Expert Systems with Applications 42 (2015) 7326–7343

Contents lists available at ScienceDirect

Expert Systems with Applications

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

Practical implementation of knowledge-based approaches for steam-assisted gravity drainage production analysis

Zhiwei Ma [a], Juliana Y. Leung [a,*], Stefan Zanon [b], Peter Dzurman [b]

[a] School of Mining and Petroleum Engineering, Department of Civil and Environmental Engineering, University of Alberta, Edmonton, Alberta T6G 2W2, Canada
[b] Nexen Energy ULC, 801 – 7th Avenue S.W., Calgary, Alberta T2P 3P7, Canada

ARTICLE INFO

Article history:
Available online 29 May 2015

Keywords:
Statistical analysis
Neural networks
Model uncertainties
Steam-assisted gravity drainage
Petroleum engineering
Data mining
Production forecast

ABSTRACT

Quantitative appraisal of different operating areas and assessment of uncertainty due to reservoir heterogeneities are crucial elements in optimization of production and development strategies in oil sands operations. Although detailed compositional simulators are available for recovery performance evaluation for steam-assisted gravity drainage (SAGD), the simulation process is usually deterministic and computationally demanding, and it is not quite practical for real-time decision-making and forecasting. Data mining and machine learning algorithms provide efficient modeling alternatives, particularly when the underlying physical relationships between system variables are highly complex, non-linear, and possibly uncertain.

In this study, a comprehensive training set encompassing SAGD field data compiled from numerous publicly available sources is analyzed. Exploratory data analysis (EDA) is carried out to interpret and extract relevant attributes describing characteristics associated with reservoir heterogeneities and operating constraints. An extensive dataset consisting of over 70 records is assembled. Because of their ease of implementation and computational efficiency, knowledge-based techniques including artificial neural network (ANN) are employed to facilitate SAGD production performance prediction. The principal components analysis (PCA) technique is implemented to reduce the dimensionality of the input vector, alleviate the effects of over-fitting, and improve forecast quality. Statistical analysis is performed to analyze the uncertainties related to ANN model parameters and dataset. Predictions from the proposed approaches are both successful and reliable. It is demonstrated that model predictability is highly influenced by model parameter uncertainty. This work illustrates that data-driven models are capable of predicting SAGD recovery performance from log-derived and operational variables. The modeling approach can be updated when new information becomes available. The analysis presents an important potential to be integrated directly into existing reservoir management and decision-making routines.

1. Introduction

SAGD is one of the most important thermal enhanced oil recovery technologies for producing heavy oil. A pair of horizontal wells, including an injection well and a production well that are located a few meters apart, is drilled into the reservoir. High-pressure steam is injected to reduce the bitumen viscosity. The heated crude oil then drains along the steam chamber edge into the production well by gravitational force. Evaluation of SAGD performance has been widely studied involving experiments (Akin & Bagci, 2001; Bagci, 2006; Shin & Polikar, 2006) and numerical simulation (Chang, Ivory, & Tunney, 2012; Chow & Butler, 1996; Egermann, Renard, & Delamaide, 2001; Fatemi, 2009; Siu, Nghiem, Gittins, Nzekwu, & Redford, 1991). However, it is often impossible to reproduce all conditions and heterogeneities at the field scale in lab-scale models, while numerical flow simulations usually provide only approximate solutions to recovery responses, as numerous simplifications and assumptions must be invoked. The modeling process itself is also quite time-consuming, limiting its application in field-scale analysis involving multiple wells. Despite the availability of large amount of production and reservoir data from different producing fields, practical application of knowledge-based models for reliable SAGD analysis and prediction is lacking.

Knowledge-based, or data-driven, modeling techniques, which entail comprehensive data analysis and implementation of machine learning methods for system forecast, provide an

* Corresponding author. Tel.: +1 (780) 492 3338.
E-mail addresses: zma2@ualberta.ca (Z. Ma), juliana2@ualberta.ca (J.Y. Leung), stefan.zanon@nexencnoocltd.com (S. Zanon), peter.dzurman@nexencnoocltd.com (P. Dzurman).

http://dx.doi.org/10.1016/j.eswa.2015.05.047
0957-4174/

Nomenclature

Symbols

B	bootstrap sample size
b_i	bias term of the neuron i
d_{sh_inj}	distance from the shale layer to the injection well
$\mathbf{d}$	data vector
f	regression function
h	pay zone thickness
h_{sh}	shale thickness
k	number of test samples for neural network
m	total number of node of *jth* ANN layer
m_{inj}	number of production wells in the well pair
m_{prod}	number of production wells in the well pair
N	total number of data records
N_e^{inj}	effective number of injection wells in certain well pair
N_e^{prod}	effective number of production wells in certain well pair
N_r	number datasets sampled with parametric bootstrapping
N/G	net-to-gross ratio
n	number of folds for n-fold cross validation
P	probability
PC	principal components
PS	principal scores
R^2	coefficient of determination
S_w	water saturation
T_i^{inj}	effective injection time of *ith* injection well during the total production period of current well pair
T_i^{prod}	effective production time of *ith* production well during the total production period of current well pair
T_{total}	total production period of a given well pair
$w_{i,j}$	weight between node i in current layer and the nodes j in previous layer
$\mathbf{w}$	model parameter vector
$\overline{X}_j$	average value of the *jth* dimension of the original dataset
X_{ij}	variable in the original dataset
x	input value need to be normalized
x^N	normalized data
x_i	input value of node i in the current layer
x_{max}	maximum value of original data before normalization
x_{min}	minimum value of original data before normalization
$\mathbf{x}_n$	input vector
y_j	output value of node j in last layer
$\mathbf{y}_n$	output vector
Z_{ij}	mean-adjusted variable

Greek letters

ϕ	porosity

Acronyms

ANN	artificial neural network
COP	cumulative oil production
CSI	cumulative steam injection
EDA	exploratory data analysis
GA	genetic algorithm
GR	gamma ray logging
LM-BP	Levenberg–Marquardt backpropagation
MLP	multi-layer perceptron
MSE	mean squared error
OPR	oil production rate
PCA	principal component analysis
PSO	particle swarm optimization
RT	true resistivity logging
SAGD	steam assisted gravity drainage
SI	shale index
SLP	single-layer perceptron
SP	Spontaneous Potential logging

attractive alternative for the purposes of recovery performance prediction and uncertainty assessment, particularly when dealing with high-dimensional data space consisting of large number of operational and geological parameters. In this work, neural network is employed to analyze SAGD production performance. Despite its recent implementation as viable proxy for recovery prediction in design optimization (Popa, Cassidy, & Mercer, 2011; Queipo, Goicochea, & Pintos, 2002), its employment in data-mining frameworks for assessing heavy oil production in heterogeneous reservoirs is lacking (Ahmadloo et al., 2010; Popa et al., 2011). In particular, applied data-driven models involving actual field data from the McMurray bitumen deposits are rare. The dataset is compiled from extensive field data analysis including logging interpretation and production analysis. PCA is applied to reduce the dimensionality of the input vector, alleviate the effects of over-fitting, and improve forecast quality. Uncertainty assessment of data-driven models is another area that is less explored in the applied expert systems literature, particularly when reservoir-engineering data is involved. In this work, various sources of uncertainty including input data uncertainty, model parameter uncertainty, and model outcome uncertainty, are quantified using Monte Carlo and bootstrapping approaches. In this application, data uncertainty primarily derives from inaccurate and incorrect data, limited number of records in dataset and imprecise (indefinite) analysis criteria. Model parameter uncertainty is common with most data-driven modeling techniques like ANN, whose training can be posed as an under-determined inverse problem with non-unique solutions.

The first objective of this research is to identify a description of pertinent predicting parameters (geologic, fluid, and operating) in relation to SAGD performance prediction to improve the predictability and accuracy of these models; EDA method is applied to extract a dataset from field data assembled from various public sources. The second objective is to demonstrate the potential of customizing applied knowledge-based modeling approaches for actual field data in providing practical tools suitable for SAGD performance prediction. In this work, ANN modeling approach is applied to predict cumulative production from a number of log-derived input attributes descriptive of both reservoir heterogeneities and operational conditions. The final objective is to propose new workflows to properly quantify and assess the uncertainties in data, model, and output. An important contribution of this work is that it demonstrates the feasibility of employing data-driven models for SAGD analysis using a realistic field dataset, a subject matter that insufficiently explored in the literature. Considering that many important data such as bottom-hole pressures, fluid properties, permeability, multi-phase flow functions, and thermal conductivities are commonly unavailable and, hence, are missing in the dataset, this work demonstrates how practical knowledge-based techniques can be used to construct data-driven models capable of predicting SAGD recovery performance from log-derived and operational variables. In addition, a novel uncertainty analysis workflow is implemented to quantify the impacts of individual uncertainty on the final model predictions.