to “learn” the differences using rate-decline characteristics. However, because of the limited availability of field data, we use a forward-flow-simulation approach in this study.

The overall strategy is to simulate multiple realizations of gas-flow rate from a hydraulically fractured shale-gas well and assign class labels depending on post-refracturing performance. In the following subsections, we discuss the flow-simulation model and the reservoir characteristics that contribute to restimulation success.

Flow-Simulation Model. In this paper, we use a probabilistic classification approach to develop a set of rules for identifying restimulation candidates, using rate-decline characteristics. In other words, in contrast to existing data-driven approaches that explicitly forecast flow rate after restimulation, the proposed classification framework returns a categorical prediction (favorable or unfavorable candidate) and an accompanying probability of restimulation success. Therefore, in generating well data for training, this enables us to implement less-detailed flow-simulation models that need only capture the aggregate rate response to fracture and reservoir properties in the productive region surrounding hydraulically fractured wells.

On this basis, the flow-simulation model used in this study features one multistage hydraulically fractured shale-gas well within a 307-acre reservoir. This model is sized to represent the areal extent of a shale-gas reservoir of midrange productivity (Kulga 2014). We have used a $61 \times 61 \times 5$ Cartesian grid with the x - and y -dimension fixed at 60 ft, whereas the vertical grid dimensions have been varied between 20 and 60 ft. Within this reservoir model, we randomly sample from a set of uncertainty parameters related to fracture, fluid, reservoir, and operational properties to generate an ensemble of well profiles with realistic shale-gas rate-decline characteristics.

Flow in naturally fractured media is modeled in a commercial simulator using the dual-permeability approach with the Gilman and Kazemi (1988) shape factor. In line with the stimulated-reservoir-volume (SRV) approach (Mayerhofer et al. 2006), we assume that the majority of reservoir productivity comes from a network of hydraulic fractures and rejuvenated natural fractures in the vicinity of the wellbore. Fig. 2 depicts the SRV geometry we have specified to model flow within the productive region around the horizontal well.

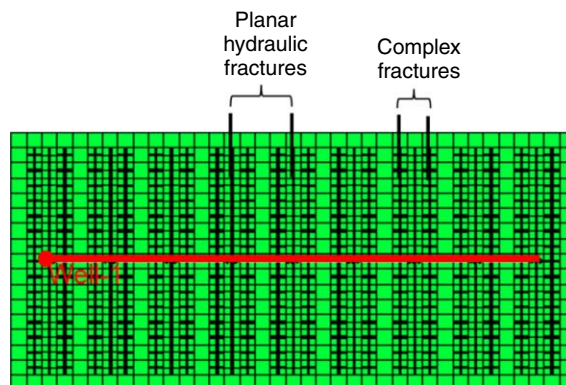


Fig. 2—SRV representation in the flow simulator. The productive region around the well is modeled by assuming equally spaced planar hydraulic fractures, each flanked by a network of rejuvenated natural fractures.

In Fig. 2, the area surrounding the wellbore has been modeled as a series of planar hydraulic fractures, each flanked by a network of rejuvenated natural fractures. We have assumed equally spaced hydraulic fractures with symmetry about the horizontal wellbore. Both the hydraulic and rejuvenated fractures around the wellbore have been defined using $5 \times 5 \times 1$ logarithmic grid refinement around the fractures, which is intended to accurately capture the fluid flow within fractures, within the shale matrix, and between fracture and shale, as proposed by Rubin (2010). Under this framework, a 1-ft grid cell (central block in refined grid) is used to mimic the effect of a 0.001-ft fracture, and the Forchheimer number is scaled to reflect this assumption. To generate an ensemble of well-production profiles, we have randomly sampled uncertainty parameters related to the SRV fracture characteristics, as displayed in **Table 1**.

Property	Minimum	Maximum
Hydraulic-fracture half-length (ft)	270	570
Hydraulic-fracture spacing (ft)	120	300
Number of hydraulic-fracture stages	5	9
Hydraulic-fracture conductivity (md-ft)	10	30
Reduction in hydraulic-fracture-tip conductivity (%)	25	90
Complex fracture-conductivity reduction (%)	60	75
Natural-fracture-spacing reduction in SRV (%)	0	90
Pressure-dependent fracture-conductivity-decline constant, d_f	10^{-3}	10^{-5}

Table 1—Flow-simulation parameters related to fracture characteristics in the SRV region.

In Table 1, all parameters are sampled from uniform distributions, with the uncertainty ranges calibrated using publicly available field-production data from West Virginia Marcellus Shale gas wells.

The physical extent of the SRV region is dictated by the hydraulic-fracture half-length, hydraulic-fracture spacing, and the number of hydraulic-fracture stages. The conductivity at the tip of the fracture is allowed to decrease with increasing distance from the wellbore, using a percentage reduction factor. This parameter serves as a proxy to account for mechanisms such as proppant crushing and fluid leakoff. Likewise, complex fracture conductivity is expressed as a percentage of induced-hydraulic-fracture conductivity. Apart from this, we have included a fractional parameter that accounts for the reduction of natural-fracture spacing in the SRV region.

Finally, to capture pressure-dependent conductivity reduction in the induced and complex fractures, as proposed in Eshkalak et al. (2014), we have applied the Chin et al. (2002) correlation,

$$k_f(\Delta P) = k_{fi} e^{-d_f \Delta P}, \dots \dots \dots (1)$$

where d_f denotes pressure-dependent permeability-decline rate (i.e., fracture-closure rate), k_{fi} represents initial fracture permeability, and ΔP is the pressure drop from the initial reservoir pressure. The induced and complex fractures are each assigned different rates of permeability reduction. Because we use constant fracture width in the flow model, d_f is referred to as a conductivity-decline constant in Table 1.

We have randomly sampled 15 additional reservoir-uncertainty parameters as inputs to the flow-simulation model. These include diffusion and sorption effects to account for physical phenomena specific to shale-gas reservoirs. In addition, well production is constrained using a constant-pressure specification. The ranges of values from which these parameters have been sampled are shown in Table 2.

Parameter	Minimum	Maximum
Formation thickness (ft)	100	300
Non-SRV fracture spacing (ft)	0.1	10
Non-SRV fracture conductivity (md-ft)	0.0001	0.05
Shale porosity (%)	3	13
Shale permeability (md)	1.00×10^{-3}	1.00×10^{-4}
Shale compressibility (1/psi)	1.00×10^{-6}	3.00×10^{-6}
Diffusivity (cm^2/s)	2.00×10^{-5}	8.00×10^{-5}
Langmuir volume (scf/ton)	50	250
Langmuir pressure (psi)	500	1,500
Bulk density (g/cm^3)	2.1	3
Initial gas saturation	0.65	0.85
Initial pressure (psi)	2,000	8,000
Reservoir depth (ft)	5,000	1,000
Initial reservoir temperature ($^{\circ}\text{F}$)	100	350
Flowing bottomhole pressure (psi)	250	1,000

Table 2—Additional reservoir and operational parameters.

Restimulation-Candidate Characterization. Using the flow simulator, the next step is to generate multiple gas-production profiles corresponding to both favorable- and unfavorable-restimulation candidates to serve as input to our classifier-training algorithm. This is achieved using the following steps.

1. Sample multiple combinations of the fracture, reservoir, and operational parameters in Tables 1 and 2.
2. Put the well under constant-pressure production for a fixed number of years before simulating a restimulation event.
3. Compute post-refracturing-performance metrics using the rate response of 1 year after restimulation.
4. Assign a categorical label for a favorable candidate (1) or an unfavorable candidate (-1), depending on the observed production characteristics. Using these labels, we then present the prerefracturing production data to the classification algorithm and train it to predict without having access to the post-refracturing outcome.

As discussed in Jayakumar et al. (2013), there are a number of restimulation-treatment techniques for multistage hydraulically fractured horizontal wells. Fig. 3 depicts the approach we have implemented to model restimulation treatment. As shown in Fig. 3, we insert new hydraulic fractures between existing planar stages to increase contact with the formation within the SRV region. In the absence of detailed fracture-model inputs, we assume the new hydraulic fractures have the same conductivity, complex fracture geometry (Fig. 2), and related fracture properties (Table 1) as the existing hydraulic fractures. This is dependent on an assumption of unchanged operational procedures and rock properties.

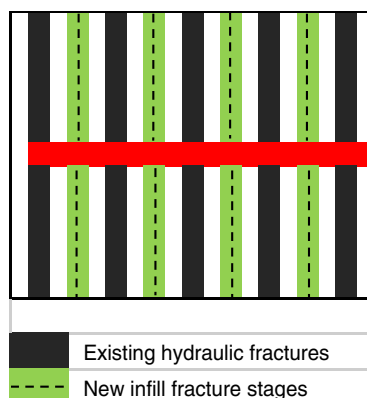


Fig. 3—Modeling of infill restimulation treatment, adapted from Jayakumar et al. (2013). The black slabs represent existing hydraulic-fracture stages, whereas the green slabs are the new infill refracturing stages inserted between planar fractures.

Given an ensemble of restimulated-well-production profiles, the next step is to define favorable- and unfavorable-restimulation-candidate examples to provide a basis for the classification algorithm to learn patterns associated with each category. To achieve this,

we fit an empirical decline model (Arps 1945) to the initial-rate data and forecast the gas rate assuming no restimulation. Using this information, we then estimate the percentage increase in cumulative production 1 year after restimulation, relative to the estimated production in the absence of restimulation. With this approach, we generate a histogram of percentage increase in cumulative production 1 year after restimulation, and apply a threshold at the 50th percentile. This procedure is depicted in Fig. 4.

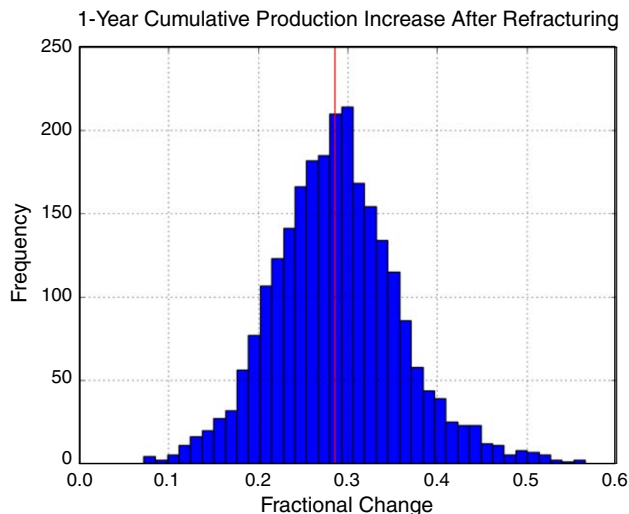


Fig. 4—Distribution of cumulative-production increase 1 year after restimulation, with restimulation taking place after 2 years of production. The vertical line represents the 50th percentile (0.285). All wells with less than 0.285 increase in production are considered unfavorable candidates for this study, and vice versa.

In Fig. 4, all wells exhibiting cumulative production improvement greater than 28.5% are deemed good restimulation candidates. Note that the use of a 50th-percentile threshold is arbitrary in this case, and more care can be taken in defining the threshold to meet specific economic objectives. Fig. 5 displays extreme examples of unfavorable and favorable candidates assigned using this criterion for wells refractured after 2 years.

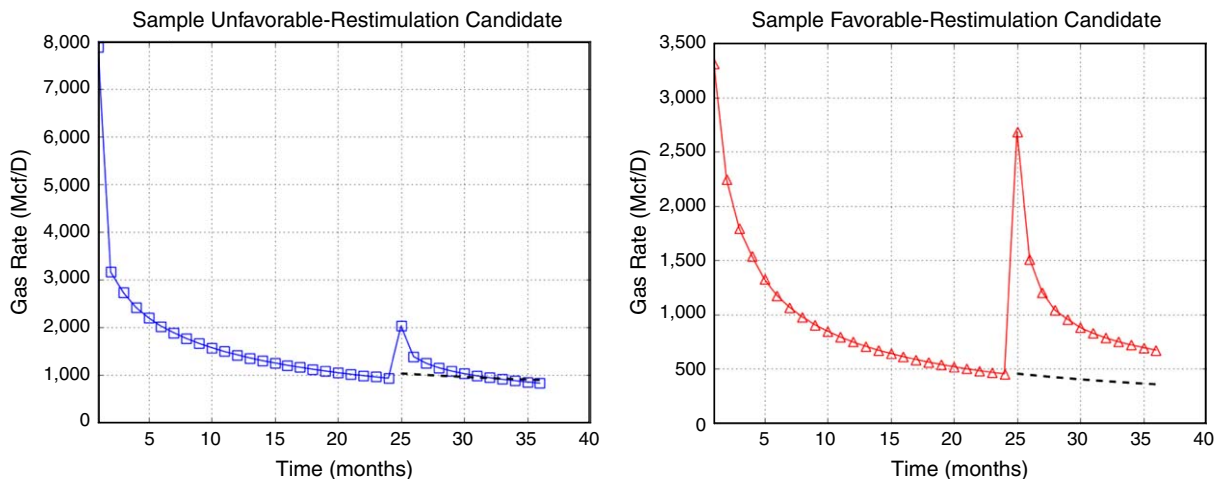


Fig. 5—Examples of (left) unfavorable and (right) favorable restimulation candidates depending on 1-year cumulative-production increase. Solid curves represent the gas-rate data, whereas the dashed curves represent the extrapolated production.

To capture the operational-rate fluctuations typical of field-production data, we have introduced perturbations (in percent) to the SRVs, which are drawn from a Gaussian distribution with a mean of zero and standard deviation of 5%. This serves to provide a more-realistic and more-challenging problem for the pattern-recognition algorithm, compared with the direct use of clean synthetic rate responses like those shown in Fig. 5. Fig. 6 displays sample input production data after incorporating rate fluctuations.

In Fig. 6, we see that the simulated flow rate at each time instance has been perturbed, whereas the overall decline behavior is retained. In line with the criteria for restimulation success depicted in Fig. 4, the well profiles labeled as unfavorable candidates show significantly steeper decline in production after restimulation. In other words, they have been labeled as unfavorable because they fall to less than the 50th-percentile threshold for cumulative production enhancement (28.5%).

Within the defined framework, what specific combination of flow-model parameters contributes to restimulation success? We have first performed a manual inspection of univariate and bivariate flow-parameter distributions used in simulating the training data, and we have observed no discernable distinction between the distributions of favorable and unfavorable candidates. This suggests that there are no individual or simple combinations of physical parameters that can be modified to guarantee restimulation success. Thus, we need to consider a higher-order interrelationship between these variables. To establish a link between the observed outcome of restimulation

and the influencing reservoir properties, we have therefore performed a multivariate analysis of the input variables used in generating the training data set. This has been achieved by use of principal-component analysis to determine the direction in multidimensional space along which we can observe maximal separation between the multivariate distributions of favorable and unfavorable restimulation candidates, depending on 1-year cumulative-production improvement. This is depicted in Fig. 7.

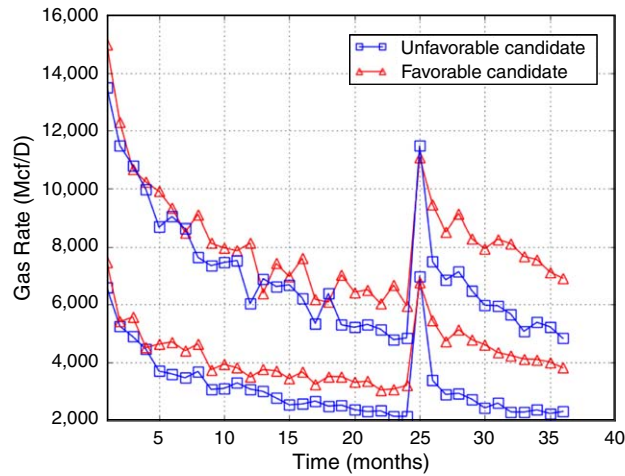


Fig. 6—Sample training data with perturbed flow rates. The unfavorable candidates display a more-dramatic decline in flow rate after restimulation.

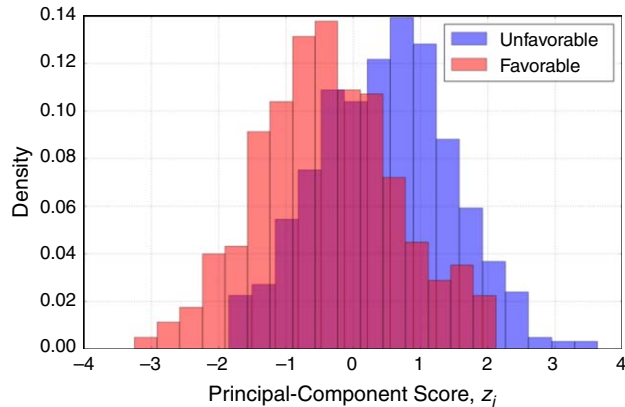


Fig. 7—Distributions of favorable and unfavorable candidates when flow-simulation parameters are projected along the direction of maximal separation in parameter space. The top- and bottom-25th percentiles of favorable and unfavorable candidates are displayed.

In Fig. 7, the fracture, reservoir, and operational parameters in Tables 1 and 2 have been projected along the direction of maximum separation between favorable and unfavorable restimulation candidates in parameter space. The blue and red histograms represent the distribution of principal-component scores for unfavorable and favorable candidates, respectively. Fig. 7 shows the wells in the top- and bottom-25th percentiles of cumulative-production increase (total of 1,250 wells), depending on cumulative-production enhancement. The principal-component score z_i for well i is given by

$$z_i = \sum_{\forall k} l_k \psi_{ki}, \dots \dots \dots (2)$$

where ψ_{ki} represents the flow-simulation parameters, which have been standardized to eliminate the influence of the relative magnitude of each individual parameter (i.e., scaled to yield zero mean). In the summation, l_k is the principal-component-vector-loading coefficient associated with each flow-simulation parameter k . The “loadings” for the flow-simulation parameters used in generating the training data are shown in Table 3. In multivariate flow-parameter space, the loadings shown in Table 3 are the coordinates of the vector along which we observe the largest separation between favorable and unfavorable restimulation candidates, as depicted in Fig. 7. These loadings are sorted in descending order of absolute magnitude. Under the current framework, the significance of Table 3 is that it summarizes the relative influence of the fracture, reservoir, and operational parameters in separating favorable and unfavorable restimulation candidates. At the top of the list, we see that the rate of pressure-dependent hydraulic-fracture closure exerts the highest influence in distinguishing between both categories, in addition to initial gas saturation, hydraulic-fracture-tip conductivity, and spacing. On the other end, the loading-vector coefficients suggest that the restimulation outcome is largely uninfluenced by factors such as reservoir depth, shale-matrix permeability, pressure-dependent closure of complex fractures, and number of hydraulic-fracture stages.

In Fig. 7, we can observe that favorable candidates tend to have more negative principal-component scores, whereas unfavorable candidates tend to have more positive scores. Although caution must be taken in interpreting the signs of the loadings, this suggests that on aggregate, larger values of parameters with negative loading coefficients in Table 3 would lead to increased chance of restimulation

success. However, there are multiple combinations of properties that can yield similar principal-component scores, when we compute the sum of products shown in Eq. 2. This illustrates one potential pitfall of production-data-analysis models and techniques that take into account only a subset of these parameters. In addition, the majority of these input parameters are impossible or expensive to infer accurately in practice. However, given that restimulation success can be quantified depending on rate response, it is clear the rate data directly hold information about the aggregate effect of different combinations of parameters and how their influence is manifested in the outcome of restimulation. For this reason, in the next section, we propose an alternative approach that decomposes production data into a collection of subtle rate-decline characteristics to serve as input to a classification model.

Parameter k	Loading l_k
Hydraulic-fracture-closure rate	0.509
Initial gas saturation	0.432
Hydraulic-fracture-tip conductivity	-0.368
Hydraulic-fracture spacing	-0.286
Langmuir volume	0.213
Diffusivity	-0.204
Shale porosity	-0.204
Reservoir thickness	0.201
Flowing bottomhole pressure	-0.179
Hydraulic-fracture half-length	-0.174
Hydraulic-fracture conductivity	-0.169
Shale compressibility	0.146
Langmuir pressure	-0.132
Initial reservoir temperature	-0.098
Complex fracture conductivity	0.094
Complex fracture spacing	0.078
Initial reservoir pressure	-0.069
Bulk density	0.069
Natural-fracture conductivity	0.047
Natural-fracture spacing	-0.042
Reservoir depth	0.016
Shale permeability	-0.008
Complex fracture closure rate	0.003
Number of hydraulic-fracture stages	-0.001

Table 3—Principal-component loadings for training data. Loadings are sorted in descending order of absolute magnitude.

Finally, to explore the sensitivity of the classification performance to different restimulation-candidate-selection criteria, we have considered the following two additional metrics:

- Restimulation-candidate selection that is dependent on instantaneous percentage increase in flow rate post-refracturing.
- Change in decline rate after restimulation. This is determined by fitting the Arps (1945) hyperbolic decline model to the data before restimulation and after restimulation, and defining the restimulation success depending on the percentage increase in the hyperbolic decline constant.

Table 4 summarizes the criteria used in assigning favorable- and unfavorable-candidate labels to the training production data. As shown, we have also considered two different options for restimulation timing: 2 and 3 years of production before treatment.

Timing of Refracturing Criteria	2 Years			3 Years		
	Minimum	Maximum	Threshold (P50)	Minimum	Maximum	Threshold (P50)
Change in 1-year cumulative production	0.0724	0.565	0.285	0.0390	0.634	0.287
Change in instantaneous rate	0.595	10.135	1.527	0.601	13.055	1.602
Change in decline rate	-0.429	2.767	0.637	-0.133	2.886	0.944

Table 4—Distribution of restimulation-candidate-selection criteria computed from training data. All values are expressed as proportions (unitless).

Production-Data-Classification Framework

In this section, we describe the proposed pattern-recognition framework for analyzing well-production data, motivated by techniques in face detection (Viola and Jones 2001). The overall aim is to develop a fast and robust probabilistic binary classification model that uses

$$x_{ij}(\tilde{q}_i; t, \Delta t, z) = \begin{cases} \sum_{t'=t}^{t+\frac{\Delta t}{2}-1} \tilde{q}_i(t') - \sum_{t'=t+\frac{\Delta t}{2}}^{t+\Delta t-1} \tilde{q}_i(t'), & (z=1), \dots \dots \dots (4a) \\ \sum_{t'=t}^{t+\frac{\Delta t}{3}-1} \tilde{q}_i(t') - \sum_{t'=t+\frac{\Delta t}{3}}^{t+\frac{2\Delta t}{3}-1} \tilde{q}_i(t') + \sum_{t'=t+\frac{2\Delta t}{3}}^{t+\Delta t-1} \tilde{q}_i(t'), & (z=2), \dots \dots \dots (4b) \end{cases}$$

where

$$\Delta t = \begin{cases} 2, 4, 6, \dots, t_f, & (z=1), \\ 3, 6, 9, \dots, t_f, & (z=2). \dots \dots \dots (5) \end{cases}$$



Feature Template, z	Geometry
1	
2	

Table 5—Features used for production-data classification. To compute a feature score for each geometry, the sum of rates in the white region is subtracted from the sum of rates in the gray region. These templates are shifted and scaled throughout the well profile, and a feature score is computed in each case.

In Eqs. 4 and 5, the index $j = 1, \dots, p$ represents a unique combination of feature geometry, time, and scale. For example, we can extract a total of $p = 236$ feature combinations from a 2-year well profile with monthly sampled data (i.e., $t_f = 24$).

Intuitively, these texture-based features adapted from image/object-detection applications can be considered analogous to rate derivatives, which have traditionally been used in rate-transient analysis to infer flow-regime information from well data. In other words, using basic forward-difference approximations,

$$\left. \frac{d\tilde{q}_i}{dt'} \right|_{t'} \approx \frac{\tilde{q}_i(t') - \tilde{q}_i(t' + \Delta t')}{\Delta t'}, \dots \dots \dots (6)$$

$$\left. \frac{d^2\tilde{q}_i}{(dt')^2} \right|_{t'} \approx \frac{\tilde{q}_i(t') - 2\tilde{q}_i(t' + \Delta t') + \tilde{q}_i(t' + 2\Delta t')}{(\Delta t')^2}. \dots \dots \dots (7)$$

Comparing Eq. 6 with Eq. 4a, we can see that features computed using the two-rectangle template ($z = 1$) are related to the first derivative of normalized flow rate. Furthermore, if we consider the summation terms in Eq. 4a as upscaling operations that coarsen the sampling interval from one unit of t to a larger timestep of $\frac{\Delta t}{2}$, then Eq. 6 collapses to Eq. 4a because $\Delta t' = 1$ at this coarser scale. This means that when we compute multiple combinations of these features for a given well profile at different t and Δt , we are effectively decomposing the signal into a series of rate derivatives at multiple time periods and scales. Importantly, this implies that the two-rectangle features carry information about flow regimes experienced over the productive life of the well. Similarly, if we compare Eq. 7 with Eq. 4b, we see that the three-rectangle features ($z = 1$) are analogous to the second derivative of the normalized rate. This attribute captures the curvature characteristics at different time periods and scales within the well data. One advantage of considering each of these features at multiple scales is that it helps to mitigate the effect of instantaneous fluctuations in rate or measurement error, making the method robust to noise.

In big-data applications involving a significant amount of finely sampled well data, it becomes more important to minimize the computational cost incurred in calculating these feature scores. To achieve this, we can take advantage of the simple geometries of these feature templates. In other words, we can precompute the cumulative normalized rate for each well at each time instance using

$$\tilde{Q}_i(t) = \sum_{t'=0}^t \tilde{q}_i(t'). \dots \dots \dots (8)$$

The summation $\tilde{Q}_i(t)$ is calculated for all $t = 1, \dots, t_f$. After precomputing $\tilde{Q}_i(t)$ for each well, we can then easily determine all feature scores at each time t and scale Δt using

$$x_{ij}(t, \Delta t, z) = \begin{cases} 2\tilde{Q}_i\left(t + \frac{\Delta t}{2} - 1\right) - \tilde{Q}_i(t - 1) - \tilde{Q}_i(t + \Delta t - 1), & (z=1) \\ 2\tilde{Q}_i\left(t + \frac{\Delta t}{3} - 1\right) - \tilde{Q}_i(t - 1) - 2\tilde{Q}_i\left(t + \frac{2\Delta t}{3} - 1\right) + \tilde{Q}_i(t + \Delta t - 1), & (z=2). \dots \dots \dots (9) \end{cases}$$

In other words, the summation operations in Eqs. 4a and 4b have been represented as a set of basic arithmetic operations in Eq. 9, thus minimizing the computational expense incurred through unnecessary iterations. Note that $\tilde{q}_i(0) = \tilde{Q}_i(0) = 0$. By translating each

feature template in time at different scales, the final output from this step is an exhaustive set of feature scores $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]$ for each well $i = 1, \dots, N$, where p is the total number of feature combinations available within each production profile.

Statistical Framework. Given a data set made of p Haar-like features scores extracted from N wells, the next step is to define a statistical basis for identifying restimulation candidates. In a manner similar to Viola and Jones (2001), we have framed this problem using a binary classification approach, and extended the methodology to extract probability estimates. In other words, instead of explicitly forecasting a real-valued flow rate after restimulation, the predictive model generates a categorical response (favorable or unfavorable restimulation candidate) along with an estimated probability of refracturing success.

The engine of the binary classification framework used in the Viola and Jones (2001) procedure is a statistical learning method known as AdaBoost. The algorithm takes in a list of feature scores $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}]$ and a categorical outcome vector $y_i \in \{-1, 1\}$ for all training well-production data $i = 1, \dots, N$ and features $j = 1, \dots, p$. In the vector y_i , -1 and 1 are categorical labels that denote unfavorable and favorable candidates, respectively. For each unique feature j , we can collate the feature scores computed from all training data and use y_i to develop a simple binary classification model depending on a single threshold b_j . This is known as a single-node binary decision tree (decision stump), as depicted in Fig. 9. The threshold b_j is selected by minimizing an impurity function known as the Gini index (Breiman et al. 1984), which serves to minimize the number of misclassifications with respect to the training data. This fitting procedure can be performed using common statistical software packages.

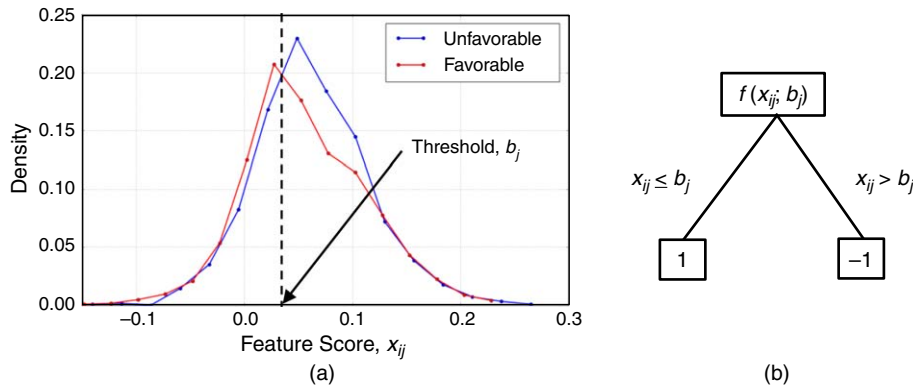


Fig. 9—(Left) Distribution of scores for feature $j = 1$ computed with respect to all wells in the training set. b_j is the optimal threshold that minimizes overlap between favorable and unfavorable categories. If x_{ij} (the feature score computed with respect to well i and feature index j) is greater than threshold b_j , the well is assigned a categorical label of -1 (unfavorable candidate), and vice versa. (Right) Depiction of the decision-stump-classification model.

In Fig. 9, we denote the decision stump trained using feature j as $f(x_{ij}; b_j) \in \{-1, 1\}$. All data with feature scores greater than b_j are predicted as unfavorable candidates, and vice versa. Note that the inequality can be reversed depending on the distribution of feature scores. For each candidate feature $j = 1, \dots, p$, we can quantify the performance of each decision stump by computing the weighted classification error ϵ_j with respect to the training data using (Hastie et al. 2009)

$$\epsilon_j = \frac{\sum_{i=1}^N w_i \cdot I[f(x_{ij}) \neq y_i]}{\sum_{i=1}^N w_i} \dots \dots \dots (10)$$

In Eq. 10, ϵ_j lies between zero and unity and w_i represents the weights assigned to each well-production profile, which are equal by default. $I[f(x_{ij}) \neq y_i]$ is an indicator function that returns zero if the prediction matches the true categorical label and returns unity in the case of a mismatch. As is visible in Fig. 9a, there is a significant amount of overlap between the distributions of favorable and unfavorable candidates, which would lead to a large number of misclassifications. Because of the simplistic nature of these single-threshold classification rules, they are in practice expected to yield only slightly better predictions than a random guess (Friedman et al. 2000). For this reason, they are referred to as “weak classifiers.” In contrast, the AdaBoost algorithm uses a sequential fitting procedure to combine or “boost” the poor predictions from these individual decision stumps into a more-accurate classifier. The AdaBoost training procedure is summarized in Table 6.

As shown in Table 6, for a given number of iterations $m = 1, \dots, M$, the overall strategy is to train one individual decision stump $f(x_{ij})$ for each of p available features; compute the associated weighted error ϵ_j using Eq. 10; and then select the weak classifier f_m that minimizes the weighted classification error ϵ_m at the current iteration m . The weighted error is used to compute a “voting weight” α_m , which is indirectly proportional to the error. This weak classifier is then reapplied to the training data, and misclassified examples are assigned higher weights for the next iteration, $w_i^{(m+1)}$. The final outcome of the training procedure is a list of the weak classifiers f_m selected after each iteration and their corresponding “voting weights” α_m for $m = 1, \dots, M$, which are linearly combined to return a categorical prediction as

$$\hat{y} = \text{sign} \left[G(\mathbf{x}_i) = \sum_{m=1}^M (\alpha_m f_m) - \theta \right] \dots \dots \dots (11)$$

In Eq. 11, $G(\mathbf{x}_i)$ returns a real number between $-\infty$ and ∞ . As indicated, the sign of the result gives the predicted category $\hat{y} \in \{-1, 1\}$, where 1 and -1 denote favorable and unfavorable candidates, respectively. In other words, if we evaluate $G(\mathbf{x}_i)$ for any arbitrary well-production profile, a negative value will be assigned to unfavorable candidates and a positive value will be assigned to favorable candidates. The variable θ is a tuning parameter that is zero by default, and will be discussed later. The intuitive explanation of Eq. 11 is that each weak classifier f_m has received a weight α_m that is proportional to its individual accuracy in distinguishing between

favorable and unfavorable candidates. In this way, the predictions from M weak classifiers are combined by means of a “committee vote” (Freund and Schapire 1996). The AdaBoost fitting procedure shown in Table 6 can be performed using common open-source statistical software.

Input: N training data
<ol style="list-style-type: none"> 1. $N/2$ examples each of favorable and unfavorable cases 2. Feature scores, $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{ip}] \forall i = 1, \dots, N$ 3. Categorical labels, $y_i = \{-1, 1\}$ (i.e., unfavorable and favorable cases, respectively)
Procedure:
<ol style="list-style-type: none"> 1. Begin with equal weights for first iteration, $w_i^{(m)} = 1 / N$ 2. For each iteration m, <ol style="list-style-type: none"> a. Normalize weights, $W_i^{(m)} = \frac{w_i^{(m)}}{\sum_{i=1}^N w_i^{(m)}}$ b. Determine weak classifier f_m that minimizes error, $\epsilon_m = \frac{\sum_{i=1}^N W_i^{(m)} I(f_m \neq y_i)}{\sum_{i=1}^N W_i^{(m)}}$ c. Compute expansion coefficient for current iteration, $\alpha_m = \ln\left(\frac{1 - \epsilon_m}{\epsilon_m}\right)$ d. Upweight the misclassified data for the next round, $w_i^{(m+1)} = w_i^{(m)} \exp[\alpha_m I(f_m \neq y_i)]$ 3. Final classifier prediction, $\hat{y} = \text{sign}\left[\sum_{m=1}^M (\alpha_m f_m) - \theta\right]$

Table 6—AdaBoost classification algorithm modified from Hastie et al. (2009).

There are a number of properties that make AdaBoost a model for our problem. First, the training algorithm implements a weighting procedure (Step 2d in Table 6) in which misclassified examples are assigned higher importance for the next iteration. It can be shown that these weights serve to minimize an exponential loss function, which is monotonic and smooth. For this reason, AdaBoost is empirically considered to be robust to overfitting (Friedman et al. 2000), in contrast to methods such as ANN and support-vector machines. Second, we can typically build a strong AdaBoost classifier using $M < p$ iterations. In other words, it functions as a natural feature-selection process, in which the more-discriminative features receive higher voting weights α_m . This greatly mitigates the computational cost associated with calculating all p features during training and also while assessing unseen candidate wells, which becomes more significant with larger data sets. Third, the additive model $G(\mathbf{x}_i)$ can be probabilistically interpreted. Following a log odds-ratio (logit) model or its inverse, the antilogit probabilistic model, the probability of an arbitrary data sample belonging to either category (1 or -1) can be expressed as (Friedman et al. 2000)

$$P(y = 1 | \mathbf{x}_i) = \frac{e^{G(\mathbf{x}_i)}}{e^{G(\mathbf{x}_i)} + e^{-G(\mathbf{x}_i)}}, \dots \dots \dots (12)$$

$$P(y = -1 | \mathbf{x}_i) = \frac{e^{-G(\mathbf{x}_i)}}{e^{-G(\mathbf{x}_i)} + e^{G(\mathbf{x}_i)}}. \dots \dots \dots (13)$$

Using this information, we can therefore scan an ensemble of production profiles and estimate the probability that each well is either a favorable or an unfavorable candidate. This provides a means to quantify the uncertainty in predicting each well to be a favorable or an unfavorable restimulation candidate. If the probabilities are not very discriminatory, the procedure would suggest that the observed production information is inadequate for making reservoir-development decisions, and other sources of information need to be investigated to lead to more-robust decisions. In addition, by using these probability estimates, we have a basis for ranking candidate wells.

The final step in the overall training framework for the production-data classifier is the cascade-training procedure (Viola and Jones 2001). The overall idea is to train a series of AdaBoost models that are increasingly more discriminative. In other words, in each successive stage, more iterations are dedicated to improving the overall classification accuracy. In this way, wells that display strongly unfavorable characteristics are weeded out early in the process, which means that we only need to compute a much smaller subset of Haar-like features. This has the potential to lead to significant time savings in applications involving massive data sets. The cascaded classifier structure is depicted in Fig. 10.

In Fig. 10, during the cascade-prediction protocol for any arbitrary well profile, we run the AdaBoost model in each successive stage by evaluating Eq. 11. At each stage, we only proceed to the next stage if a value of unity is predicted. The probability of restimulation success can be estimated at the stage of termination using Eq. 12.

The cascade model shown in Fig. 10 is trained by tuning the true-positive and false-positive classification ratios in each successive AdaBoost model to impart a desired level of accuracy in predicting favorable and unfavorable candidates. In the binary classification literature, the true-positive ratio represents the proportion of data that belong to the favorable category (1) and are correctly predicted. On the other hand, the false-positive ratio refers to the proportion of data that belong to the unfavorable category (-1) and are erroneously predicted as members of the favorable category. In Eq. 11, when we take the sign of $G(\mathbf{x}_i)$ as a basis for assigning data to classes -1 or 1, an AdaBoost binary classification threshold of zero is implied. The θ parameter, which is zero by default, can therefore be decreased to ensure more training data are assigned to the favorable class (i.e., increase true-positive ratio) and vice versa. In the limit, if we set $\theta = -\infty$, all data will be assigned to the favorable class, with true-positive and false-positive ratios both equaling unity. In the cascade-training process, we take advantage of θ in tuning each successive AdaBoost model to accept increasingly fewer false-positive classifications while simultaneously maximizing the accuracy of true-positive classification.

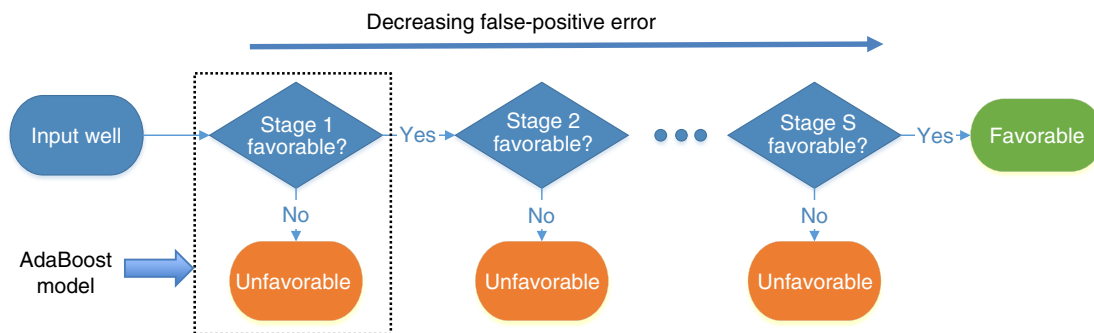


Fig. 10—Cascaded-classifier-prediction protocol, modified from Viola and Jones (2001). Each cascade stage contains one AdaBoost model. Each successive stage is increasingly more discriminative (i.e., lower false-positive error). If at any stage an unfavorable candidate is predicted, the process is terminated. The probability of restimulation success can then be estimated at the stage of termination using Eq. 12.

The pseudocode for the cascaded classifier-training algorithm is presented in **Table 7**. In the table, F_{ps} and D_{ps} are user-defined multiplicative factors used to control the false-negative and false-positive ratios per cascade stage, whereas the optional parameter n is used to control the complexity of the AdaBoost model trained in each stage. In addition to improved computational efficiency, this approach has the potential to serve as a flexible training procedure that can be modified to prioritize the classification accuracy of favorable or unfavorable wells depending on the economic objectives.

Input:

1. Pool of K training data with class labels, where $K \gg N$
 - a. Training/testing-data split ratio, N/V , where $N + V = 1$
 2. User-defined objectives
 - a. Maximum false-positive ratio per stage, F_{ps}
 - b. Minimum true-positive ratio per stage, D_{ps}
 - c. Number of features per stage multiplier, n (optional)
-

Procedure:

1. Randomly select $N/2$ favorable-training cases and V test samples from pool
 2. Initialize $F_0 = 1$, $D_0 = 1$, $s = 0$
 3. While $F_s > 1 - D_s$,
 - a. Increment stage $s = s + 1$; initialize number of iterations, $m = 0$
 - b. While $F_s > F_{ps} \times F_{s-1}$ or $m \leq n \times s$,
 - i. Begin new AdaBoost iteration, $m = m + 1$
 - ii. Train AdaBoost classifier with selected training samples
 - iii. While $D_s \leq D_{ps} \times D_{s-1}$,
 - Decrease the threshold θ of the classifier at current stage s
 - c. Evaluate new cascade classifier on negative examples in the training pool and select $N/2$ false positives as the new unfavorable-training cases
-

Table 7—Cascade-training procedure adapted from Viola and Jones (2001).

Results and Discussion

In this section, we present the results of model training and evaluation for the proposed production-data-classification algorithm. Next, we compare this with the performance of a type-curve approach in identifying restimulation candidates. Finally, we show the results of a field-validation study using the proposed pattern-recognition approach.

Classifier Training and Evaluation. Using the generated training production data along with categorical labels denoting favorable (1) and unfavorable (−1) restimulation candidates, we have applied the proposed Viola and Jones (2001) framework to train a production-data-classification model. The training and evaluation procedure involves the following steps:

1. Given a collection of training data, compute feature scores for all combinations of Haar-like feature geometry, scale, and starting time within the individual well profiles.
2. Train the AdaBoost cascade model with a random sample of N well profiles, using feature scores computed with respect to the data before restimulation.
3. Assess the model predictive performance by scanning V wells and comparing the predicted restimulation outcome with the known outcome.

Table 8 displays the cascade-training parameters we have implemented.

Number of well-production profiles used for training, N	400
Number of well-production profiles used for testing, V	100
True-positive ratio per stage, D_{ps}	0.97
False-positive ratio per stage, F_{ps}	0.90
Features per stage multiplier, n	10

Table 8—Cascaded-classifier-training parameters.

As previously discussed, in defining restimulation success, we have considered three separate candidate-selection criteria. In each case, to capture the effect of the timing of refracturing, we have used either 2 or 3 years of production data before the restimulation event. After the training process, the predictive performance of each of these classifiers has been tested by assessing a synthetic test set containing $V = 100$ well-production profiles, which were not available to the training algorithm. **Table 9** summarizes the results of the training exercise for the production-data classifier and their predictive performance with respect to synthetic data.

Criteria	2 Years			3 Years		
	Accuracy (%)	Cascade Stages	Training Time (seconds)	Accuracy (%)	Cascade Stages	Training Time (seconds)
Change in 1-year cumulative production	83	13	118	81	11	215
Change in instantaneous rate	78	13	375	76	14	831
Change in decline rate	84	10	115	82	13	422

Table 9—Results for production-data-classifier training and performance with respect to synthetic test data.

The results in Table 9 show that regardless of the restimulation-candidate-selection criteria, we are able to distinguish between favorable and unfavorable candidates with accuracy of 76 to 84%. By comparison, a monolithic (single-stage) AdaBoost classifier yields an overall classification accuracy of 76% (for 200 iterations) in the case involving 2 years of production, depending on the cumulative-production-improvement criterion. The results in Table 9 suggest that improvement in instantaneous rate after restimulation is the least reliable indicator of restimulation success, as highlighted by the lower classification accuracy for both 2- and 3-year cases.

In Table 9, we see that the trained classifiers have 10 to 14 cascade stages. We can observe that the instantaneous-rate-change criterion requires more training time for both 2- and 3-year cases. This can be attributed to that fact that more AdaBoost iterations are required in each cascade stage to meet the convergence criteria shown in Step 3b of Table 7. This is further evidence that instantaneous increase in flow rate after restimulation is the least discriminative metric for restimulation success. Nevertheless, in all cases, although training might potentially be computationally intensive, the final trained classifier is able to provide predictions on the test set in near real time. For example, the classifier derived from the cumulative-production-improvement criterion can make predictions on 100 samples of 2-year well production in 0.35 seconds. All reported run times are derived from a 64-bit system with 3.4-GHz processor and 32 GB RAM.

As previously discussed, only a subset of all Haar-like feature combinations within a well-production profile is used in the AdaBoost classification model within each cascade stage. In other words, the sequential fitting procedure described in Table 6 is aimed at selecting the most discriminative features that best enable us to distinguish between restimulation candidates. Therefore, by looking at the features selected by the AdaBoost model at each stage, we can gain additional insight into the mode of operation of the proposed classification scheme. **Tables 10 through 12** summarize the features selected by the trained model that uses 2 years of data before the restimulation event in classifying wells on the basis of cumulative production improvement.

First, from Table 10, we see that within all cascade stages, the majority of features selected are two-rectangle templates ($z = 1$), with relatively fewer occurrences of three-rectangle templates ($z = 2$). This suggests that features relating restimulation success to first-order rate derivatives at different scales and time periods, which help to highlight changes in flow regimes, are more discriminative than those that consider curvature-related information. Second, from Table 11, we see that across all cascade stages, features evaluated within the first 6 months of production have been selected more frequently than those evaluated afterward. This suggests that the most-informative rate-decline characteristics useful in identifying restimulation candidates are available at early periods in the life of the well. Third, as shown in Table 12, features with maximum scale of 6 months offer the most discriminative information across different cascade stages. Overall, larger features evaluated at later periods tend to offer limited useful information in discerning between favorable and unfavorable restimulation candidates.

Comparison With Type-Curve Approach. To compare the performance of our classification model with that of traditional approaches, we have analyzed our synthetic well data using the trilinear-flow solution proposed by Brown et al. (2011). We have selected the trilinear model as a representative analytical solution among a number of existing type-curve methods developed specifically for multistage hydraulically fractured shale wells. In **Fig. 11a**, we have plotted the prerefracturing production data on dimensionless axes according to the trilinear analytical solution. The characteristic dimensionless quantity P_{wD} is proportional to the rate-normalized pseudopressure drop, and takes into consideration several well, reservoir, fluid, and dual-permeability fracture parameters.

In **Fig. 11a**, we have plotted the rate data on $1/P_{wD}$ vs. t_D . We take the inverse of P_{wD} to put the data in terms of pressure-normalized rate. In addition, t_D has been computed using material-balance pseudotime. These steps, as are standard in rate-transient analysis, serve to account for nonconstant rate and nonlinear gas properties as a function of pressure. To plot the well data, information on hydraulic-fracture properties and geometry, fluid and reservoir properties, and pressure information are required. In addition, given that the solution is developed in Laplace space, a numerical technique such as the Stehfest (1970) algorithm is required to represent the results in the time domain.

Stage	Ratio of Two-Rectangle Features ($z = 1$)	Ratio of Three-Rectangle Features ($z = 2$)
1	1.00	0.00
2	0.67	0.33
3	0.75	0.25
4	0.88	0.13
5	0.86	0.14
6	0.88	0.13
7	0.94	0.06
8	0.90	0.10
9	0.88	0.13
10	0.94	0.06
11	0.67	0.33
12	0.75	0.25
13	1.00	0.00
Average	0.85	0.15

Table 10—Proportions of features selected in each cascade stage, depending on the feature template.

Stage	Ratio of Features Evaluated at $t = [1,6]$ Months	Ratio of Features Evaluated at $t = [7,12]$ Months	Ratio of Features Evaluated at $t = [13,18]$ Months	Ratio of Features Evaluated at $t = [19,24]$ Months
1	0.44	0.11	0.22	0.22
2	0.67	0.00	0.00	0.33
3	0.25	0.38	0.38	0.00
4	0.40	0.23	0.20	0.18
5	0.43	0.14	0.29	0.14
6	0.50	0.38	0.13	0.00
7	0.29	0.26	0.35	0.10
8	0.47	0.20	0.27	0.07
9	0.63	0.00	0.38	0.00
10	0.35	0.35	0.24	0.06
11	1.00	0.00	0.00	0.00
12	0.00	0.75	0.00	0.25
13	0.25	0.00	0.50	0.25
Average	0.44	0.21	0.23	0.12

Table 11—Proportions of features selected in each cascade stage, depending on the time instance t at which the features are evaluated.

One potential work flow using this type-curve approach would be to select restimulation candidates depending on the qualitative rate-related criteria inferable from the dimensionless axes, such as cumulative production. In other words, the wells with similar cumulative production profiles observed on the dimensionless axes would be grouped together and classified as favorable or unfavorable candidates. For example, on this basis, curves clustered toward the top right might be considered favorable candidates, whereas those located toward the bottom left might be considered unfavorable candidates. However, this approach can potentially lead to misclassifications and consequently missed opportunities, as evidenced by the significant overlap between unfavorable and favorable candidates in Fig. 11a. To illustrate this, we consider the two highlighted wells in Fig. 11a and observe their post-refracturing time-domain response in Fig. 11b. Judging from the steeper production decline shown by the lower curve (blue) in Fig. 11b after refracturing, we see that the higher curve (red) has been missed as a favorable restimulation candidate.

We can investigate the combination of fracture, reservoir, and operational parameters that have led to these misclassifications by revisiting the previously discussed principal-component-analysis results. The flow-simulation parameters corresponding to both wells and the resulting principal-component data are displayed in Table 13. Overall, the properties used in simulating the false-positive well in Fig. 11b (blue) yield a principal-component score of -0.437 , whereas those used in simulating the false-negative well (red) yield a score of 1.398 . Looking at Fig. 7, we see that these scores both fall within the overlapping region between favorable and unfavorable candidates, which is characterized by high uncertainty in determining the outcome of restimulation. This can be attributed to the counterbalancing effect of competing reservoir influences. Specifically, if we consider the flow-model parameters used in simulating the

false-positive case, we see that it is characterized by relatively tight hydraulic-fracture spacing and low reservoir pressure, according to the uncertainty ranges displayed in Tables 1 and 2. However, these are mitigated by factors such as a relatively gentle rate of pressure-dependent hydraulic-fracture closure, high hydraulic-fracture conductivity, and half-length. As shown in Table 13, when we evaluate the net weighted sum of these competing reservoir influences (using Eq. 2) along the vector of maximal separation between favorable and unfavorable candidates in multivariate parameter space, the aggregate effect approaches net zero (i.e., $z_i = -0.437$). Similarly, in the false-positive case, the influences of high reservoir pressure and half-length are counteracted by poor completion, evidenced by poor hydraulic-fracture spacing and a more-dramatic rate of pressure-dependent hydraulic-fracture closure.

Stage	Ratio of Features With Scale $\Delta t = [1,6]$ Months	Ratio of Features With Scale $\Delta t = [7,12]$ Months	Ratio of Features With Scale $\Delta t = [13,18]$ Months	Ratio of Features With Scale $\Delta t = [19,24]$ Months
1	0.44	0.11	0.22	0.22
2	0.67	0.00	0.00	0.33
3	0.25	0.38	0.38	0.00
4	0.40	0.23	0.20	0.18
5	0.43	0.14	0.29	0.14
6	0.50	0.38	0.13	0.00
7	0.29	0.26	0.35	0.10
8	0.47	0.20	0.27	0.07
9	0.63	0.00	0.38	0.00
10	0.35	0.35	0.24	0.06
11	1.00	0.00	0.00	0.00
12	0.00	0.75	0.00	0.25
13	0.25	0.00	0.50	0.25
Average	0.44	0.21	0.23	0.12

Table 12—Proportions of features selected in each cascade stage, depending on the feature scale Δt .

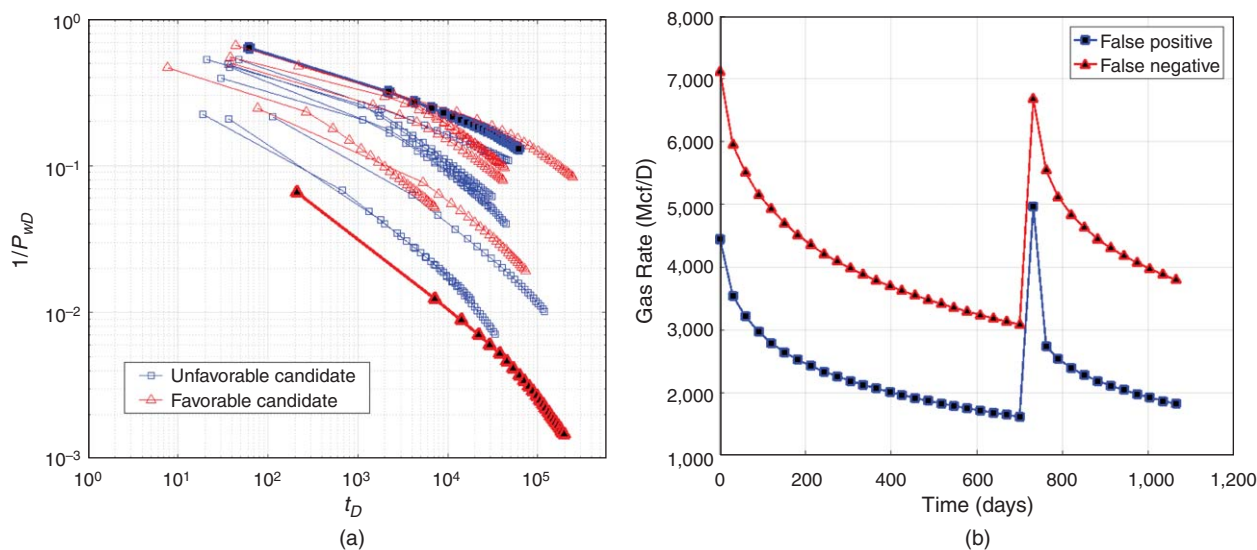


Fig. 11—Restimulation-candidate selection using type-curve analysis (i.e., trilinear-flow solution) (Brown et al. 2011). (Left) Well data plotted on dimensionless axes. (Right) Examples of wells that would be misclassified if restimulation candidates were selected depending on qualitative criteria observable on dimensionless axes, such as cumulative production.

Overall, these results demonstrate that the combination of reservoir and well parameters and their influence on the rate response can render it difficult to gauge good restimulation candidates solely using traditional data-analysis methods such type-curve approaches. In other words, we are not able to resolve the influence of competing reservoir effects on phenomena such as restimulation success by visualizing rate data on dimensionless axes, such as those that take into account a limited set of fracture, fluid, and reservoir properties. On the other hand, by decomposing the raw production data into a set of attributes that are related to rate derivatives at multiple time periods and scales within the well life, we are able to train a probabilistic classification model to better capture the distinction between favorable and unfavorable candidates. More-detailed modeling to incorporate geomechanics and a wider array of well-completion strategies is required to completely understand the full range of reservoir parameters that affect restimulation-candidate selection. However, these results depict the manner in which competing reservoir influences can be masked, and show that we are able to better capture the multivariate separation between restimulation candidates using flow-rate-derived features.

Parameter k	False-Positive Case		False-Negative Case	
	Value	$l_k \psi_{ki}$	Value	$l_k \psi_{ki}$
Hydraulic-fracture-closure rate	3.37	-0.548	4.19	0.169
Initial gas saturation	0.797	0.343	0.801	0.377
Hydraulic-fracture-tip conductivity	19.38	-0.518	5.56	-0.091
Hydraulic-fracture spacing	120	0.378	120	0.378
Langmuir volume	86.04	-0.232	54.56	-0.348
Diffusivity	5.27×10^{-5}	-0.024	6.26×10^{-5}	-0.142
Shale porosity	8.89×10^{-2}	-0.065	4.26×10^{-2}	0.264
Reservoir thickness	253	0.187	276	0.268
Flowing bottomhole pressure	490.34	0.107	451.83	0.139
Hydraulic-fracture half-length	570	-0.253	510	-0.151
Hydraulic-fracture conductivity	28.24	-0.239	11.90	0.239
Shale compressibility	2.57×10^{-6}	0.145	2.25×10^{-6}	0.062
Langmuir pressure	580.28	0.186	586.41	0.183
Initial reservoir temperature	325.02	-0.136	149.35	0.102
Complex fracture conductivity	11.05	0.148	3.27	-0.105
Complex fracture spacing	0.519	-0.110	1.561	0.048
Initial reservoir pressure	2455.30	0.102	6462.52	-0.060
Bulk density	2.955	0.106	2.615	0.014
Natural-fracture conductivity	4.43×10^{-3}	-0.065	3.02×10^{-2}	0.019
Natural-fracture spacing	2.948	0.031	2.184	0.042
Reservoir depth	8825	0.014	5840	-0.018
Shale permeability	3.007×10^{-4}	0.008	1.856×10^{-4}	0.011
Complex fracture closure rate	4.26	0.001	4.07	0.000
Number of hydraulic-fracture stages	7	0.000	9	-0.001
Principal-component score, z_i	-	-0.437	-	1.398

Table 13—Principal-component-analysis data corresponding to false-positive and false-negative cases identified in type-curve analysis. $l_k \psi_{ki}$ represents the product of the standardized flow parameter and the corresponding principal-component loading (see Eq. 2).

Validation With Field Data. Finally, we have validated the proposed approach using publicly available field production data. To achieve this, we have considered 17 hydraulically fractured gas wells from Freestone County within the Haynesville/Bossier shale play in east Texas. The sample size is limited by the rigorous screening criteria that we have applied to ensure that the field data match the simulated conditions as closely as possible. Specifically, each selected case is a hydraulically fractured horizontal shale-gas well that has undergone restimulation within the first 2 to 3 years of production. Each restimulation event has been corroborated by completion reports that show evidence of reperforation within the same formation. Also, to limit the effect of variable completion procedures, the selected wells have been constrained to the same operator within the same county. The production data from the wells used in this study are displayed in Appendix A.

To quantitatively define whether each of these wells underwent successful restimulation, we have extrapolated the flow rate from the primary production period using the Arps (1945) hyperbolic model and estimated the percentage increase in cumulative production 1 year after restimulation. On this basis, the field-production profiles shown in Appendix A have been defined as favorable restimulation candidates, whereas those in Appendix A have been defined as unfavorable candidates for restimulation. These class assignments have been determined using the thresholds specified in Table 4.

Next, we have attempted to predict suitability for restimulation treatment in each of these wells by making predictions solely using the gas-rate data before the refracturing events. To achieve this, we have assessed each of these wells using the cascaded classifiers that were trained with the cumulative production improvement 1 year after restimulation. As previously mentioned, the selected wells were each refractured after approximately 2 or 3 years of production. Wells that were refractured close to the second year of production have been assessed using the trained 2-year classifier, whereas those refractured at approximately the third year have been assessed using the trained 3-year classifier. Because of the limited availability of field examples, missing production data have been estimated using the Arps (1945) empirical relationship where necessary. For example, because Well 7 in Appendix A has 21 months of available data, we have extrapolated the final 3 months of production in order to provide 2 complete years of production to the algorithm, so as to maintain consistency with the trained 2-year classifier. In addition, given that the field data are reported in terms of monthly production volumes, we have used the Arps (1945) empirical relationship to estimate the instantaneous daily production at the start of each month and to maintain uniformity with the output of the flow simulator. To account for the time-varying hyperbolic decline constant, we have achieved this by fitting the field data in intervals of 6 months.

Table 14 displays the results of evaluating the field cases using the trained classifiers. The assumed timing of restimulation for each field case, the class labels we have assigned depending on the increase in cumulative production after refracturing, the predictions made by the cascaded classifier, and the estimated probability of restimulation success are shown. From these results, we see that 12 of 17 of the field-production profiles have been correctly classified using the defined criteria. In other words, using only gas-rate data before restimulation, we are able to correctly predict the outcome of restimulation with approximately 71% accuracy. This is a statistically significant result, according to the single-proportions t -test (p -value = 0.02). Overall, these results demonstrate the merit of the proposed

classification approach in analyzing production data from hydraulically fractured gas wells to assist reservoir management and decision making. In contrast to existing approaches that aim to directly predict gas production after restimulation, this classification method returns a categorical outcome and an accompanying probability of restimulation success or failure. The computation time required to assess all 17 field case was instantaneous. The estimated probabilities of fracture success reported in Table 14 are derived from the multivariate distribution of feature scores using the synthetic training-data set. In other words, they are influenced by the similarity between the rate-decline characteristics exhibited by the field data and the synthetic examples leading up to the restimulation event. The predictive performance of the trained models therefore suggest overall agreement between the synthetic rate profiles used in training and the shale-gas production observed in these field cases.

Well	Assumed Timing of Restimulation (years)	Change in 1-Year Cumulative Production	Assigned Category	Predicted Category	Estimated Probability of Restimulation Success
1	3	0.170	Unfavorable	Unfavorable	0.32
2	3	0.224	Unfavorable	Favorable	0.92
3	2	0.462	Favorable	Unfavorable	0.31
4	3	0.451	Favorable	Favorable	0.92
5	2	0.466	Favorable	Favorable	0.97
6	3	0.288	Favorable	Favorable	0.98
7	2	0.680	Favorable	Favorable	0.97
8	3	0.405	Favorable	Favorable	0.92
9	2	0.256	Unfavorable	Unfavorable	0.31
10	3	0.498	Favorable	Unfavorable	0.36
11	2	0.374	Favorable	Favorable	0.97
12	3	0.373	Favorable	Favorable	0.93
13	3	0.382	Favorable	Favorable	0.78
14	2	0.665	Favorable	Unfavorable	0.31
15	3	0.281	Unfavorable	Unfavorable	0.36
16	3	0.136	Unfavorable	Favorable	0.98
17	2	0.605	Favorable	Favorable	0.98

Table 14—Results of field study to validate the classification approach.

From the results in Table 14, we can observe that when we run our classification model on field data, 7 of 10 wells restimulated at approximately the 3rd year have been accurately predicted, whereas 5 of 7 of the wells restimulated after 2 years have been accurately predicted. This is consistent with the assessment of synthetic data in Table 9, where we see slightly better performance in predicting 2-year cases than 3-year cases. **Table 15** displays two confusion matrices that describe the classification performance of the trained cascades in assessing both synthetic and field data, depending on the criteria of change in cumulative production 1 year after restimulation.

The upper portion of Table 15 summarizes the model predictions derived from the synthetic test set for both 2- and 3-year cases. We see that in the cases of misclassification, there is a slightly greater tendency to predict favorable restimulation candidates as unfavorable than unfavorable candidates as favorable [i.e., the number of misclassifications in the lower-left corner of each confusion matrix shown in Table 15 (19; 3) is slightly greater than those in the upper right corner of each confusion matrix (17; 2)]. This implies that the cascade is inherently risk-averse in that it is more inclined to allow a practitioner to bypass reserves (false negative or Type II error) than to lose money on poor performance after refracturing treatment (false positive or Type I error). This behavior is a consequence of the continuation criterion used in the cascade-training procedure (Table 7), which is given by

$$F_s > 1 - D_s \dots \dots \dots (14)$$

Synthetic Test Data		
Actual	Predicted	
	Unfavorable	Favorable
Unfavorable	83	17
Favorable	19	81

Field Data		
Actual	Predicted	
	Unfavorable	Favorable
Unfavorable	3	2
Favorable	3	9

Table 15—Confusion matrices summarizing classifier error in predicting favorable and unfavorable candidates using synthetic test data (upper) and field data (lower).

The inequality in Eq. 14 means that cascade training is continued until the false-positive-classification ratio F_s is less than or equal to unity minus the true-positive ratio D_s . This criterion is designed to provide balanced allocation of error between both favorable and unfavorable categories. However, with a smaller number of training- and testing-data samples, there is a greater tendency toward convergence at a value of F_s slightly lower than $1 - D_s$. This factor can be mitigated by using more data for cascade training and testing. In addition, we can adjust Eq. 14 as

$$F_s > 1 - D_s + \beta, \dots \dots \dots (15)$$

where β is a positive or negative fraction. A positive value would lead to increased emphasis on correctly classifying favorable candidates at the expense of error in predicting unfavorable candidates, whereas a negative value would have the opposite effect. This can serve as a flexible tuning parameter to train the cascaded classifier depending on various economic objectives.

We can observe some disparities in predictive performance between the synthetic and field cases. For example, in the lower portion of Table 15, we see that two of five unfavorable wells have been misclassified (false-negative error of 40%) compared with a false-positive ratio of 25%. This contrasts with the results we observe in the upper portion of Table 15, where the false-positive error is slightly higher. This can likely be attributed to the small sample size of unfavorable candidates in the field-data pool (five wells). In other words, if we consider additional field data, the overall classification accuracy is likely to be more similar to that reported with the synthetic cases. In addition, looking more closely at the estimated probabilities of restimulation success in Table 14, we see that although correctly classified restimulation candidates have high probability, the probability estimates for misclassified wells are comparatively high. One potential source of error is the estimation of multiple months of production using simple empirical relationships, which might smooth out important decline characteristics and flow-regime changes that are useful in distinguishing between favorable and unfavorable candidates. Overall, the results of this field study are likely to improve with more-complete knowledge of reservoir and completion information, as well as more-detailed flow modeling that is calibrated to the specific field.

Conclusions

In this paper, we have proposed a novel probabilistic classification approach for analyzing massive production data sets for hydraulically fractured shale gas wells to yield information on the effect of near-wellbore fracture characteristics on well productivity. We have demonstrated the viability of this classification approach for a use case in restimulation-candidate selection. The results show that we are able to distinguish refracturing candidates in the presence of erratic rate fluctuations, using a cascaded AdaBoost classifier algorithm that provides both categorical predictions on restimulation success and probability estimates. This serves to quantify the uncertainty associated with the classifier predictions.

Adapted from face-detection methods, we use a series of texture-based attributes called Haar-like features to characterize well data as input to the classification algorithm. This data-driven approach is dependent on the understanding that the rate-decline characteristics invariably reflect the multivariate combination of fracture, fluid, and reservoir properties that contribute to restimulation success. We have shown that these Haar-like features are analogous to first and second derivatives of flow rate. Therefore, by using these features as input to our cascaded AdaBoost classification model, we have developed a fast and robust statistical model that is sensitive to flow-regime changes as well as curvature characteristics at multiple scales and time periods within the life of the well.

To serve as training examples for the proposed classification algorithm, we have generated production data from restimulated wells using a forward-flow-simulation approach, and quantified restimulation success using numerical metrics such as cumulative-production improvement 1 year after restimulation. By considering the multivariate distribution of input flow-simulation parameters, we have estimated the relative importance of physical parameters that influence restimulation success. Through analysis of the principal-component loadings, we have shown that multiple combinations of physical parameters have the potential to counteract one another, thereby making it difficult to distinguish between restimulation candidates using the type-curve approach considered. However, by capturing a series of subtle rate fluctuations embedded in Haar-like features, which are related to rate derivatives and curvature attributes at different time periods and scales, the proposed approach is able to distinguish between restimulation candidates to a high degree of accuracy for practical purposes.

It is important to emphasize that the use of simulated data in this study is because of a lack of available field data with known reservoir properties and completion information. However, with access to field-production information, the proposed method can easily also be applied to train a classification model capable of distinguishing between wells that exhibit any two different criteria of interest, and to develop fast and robust models to analyze future wells. One potential limitation of this approach for use with field data is the dependence of the Haar-like features on the duration and sampling interval of the production data, because they are parameterized by time and scale. This confines the application of our methodology to regularly sampled production data. The method proposed in this work is not intended to serve as a replacement for traditional simulator-based analyses. On the contrary, it can serve as a complement to stochastic history-matching procedures to screen candidate reservoir models that match specific physical criteria.

Finally, the method proposed in this research addresses challenges related to big data from hydraulically fractured wells within fractured shale reservoirs by developing a flexible probabilistic screening tool that is specifically geared toward statistical robustness, computational efficiency, and scalability required for applications involving massive data sets. In other words, although we have demonstrated this approach within the context of restimulation-candidate selection, it can be easily modified to address other reservoir management/development decisions by harnessing the rate-related information embedded within massive collections of production data. Although the training procedure might potentially be relatively computationally intensive, the assessment of production data using the trained model approaches real time in all cases we have considered. This renders the method amenable to applications requiring real-time analysis of well performance, which has the potential to affect the management of massive data sets obtained from shale-gas fields. This helps to facilitate a transition toward work flows involving continuous analysis of well data to bolster confidence and mitigate uncertainty in reservoir management and decision making.

Nomenclature

- b = Arps (1945) hyperbolic decline constant
- b_j = threshold for binary classification in weak classifier trained on feature index j
- d_f = pressure-dependent fracture-conductivity-decline constant, 1/psi
- D_{ps} = minimum true-positive ratio per stage in cascade training
- $f(x_{ij})$ = binary decision stump corresponding to feature j

F_{ps} = maximum false-positive ratio per stage in cascade training
 G = AdaBoost classifier model
 k_f = fracture permeability, md
 n = number of features per stage multiplier in cascaded training
 N = number of production-data profiles used in training
 p = total number of feature combinations
 q_i = gas-flow rate for well i , Mcf/D
 \tilde{q}_i = standardized gas-flow rate for well i
 Q_i = standardized cumulative gas-flow rate for well i
 t = time, days
 t_f = total producing time, days
 V = number of production-data profiles used in testing
 w_i = weight assigned to data from well i during AdaBoost training iteration
 \mathbf{x}_i = row vector that contains the features scores associated with each well i
 x_{ij} = Haar-like feature score computed using production data from well i depending on feature j
 y_i = true category of well i
 z = feature-template indicator
 α_m = expansion coefficient in AdaBoost model at iteration m
 β = tuning parameter for cascade-training-termination criteria
 ε_m = classification error associated with feature j at stage m
 θ = AdaBoost model-classification threshold

Acknowledgments

Support from the National Science Foundation (Grant No. 1546553) is gratefully acknowledged.

References

- Anderson, D. M., Nobakht, M., Moghadam, S. et al. 2010. Analysis of Production Data From Fractured Shale Gas Wells. Presented at the SPE Unconventional Gas Conference, Pittsburgh, Pennsylvania, 23–25 February. SPE-131787-MS. <https://doi.org/10.2118/131787-MS>.
- Arps, J. J. 1945. Analysis of Decline Curves. *Trans. AIME* **160** (1): 228–247. SPE-945228-G. <https://doi.org/10.2118/945228-G>.
- Aulia, A., Rahman, A., and Velasco, J. J. Q. 2014. Strategic Well Test Planning Using Random Forest. Presented at the SPE Intelligent Energy Conference & Exhibition, Utrecht, The Netherlands, 1–3 April. SPE-167827-MS. <https://doi.org/10.2118/167827-MS>.
- Bello, R. O. and Wattenbarger, R. A. 2010. Multi-Stage Hydraulically Fractured Horizontal Shale Gas Well Rate Transient Analysis. Presented at the North Africa Technical Conference and Exhibition, Cairo, 14–17 February. SPE-126754-MS. <https://doi.org/10.2118/126754-MS>.
- Breiman, L., Friedman, J., Stones, C. J. et al. 1984. *Classification and Regression Trees*. Boca Raton, Florida: CRC Press.
- Brown, M., Ozkan, E., Raghavan, R. et al. 2011. Practical Solutions for Pressure-Transient Responses of Fractured Horizontal Wells in Unconventional Shale Reservoirs. *SPE Res Eval & Eng* **14** (6): 663–676. SPE-125043-PA. <https://doi.org/10.2118/125043-PA>.
- Chin, L. Y., Raghavan, R., and Thomas, L. K. 2002. Fully Coupled Geomechanics and Fluid-Flow Analysis of Wells With Stress-Dependent Permeability. *SPE J.* **5** (1): 32–45. SPE-58968-PA. <https://doi.org/10.2118/58968-PA>.
- Clarkson, C. R. 2013. Production Data Analysis of Unconventional Gas Wells: Review of Theory and Best Practices. *Int. J. Coal Geol.* **109–110** (1 April): 101–146. <https://doi.org/10.1016/j.coal.2013.01.002>.
- Eshkalak, M. O., Aybar, U., and Sepehmoori, K. 2014. An Integrated Reservoir Model for Unconventional Resources, Coupling Pressure Dependent Phenomena. Presented at the SPE Eastern Regional Meeting, Charleston, West Virginia, USA, 21–23 October. SPE-171008-MS. <https://doi.org/10.2118/171008-MS>.
- Feblowitz, J. 2013. Analytics in Oil and Gas: The Big Deal About Big Data. Presented at the SPE Digital Energy Conference, The Woodlands, Texas, 5–7 March. SPE-163717-MS. <https://doi.org/10.2118/163717-MS>.
- Freund, Y. and Schapire, R. E. 1996. Experiments With a New Boosting Algorithm. *Proc.*, 13th International Conference on Machine Learning, Bari, Italy, 3–6 July, 148–156.
- Friedman, J., Hastie, T., and Tibshirani, R. 2000. Additive Logistic Regression: A Statistical View of Boosting. *Annal. Stat.* **28** (2): 337–407.
- Gilman, J. and Kazemi, H. 1988. Improved Calculations for Viscous and Gravity Displacement in Matrix Blocks in Dual-Porosity Simulators. *J Pet Technol* **40** (1): 60–70. SPE-16010-PA. <https://doi.org/10.2118/16010-PA>.
- Hastie, T., Tibshirani, R., and Friedman, J. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York City: Springer.
- Jayakumar, R., Rai, R., Boullis, A. et al. 2013. A Systematic Study for Refracturing Modeling Under Different Scenarios in Shale Reservoirs. Presented at the SPE Eastern Regional Meeting, Pittsburgh, Pennsylvania, 20–22 August. SPE-165677-MS. <https://doi.org/10.2118/165677-MS>.
- Kulga, I. B. 2014. *Analysis of the Efficacy of Carbon Dioxide Sequestration in Depleted Shale-Gas Reservoirs*. PhD dissertation, Pennsylvania State University, University Park, Pennsylvania (June 2014).
- Mayerhofer, M. J., Lolon, E. P., Youngblood, J. E. et al. 2006. Integration of Microseismic-Fracture-Mapping Results With Numerical Fracture Network Production Modeling in the Barnett Shale. Presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 24–27 September. SPE-102103-MS. <https://doi.org/10.2118/102103-MS>.
- Mohaghegh, S., Reeves, S., and Hill, D. 2000. Development of an Intelligent Systems Approach for Restimulation Candidate Selection. Presented at the SPE/CERI Gas Technology Symposium, Calgary, 3–5 April. SPE-59767-MS. <https://doi.org/10.2118/59767-MS>.
- Moore, L. P. and Ramakrishnan, H. 2006. Restimulation: Candidate Selection Methodologies and Treatment Optimization. Presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 24–27 September. SPE-102681-MS. <https://doi.org/10.2118/102681-MS>.
- Ozkan, E., Brown, M. L., Raghavan, R. S. et al. 2009. Comparison of Fractured Horizontal-Well Performance in Conventional and Unconventional Reservoirs. Presented at the SPE Western Regional Meeting, San Jose, California, 24–26 March. SPE-121290-MS. <https://doi.org/10.2118/121290-MS>.
- Patri, O. P., Panangadan, A. V., Chelmiss, C. et al. 2014. Predicting Failures From Oilfield Sensor Data Using Time Series Shapelets. Presented at the SPE Annual Technical Conference and Exhibition, Amsterdam, 27–29 October. SPE-170680-MS. <https://doi.org/10.2118/170680-MS>.
- Roussel, N. P. and Sharma, M. M. 2011. Selecting Candidate Wells for Refracturing Using Production Data. Presented at the SPE Annual Technical Conference and Exhibition, Denver, 30 October–2 November. SPE-146103-MS. <https://doi.org/10.2118/146103-MS>.

Rubin, B. 2010. Accurate Simulation of Non Darcy Flow in Stimulated Fractured Shale Reservoirs. Presented at the SPE Western Regional Meeting, Anaheim, California, 27–29 May. SPE-132093-MS. <https://doi.org/10.2118/132093-MS>.

Seemann, D., Williamson, M., and Hasan, S. 2013. Improving Reservoir Management Through Big Data Technologies. Presented at the SPE Middle East Intelligent Energy Conference and Exhibition, Manama, Bahrain, 28–30 October. SPE-167482-MS. <https://doi.org/10.2118/167482-MS>.

Shelley, R. F. 1999. Artificial Neural Networks Identify Restimulation Candidates in the Red Oak Field. Presented at the SPE Mid-Continent Operations Symposium, Oklahoma City, Oklahoma, 28–31 March. SPE-52190-MS. <https://doi.org/10.2118/52190-MS>.

Stehfest, H. 1970. Algorithm 368: Numerical Inversion of Laplace Transforms. *Commun, ACM* **13** (1): 47–49. <https://doi.org/10.1145/361953.361969>.

Tavassoli, S., Yu, W., Javadpour, F. et al. 2013. Selection of Candidate Horizontal Wells and Determination of the Optimal Time of Refracturing in Barnett Shale (Johnson County). Presented at the SPE Unconventional Resources Conference Canada, Calgary, 5–7 November. SPE-167137-MS. <https://doi.org/10.2118/167137-MS>.

Viola, P. and Jones, M. 2001. Rapid Object Detection Using a Boosted Cascade of Simple Features. *Proc., 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Kauai, Hawaii, 8–14 December, Vol. 1, 1-511–1-518.

Appendix A—Favorable-Candidate Wells Used in Field-Validation Study

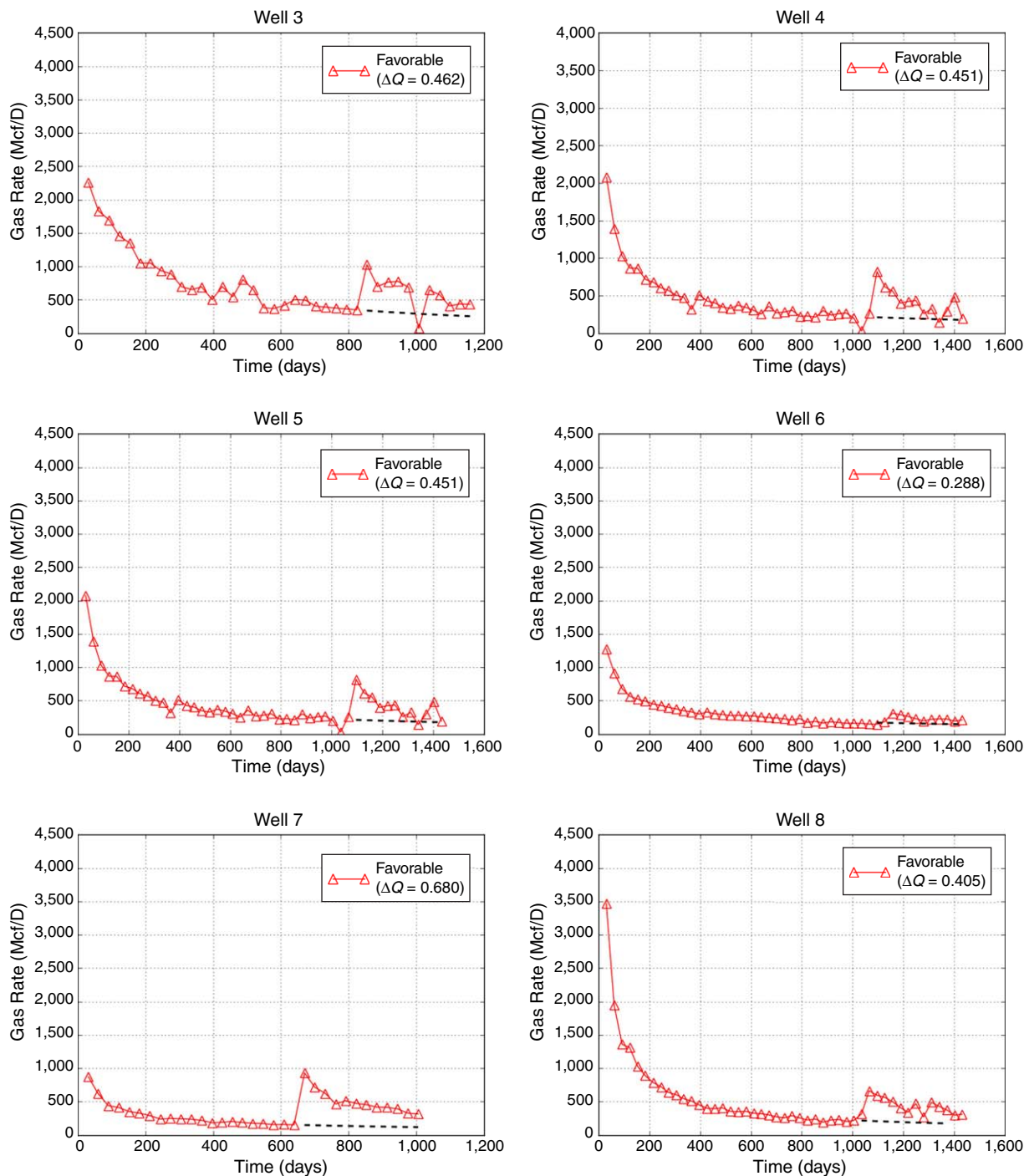


Fig. A-1—Gas-rate profiles for favorable candidate wells used in the field-validation study. The dashed lines represent the estimated rate trend in the absence of restimulation. The fractional increase in cumulative production (ΔQ) is estimated relative to the dashed line.

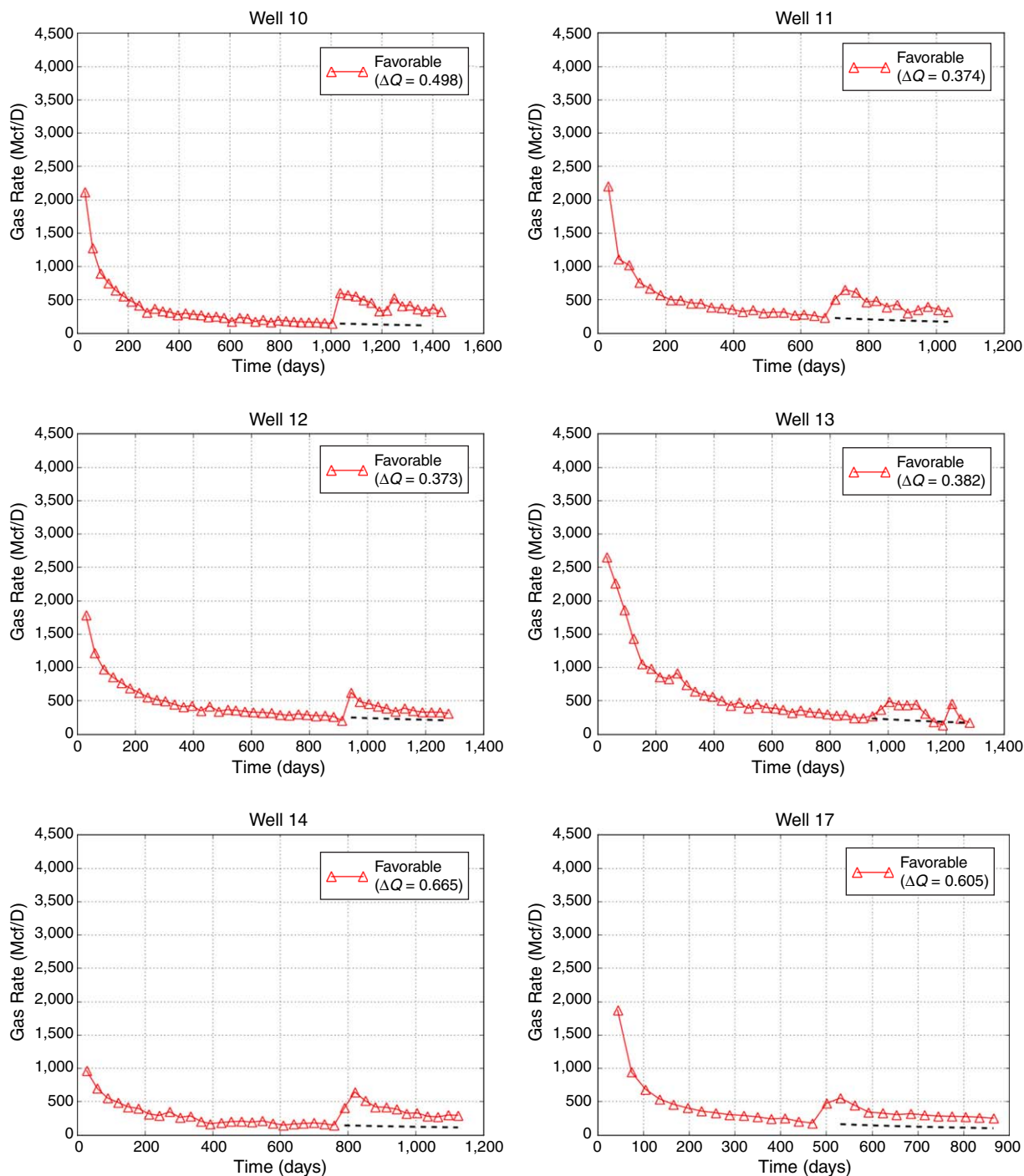


Fig. A-1 (continued)—Gas-rate profiles for favorable candidate wells used in the field-validation study. The dashed lines represent the estimated rate trend in the absence of restimulation. The fractional increase in cumulative production (ΔQ) is estimated relative to the dashed line.

Appendix B—Unfavorable-Candidate Wells Used in Field-Validation Study

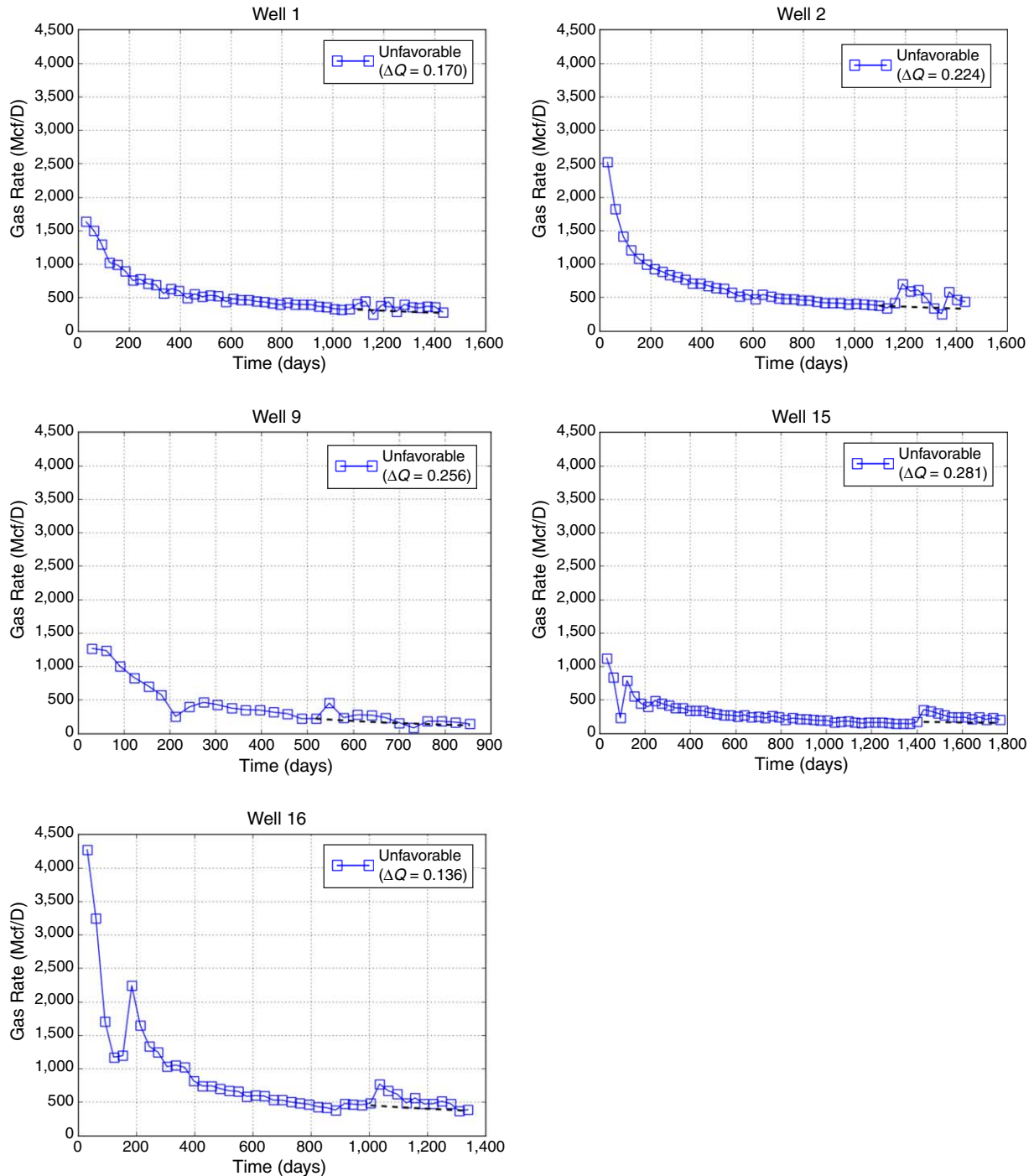


Fig. B-1—Gas-rate profiles for unfavorable candidate wells used in the field-validation study. The dashed lines represent the estimated rate trend in the absence of restimulation. The percent increase in cumulative production (ΔQ) is estimated relative to the dashed line.

Egbadon Udegbe is a PhD-degree candidate in energy and mineral engineering at Pennsylvania State University. His research interests are centered on the application of techniques in statistical learning and computer vision to aid in the analysis of subsurface big data from fractured reservoirs. Udegbe holds a bachelor's degree in electrical engineering.

Eugene Morgan is an assistant professor of petroleum engineering in the John and Willie Leone Family Department of Energy and Mineral Engineering at Pennsylvania State University. His current research interests include big-data analytics for the oil and gas industry, as well as stochastic inversion techniques for seismic-survey data. Morgan holds PhD and master's degrees in civil and environmental engineering from Tufts University. He is a member of SPE.

Sanjay Srinivasan is a professor of petroleum and natural-gas engineering at Pennsylvania State University and holds the John and Willie Leone Family Chair in Energy and Mineral Engineering. His primary research focus is in the area of petroleum-reservoir characterization and improved management of reservoir-recovery processes. Some of the algorithms and methods that Srinivasan has pioneered have been applied for early appraisal of ultradeepwater plays in the Gulf of Mexico and for characterizing natural-fracture networks in conventional and unconventional reservoirs.