geoconvention Virtual Event September 13-15 2021

Feature extraction of petrophysical log data for machine learning-based SAGD performance prediction

*Namhwa Kim[1)], Kyungbook Lee[2)], Hyundon Shin[1)]**

1) Department of Energy Resources Engineering, Inha university, Republic of Korea

2) Department of Geoenvironmental Sciences, Kongju national university, Republic of Korea

Summary

The steam-assisted gravity drainage is the most implemented in-situ thermal recovery process to produce bitumen at Athabasca oilsands in Canada. The SAGD performance is closely related to reservoir quality and heterogeneity. The numerical reservoir simulation has been used to evaluate oilsands reservoir. However, it requires lots of static and dynamic data to generate 3D reservoir model and its forward numerical simulation. In this study, we proposed a feature extraction method of petrophysical log data for machine learning application to predict SAGD performance. The collected dataset had small number of samples to treat high dimensionality of log data at once. Therefore, feature extraction was implemented to extract relevant features that reflect reservoir payzone quality and shale barrier effectively. Multi-Layer Perceptron, which is a most popular artificial neural network in machine learning, was utilized to learn underlying patterns between extracted features and cumulative oil production for seven year operation.

Theory

The numerical simulation method with three-dimensional reservoir model has been used to predict SAGD performance. Although the simulation method is based on physics-based fluid flow as well as spatial distribution of reservoir parameter, it needs to analyze various field data to construct 3D reservoir model. It takes much time and labour for modelling work because it needs to deal with various field data (Amirian et al., 2015; Wang et al., 2019).

The ML applications have aroused focus for SAGD performance prediction (Akbilgic et al., 2015; Amirian et al., 2015; Ma et al., 2017). However, the previous studies have common thing to be improved: The previous studies averaged the heterogeneous reservoir parameters within pay zone to use them as representatives for reservoir quality. It makes the input features cannot reflect the heterogeneity.

Method

The workflow of this study consists of three parts (Fig. 1): Data collection, data pre-processing, and feature extraction method optimization. A total 63 pairs of well logs and production data were acquired with several constraints from six various SAGD fields in Athabasca oil sands area, Canada. The SAGD production wells which have more than seven years production history were selected to calculate cumulative oil production for seven-year operation (CUMOIL7). Also, the closest exploration well within SAGD drainage boundary (50 m away from SAGD horizonal well) was chosen to define a relationship between well log data and SAGD performance. For petrophysical log data, gamma ray (GR), deep resistivity (RT), and neutron porosity (NPHI) logs were selected to reflect reservoir quality. Then, several cut-off criteria were used for log data to discriminate payzone and shale barrier (SB) in 30 m vertical interval from the SAGD producer (Fig. 2).

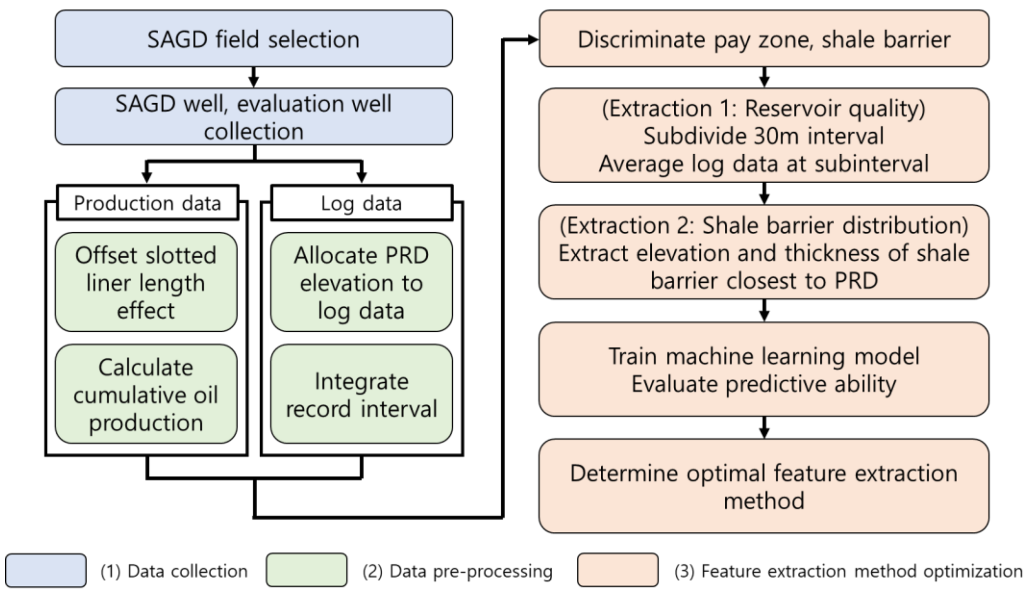
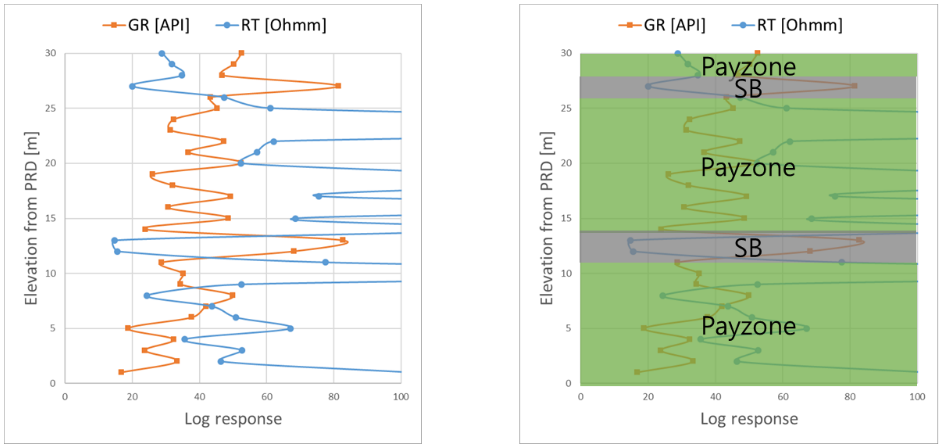


Fig. 1 Workflow of this study



(a) GR and RT log responses (b) Pay zone and shale barrier categorization

Fig. 2 Pay zone and shale barrier discrimination from log responses

The feature extraction was implemented to obtain payzone quality and vertical distribution of SB from log data. For the payzone quality, thickness (H), averaged GR, RT, NPHI of payzone in 30m pay interval were extracted (Fig. 3). However, it needed to divide 30m interval into several number of subinterval groups, because the reservoir parameters at each elevation from production well (PRD) has different effect on SAGD performance. For the vertical distribution of SB, elevation from PRD and thickness of two shale barrier, which are closest to PRD, were extracted. Because the closer SB to PRD has more detrimental effect for steam chamber expansion.

Also, the longer subintervals extracts smaller number of input features from 30-meter interval, and it would be hard to reflect heterogeneity of reservoir. Therefore, the dimension reduction of log data was examined by sensitivity analysis on subinterval length (Table 1). The base case used log responses at each 1-meter elevation for 30-meter interval. Finally, the MLP neural network were trained using extracted features to predict CUMOIL7.

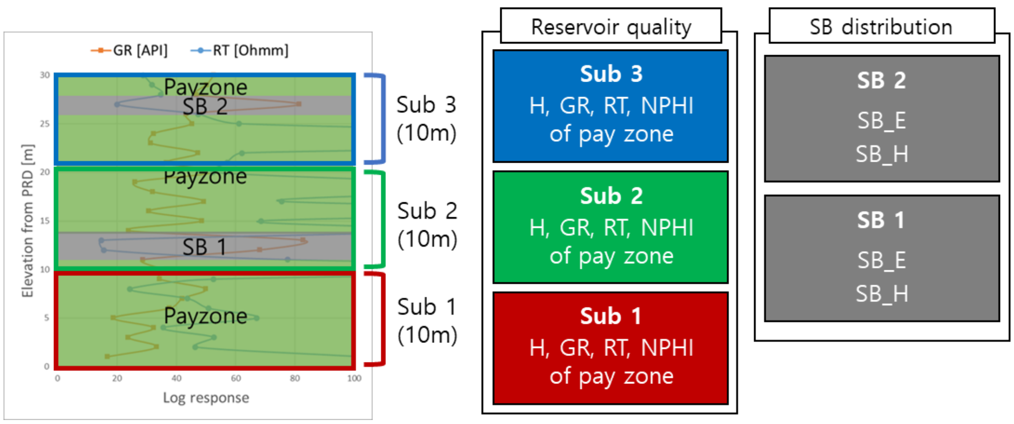


Fig. 3 An example of feature extraction by 10-meter subintervals for reservoir quality and shale barrier distribution

Table 1. Summary of feature extraction application to log data

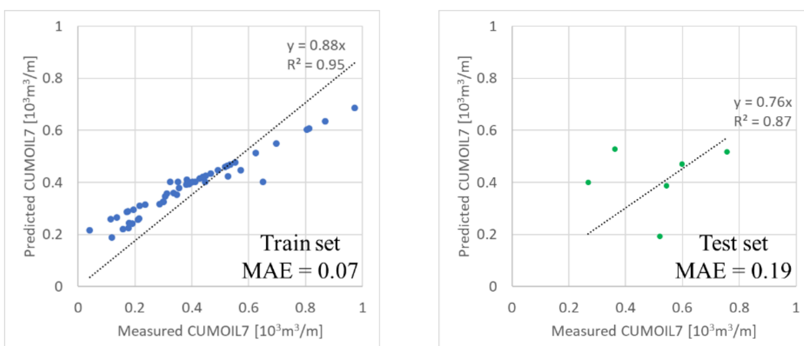
Case	Pay zone quality			Shale barrier		No. of total features	Heterogeneity reflection
	Subinterval	Extracted features	No. of features	Extracted features	No. of features		
Base	-	-	-	-	-	90	
5m	[1~5]	H, GR, RT, NPHI	4	SB_E, SB_H of SB1, SB2	4	28	↑ High
	[6~10]		4				
	[11~15]		4				
	[16~20]		4				
	[21~25]		4				
	[26~30]		4				
10m	[1~10]		4		4	16	
	[11~20]		4				
	[21~30]		4				
15m	[1~15]		4		4	12	Low ↓
	[16~30]		4				
30m	[1~30]		4		4	8	

Results

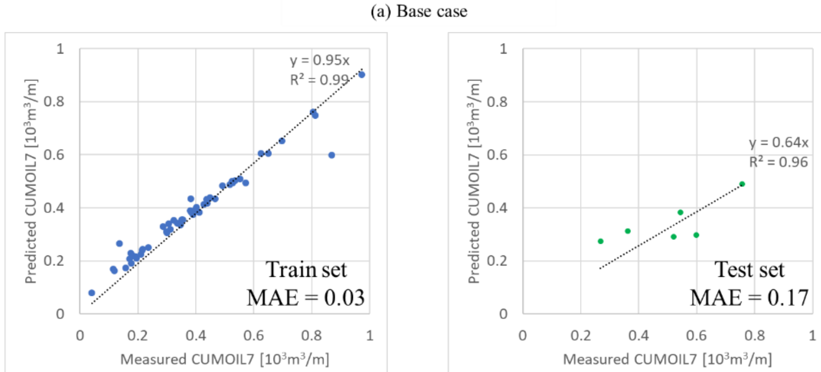
The predicted performance was evaluated by the mean absolute error (MAE; Eq. 1), which was calculated between measured CUMOIL7 and predicted CUMOIL7. The results showed that the base case, which did not implement feature extraction for log data, had the highest MAE of 0.19 for test set (Fig. 4). It means that the MLP could not capture a pattern between input and target feature correctly, because of excessive number of features than dataset size. On the other hand, the feature extraction cases showed lower MAE than that of base case.

$$MAE = \frac{1}{n}\sum |CUMOIL7_{measured} - CUMOIL7_{predicted}| \quad \text{(Eq. 1)}$$

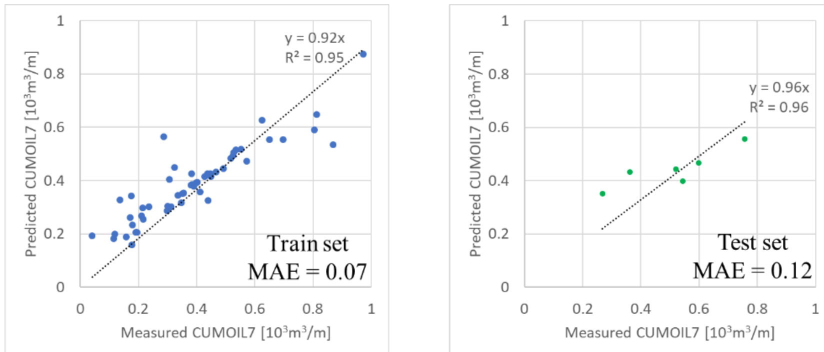
Especially, a trend was found between subinterval length and MAE value. The MAE of test set decreased by increasing subinterval length starting from base case. Then MAE of test set increased after 10m length case (Fig. 5). This result means that the feature extraction needs to find a balance point, which preserves heterogeneity of log data as well as reduces dimension of log data.



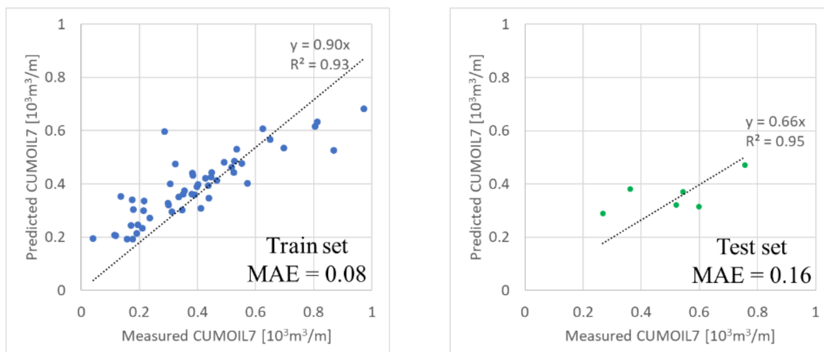
(a) Base case



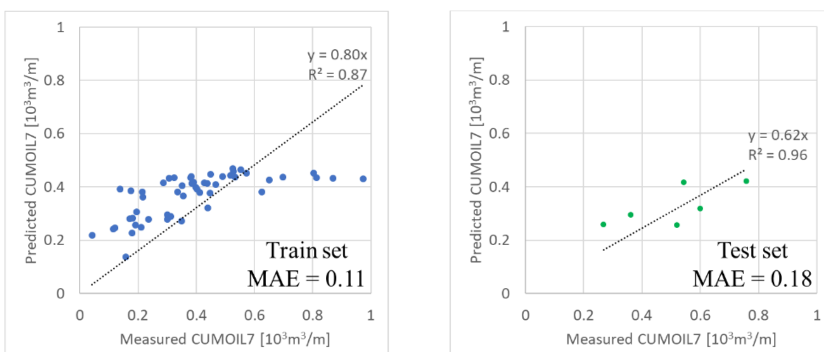
(b) 5-meter subinterval case



(c) 10-meter subinterval case



(d) 15-meter subinterval case



(e) 30-meter subinterval case

Fig. 4 Prediction results of train set and test set for each case

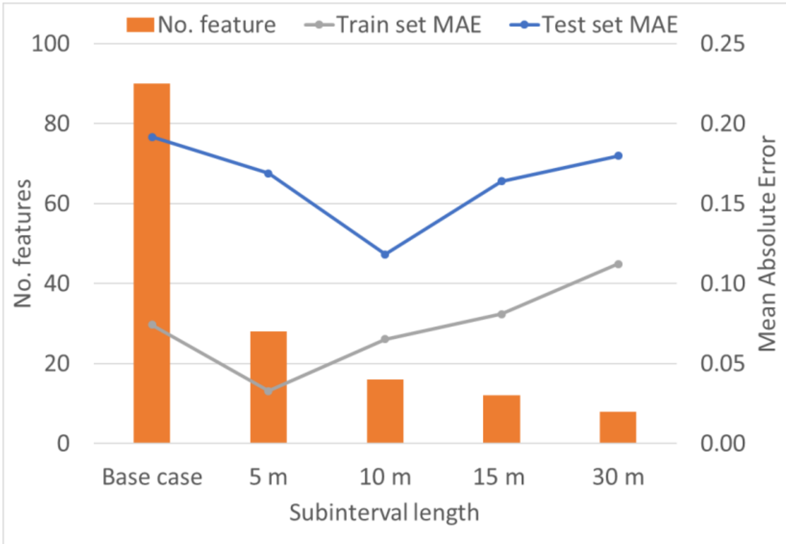


Fig. 5 Comparison of prediction results and number of input features

Acknowledgements

This study was supported by the Demonstration-scale Offshore CO2 Storage Project in Pohang Basin, Korea (20162010201980). Also, the authors thank to the Korea Institute of Geoscience and Mineral Resources for providing SAGD field data.

References

Akbilgic, O., Zhu, D., Gates, I. D., & Bergerson, J. A. (2015). Prediction of steam-assisted gravity drainage steam to oil ratio from reservoir characteristics. *Energy*, *93*, 1663–1670. https://doi.org/10.1016/j.energy.2015.09.029

Amirian, E., Leung, J. Y., Zanon, S., & Dzurman, P. (2015). Integrated cluster analysis and artificial neural network modeling for steam-assisted gravity drainage performance prediction in heterogeneous reservoirs. *Expert Systems with Applications*, *42*(2), 723–740. https://doi.org/10.1016/j.eswa.2014.08.034

Ma, Z., Leung, J. Y., & Zanon, S. (2017). Practical Data Mining and Artificial Neural Network Modeling for Steam-Assisted Gravity Drainage Production Analysis. *Journal of Energy Resources Technology, Transactions of the ASME*, *139*(3), 1–10. https://doi.org/10.1115/1.4035751

Wang, S., Chen, Z., & Chen, S. (2019). Applicability of deep neural networks on production forecasting in Bakken shale reservoirs. *Journal of Petroleum Science and Engineering*, *179*(March), 112–125. https://doi.org/10.1016/j.petrol.2019.04.016