Using Machine Learning Algorithms to Solve Polymer Flooding Problem

Authors T. Imankulov[1,2], Y. Kenzhebek[1], E. Makhmut[1], D. Akhmed-Zaki[2]
View Affiliations

Publisher: European Association of Geoscientists & Engineers
Source; ECMOR 2022, Sep 2022, Volume 2022, p.1 - 9
DOI: (Digital Object Identifier) https://doi.org/10.3997/2214-4609.202244056

« Previous | T o C | Next »

Abstract

Summary

The application of different methods of machine learning in the oil and gas industry is becoming relevant. The data–driven approach makes it possible to build excellent oil prediction models to increase oil recovery. This article discusses machine learning algorithms for solving the problem of polymer injection into an oil reservoir. The problem of supervised learning, which is one of the classes of machine learning problems, is considered. Our problem belongs to the class of regression problems in terms of machine learning methods. The considered generated data from the mathematical model were used for the training and test set. To build a machine learning model, such parameters of the oil production problem as porosity, viscosity of the oil phase, polymer injection ratio, absolute rock permeability and oil recovery factor were used. The value of the oil recovery factor was chosen as the output parameter of the regression model. Over 350 thousand generated data were applicated to implement multiple regression methods. Different regression algorithms were developed and tested, and it was also found that for our synthetic data, the considered models train quite well and predict the value of the oil recovery factor. An artificial neural network with multiple hidden layers with optimally selected hyperparameters was built. The hyperparameters of the artificial neural network were optimally selected. To prevent overfitting, the early stopping function was used, where training stops at the number of epochs without improvement. The tensorflow deep learning library was used to implement regression algorithms for predicting the oil recovery factor. In this way, it is supposed that the data-based different regression algorithms reviewed in the article can be valuable for predicting the oil recovery factor using engineering data from diffrerent oil fields during processing stages.

Most Read This Month

OA An Overview of Carbon Capture and Storage and its Potential Role in the Energy Transition

Authors Joseph M. English and Kara L. English
pp.: 35–40 (6)

OA The future of work in exploration: A new model to bring people, technology and data together

By Paul Genberg

pp.: 81–84 (4)

OA Carbon Capture and Storage Potential in Ireland — Returning Carbon Whence It Came

Authors Joseph M. English and Kara L. English
pp.: 35–43 (9)

OA Deep neural network application for 4D seismic inversion to changes in pressure and saturation: Optimizing the use of synthetic training datasets

Authors Gustavo Côrte, Jesper Dramsch, Hamed Amini and Colin MacBeth
pp.: 2164–2185 (22)

OA The principles of helium exploration

Authors Diveena Danabalan, Jon G. Gluyas, Colin G. Macpherson, Thomas H. Abraham-James, Josh J. Bluett, Peter H. Barry and Chris J. Ballentine

Well Services Operations in Offshore Unmanned Platform; Challenges and Solutions

Authors O. Alyafei and J. Foster

OA Constraining basin thermal history and petroleum generation using palaeoclimate data in the Piceance Basin, Colorado

Authors Yao Tong, Daniel E. Ibarra, Jeremy K. Caves, Tapan Mukerji and Stephan A. Graham

Shallow Shear Wave Reflections: What Can the Amplitude Tell Us and How?

By Ranajit Ghose

+ More

Introduction

The application of chemical flooding in oil reservoirs has become an interesting subject of research in recent times. In the oil recovery process, when waterflooding reaches its recovery capacity limit, chemical methods are applied to the reservoirs as a tertiary method. Various chemicals are injected into reservoirs to change the flow of fluids and improve sweep efficiency. For the process of enhanced oil recovery, various methods have been developed, such as polymer, surfactant-polymer, alkaline-surfactant-polymer flooding and others. Currently, polymer flooding is considered one of the technical and commercially proven methods in oil recovery processes. The main advantage of such chemical methods is that these methods can be operated continuously in combination with waterflooding without the use of any complex integrated systems.

There are several works related to carbon dioxide injection in oil production processes (Mohammadi et al., 2014; Kashkooli et al., 2022). The work (Mohammadi et al., 2014) considers the injection of CO^2 into oil reservoirs using neural networks and mathematical models to predict the oil recovery factor. The authors used 252 datasets that were generated from a mathematical model to train a regression model and a three-layer neural network. As a result, the authors found that the forecast of the neural network during CO2 injection is wrong by about 0.396% and is better consistent with real data than the regression model. Another paper explored the application of neural networks for multi-criteria optimization of the oil production process with alternating water and CO^2 (Menad and Noureddine, 2019). The authors built three multilayer perceptron models using the LMA, BR, and SGC algorithms. As a result, it was revealed that the proposed Levenberg-Marquardt algorithm predicts output data with high accuracy and with minimal time.

The authors (Chakra et al., 2013) explore the application of a higher-order neural network to forecasting cumulative oil production. The authors modeled a higher-order neural network with insufficient field data from fields in India. As a pre-processing, the authors reduced the noise in the data and performed a cross-correlation analysis to select the optimal input parameters of the neural network. A single computational layer was used with different test cases where the number of neurons in the hidden layer varied between 1 and 10. As a result, the authors showed that using a higher order neural network, the model predicts the cumulative oil production with high accuracy, where the percentage of absolute mean error was 3.99%.

A study (Alkhatib et al., 2013) examined the application of the Least Squares Monte Carlo (LSM) method using surfactant flooding. The authors considered the impact of uncertainty on the technical and economic parameters of oil production. 3D homogeneous and 2D heterogeneous reservoir models were chosen as models. As a result, the authors found that the Monte Carlo least squares method is able to obtain flexibility values and mitigate circumstances in the face of uncertainty when implemented by surfactants. In addition, the results of the flexibility of homogeneous and heterogeneous models are shown.

The following papers discuss the valuation of information (VOI) for various reservoir applications. The authors of the work (Barros et al., 2015) considered the assessment of the value of information in a closed cycle of management (CLRM) for assessing the added value in the face of geological uncertainty. The authors applied elements of history matching and optimization from the framework developed by CLRM with the theory of value of information, considering three examples in the problem of optimal reservoir management. However, according to the authors, the computational load of the method under consideration is significantly large, which should be optimized. The authors (Dutta et al., 2019) proposed regression models for VOI analysis for underground energy applications. The study used PLSR, PCR and Random Forest regression models, where RF gives less uncertainty than other methods. As a result, the authors found that the use of regression models to effectively estimate the value of information would significantly reduce the computational experiments of VOI analysis for spatial decision and data collection problems. In a similar work (Tadjer et al., 2021), the authors proposed an alternative to assessing the value of information for making optimal decisions in reservoir management. To reduce computational costs, an alternative method has been applied which is called approximate

dynamic programming. The authors tested the performance of LSM, PCR, PLC and neural network regression methods to solve the optimal start of the polymer flooding process in reservoirs. In conclusion, the authors believe that in order to enhance oil recovery for 2D and 3D models, ML regression methods can be used to determine the uncertainty of the state and economic parameters, which in turn will contribute to making the right decisions.

There are many works related to the application of chemical and polymer flooding in oil reservoirs (Ahmadi and Pournik, 2016; Jahani-Keleshteri, 2017). One such study (Ahmadi and Pournik, 2016) examines the application of least squares support vector machine (LSSVM) to solve the problem of chemical flooding in oil reservoirs. 202 datasets were used for training, where each has 7 chemical flooding inputs and 2 outputs. Surfactant concentration parameters and surfactant sample size are of high importance relative to other parameters. As a result, the authors developed a predictive model for monitoring oil recovery and net present value. In another similar work (Jahani-Keleshteri, 2017), where the authors applied the LSSVM method to predict distillate consumption in the process of steam distillation of oil production.

The authors' study (Ebaga-Ololo and Chon, 2017; Larestani et al., 2022) reviewed artificial neural network polymer flood performance prediction and surfactant polymer flood performance using cascaded neural networks and gradient boosting. As a result of the sensitivity analysis, the authors found that the most influencing factor on the output parameter of the oil recovery factor was the size of the polymer plug, the concentration of the surfactant, and the size of the surfactant portion. The authors (Le Van and Chon, 2016) proposed a neural network approach to evaluate the performance of alkaline surfactant polymer flooding in viscous oil reservoirs using a simulation dataset. The input parameters of the neural network consisted of 13 neurons with different parameters of two types of polymer and 8 neurons in the hidden layer, and the oil recovery coefficient was taken at the output. In conclusion, the authors argue that the predictive oil recovery factor developed by the ANN for economic analysis helps to flexibly adapt to changes in the oil price.

The authors' study (Amirian et al., 2018) describes a polymer flooding process for heavy oil. And studies (Daripa and Dutta, 2017; Dang et al., 2018) have reviewed surfactant polymer flood simulations using a hybrid method and hybrid EOR processes in low salinity ASP flooding.

In the work (Kang et al., 2016), the authors devoted to the use of an artificial neural network to assess the viscosity of polymer solutions. As is known, the viscosity of the injected polymer is the most important property for enhanced oil recovery, which was tested in this work with FlopaamTM 3330S, FlopaamTM 3630S and AN-125 solutions. Temperature, polymer concentration, NaCI, Ca^{2+} and Shear rate concentration were taken as input parameters. The test results showed that the predicted results of the ANN are in excellent agreement with the measured values of the used solutions, and the estimation of the viscosity according to the Carro model in some cases is not accurate due to changes in the influencing parameters.

Method and/or Theory

The aim of this study is to apply machine learning methods to solve the problem of polymer flooding. The dataset was generated from the Buckley-Leverett mathematical model by simulation for various values of oil parameters. The Buckley-Leverett model is described as follows:

$$m\frac{\partial s}{\partial t} + div\vec{v}_1 = 0 \tag{1}$$

$$-m\frac{\partial s}{\partial t} + div\vec{v}_2 = 0 \tag{2}$$

$$\vec{v}_i = -K_0 \frac{f_i(s)}{\mu_i} \nabla P \tag{3}$$

where K_0 – permeability tensor, m – porosity, s – water saturation, f_i, μ_i – relative phase permeabilities and viscosities of liquids of the corresponding phases, which are dependences of the following form:

$$f_1(s) = s^{3.5}; \; f_2(s) = (1-s)^{3.5}; \tag{4}$$

The equation for the distribution of polymer concentration:

$$m \frac{\partial}{\partial t}(c_p s) + div(v_1 c_p) = div(D_p s \nabla c_p) \tag{5}$$

where c_p – polymer concentration in water, D_p – diffusion coefficient.

Data collection

The dataset used in this study was derived from a Buckley-Leverett mathematical model in which polymer flooding was simulated. The dataset was generated for various input oil production parameters. The original dataset contains 4800 data. Each data had 4 inputs: porosity, oil phase viscosities, rock absolute permeability, polymer injection ratio. In addition, the oil recovery factor was selected at the output. The ranges and amount of data generated are shown in Table 1.

Parameter	Type	Amount of variation	Min	Max	Mean
Polymer concentration	Input	6	0.05	0.3	0.175
Porosity	Input	20	0.11	0.289	0.167
Viscosity	Input	20	0.105	0.485	0.243
Absolute permeability	Input	2	0.01	0.095	0.045
Recovery factor	Output		0.15	0.86	0.5

Table 1 *Description and statistical analysis of implemented data samples*

Machine Learning Methods

This paper considers the use of such regression methods as multiple linear regression, support vector regression and random forest regression. There are various types of regression algorithms such as linear regression, ridge regression, polynomial regression, lasso regression, etc., each suitable for different purposes. Linear regression is one of the simplest and most commonly used forms of ML regression algorithms. The equation for linear regression looks like this:

$$y = mx + b$$

where m is the slope of the line, b is the point of intersection, x is the independent variable (input feature), y is the dependent variable (output feature). In linear regression, "least squares" is used to find the slope and y-intercept. Least squares minimizes the square of the distance from each data point to the fitted line. For multiple linear regression (MLR) models with multiple input features, you can use the following equation:

$$y = m_1 x_1 + m_2 x_2 + \cdots + m_n x_n + b$$

When using a multiple regression model, it is assumed that there is a linear relationship between the input and output features. In addition, it is assumed that the independent variables (input features) are

continuous, not discrete. Most of the problems associated with petroleum engineering have a non-linear relationship between input and output features. Therefore, this algorithm should be used with caution when solving some nonlinear problems.

Support vector regression is a type of supervised learning algorithm that is used to predict discrete values. The Support Vector Regression (SVR) function is very similar to how the SVM method works. The main idea of SVR is to find the most suitable line. In SVR, the best matching line is the hyperplane with the maximum number of points (Belyadi and Haghighat, 2021).

Random forest regressor (RFR) is another powerful supervised machine learning algorithm that can be used for both regression and classification problems. A random forest is an ensemble of decision trees, or it can be thought of as a forest of decision trees. Because random forest combines many decision tree models into one, it is known as an ensemble algorithm. Combining multiple decision trees into a single model is, in fact, the fundamental concept of using a random forest. Predictions made by a single decision tree may not be accurate, but when combined, the prediction will be closer to the average. The reason a random forest is usually more accurate than a single decision tree is that much more knowledge is included from many predictions. For regression problems, random forest uses the mean of the decision trees for the final prediction.

To evaluate the regression models, the metrics were chosen as the correlation coefficient R^2, the mean square error MSE , the root mean squared error RMSE and the mean absolute error (MAE). One of the reasons why MSE is widely used in real ML applications is that larger errors are reported more when using MSE as an objective function compared to MAE.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - y_{pred}^{(i)})^2$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y^{(i)} - y_{pred}^{(i)}\right|$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - y_{pred}^{(i)})^2}$$

Results and Discussion

As mentioned earlier, the data set contains 4800 data. The dataset was divided into a training and test set. For training, 70% of the data was used, and for the test, the remaining 30% of the data. The dataset has been tested using algorithms such as multiple linear regression, random forest regression (RFR), and support vector regression (SVR). Figure 1 shows the result of predicting the oil recovery factor for the linear regression method.

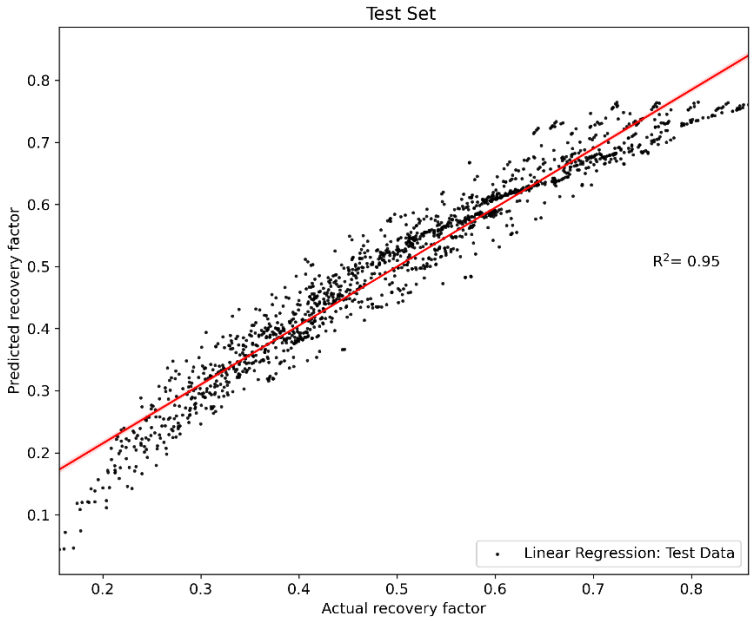

Figure 1 *Predicted results for recovery factor by multiple linear regression for test set.*

The following Figure 2 shows the result for the Random Forest Regressor method.

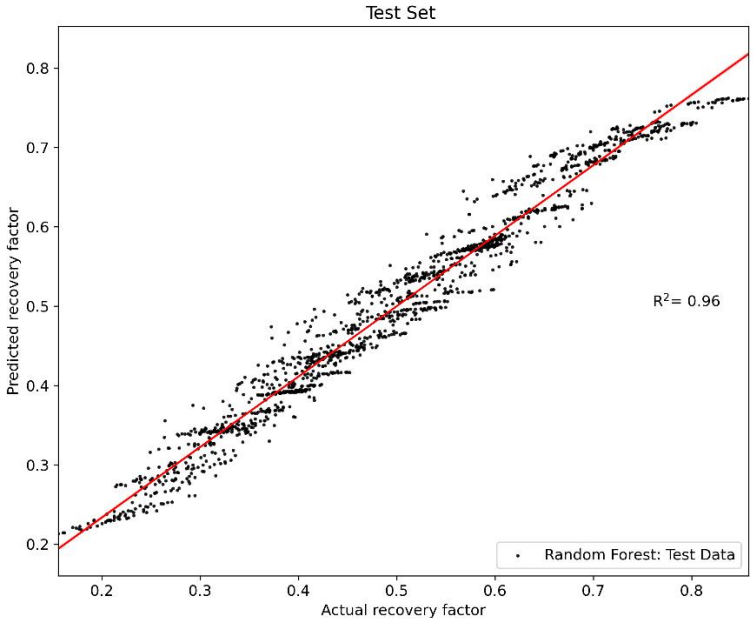

Figure 2 *Predicted results for recovery factor by Random Forest for test set.*

The following figure 3 shows the result for the SVR method.

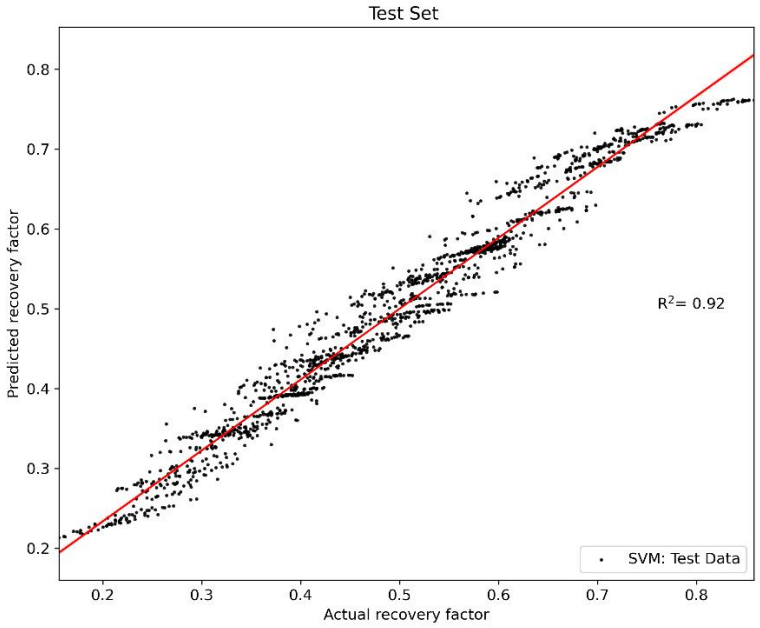

Figure 3 *Predicted results for recovery factor by SVR for test set.*

Finally, Table 2 presents a comparison between the regression models used in terms of statistical errors.

Statistical parameters	MLR	Random Forest Regressor (RFR)	Support Vector Regressor (SVR)
Training Correlation coefficient (R^2)	0.948	0.971	0.935
Test Correlation coefficient (R^2)	0.946	0.963	0.917
Mean square error (MSE)	0.0013	0.0009	0.002
Mean absolute error (MAE)	0.0284	0.023	0.0373

Table 2 *Statistical parameters of proposed approaches in predicting the efficiency of polymer flooding.*

As can be seen from Table 2, the Random Forest Regressor model has the highest accuracy from a statistical point of view.

Conclusions

This article was devoted to the application of classical machine learning methods to solve the problem of polymer flooding based on polymer concentration, viscosity, porosity and absolute permeability. A dataset consisting of 4800 data was generated by simulation using the Buckley-Leverett mathematical model for various values of oil production parameters. As a result, multiple linear regression models, random forest regression, and support vector regression models were presented for oil recovery factor prediction. The results of the comparison showed that the RFR model gives higher accuracy than other methods. This model predicted the oil recovery factor with an error of about 3.7% in terms of the coefficient of determination. An estimate was also made where MAE error rates are 2.8% for MLR, 3.7% for SVR, and 2.3% for the RFR method.

Acknowledgements

This research was funded by the Science Committee of the Ministry of Education and Science of the Republic of Kazakhstan (Grant No. AP09260564).

References

Ahmadi, M., Pournik, M. [2016]. A predictive model of chemical flooding for enhanced oil recovery purposes: Application of least square support vector machine. *Petroleum*, Volume 2, Issue 2, pp. 177-182, ISSN 2405-6561. https://doi.org/10.1016/j.petlm.2015.10.002.

Alkhatib, A., Babaei, M., King R. [2013]. Decision Making Under Uncertainty: Applying the Least-Squares Monte Carlo Method in Surfactant-Flooding Implementation. *SPE J.*, 18: 721–735. https://doi.org/10.2118/154467-PA

Amirian, E., Dejam, M., Chen, Z. [2018]. Performance forecasting for polymer flooding in heavy oil reservoirs. *Fuel*, Volume 216, pp. 83-100, ISSN 0016-2361. https://doi.org/10.1016/j.fuel.2017.11.110

Barros, E. G., Leeuwenburgh, O., Hof, P. M., Jansen, J. D. [2015]. Value of Multiple Production Measurements and Water Front Tracking in Closed-Loop Reservoir Management. *Paper presented at the SPE Reservoir Characterisation and Simulation Conference and Exhibition, Abu Dhabi*, UAE, September 2015. https://doi.org/10.2118/175608-MS

Belyadi, H., Haghighat, A. [2021]. Machine Learning Guide for Oil and Gas Using Python: Chapter 5 - Supervised learning. Gulf Professional Publishing, pp. 169-295, ISBN 9780128219294. https://doi.org/10.1016/B978-0-12-821929-4.00004-4.

Chakra, C., Song, K., Gupta, M., Saraf, D.N. [2013]. An innovative neural forecast of cumulative oil production from a petroleum reservoir employing higher-order neural networks (HONNs). *Journal of Petroleum Science and Engineering*, Volume 106, pp. 18-33, ISSN 0920-4105. https://doi.org/10.1016/j.petrol.2013.03.004.

Dang, C., Nghiem, L., Nguyen, N., Yang, C., Chen, Z., Bae, W. [2018]. Modeling and optimization of alkaline-surfactant-polymer flooding and hybrid enhanced oil recovery processes. *Journal of Petroleum Science and Engineering*, Volume 169, pp. 578-601, ISSN 0920-4105. https://doi.org/10.1016/j.petrol.2018.06.017.

Daripa, P., Dutta, S. [2017]. Modeling and simulation of surfactant–polymer flooding using a new hybrid method. *Journal of Computational Physics*, Volume 335, pp. 249-282, ISSN 0021-9991. https://doi.org/10.1016/j.jcp.2017.01.038.

Dutta, G., Mukerji, T., Eidsvik, J. [2019]. Value of information analysis for subsurface energy resources applications. *Applied Energy*, Volume 252, 113436, ISSN 0306-2619. https://doi.org/10.1016/j.apenergy.2019.113436.

Ebaga-Ololo, J., Chon, B. H. [2017]. Prediction of Polymer Flooding Performance with an Artificial Neural Network: A Two-Polymer-Slug Case. *Energies* 10, no. 7: 844. https://doi.org/10.3390/en10070844

Jahani-Keleshteri, Z. [2017]. A robust approach to predict distillate rate through steam distillation process for oil recovery. *Petroleum Science and Technology*, 35:5, 419-425. https://doi.org/10.1080/10916466.2016.1261160

Kang, P., Jong-Se, L., Chun, H. [2016]. Artificial Neural Network Model to Estimate the Viscosity of Polymer Solutions for Enhanced Oil Recovery. *Applied Sciences* 6, no. 7: 188. https://doi.org/10.3390/app6070188

Kashkooli, S.B., Gandomkar, A., Riazi, M., Tavallali, M.S. [2022]. Coupled optimization of carbon dioxide sequestration and CO2 enhanced oil recovery. *Journal of Petroleum Science and Engineering*, Volume 208, Part E, 109257, ISSN 0920-4105. https://doi.org/10.1016/j.petrol.2021.109257.

Larestani, A., Mousavi, S.P., Hadavimoghaddam, F., Ostadhassan, M., Hemmati-Sarapardeh, A. [2022]. Predicting the surfactant-polymer flooding performance in chemical enhanced oil recovery: Cascade neural network and gradient boosting decision tree. *Alexandria Engineering Journal*, Volume 61, Issue 10, pp. 7715-7731, ISSN 1110-0168. https://doi.org/10.1016/j.aej.2022.01.023.

Le Van, S., Chon, B.H. [2016]. Artificial Neural Network Model for Alkali-Surfactant-Polymer Flooding in Viscous Oil Reservoirs: Generation and Application. *Energies* 9, no. 12: 1081. https://doi.org/10.3390/en9121081

Menad, N.A., Noureddine, Z. [2019]. An efficient methodology for multi-objective optimization of water alternating CO2 EOR process. *Journal of the Taiwan Institute of Chemical Engineers*, Volume 99, pp. 154-165, ISSN 1876-1070. https://doi.org/10.1016/j.jtice.2019.03.016.

Mohammadi, M., Kouhi M., Mohebbi A. [2014]. Prediction of Oil Recovery Factor in CO2 Injection Process. *Petroleum Science and Technology*, 32:17, 2093-2101. https://doi.org/10.1080/10916466.2012.743563

Tadjer, A., Bratvold, R., Hong, A., Hanea, R. [2021]. Application of machine learning to assess the value of information in polymer flooding. *Petroleum Research*, Volume 6, Issue 4, pp. 309-320, ISSN 2096-2495. https://doi.org/10.1016/j.ptlrs.2021.05.006.