SPE-189205-MS

Decline Curve Analysis for Production Forecasting Based on Machine Learning

Yunan Li, Texas A&M University; Yifu Han, University of Oklahoma

This paper was prepared for presentation at the SPE Symposium: Production Enhancement and Cost Optimisation held in Kuala Lumpur, Malaysia, 7-8 November 2017.

Abstract

A new method is developed by using machine learning technique to forecast single well production in both conventional and unconventional reservoirs. Unlike investigating and analyzing existing wells production rate and decline curve, this method predicts new well production rate according to reservoir properties such as matrix permeability, porosity, formation pressure and temperature, as well as hydraulic fracture parameters including fracture half length, fracture width and fracture conductivity.

In this paper, an inversion scheme is coupled with decline curve model, Logistic Growth Model (LGM), to obtain a set of decline curve parameters by fitting with production data. Both the Principal Component Analysis (PCA) and sensitivity study are applied to analyze the variance and identify key factors that influence production rate from reservoir and hydraulic fracture parameters. The sensitivity analysis results and scree plot from PCA serve as references to select key factors. Lastly, Neural Network (NN) technology is applied to investigate the pattern and correlation of selected reservoir and hydraulic fracture parameters and decline curve parameters. Therefore, the NN model can be applied to forecast production rate for a new well according to given reservoir and hydraulic information.

There is a good agreement between the available production data and decline curve model predicated production data based on the inverted decline curve model parameters. The scree plot and bi-plot generated by PCA provide the weight percentage of each component and help to identify factors that should be considered. Field production data is used to verify the feasibility of this method. This field case study is conducted by fitting the predicted production data (decline curve) based on NN model with field production data. The Mean Squared Estimation (MSE) of NN model is 0.013 Mscf/D and the overall R value is 0.917. This indicates that NN model is reliable to study the dataset and provide proper production (decline curve) prediction. The results illustrate that the predicted production data (decline curve) has good accuracy.

This paper proposes a statistical way for production forecasting based on machine learning. Instead of forecasting future production of existing wells, it provides meaningful reference for the evaluation of a new well and decision making.

Introduction

Machine learning is a statistical technology to get computers output expected results without being explicitly programmed. Machine learning is based on big-data without explicitly programming, therefore, it saves programming time and avoids various ideal scenarios restrictions, and it also improves the output prediction accuracy. Machine learning method such as Artificial Neural Network (ANN) is applied in oil and gas industry for production forecasting (Cao et al., 2016).

Decline curve analysis (DCA) has been applied for production rate analysis and forecasting for a long time. Arps decline model (Arps, 1945) is the most popular one, and various models are developed for different reservoirs. In our research work, we have selected Logistic Growth model (Clark et al., 2011) to describe oil production rate data. This model predicts accurate oil production for various geological formations especially for unconventional reservoirs such as shale gas or tight gas reservoirs.

The goal of this work is to apply machine learning technique for single well production forecasting in unconventional reservoirs. First of all, important geological factors such as permeability, porosity and hydraulic fracture factors for oil recovery are identified, and then decline curves are investigated by using inversion technique to get parameters in decline curve model. After obatining parameters in decline curve model, the predicted production profile can be output. Lastly, NN has been applied to find the correlations of identified geological and hydraulic fracture factors impacting oil recovery and parameters in decline curve model, therefore, a given set of geological factors can be input into the NN and then a predicted oil production profile can be output. With rock cores and known hydraulic fractures and reservoir properties information, NN can make good predictions of oil production profile.

According to Darcy's equation, the flow velocity depends on matrix permeability and formation pressure. The formation pressure provides the driving force for fluid flow, and matrix permeability measures the drag force for fluid flow. In addition, porosity is also important because it decides the oil storage capacity deep in the reservoir. Fluid viscosity is not considered because we have both oil and gas fileds data. Formation temperature needs to be considered as key factor since it is a very important factor for formation pressure, thus influenes the production (Hawkes et al., 2000). The formation temperature also influences the imbibition and fluid mobility. Young's modulus and Poisson's ratio help to describe the rock mechanics characteristics. Together with rock compressibility factor, they can describe the formation rock features, which is also related to the production forecasting. In the end, well radius is also considered from production engineer perspective since different production pipe diameters influence the production rate. We performed PCA and sensitivity analysis to evaluate the relative importance of parameters mentioned above. The top 7 important factors would be taken for the NN studies.

Methodology

The research work illustrates a promising methodology for production forecasting. By applying neural network, the hidden correlation and pattern of reservoir properties and production profile could be investigated. One of the advantages of this method is that it avoids to make hypothesis. For reservoir simulators, limitations and hypothesis need to be set to make an ideal case representing real problem, even though the accuracy is high enough. However, uncertainties exist in each idealized mathematical model. It would be better to input geological and fracture data and production data first, then use NN to find hidden pattern to output decline curve model parameters. Figure 1 illustrates the workflow of this method. Sensitivity study and PCA are utilized to select appropriate parameters for NN input. At the same time, the inversion method is applied to fit production profile. Therefore, a new set of reservoir and hydraulic fracture parameters can be input to the NN algorithm and make predictions on production profile. The prediction provides a rough estimation of the decline curve trend, which could assist to evaluate the formation quality.

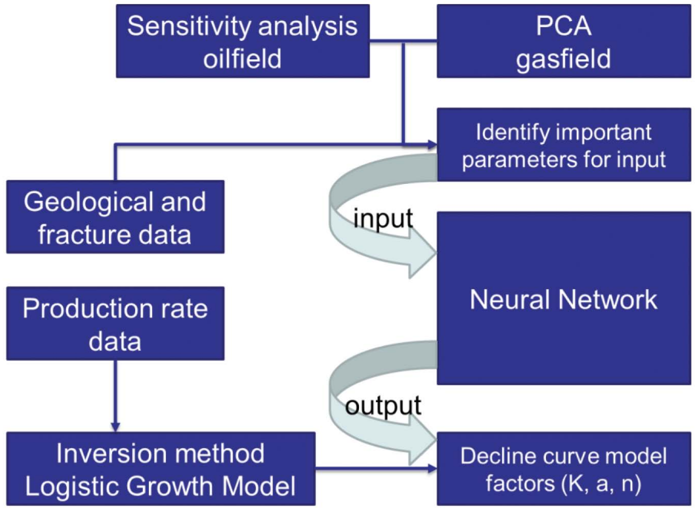


Figure 1—Overall workflow of the research work

Sensitivity Study

In this paper, sensitivity study is conducted to evaluate the impact of reservoir and hydraulic fracture parameters on production to provide solid support for NN training. From which, we are able to see the relative importance of all identified parameters. The benchmark scenario is a sandstome oilfield with permeability of 1.5 md and porosity of 10.4%. There is no vertical fluid flow considered. The Young's Modulus is 1.7×10^4 psi and the Poisson's Ratio is 0.3. The producer's bottom hole pressure (BHP) is 75 psi and there is no skin factor. A systemetic reservoir simulation could be performed to quantitatively evaluate all identified parameters.

Table 1 indicates sensitivity analysis and the variance of factors which are input parameters for NN. For reservoir simulation, cumulative oil production is set as objective and all other parameters except for the parameters listed in Table 1 are the same for all scenarios. For each simulation, we only change the value of one parameter, and for each parameter, we have two scenarios which are high case and low case. Therefore, we can obtain the variance of each parameter and cumulative oil production. In order to compare the variance of each parameter in the same scale, we normalized the oil production variance by dividing the variance of each parameter. The variance could be determined by the difference of high case and low case values normalized by the base case values. The sensitivity analysis histogram in Figure 2 illustrates the relative importance of all parameters.

Table 1—Factors and variance for sensitivity study

Factors	No.	Variance
Matrix permeability	1	0.59
Fracture conductivity	3	0.83
Porosity	2	0.09
Formation pressure	5	1.53
Fracture half length	4	1.83
Young's modulus	7	0.38
Formation temperature	6	1.21
Rock compressibility	9	0.01
Poisson ratio	8	0.34
Well radius	10	0.07

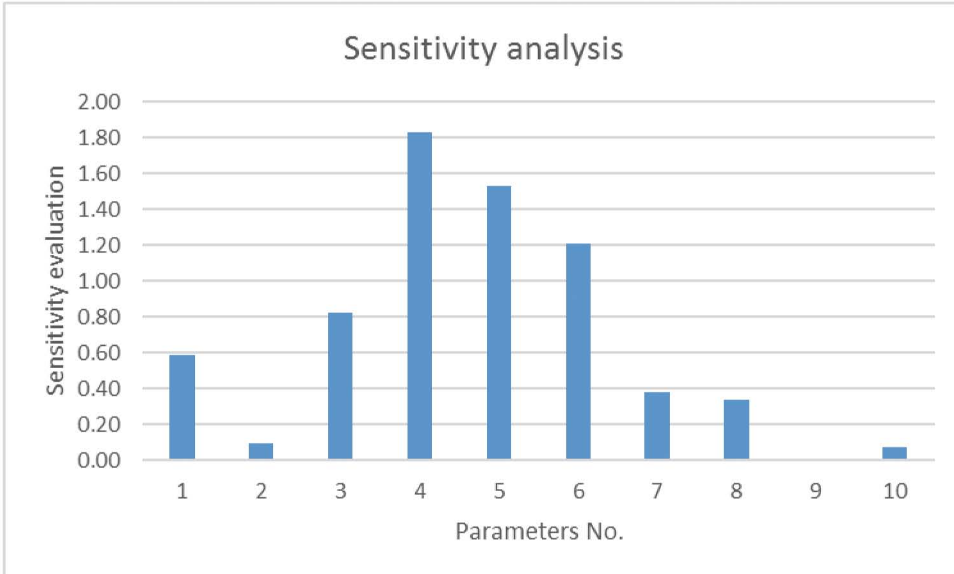


Figure 2—Evaluation of identified parameters

The results of sensitivity analysis demonstrate that fracture properties such as fracture half length and fracture conductivity and formation pressure and temperature are important factors. It has been stated that hydraculic fracture is a treatment to enhance oil recovery especially for low permeability formaitons (Zhao et al., 2010), so it is necessary to consider them as NN input. Rock matrix permeability and rock mechanics properties can also influence oil recovery. Well radius can also influence cumulative oil production, but its effect is negligible compared with other factors. The top 7 important factors are selected as the input of NN in the following part.

Principal Component Analysis

PCA is a statistical procedure converting observation vector of possibly correlated variables into vector of linearly uncorrelated variables. PCA is a widly used technique in history matching (Chen et al., 2014). We applied PCA to identify key factors for gasfields from reservoir and hydraulic fracture parameters and compare the selected factors with these of sensitivity analysis. We selected almost the same factors as in sensitivity study except for fracture width and gas specific gravity which we didn't include in sensitivity study, and Young's modulus and Poisson ratio are not taken into account in PCA.

Since reservoir fluid properties play an important role in oil recovery, we need to evaluate oil field and gas field separately. Then, we can compare the results and select the important factors in both cases for further analysis. For example, we need to consider geomechanics parameters such as Young's Modulus and Poisson's Ratio for oilfields because the shrinked porosity through production provides an additional driving force to displace oil out of porous media. For gasfields, this effect is not significant because of the gas properties. This step ensures to select the common feature parameters as input of the NN system.

$$Outlier < Q1 - 1.5 * IQR \tag{1}$$

$$Outlier > Q3 + 1.5 * IQR \tag{2}$$

$$IQR = Q3 - Q1 \tag{3}$$

where IQR is interquartile range, which is the measure of statistical dispersion, being equal to the difference between 75% and 25% percentiles.

Figure 3 is a boxplot which shows the variance and distribution of input parameters. From the boxplot, we could compare the variation range of each identified parameter and eliminate outliers for further analysis.

Boxplot is a very helpful way to display detected outlier points in a dataset (Saleh et al., 2014). There are two outliers for permeability and one outlier for porosity. The permeability has larger variation because we consider permeability in both conventional reservoir and unconventional reservoirs. In addition, in order to take fracture properties in different scenarios into account, the variation of fracture half length, and fracture conduvtivity is also large. Figure 4 and Figure 5 describe the distribution of sample points in 3-D coordinate system in which the axises are in the direction of the first three principal components. In order to accurately perform statistic analysis, we have to make all parameters under the same scale. Therefore, we normalized the feature matrix by dividing the average of each colume in order to show the diviation of each point regarding the average. Therefore, the normalized principle components axises have changed and the data points are shifted to be around (0,0,0). Equation 4, 5 and 6 illustrates the method of Z-score normalization.

$$x_z = \frac{x - \mu}{\sigma} \tag{4}$$

$$\mu = \frac{1}{N} \times \sum_{i=1}^{N} x_i \tag{5}$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2} \tag{6}$$

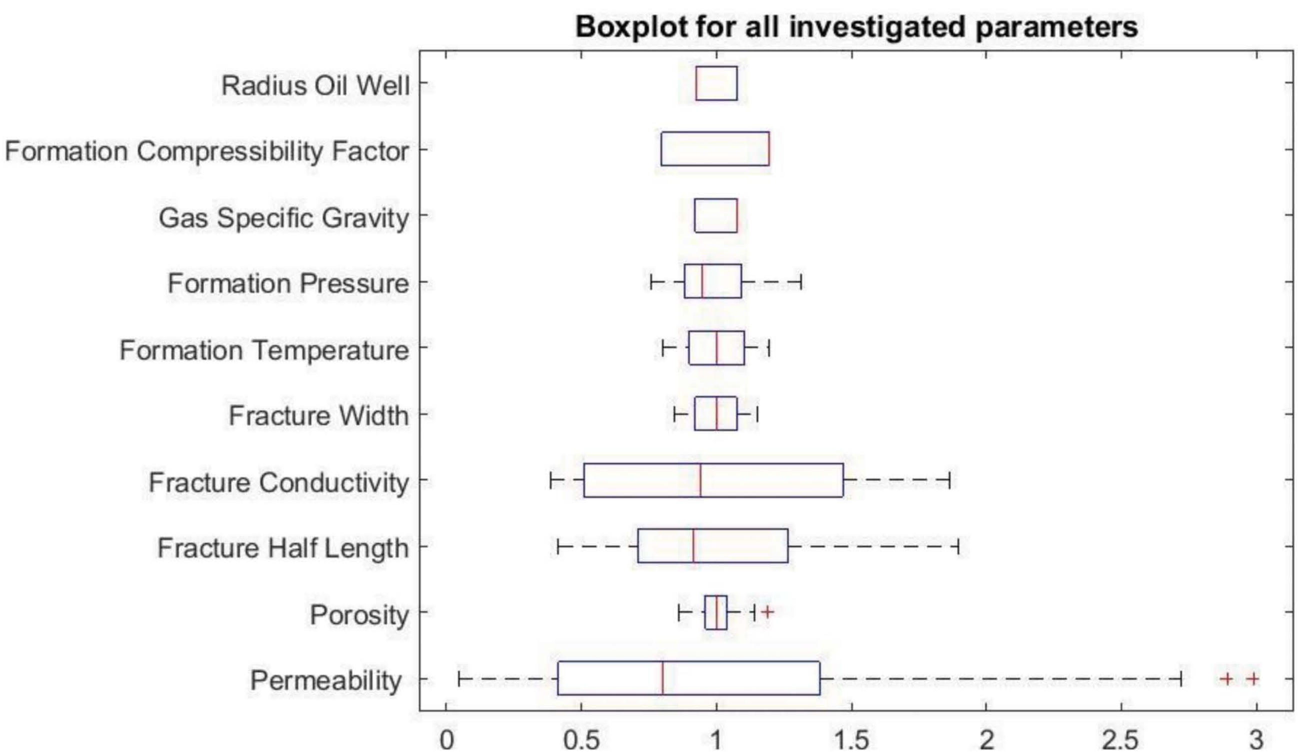


Figure 3—Boxplot for all investigated parameters

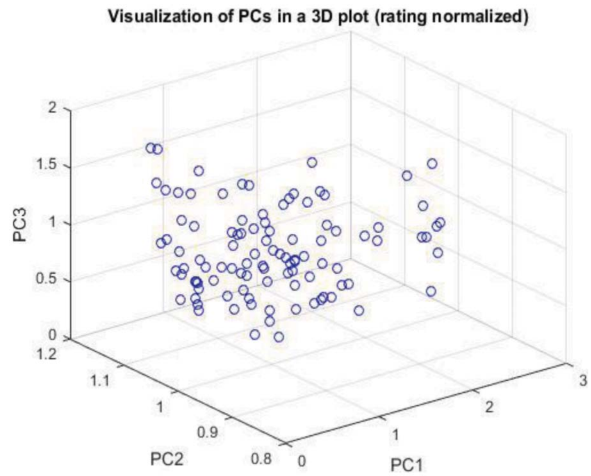


Figure 4—Visualization of PCs in a 3D plot

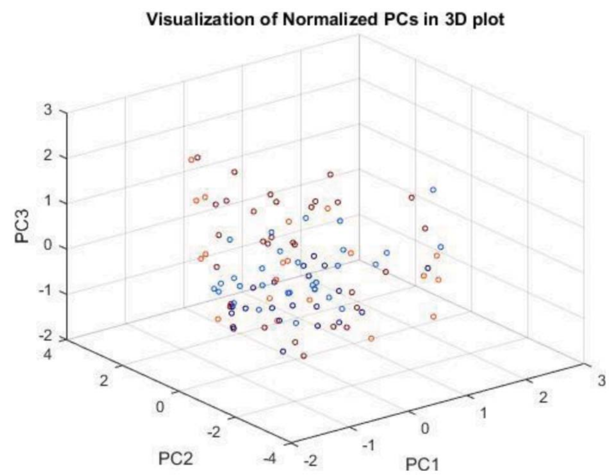
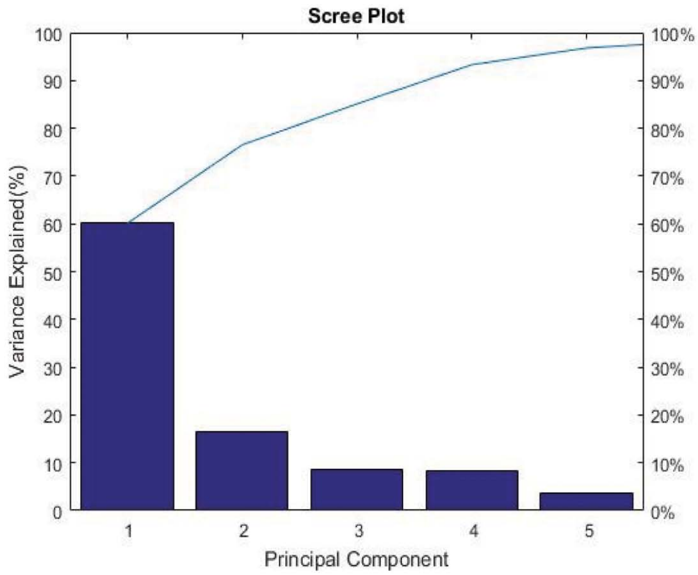


Figure 5—Normalized PCs visualization in a 3D pot

where μ is the mean of the entire dataset and σ is the standard deviation of the dataset.

It is necessary to use scree plot to limit the number of factors to be extracted from dataset (Kazakov et al., 2011). Figure 6 is scree plot which indicates the first three principle components which are able to represent over 85% of the entire dataset, so it is meaningful to take the first 3 components for further analysis. The percentage of each component is listed in the following table, ranking from the 1st principle component at the top to the 9th principle component at the bottom. Obviously, it is not necessary to have the low percenrage components so we took the first three principle components for further analysis.

$$PC\% = \sqrt{\frac{\sum_{i=1}^{K} \sigma_i^2}{\sum_{i=1}^{n} \sigma_i^2}} \tag{7}$$



60.09%
16.42%
8.19%
3.50%
1.48%
1.40%
0.31%
1.02e-26%
3.37e-31%

Figure 6—Scree plot for the principle components

Figure 7 can help to visualize the relative importance of each identified parameter regarding to the most important three principle components. Each point could be projected to the three axises, and the corresponding values show the positive or negative relations. Therefore, we could identify permeability, porosity, fracture width, fracture half length, fracture conductivity, formation pressure, and formation temperature as important inputs for NN analysis.

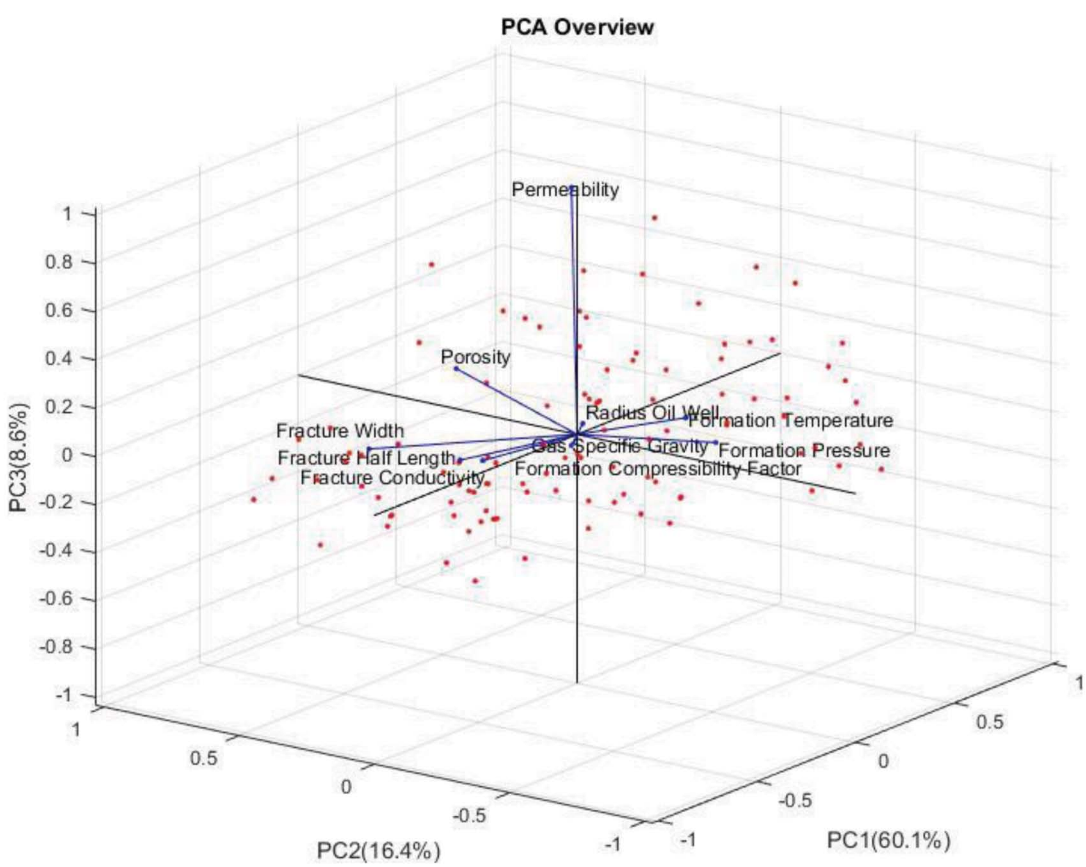


Figure 7—Principle component analysis overview for all factors

Logistic Growth Decline Curve Model

The extremely low permeability oil and gas reservoirs have gained lots of attention in the recent years. Production forecasting is chanllenging in these reservoirs since traditional reservoir models which previously yielded reasonable production prediction and reserves estimates are not applicable. Traditional decline curve model such as Arps' can't be applied in unconventional reservoirs such as shale gas or tight gas revervoirs. Various new decline curve models have been proposed to improve the accuracy and consistency of reserve estimates. Doung (2010) proposed a new approach for production rate prediction and expected ultimate recovery in tight and fracture flow dominated gas wells. Mishra et al. (2014) developed a new method to calculate the parameters in hyperbolic decline curve model and logistic growth model was firstly applied for decline curve analysis by Clark et al. (2011). However, none of the new proposed models have gained widespread industry acceptance. The logisctic model does not extrapolate to non-physical values and can also incorporate known volumetric quantities of hydrocarbon into the prediction to constrain the reserve estimate (Clark et al., 2011).

Logistic growth models are widely used in various fields by different forms and in order to predict oil and gas well production Clark et al. (2011) proposed the altered form of logistic growth model after empirical analysis. The logistic growth model is a growth equation and in this case the growth is cumulative oil or gas production. The proposed logistic growth model is very flexible to model numerous extremely low permeability oil and gas wells decline curve behaviors.

$$Q(t) = \frac{Kt^n}{a + t^n} \tag{8}$$

where Q is cumulative production rate, t is time, K is carrying capacity, a is constant and n is hyperbolic exponent.

The derivative of cumulative oil or gas production with respect to time can be taken to obtain oil or gas production rate:

$$q(t) = \frac{dQ}{dt} = \frac{Knat^{n-1}}{(a + t^n)^2} \tag{9}$$

where q is production rate.

There are three parameters in logistic growth decline curve model: the carrying capacity K, constant a and hyperbolic exponent n. The carrying capacity is the primary depletion oil and gas recoverable amount or estimated ultimate recovery for the well under investigated regardless of time or economic constraints. The constant a is proved to be the time when half of carrying capacity is reached (Clark et al., 2011). The hyperbolic exponent n controls decline curve steepness and high value of n indicates gradual decline during production time while small value of n stands for steep decline during a short period at high production rate before stabilizing at low production rate. The value of hyperbolic exponent n is generally between 0 and 1 and can also exceed 1 for wells whose initial production rate is not their peak production rate.

Decline Curve Model Inversion

The second step in the workflow is to perform inversion technique to get estimations of parameters in logistic growth decline curve model. We selected production data of 100 oil wells including unconventional reservoir field production data and synthetic production data generated by reservoir simulation. The goal of this procedure is to fit the logistic growth decline curve model to the cumulative oil production data which is easier to match compared with production rate data. The cumulative oil production data which is the integral of the production rate is smoother than production rate and has less outliers. Figure 8 shows the good agreement of the cumulative oil production data and modeled cumulative oil production data based on the inverted logistic growth model parameters. The results indicate that the EUR for this well regardless of economic constraints is around 300 MBbl and after 33 months of production half of the carrying capacity was reached.

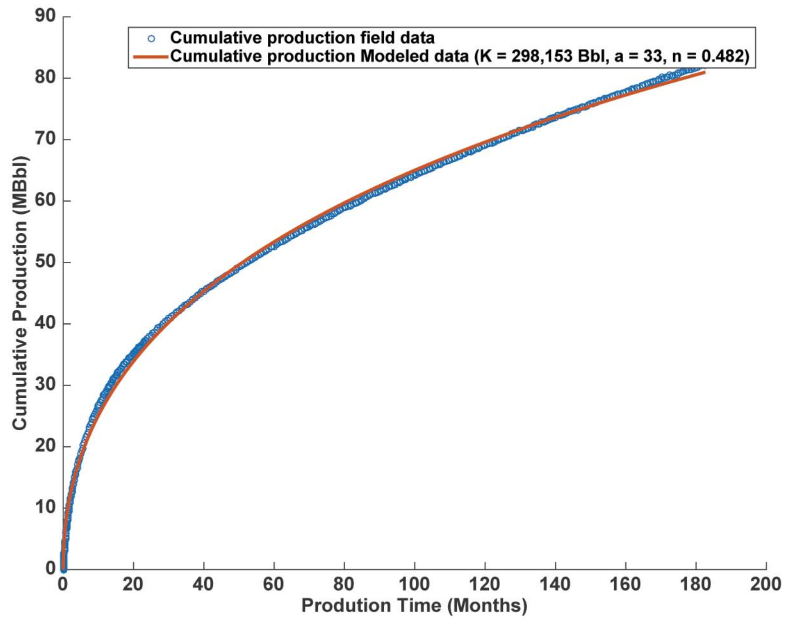


Figure 8—Curve fit for decline curve mathematical model

Figure 9 shows the distribution of estimated carrying capacity K for the 100 wells and the mean carrying capacity value is close to majority of the values and there are only few high EUR value wells.

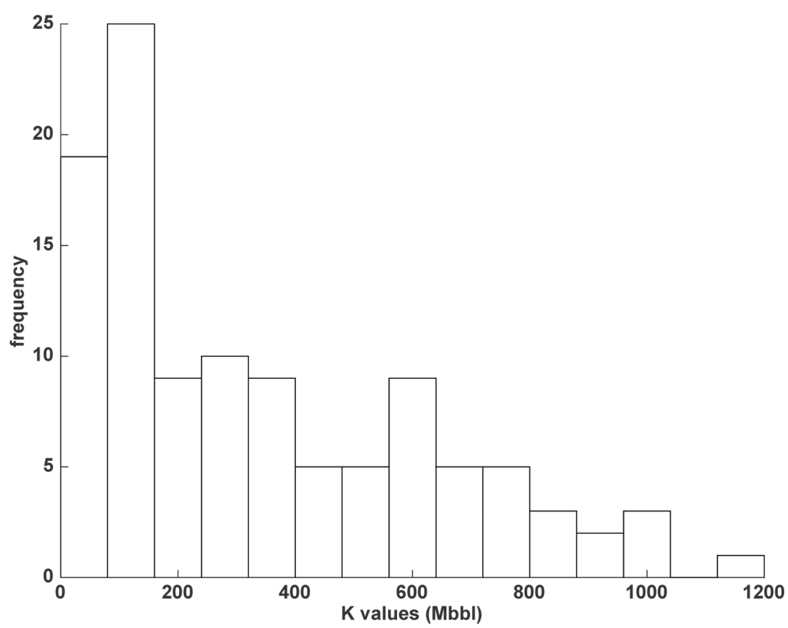


Figure 9—K values frequency distribution

The distribution of constant a values is shown in Figure 10 and is basically Gaussian shape. As a result of various reservoir and hydraulic fracture properties and operating conditions, the minimum a value is less than 10 months and maximum is close to 200 months, which indicates the production rate is gradual and it takes long time to recovery half of its oil. The distribution of hyperbolic exponent n is bimodal as illustrated in Figure 11.

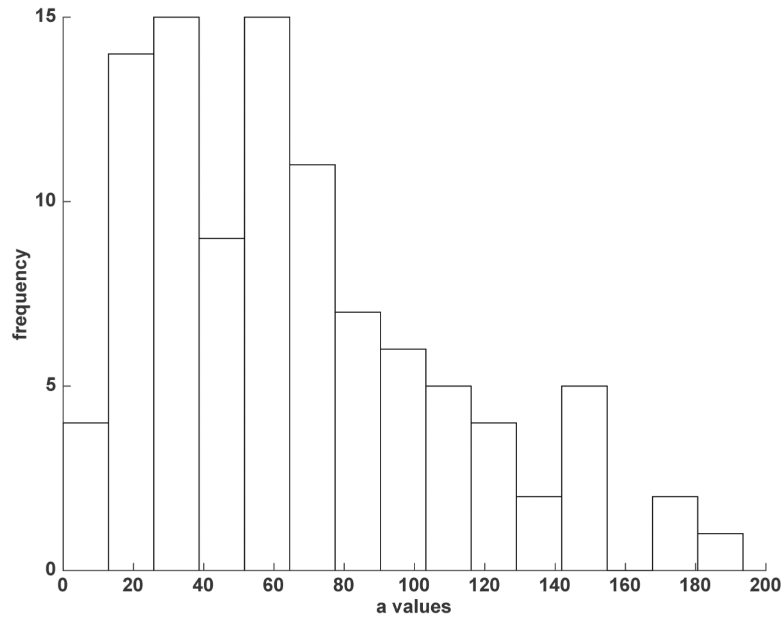


Figure 10—a values frequency distribution

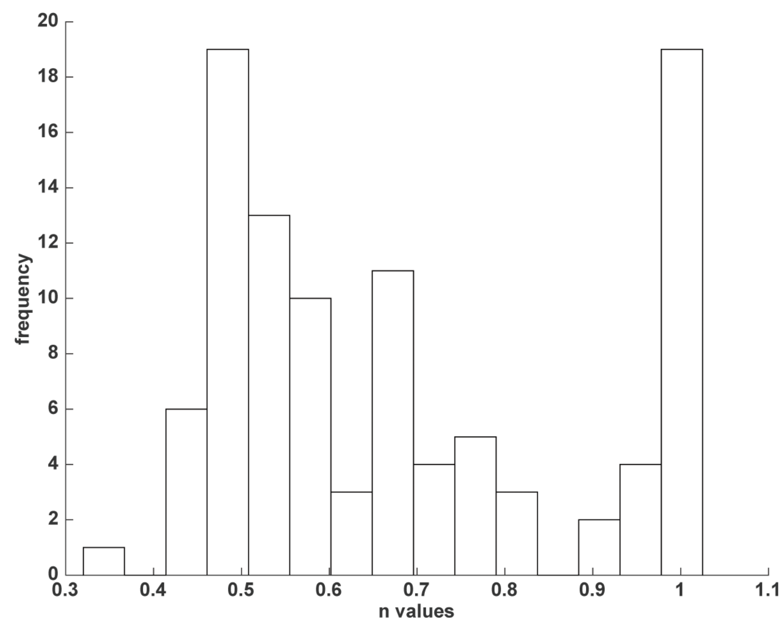


Figure 11—n values frequency distribution

Neural Network Regression

The structure of NN is as shown in Figure 12 above. Studies on NN have been motivated by the imitation of human brain operations (Haykin, 1994). There are 7 selected inputs in NN and 50 layers hidden for calculation and the outputs are the decline curve model factors. For each single case, the input is a column matrix composed of permeability, porosity, fracture width, fracture half length, fracture conductivity, formation pressure, and formation temperature from the top to the bottom. The output is also a column matrix with K, a, n listing from top to the bottom.

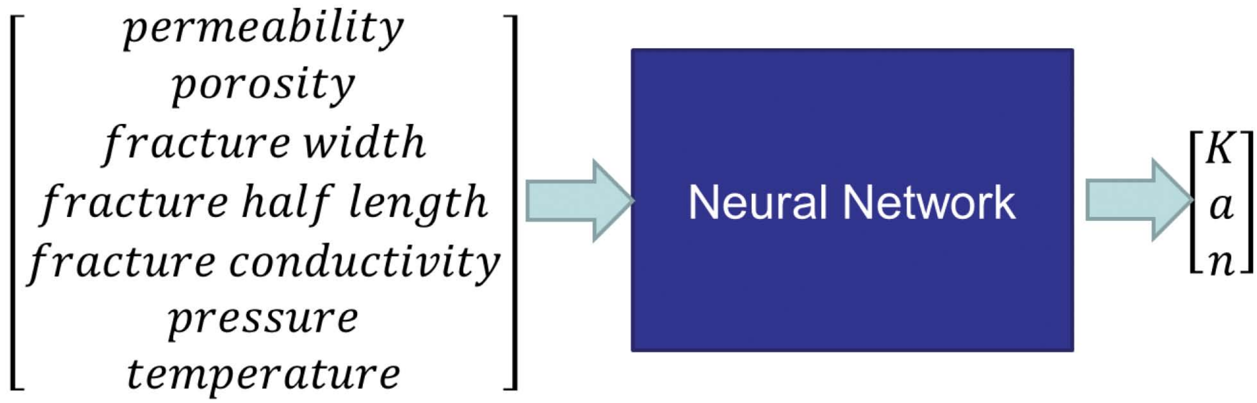


Figure 12—Neural network regression workflow

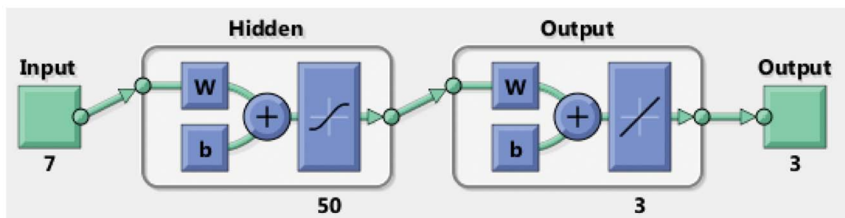


Figure 13—Neural network structure

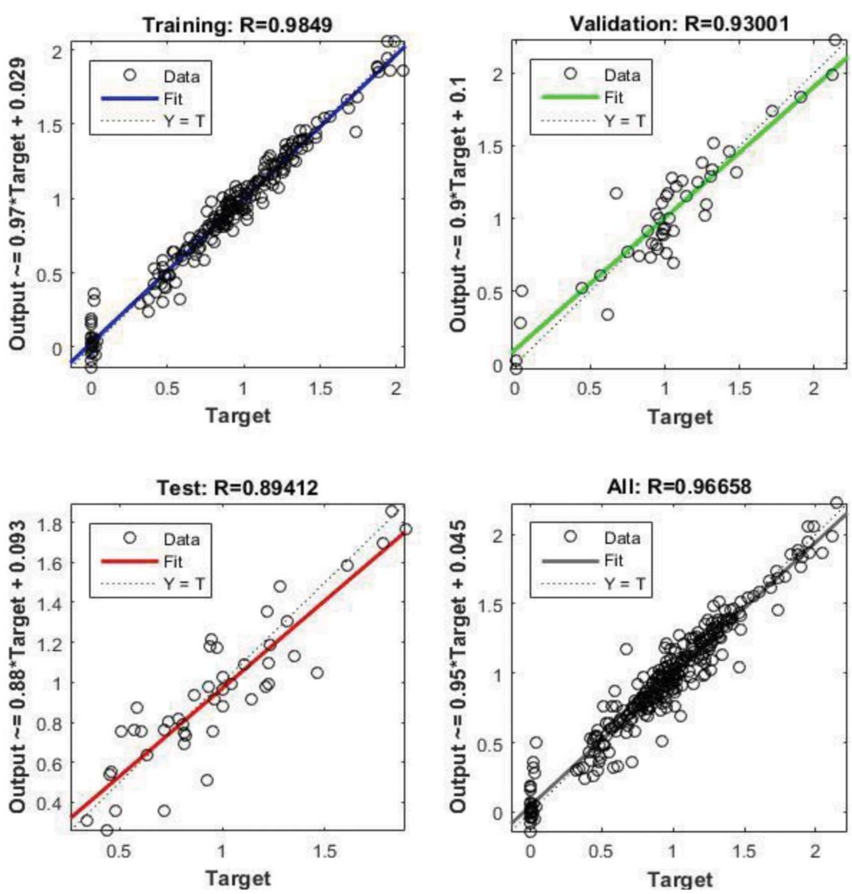


Figure 14—Results for neural network regression

For NN Regression, we have used 70% of the data as the training set, 15% for validation, and 15% for testing. According to the regression situations, we find the training set has the best regression. Even thougn the testing set deviates a little bit, it still close to 0.9, and the overall regression effect is good. The overall mean squared error (MSE) is 0.013 Mscf/D and the correlation coefficient (R) is 0.92. So this network could be counted for real case study. It is important to have the network trained so that whenever we have a new input matrix, we can easily output the decline curve model factors in order to mathematically describe the forecasted production rate.

Case study

The selected field data for verification is from a single phase oil well with vertical finite fractures. The hydraulic fracturing was performed for this vertical well. The formation permeability ranges from 1.5 md to 0.1 md, with porosity ranging from 8% - 12%. The formation thickness is around 26 - 36ft. There are no skin factors for this specific well. The oil has an API of 45. The formation temperature and pressure are 154 F and 2000 psi, respectivley. The well is produced with a constant BHP at the beginning, then shuts in for a week after 3 months production. After which, the well is kept produced with constant BHP control for the remaining days.

With the trained NN, we only need to input the formation properties corresponding with the required matrix format. Then, the output matrix could be automatically calculated including the values for K, a, and n. With these parameters, we can use the LGM to describe the decline curve. The dimensionless process is performed by deviding the maximum value of the whole colume matrix to make the normalized production rates within the range of 0 and 1.

In this case, we need to normalize the data because different oil or gas fields have different reservoir dimension, so it is impossible to predict the actual numbers without input of reservoir volume. Instead, the forecast of decline curve trend is also very important to evaluate the formation at the initial phase. The dimensionless production rate against time type curves represent various scenarios, so it is important to forecast the trend (Clark et al., 2011). And the dimensionless production rate also provides an approach to perform Rate Transient Analysis (RTA) for more reservoir information. According to Figure 15, we can find the predicted dimensionless production rate can match with the field data well after about half years of production. However, for the initial 180 days, we could not use the mathematical model to have an increase trend at the beginning and then decrease. And the decline curve model could not represent the shut-in effects, which is also a limitation for field case matching. Since the well is shut-in for only a week, compared with more than two years production, it does not have a huge influence on the dimensionless production rate matching. Therefore, we performed cumulative dimensionless recovery to compare. In this case, we can find in Figure 16 that the predicted curve matches very well with field data, providing a great reference for engineers to evaluate this formation.

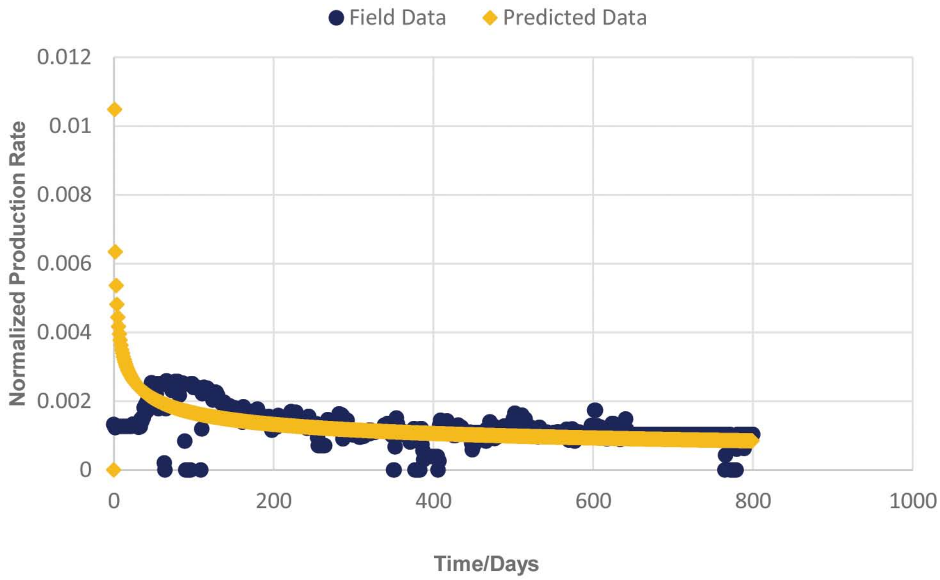


Figure 15—Dimentionless production rate data match

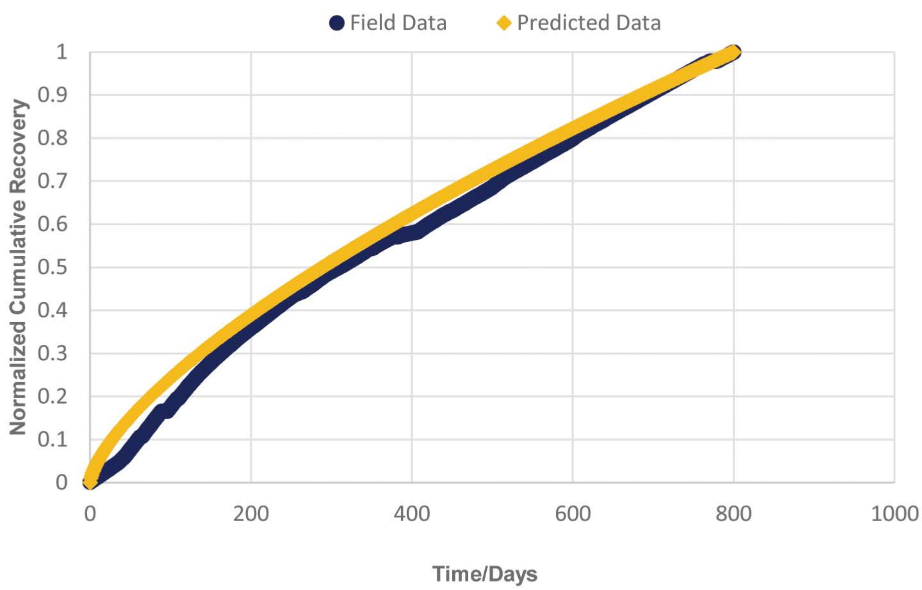


Figure 16—Dimentionless cumulative production data match

Conclusions and future work

This research work aims to indicate an approach to correlate the reservoir formation parameters to the production profile. We have made attempts through substituting the production profile by the decline curve model factors. In this way, a time based dataset could be represented simply by a column matrix. The sensitivity analysis and principle component analysis combine and yield the factors including fracture and formation parameters to be important. They are therefore the input for neural network regression. More importantly, this attempt shows a promising way to make forecast on production profile. When it comes to the initial phase of a formation, engineers only have limited knowledge on the rock properties and fracture parameters. Without even knowing more complicated parameters, people can make rough estimations by using this trained neural network algorithm. It has advantages because the neural network get trained by the real data in order to avoid uncertainties or making hypothesis. The high computational efficiency is also an advantage of this method compared with reservoir simulators, which require more information like relative permeability curves, capillary pressure curves, etc. The dimensionless production profile match in the Figure 15 and Figure 16 show the very good results for a match. The increasing trend and gradient are good indicators for the evaluation of a new formation.

As for the future work, we will do functional principle components analysis (fPCA) for a more accurate analysis. In addition, we aim to include more decline curve models such as Arps', Doung' model, etc. in the inversion step. The algorithm is able to automatically select the best decline curve model to fit the production data under various formation settings. Accordingly, the neural network can be improved by outputting various decline curve model factors in the expected column matrix. Attempts to correlate the pressure transient profile with formation parameters are also important to understand and evaluate the formation at the beginning phase.

Nomenclature

a = Logistic Growth Model parameter
K = Logistic Growth Model parameter
N = number of elements in the dataset
n = Logistic Growth Model parameter
Q = cumulative production
q = production rate
t = production time
μ = mean value
σ = standard deviation
x_i = element of a dataset
x_z = Z-score normalized elements

References

Abass, H. H., Al-Tahini, A. M., Abousleiman, Y., and Khan, M. R. 2009. New Technique to Determine Biot Coefficient for Stress-Sensitive Dual-Porosity Reservoirs. Presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, 4-7 October. SPE-124484-MS. http://dx.doi.org/10.2118/124484-MS

Arps, J. J. 1945. Analysis of Decline Curves. Published in *Petroleum Transactions, AIME*, **160** (1945): 228-247.

Cao, Q., Banerjee, R., Gupta, S., Li, J., Zhou, W., & Jeyachandra, B. 2016. Data Driven Production Forecasting Using Machine Learning. Presented at SPE Argentina Exploration and Production of Unconventional Resources Symposium, Buenos Aires, Argentina, 1-3 June. doi: 10.2118/180984-MS

Chen, C., Gao, G., Honorio, J., Gelderblom, P., Jimenez, E., and Jaakkola, T. 2014. Integration of Principal-Component-Analysis and Streamline Information for the History Matching of Channelized Reservoirs. Presented at SPE Annual Technical Conference and Exhibition, Amsterdam, The Netherlands, 27-29 October. doi: 10.2118/170636-MS

Duong, A. N. 2010. An Unconventional Rate Decline Approach for Tight and Fracture-Dominated Gas Wells. Presented at the SPE Canadian Unconventional Resources & International Petroleum Conference, Calgary, Alberta, Canada, 19-21 October. doi: 10.2118/137748-MS

Hawkes, R., Su, Z., and Leech, D. 2000. Field Data Demonstrates Thermal Effects-Important in Gas Well Pressure Buildup Tests. Presented at the SPE Canadian International Petroleum Conference, Calgary, Alberta, Canada, 4-8 June. doi: 10.2118/2000-031

Haykin, S. 1994. Neural Networks-A Comprehensive Foundation, Macmillan College Publishing Company.

Kazakov, N., and Miskimins, J. L. 2011. Application of Multivariate Statistical Analysis to Slickwater Fracturing Parameters in Unconventional Reservoir Systems. Presented at the SPE Hydraulic Fracturing Technology Conference, Woodlands, Texas, 24-26 January. doi: 10.2118/140478-MS

Mishra, S., Kelley, M., and Makwana, M. 2014. Dynamics of Production Decline from Shale Gas Reservoirs: Mechanistic or Empirical Models? Presented at the SPE Eastern Regional Meeting, Charleston, WV 21–23 October.

Saleh, L. D., Wei, M., and Bai, B. 2014. Data Analysis and Updated Screening Criteria for Polymer Flooding Based on Oilfield Data. Society of Petroleum Engineers. doi: 10.2118/168220-PA

Zhao, H., Chen, M. 2010. Extending behavior of hydraulic fracture when reaching formation interface. *Journal of Petroleum Science and Engineering*, **74** (1-2), 26-30.