Journal of Natural Gas Science and Engineering 38 (2017) 621–637

Contents lists available at ScienceDirect

Journal of Natural Gas Science and Engineering

journal homepage: www.elsevier.com/locate/jngse

New forecasting method for liquid rich shale gas condensate reservoirs with data driven approach using principal component analysis

Aaditya Khanal [a, *], Mohammad Khoshghadam [a], W. John Lee [b], Michael Nikolaou [a]

[a] Department of Chemical Engineering, Petroleum Engineering Program, University of Houston, Houston, TX, United States
[b] Department of Petroleum Engineering, Texas A&M University, College Station, TX, United States

A R T I C L E I N F O

Article history:
Received 19 August 2016
Received in revised form
5 January 2017
Accepted 9 January 2017
Available online 10 January 2017

Keywords:
Production forecasting
Principal component analysis
Reservoir simulation
Machine learning
Data driven
Type wells

A B S T R A C T

Accurate production performance evaluation and forecasting in shales during the early stages of development can play an important role in minimizing uncertainties associated with unconventional reservoirs. Given the limited reliability in forecasts from traditional decline models when applied to unconventional reservoirs, new tools to supplement the ones in use today are required to improve the accuracy of production forecasts. In this study, we present a method involving principal component analysis (PCA), which is a simple, non-parametric method of extracting relevant information from large data sets to perform production forecasting of liquid rich shale gas condensate reservoirs.

We used a comprehensive compositional reservoir model to create several iterations of synthetic production histories from liquid rich shales (LRS) wells based on Monte Carlo simulation with predefined probability distributions. Cumulative gas, gas rate, and condensate-to-gas ratio (CGR) for the simulated cases were decomposed into principal component (PC) scores and coefficients were used to recreate the original data. The dataset was cross-validated to check its ability to predict the missing production data based on PC scores and coefficients of the limited production data. Principal component analysis was further applied to the field data from several wells from Eagle Ford shale. We re-created and cross-validated the field data by using limited PC which led to good matches of the original production data. Two to three PC's were required to recreate the initial data with reasonable accuracy depending on the quality of the input data. During the validation step, we observed that some of the wells exhibited significant error which could be attributed to significantly different production profiles of those wells compared to the other wells. For simulated data, four PC was enough to yield the prediction with average error of 0.16%, 0% and 0.77% respectively for gas rate, cumulative gas and CGR respectively. For field data, three PC yielded the best prediction with average error of 1.63% and 2.98% for gas rate and oil rate respectively.

This work shows that multivariate statistics and data driven methods can be used as an important approach to complement existing tools like reservoir simulation and decline curve analysis to perform production data analysis. PCA can also be used and can generate accurate results relatively quickly. We recognize that even more rapid approximate methods will be required for routine analysis. Understanding the limitations of different approximate methods and application of methods to overcome these limitations in given circumstances should lead to optimal use of these methods.

1. Introduction

Proved reserves of US total natural gas and oil increased by 9.8% and 9.3% to 388.8 Tcf and 39.9 billion bbls respectively in November 2014 compared to the estimates in November 2013 (EIA, 2015). This record increase in proved reserves can be attributed to several technological advancements including multistage hydraulic fracturing and pad-drilling of unconventional shale reservoirs. Sustained low prices have made recovery economics more challenging and will most likely reduce the reserves in coming studies without

* Corresponding author.
E-mail address: aadityakhanal@gmail.com (A. Khanal).

http://dx.doi.org/10.1016/j.jngse.2017.01.014
1875-5100/

affecting the resource estimates (EIA, 2015). Even in this period of low prices, oil and condensates are more attractive economically due to their higher prices per BTU compared to gas.

Gas condensate or retrograde condensate reservoirs provide an attractive alternative to dry gas reservoirs because of their high liquid content. However, due to the subsurface complexity of pressure, volume and temperature (PVT) behavior for gas condensates, production forecasting and reserve estimation is difficult and often inaccurate. Gas condensate, initially a single-phase gas, undergoes retrograde condensation in the reservoir below the dew point pressure. This behavior results in more complicated multiphase flow than in black oil and dry gas reservoirs. Furthermore, the retrograde condensate forms a condensate bank near the well and reduces the overall well productivity for both oil and gas. This is one of several reasons that increases uncertainty and risk while developing these unconventional reservoirs.

A decline curve fitting method presented by Arps (1945) is still the most widely used method for both conventional and unconventional wells. For this method to be applicable, a reservoir should be in boundary dominated flow (BDF) with unchanging drainage area and constant flowing bottomhole pressure (bhp). Unconventional reservoirs, due to their low permeability, show long periods of linear transient flow which could last until they reach their economic limits. Khanal et al. (2015a,b) have analyzed several empirical decline curve methods such as Arps decline model, Duong's production decline and stretched exponential production decline models to forecast production from unconventional gas condensate reservoirs. Extensive simulation study of several reservoir fluids with different initial condensate to gas ratio (CGR) showed the importance of pressure normalized diagnostic plots to identify the flow regimes for accurate future production forecasting. Application of empirical methods without proper identification of flow regimes with diagnostic plots led to over prediction in most cases. It was also seen that multi-segmented approach, where each flow regime was modeled by separate decline model resulted in a better production forecasting.

Decline curve analysis requires sound judgement on part of the engineer, as the empirical constants like *b* (Arps model) are selected based on experience and/or analogies. This brings positive or negative bias to the forecasted hydrocarbons depending on the analyst when these methods are used. Other well-known tools for production data analysis and forecasting include straight-line methods, reservoir simulation, and history matching. Each of these tools and others has its own utility depending on available information and resources, and each often provide complementary answers to the problem faced. One such tool from multivariate statistics is principal component analysis (PCA), which can be used effectively to identify similar wells and dissimilar wells. Principal component analysis can also be used to predict the performance of wells with similar properties in conjunction with other tools such as linear regression.

2. Principal component analysis

2.1. Introduction and background

Production data analysis is an evolving field where several new techniques are applied together to develop, analyze, forecast, and evaluate the production from oil and gas reservoirs. These techniques have been applied to various types of reservoirs with varying degrees of success in the past. Several authors (Denney, 1999; Srinivasan and Ertekin, 2008; Mohaghegh, 2009; Fulford et al., 2016) have recently discussed the use of artificial intelligence techniques like neural networks and fuzzy logic to forecast production for oil and gas reservoirs.

One such method which can be used for production data analysis is PCA, which has been used extensively in geosciences to reduce redundant data and retain useful data obtained from seismic data acquisition (Saleh and de Bruin, 2000, Tingdahl and Hemstra, 2003; Guo et al., 2006; Chopra and Marfurt, 2014). It has also been used for other reservoir engineering applications such as history matching (Honorio et al., 2015; Chen et al., 2016), reservoir property estimation (Scheevel and Payrazyan, 2001; Dadashpour et al., 2011), gas flow in nano-pores and the analysis of production history for unconventional gas reservoirs (Bhattacharya and Nikolaou, 2011, 2013). Researchers have also used PCA to perform fluid composition characterization of crude oils from different depths and wells to resolve compositional changes related to the source of the oil generating sediments and its exposure to biological and/or physical weathering processes. More recently, PCA has been used to reduce uncertainty in history matching and to identify reservoir measurements that best represent the overall reservoir behavior (Bertolini and Schiozer, 2016).

Principal component analysis reduces the dimensionality of multivariate datasets by reducing the variables in a data-set into smaller number of derived variables, principal components (PC), which are linear combinations of original variables (Jolliffe, 2014). Principal Components are arranged in descending order, and the axis corresponding to the first principal component (or eigenvector) is the one along which the variance of the data is maximized, followed by the second principal component and so on (Jolliffe, 2014).

Assume a matrix X of rank m (and variables) can be represented by a matrix X' of lower rank p such that

$$\begin{aligned}[X]_{n\times m} &\approx \left[X'\right]_{n\times p}\\ &= (PC_1)v_1^T + \ \ldots\ldots\ldots + (PC_m)v_m^T\\ &\approx (PC_1)v_1^T + \ \ldots\ldots\ldots + (PC_p)v_p^T + E^p\end{aligned} \tag{1}$$

In the equations above, the PC_i are the principal components scores, v_i are the loadings or principal component coefficients which describe the systematic part of the data, and E^p is the residual matrix which describes the model and measurement errors (Shlens, 2003). The number of principal components required to accurately represent the original data depends on two factors: noise and redundancy and usually done empirically (Jolliffe, 2014).

Principal Component Analysis is performed by creating the covariance matrix, calculating the respective eigenvectors and eigenvalues of the matrix and finally ranking them based on their respective eigenvalues (Jolliffe, 2014). The eigenvectors with the greatest eigenvalues are the Principal Components of the data matrix. Principal components can also be calculated by using singular value decomposition (SVD) where a matrix M is decomposed into matrices U, S and V such that:

$$M = USV^T$$

where U and V are orthonormal vectors (vectors with unit norm and zero inner product) and S is a diagonal matrix with eigenvalues in the main diagonal (Jolliffe, 2014). The product of matrix U and S yields the PC_i represented in Equation (1) and V represents the principal component coefficients v_i.

The field of petroleum engineering is inundated with data from various sources, such as geological, production, and experimental data. In the United States, most of these data are proprietary except for a few publicly reported data sets such as rate-time data or occasional well testing data. This collection of data can be viewed as a large matrix, which, when analyzed using multi-variate statistical methods like PCA, can be reduced to a smaller matrix that retains