



Novel approach for predicting water alternating gas injection recovery factor

Lazreg Belazreg¹ · Syed Mohammad Mahmood¹ · Akmal Aulia¹

Received: 1 November 2018 / Accepted: 6 May 2019 / Published online: 20 May 2019
© The Author(s) 2019

Abstract

Water alternating gas (WAG) injection process is a proven EOR technology that has been successfully deployed in many fields around the globe. The performance of WAG process is measured by its incremental recovery factor over secondary recovery. The application of this technology remains limited due to the complexity of the WAG injection process which requires time-consuming in-depth technical studies. This research was performed for a purpose of developing a predictive model for WAG incremental recovery factor based on integrated approach that involves reservoir simulation and data mining. A thousand reservoir simulation models were developed to evaluate WAG injection performance over waterflooding. Reservoir model parameters assessed in this research study were horizontal and vertical permeabilities, fluids properties, WAG injection scheme, fluids mobility, trapped gas saturation, reservoir pressure, residual oil saturation to gas, and injected gas volume. The outcome of the WAG simulation models was fed to the two selected data mining techniques, regression and group method of data handling (GMDH), to build WAG incremental recovery factor predictive model. Input data to the machine learning technique were split into two sets: 70% for training the model and 30% for model validation. Predictive models that calculate WAG incremental recovery factor as a function of the input parameters were developed. The predictive models correlation coefficient of 0.766 and 0.853 and root mean square error of 3.571 and 2.893 were achieved from regression and GMDH methods, respectively. GMDH technique demonstrated its strength and ability in selecting effective predictors, optimizing network structure, and achieving more accurate predictive model. The achieved WAG incremental recovery factor predictive models are expected to help reservoir engineers perform quick evaluation of WAG performance and assess a WAG project risk prior launching detailed time-consuming and costly technical studies.

Keywords WAG · Recovery factor · Reservoir simulation · Machine learning

List of symbols

API	Oil gravity	MMP	Minimum miscibility pressure
D	Depth (ft)	n	Total number of observations
EOR	Enhanced oil recovery	m	Number of input vectors to the data mining method
FVF	Formation volume factor (Rbbl/stb)	OOIP	Original oil in place (MMstb)
GOR	Gas–oil ratio (SCF/STB)	P	Reservoir pressure (psi)
h_p	Reservoir thickness (ft)	P_b	Bubble point pressure
HCPVI	Hydrocarbon pore volume injected (fraction)	RF	Recovery factor (%)
K	Reservoir horizontal permeability (md)	Incr. WAG RF	Incremental WAG recovery factor
K_h	Reservoir horizontal permeability (md)	R_s	Solution gas–oil ratio
K_v	Reservoir vertical permeability (md)	T	Reservoir temperature (°C)
		WAG	Water alternating gas
		WOR	Water oil ratio
		WPVI	Water pore volume injected
		μ_o	Oil viscosity (cp)
		γ_g	Gas gravity

✉ Lazreg Belazreg
lbelazreg@gmail.com

¹ Department of Geosciences and Petroleum Engineering,
University Teknologi PETRONAS, Jalan Desa Seri Iskandar,
32610 Bota, Perak, Malaysia

Prediction model input vectors

P1	Horizontal permeability (md)
P2	Permeability anisotropy (%)
P3	API
P4	Gas gravity
P5	Water viscosity (cp)
P6	S_{org} (%)
P7	Land coefficient
P8	WAG cycle (months)
P9	Solution gas–oil ratio (m^3/m^3)
P10	WAG ratio
P11	Pore volume of injected water at WAG startup (%)
P12	Reservoir pressure (bars)
P13	Hydrocarbon pore volume of injected gas (fraction)

Introduction

With the decline in the production rate of petroleum reservoirs and increase in energy demand, E&P operators have started evaluating and implementing enhanced oil recovery (EOR) technology to extract the remaining oil after primary and secondary recoveries (i.e., waterflooding, gas injection). EOR technology has demonstrated promising results in enhancing field recovery factor and maintaining field production plateau.

Currently, the ultimate oil recovery factor is about 35% which means that two-thirds of the oil remain underground. Increasing the recovery factor from 35 to 45% would bring about 1 trillion bbl. of Oil (Labastie 2011).

WAG injection process is one the proven EOR technologies (Christensen et al. 2001). In recent years, water alternating gas (WAG) flooding process has gained an increasing interest in enhancing oil recovery worldwide. Additional economic profit from WAG process is the reduction of the required amount of gas to be injected into the reservoir (Jaber et al. 2017).

The WAG injection process in oil fields has shown an increase in the recovery factor typically ranging from 5 to 10% over water or gas injection (Christensen et al. 2001). However, the application of this technology remains limited due to the complexity of the WAG process and difficulty in quantifying the expected performance prior launching time-consuming and costly technical studies. Technical study usually starts by coreflood experiment, followed by complex reservoir model construction, and then WAG pilot test to calibrate the expected WAG performance. Complexity of the WAG process is mainly related to the WAG physical process and WAG optimization.

Afzali et al. (2018) demonstrated that the complexity of the WAG physical process is mainly related to three-phase

flow, inter-phase mass transfer, swelling, oil trapping, and water blocking by the injected gas that are not well understood by scientist and researchers.

Field-scale WAG project optimization complexity usually related to the time and a cost each of the optimization tasks takes with lack of robust and powerful technique that makes the WAG project profitable (Panjalizadeh et al. 2015; Chen et al. 2010).

Figure 1 demonstrates the typical WAG project approval workflow.

For the seek of developing WAG incremental recovery factor predictive model that forecasts WAG incremental recovery factor prior launching detailed expensive technical study, this research work was performed. This research approach was based on an integrated study that involved reservoir modeling and data mining. The study started by literature review of WAG process and data mining techniques, followed by building a thousand reservoir simulation models for WAG and waterflooding based on selected sensitivity parameters, and then building a predictive model from the reservoir modeling study input using the two selected data mining techniques which are regression and GMDH. A thousand reservoir models for waterflooding and WAG injection were developed based on full factorial design of experiment (DOE), using ten (10) input sensitivity variables. Three additional input parameters which are hydrocarbon pore volume of injected gas, reservoir pressure, and pore volume of injected water prior WAG startup were output or calculated from the reservoir models.

The selected reservoir modeling study parameters and ranges are based on literature review of published WAG pilot projects and WAG studies, plus one factor at time (OFAT) sensitivity. The list of selected sensitivity parameters are horizontal permeability, vertical permeability, oil gravity, gas gravity, water viscosity, solution gas–oil ratio, WAG ratio, WAG cycle, land coefficient, and residual oil saturation to gas.

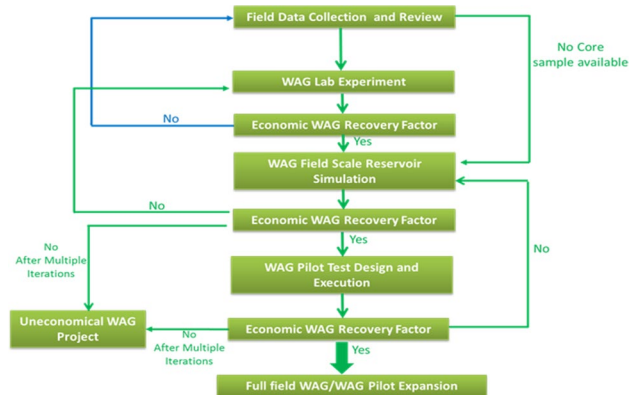


Fig. 1 Typical workflow for big scale WAG project approval

Research methodology

This research work was done through multiple steps as listed below:

- WAG literature review,
- List the parameters that demonstrated an impact on WAG recovery from the literature review and OFAT sensitivity study,
- Construct and run a thousand reservoir models for both waterflooding and WAG based on full factorial design of experiment,
- Prepare reservoir model data for data mining study. This includes the selected thirteen (13) parameters and calculated WAG incremental recovery factor. Total observations used in this research study were four thousand two hundred ninety (4290) observations,
- Select regression and group method of data handling (GMDH) methods for predictive model construction based on literature review,
- Run predictive model training using 70% of the total observations, followed by model validation using the remaining 30%.

Figure 2 summarizes the selected variables (plus injected gas volume) for WAG incremental recovery factor prediction.

Water alternating gas recovery factor and mechanisms

The overall recovery factor (efficiency) RF of any secondary or tertiary oil recovery method is the product of a combination of three individual efficiency factors as given by the following generalized expression (Tarek 2010):

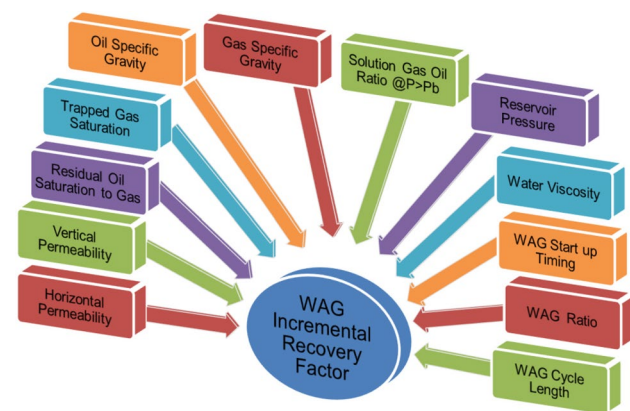


Fig. 2 Selected variables for WAG incremental recovery factor prediction

$$RF = E_D E_A E_V \quad (1)$$

where RF = the overall recovery factor (0–1). E_D = the displacement efficiency (0–1). E_A = the areal sweep efficiency (0–1). E_V = the vertical sweep efficiency (0–1).

The overall recovery factor is a function of multiple factors including fluids mobilities, injection patterns, areal and vertical heterogeneities, degree of gravity segregation, and total pore volume injected (Tarek 2010).

WAG incremental recovery factor is a result of the increase in both displacement and volumetric sweep efficiencies due to the reduction of the residual oil saturation and improvement of both areal and vertical sweep efficiencies.

Multiple research papers were published during the last decades on WAG recovery mechanisms. This includes three-phase WAG hysteresis, residual oil to gas, mobility control, and oil vaporization and swelling.

Lazreg et al. (2017) demonstrated the impact of two-phase and three-phase WAG hysteresis on WAG incremental recovery factor based on an integrated research study that incorporated findings from both lab experiments and reservoir simulation from multiple Malaysian oil fields. This technical paper illustrated that three-phase WAG hysteresis could increase WAG incremental recovery factor by 1–2% on top of secondary recovery. Skauge and Larsen (1994) demonstrated that the residual oil saturation by three-phase flow was significantly lower than the residual oil saturation from two-phase waterflooding and gas injection.

Mobility ratio is an important factor that controls volumetric sweep efficiency of gas injection process with a favorable mobility of less than one (< 1). Reduction of the mobility ratio can be obtained by increasing the gas viscosity or reducing the relative permeability of the fluids. Reduced mobility of the gas phase can be achieved by injecting water and gas alternately. It is essential to adjust the amount of water and gas to achieve the best possible displacement efficiency. Too much water will result in poor microscopic displacement, and too much gas will result in poor vertical, and possibly horizontal, sweep (Christensen et al. 2001).

Oil swelling and vaporization in the presence of oil and gas phases is one the components of the incremental WAG recovery factor. The improvement of oil recovery during gas EOR include oil swelling, gas–oil interfacial tension (IFT) reduction, oil viscosity reduction, and extraction of light and intermediate hydrocarbons for immiscible flooding to completely miscible displacement (Tunio et al. 2011; Cao and Gu 2013; Blunt et al. 1993). Chordia and Trivedi (2010) showed that when CO_2 contacts the oil, swelling occurs, causing the oil to expand and move toward the producing well. Observations suggest that when the oil and gas mix, drainage rates become higher in the oil zone, driving the excess oil toward the fractures.

Reservoir model input and selected parameters ranges for WAG recovery factor prediction

Table 1 shows the main reservoir input data for reservoir model used during this WAG study.

Reservoir properties sensitivity

Multiple research work demonstrated that reservoir permeability is one of the main factors controlling WAG performance. Yu et al. (2017) showed CO₂–water alternating flooding experiment results which indicates that it is permeability that mainly impacts the displacement efficiency of CO₂–EOR in low-permeability reservoir.

The effect of vertical segregation was studied by Jackson et al. (1985), which concluded that the relationship between permeability ratio and oil recovery rates is of inverse proportions. Laboratory investigation found that a lower kv/kh will generally result in a slightly higher recovery factor in heterogeneous reservoir due to a more dominant of vertical permeability (Tham et al. 2011).

Table 2 shows the range used for horizontal and vertical permeabilities. The ranges were selected covering wide range of oil field reservoirs.

Figures 3 and 4 demonstrate the dependency of the WAG recovery factor and performance on horizontal and vertical permeabilities.

The simulation results from horizontal permeability sensitivity demonstrated that the higher the horizontal permeability, the higher the initial oil production rate under WAG injection process; however, the ultimate WAG recovery

factor might be lower with high permeability if WAG process was not properly optimized. Gas override was one of the issues that lead to oil production loose with high gas–oil ratio (GOR) in this case.

This issue has been reported in few WAG pilots as it was demonstrated by Pritchard et al. (1990).

Sensitivity on vertical permeability in this research study demonstrated that generally the higher the vertical permeability, the higher the field recovery factor under WAG injection. This result demonstrated that gravity segregation can have positive effect on WAG performance under low to moderate reservoir permeability (i.e., average horizontal permeability of 45 md).

Fluid properties sensitivity

Based on a wide variety of reservoir conditions, a range was taken for the different input parameters of black oil PVT model. Table 3 summarizes PVT variables and ranges used in this study.

The importance of fluid properties in estimating fluid in place, understanding fluid flow in porous media, developing a reservoir, and optimizing the ultimate recovery was demonstrated by many authors and researchers (Tarek 2010; Ling and Shen 2011; Satter and Iqbal 2016; Denney 2012).

Table 1 Reservoir model input data

Basic reservoir and fluid properties			
Reservoir		Fluid	
Rock	Sandstone	Crude oil type	Light oil
Porosity (fraction)	0.149	Oil gravity	Variable
Horizontal permeability (md)	Variable	Gas gravity	Variable
Vertical permeability (md)	Variable	Solution GOR (Sm ³ /Sm ³)	Variable
Dimensions XY (m)	100 × 100	Oil viscosity (cp)	The function of oil gravity, gas gravity, initial solution GOR
Initial water saturation (fraction)	0.1	Gas viscosity (cp)	
Residual oil saturation to water (fraction)	0.25	Oil FVF (RB/STB)	
Residual gas saturation to gas (fraction)	Variable	Gas FVF (ft ³ /scf)	
Max trapped gas (fraction)	Variable	Oil and gas compressibilities (1/psi)	
Initial pressure (bar)	340	Water viscosity	Variable
Reservoir temperature (°C)	100	Water FVF (Rm ³ /Sm ³)	1
Depth (m)	3000	Water compressibility (1/bar)	4.52E–5

Table 2 Reservoir permeability sensitivity

Input variable	Min value	Max value
Horizontal permeability (md)	50	1000
Permeability anisotropy (K_v/K_h)	0.01	1

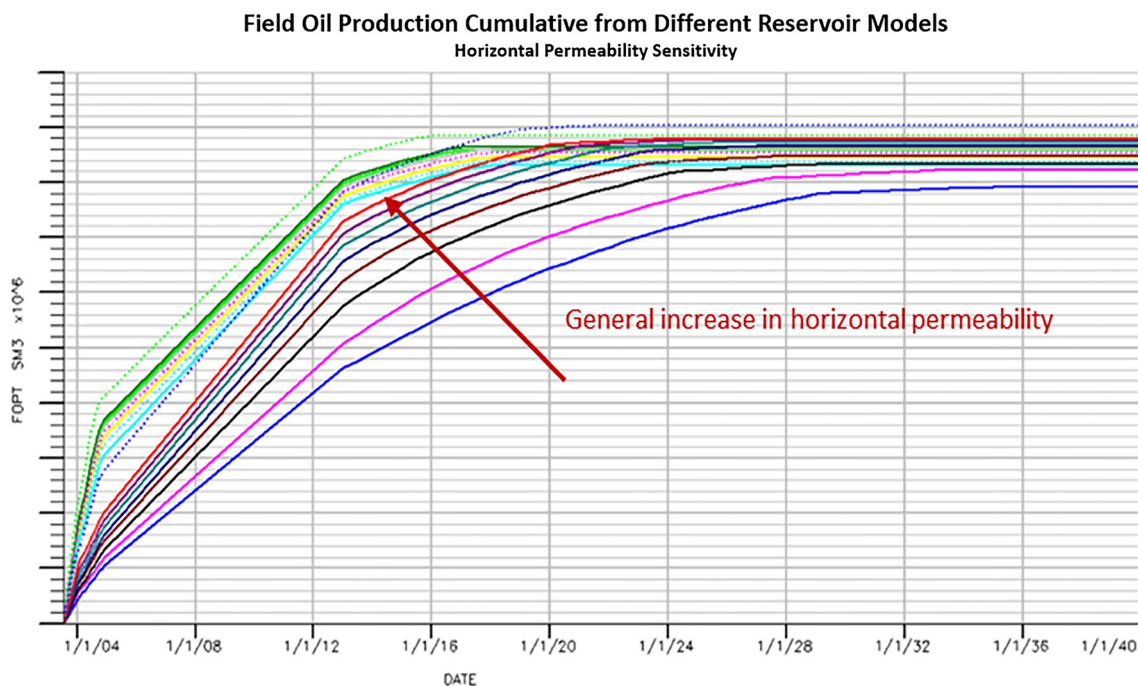


Fig. 3 Impact of reservoir horizontal permeability on WAG recovery factor

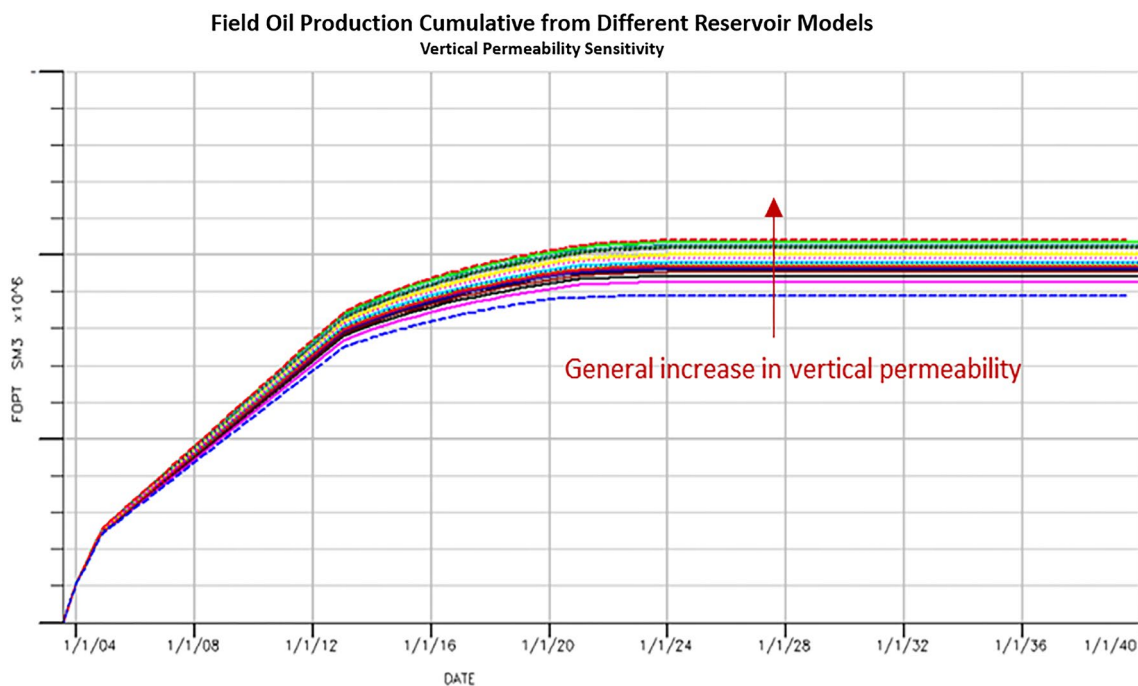


Fig. 4 Impact of reservoir vertical permeability on WAG recovery factor

Tarek (2010) demonstrated how oil viscosity affects fractional flow curve for both water-wet and oil-wet rock systems. Figure 5 shows that regardless of system wettability, higher oil viscosity shifts the curve upward leading to lower

recovery factor from waterflooding. However, higher water viscosity shifts the curve downward.

Yavuz et al. (2019) demonstrated that the lower oil density will have higher mobility and flow with low resistance,

Table 3 PVT input variables sensitivity

Input variable	Min value	Max value
Solution GOR (scf/STB)	350	2000
Oil gravity	25	50
Gas specific gravity	0.55	0.9
Water viscosity ^a (cp)	0.1	1

^aWater viscosity was sensitized independently from other water PVT data

whereas higher oil density will have lower mobility and flows with a high resistance.

The correlations used in building the different PVT used in this study tables based on PVT sensitivity parameters are summarized in “Appendix A” section.

Relative permeability sensitivity

Two parameters are sensitized in this study, land coefficient and ratio between residual oil saturation to gas and residual oil saturation to water. Land coefficient is one of the input for calculating of the trapped gas saturation that control gas

mobility during the imbibition period, while the residual oil saturation to gas indicates the additional movable oil under gas injection as compared to waterflooding.

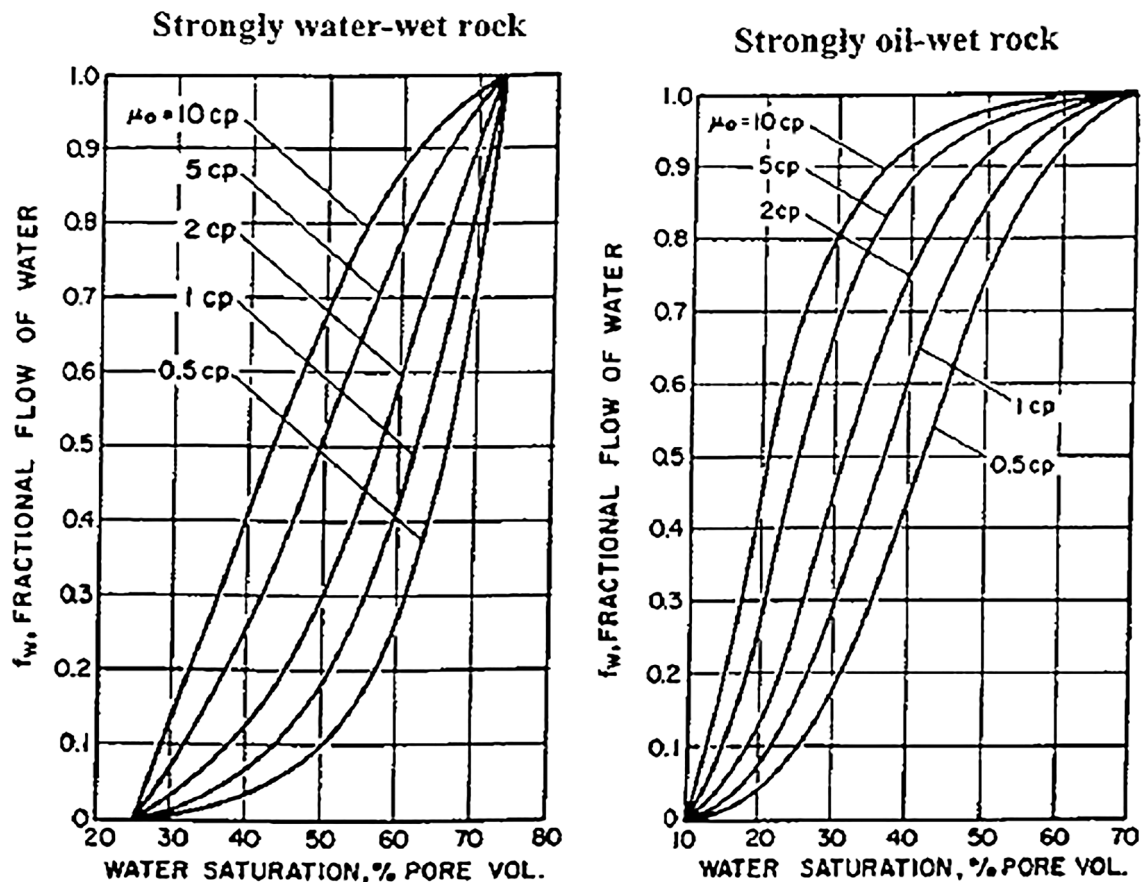
Two-phase hysteresis model was considered in this study, and hence, both drainage and imbibition curves for water–oil and gas–oil were included. Table 4 summarizes the sensitivity taken on the relative permeability curves related to trapped gas saturation and residual oil saturation to gas.

Trapped gas saturation is one of the main parameters contributing to the incremental WAG recovery factor. Efficiency of trapped gas saturation is expressed in the form of additional oil recovery as a fraction or percentage of the gas quantity that remains in the pore space during the waterflooding process (Feigl 2011).

Table 4 Relative permeability input sensitivity

Input variable	Min value	Max value
Land coefficient	1	6
Ratio S_{org}/S_{orw}	0.2	1

^a S_{orw} is 0.25 in this study

**Fig. 5** Effect of oil viscosity on fractional flow curve (Tarek 2010)

During WAG process, gas is always the non-wetting phase. Gas will then come through to the center of the reservoir pores, and water/oil will drain around the edges of the gas. In this case, low residual oil saturation is always expected. Moreno et al. (2011) reported that gas flooding promoted lower residual oil saturation than water flooding based on coreflood experiments.

Lazreg et al. (2017) has demonstrated the benefits of two-phase and three-phase hysteresis in reducing the water and gas mobilities, increasing oil mobility, and reducing the residual oil saturation as compared to water or gas flood. This process of injecting alternatively cycles of imbibition and drainage causes residual oil saturation to be lower than those of waterflooding and gas flooding.

As per Eclipse simulator manual (Eclipse Simulator Reference Manual 2014), $S_{g\text{trap}}$ is calculated as per Eq. 1

$$S_{g\text{trap}} = S_{g\text{cr}} + \frac{(S_{g\text{m}} - S_{g\text{cr}})}{(1 + C * (S_{g\text{m}} - S_{g\text{cr}}))} \quad (2)$$

where $S_{g\text{trap}}$ is the trapped gas saturation, $S_{g\text{m}}$ is the maximum gas saturation attained, and $S_{g\text{cr}}$ is the critical gas saturation.

In the other hand, theoretically speaking the lower the residual oil saturation, the higher the movable oil in the reservoir. However, Mahesh and Britt (2015), demonstrated that lower oil residual reservoirs are not always producing higher oil volumes due to other reservoir heterogeneities.

WAG injection scheme

WAG injection scheme is one of the critical parameters that control WAG injection performance, as WAG ratio, WAG cycle, WAG slug size, injection rate, WAG duration, and start-up timing (Yang et al. 2008).

WAG slug size is one of the important WAG design parameters. The ratio of water slug size to gas slug size, as one of the WAG injection scheme parameters, was found to strongly affect the trapping mechanism during the WAG flooding process (Rogers and Grigg 2000).

Another parameter is cycle length (CL) of the WAG flooding process which is one of the important parameters affecting the whole process. It was found practically to be the critical factor controlling the WAG process design (Behrouz et al. 2007). Wu et al. (2004) concluded in their study that the cycle length during miscible WAG flooding was found to be a critical factor in the WAG process design in the heterogeneous reservoir (Jaber et al. 2017).

The WAG ratio is very important parameter in WAG process design (Chen et al. 2010, Farshid et al. 2010). A WAG ratio of 1:1 is the most common in field applications (Christensen et al. 2001). However, WAG ratio strongly

depends on availability of gas to be injected and injection wells capacity (John and Reid, 2000).

The assumption of WAG startup after 10 years of waterflooding was taken in this study. Table 5 summarizes the sensitivity parameters ranges used in this study.

Figures 6 and 7 demonstrate the impact of WAG cycle length and WAG ratio on WAG performance and recovery factor.

The main conclusions from sensitivity analysis on WAG cycle length and WAG ratio are:

- Shorter WAG cycle length gave higher recovery factor which was mainly due to the improvement of injectant fluids mobilities that caused higher volumetric sweep efficiency.
- High WAG ratio accelerated the WAG recovery factor at the beginning of the WAG injection; however, continuing with high WAG ratio leads to early gas breakthrough and lose of well productivity for few reservoir models.
- Reservoir pressure and gas–oil ratio at production wells are factors that control periodic practical WAG ratio during WAG injection process.

However, low WAG cycle length can increase the logistics and operational cost significantly.

Gas utilization cost versus incremental WAG reserves from increasing gas injection ratio is a crucial factor in WAG project economics.

Reservoir simulation models construction and input

A reservoir simulation model with two producers and two injectors was selected for this WAG research study. Waterflooding was injected for the first ten (10) years followed by either water or WAG injection till end of field life. Table 1 summarizes basic reservoir model input.

A thousand reservoir simulation models were created, based on FFD design of experiment, and simulated under waterflooding and WAG injection processes till field life. The reservoir models selected parameters plus WAG incremental recovery factor were input to the two data mining techniques for prediction model training and validation.

Table 5 WAG scheme sensitivity

Input variable	Min value	Max value
WAG ratio	3:1	1:5
WAG cycle (month)	2	24
WPVI@WAG startup ^a	–	–

^aWPVI: calculated water pore volume injected at WAG start-up

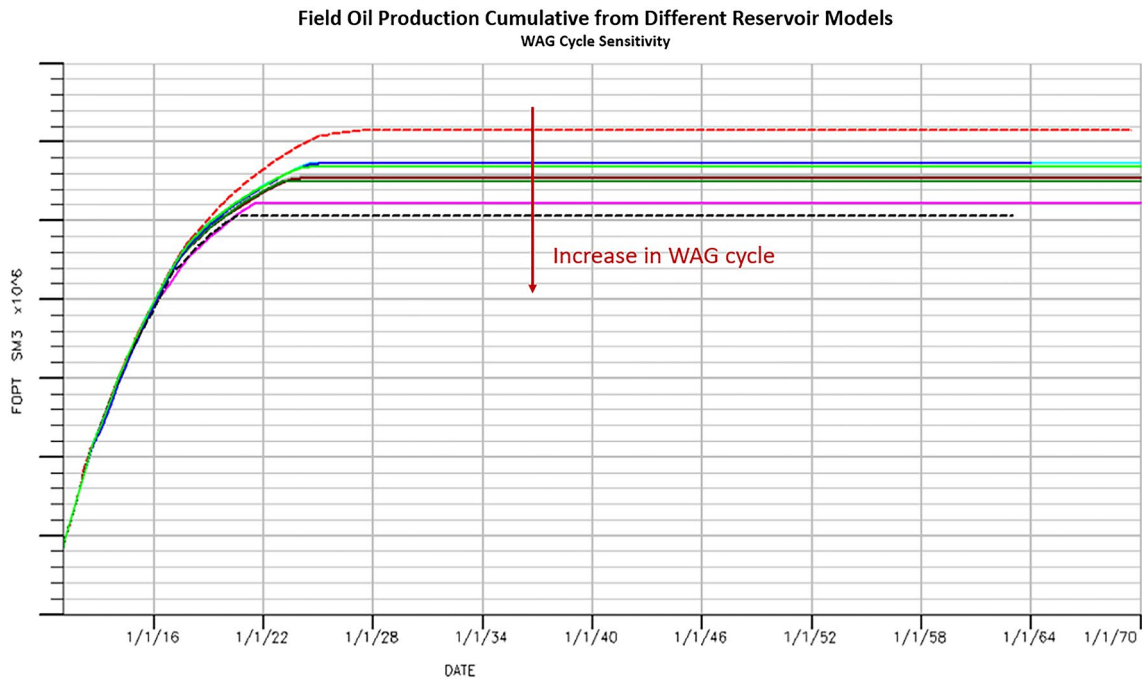


Fig. 6 Impact of WAG cycle length on WAG recovery factor

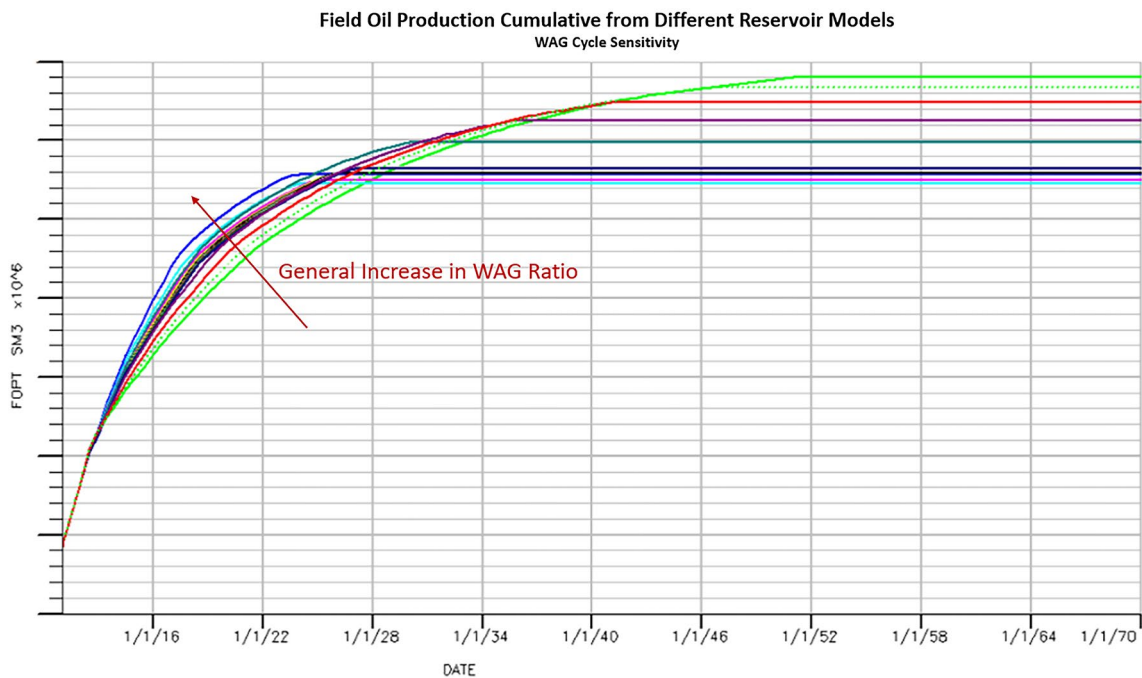


Fig. 7 Impact of WAG ratio on WAG recovery factor

Data mining techniques

Data mining (DM) is the computational process of discovering patterns in large data sets (“big data”) involving methods

at the intersection of artificial intelligence, machine learning, statistics, and database systems (James et al. 2013).

Data mining is an important part of the processes of knowledge discovery in medicine, economics, finance, telecommunication, and various scientific fields. Data mining

helps to uncover hidden information from an enormous amount of data that are valuable for the recognition of important facts, relationships, trends, and patterns (Medvedev et al. 2017).

Nowadays, DM has attracted a lot of attention in data analysis area, and it became a recognizable new tool for data analysis that can be used to extract valuable and meaningful knowledge from data (Ahmed et al. 2016).

As a highly application-driven domain, data mining has incorporated many techniques from other domains such as statistics, machine learning, pattern recognition, database and data warehouse systems, information retrieval, visualization, algorithms, high-performance computing, and many application domains (Fig. 8) (Han et al. 2012).

Statistics studies the collection, analysis, interpretation/explanation, and presentation of data. Data mining has an inherent connection with statistics. A statistical model is a set of mathematical functions that describe the behavior of the objects in a target class in terms of random variables and their associated probability distributions. Statistical models are widely used to model data and data classes (Han et al. 2012).

Machine learning is a learning method that automates the acquisition of knowledge, and it plays an important role in artificial intelligence research. An intelligent system without learning ability cannot be regarded as a real intelligent system, but the intelligent system in the past was generally lack of learning ability (Teng and Gong 2018).

Machine learning

Machine learning, by its definition, is a field of computer science that evolved from studying pattern recognition and computational learning theory in artificial intelligence. It is the learning and building of algorithms that can learn from and make predictions on data sets. These procedures operate by construction of a model from example inputs in order to

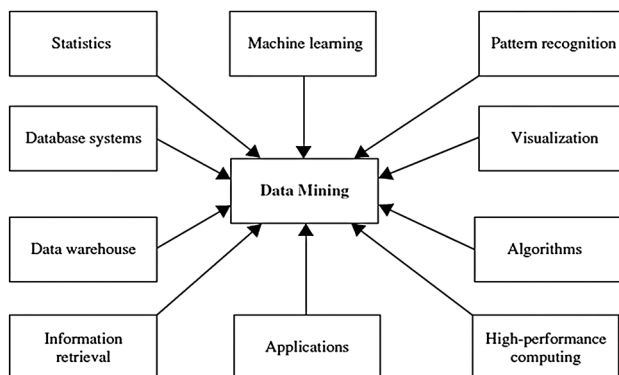


Fig. 8 Data mining adopts techniques from many domains (Han et al. 2012)

make data-driven predictions or choices rather than following firm static program instructions (Simon et al. 2016).

Machine learning involves two types of tasks:

- Supervised machine learning where the program is trained based on a pre-defined set of data, which then facilitate the program ability to get an accurate conclusion with new data.
- Unsupervised machine learning where the program is given set of data for list of vectors, and program must find relationships and patterns therein.

The most popular approaches to machine learning are artificial neural networks and genetic algorithms (Negnevitsky 2011). Artificial neural network (ANN) is computing system inspired by biological neural networks that constitute human brains. ANN is capable of approximating nonlinear functional relationships between input and output variables (Kim et al. 2018). The basic processing elements of neural networks are neurons. Neurons in ANN are characterized by a single, static, continuous-valued activation. A collection of neurons is referred to as a layer, and the collection of interconnected layers forms the neural networks (Kim et al. 2018).

Figure 9 shows a typical structure for an artificial neural network.

The development of neural networks was introduced to partly improve the modeling procedure, but their high degree of subjectiveness in the definition of some of their parameters as well as the demand of long data samples remains significant obstacles (Anastasakis and Mort 2001).

The group method of data handling (GMDH) is family of inductive self-organizing data-driven approach that requires small data samples, and it has the ability in optimizing neural network models structure objectively. GMDH technique has been used in data mining, knowledge discovery, prediction, complex system modeling, and pattern recognition (Lemke and Motzev 2016).

In this research study, two data mining techniques were used to develop WAG incremental recovery factor which are regression technique, and group method of data handling (GMDH).

Regression

Regression is a statistical technique to determine the relationship between two or more variables. It is used for predicting an output as a function of given input vectors. There are multiple types of regression techniques starting by the simplest regression technique which is linear regression and then the other advanced regression techniques.

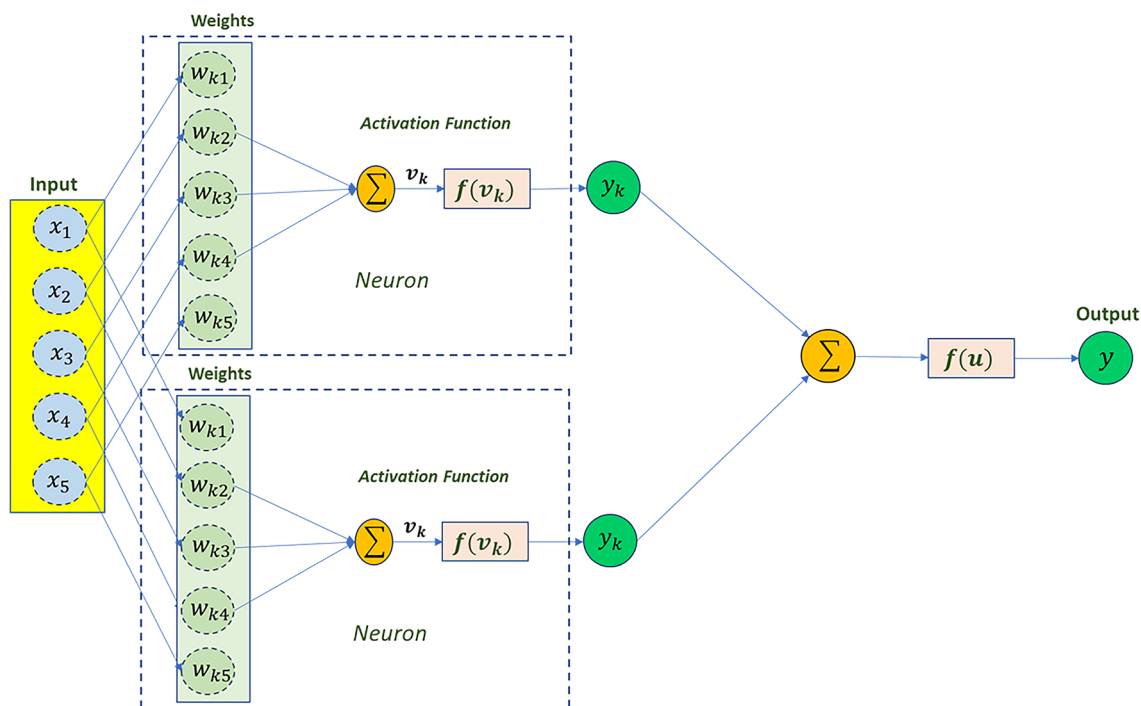


Fig. 9 Neural network typical structure

Multiple regression is a technique for modeling the association among the scalar dependent variable V and one or more descriptive variables indicated by Y . It predicts the future value of the variable with respect to other variables. $V = w_0 + w_1y_1 + \dots + w_ny_n + \epsilon$ where V implies the dependent variable, $w_0 - w_n$ implies the coefficients, $y_1 - y_n$ implies the independent variables, and ϵ implies the random error (Bini et al. 2016).

Group method of data handling (GMDH)

GMDH was developed to produce a model by looking only at input data and the desired output (Semenov et al. 2010). GMDH is a supervised feed-forward networking model in which the original input vectors are used to generate the initial layer of the network, with each subsequent layer feeding its outputs to the next layer. The model's underlying concept resembles animal evolution or plant breeding, as it adheres to the principle of natural selection. The multilayer criterion preserves superior networks for successive generations, eventually yielding an optimal network (Tsong-Min and Pei-Hwa 2016).

The topology of the GMDH network is determined using a layer-by-layer pruning process based on the pre-defined criterion of what are the best nodes at each layer. Farlow (1981) recognized that many types of mathematical models require the modeler to know the system variables that may generally be very difficult to find. The modeler will be

forced to guess these variables; this guess not only is time-consuming but also produces unreliable prediction models.

GMDH uses an iterative polynomial regression procedure to synthesize any model. The polynomial regression equations can produce a high-order polynomial model using effective predictors. Farlow (1981) started by computing the quadratic polynomial regression equation:

$$y = a + bx_i + cx_j + dx_i^2 + ex_j^2 + fx_ix_j$$

where y is the output sample, (x_i, x_j) is a pair of input samples; and $a, b, c, d, e,$ and f are the polynomial coefficients to be determined by the training data set.

The matrix of input variables x_{ij} with m predictors, n observations, and the output variable y_i is defined below.

$$\begin{pmatrix} x_{11} & \dots & x_{m1} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ x_{1n} & \dots & x_{mn} \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

During the training, GMDH will use the input matrix of n observations and $m+1$ input variables (m x_{ij} independent variables and one dependent variable Y_i).

The training iterations will start by taking all the independent variables (two columns at a time) and then constructing the quadratic regression polynomial (equation) that best fits the dependent variables. Each pair of input

vectors will form a final quadratic regression polynomial equation. The first layer is constructed using m independent variables and the dependent variable for form $k = m(m - 1)/2$ regression polynomials. New variables ($z_{1n}, z_{2n}, \dots, z_{kn}$) that describe better the dependent variable will be input to the second layer, and so on. Less effective variables will be eliminated using either regulatory criterion or root mean squared error.

Training iteration will continue to produce new variables that describe the solution better than the previous ones until the minimum error value of the current layer is greater than the previous one. Figures 10 and 11 show typical layout and vectors matrix of GMDH method.

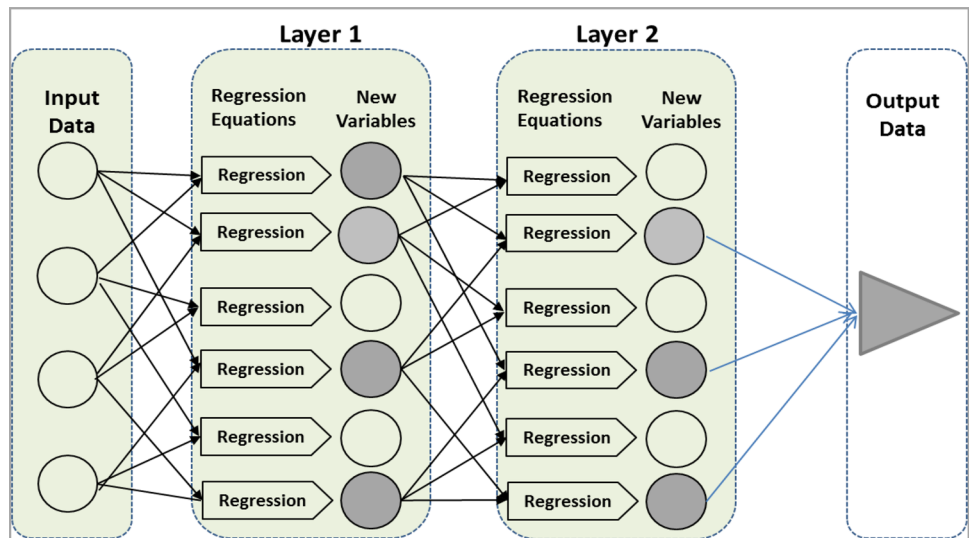
GMDH technique has been used in data mining, knowledge discovery, prediction, complex system modeling, and pattern recognition (Lemke and Motzev 2016).

The following criterion was used to measure the error between actual and predicted WAG incremental recovery factor.

Mean squared error

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - F(x_i))^2}{n}}$$

Fig. 10 GMDH method layout



$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & \dots & \dots \\ \dots & \dots & x_{ij} & \dots & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{n1} & \dots & \dots & \dots & x_{nm} \end{bmatrix} \Rightarrow Z = \begin{bmatrix} z_{11} & z_{12} & \dots & \dots & z_{1,m(m-1)/2} \\ z_{21} & z_{22} & \dots & \dots & \dots \\ \dots & \dots & \dots & z_{ij} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_{n1} & \dots & \dots & \dots & z_{n,m(m-1)/2} \end{bmatrix} \dots \Rightarrow Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}$$

$$y_i = a_{pq} + b_{pq}x_{ip} + c_{pq}x_{iq} + d_{pq}x_{ip}^2 + e_{pq}x_{iq}^2 + f_{pq}x_{ip}x_{iq}$$

Fig. 11 GMDH method algorithms

$$MAE = \frac{\sum_{i=1}^n |y_i - F(x_i)|}{n}$$

where n : number of observations, y_i : vector of observed values, $F(x_i)$: vector of predicted values, S^2 : variance

GMDH external criterion

It is regularity criterion used to test the model adequacy. It evaluates the output of each new neuron in the GMDH network using the pre-defined regulatory criterion (i.e., root mean square error (RMSE) between the predicted and actual outputs of a neuron). Neurons that fulfill the regulatory criterion will survive and are used as input to the next layer, and neurons that do not fulfill the criterion will be discarded.

Building GMDH model procedure

The steps in building a GMDH model are:

Step 1 Divide the input data into training and test sets

The input data are divided into training and test sets. The training set data are used to train the model and estimate

certain characteristics of the nonlinear system, and the test set is then used to validate the model and determine the complete set of characteristics.

Step 2 Generate new variables in each layer

New variables (neurons) for each layer are generated from the combinations of input variables. The number of combinations is given by:

$C_r^m = \frac{m!}{r!(m-r)!}$, where m is the number of input vectors and r is usually set to two (Farlow 1981).

With $m=2$, new variables count as per previous equation is $C_2^m = \frac{m(m-1)}{2}$

Step 3 Optimization principle for elements in each layer

Regression analysis is applied to the training data to calculate the optimum partial descriptions of the nonlinear system.

Usually root mean square (RMS) is used as an external criterion (index) to screen out underperforming neurons in each layer. RMS is defined as:

$$RMS = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

where \hat{y}_i is the actual output, y_i is predicted output, and n is number of observations.

RMSE balance is the criterion used in this study.

Step 4 Stopping rule for the multilayer structure generation

By comparing the index value of the current layer with that of the next layer to be generated, further layers are prevented from being developed if the index value does not improve or falls below a certain objective default value; otherwise, steps 2–3 are repeated until the value matches the limited condition set above.

Difference between GMDH and neural networks

The major differences of the two techniques can be summarized in the ability of GMDH to objectively select the optimum model, avoid model overfitting problems, and select the most relevant input variables. Anastasakis and Mort (2001) summarized features of both NN and self-organizing modeling in a variety of categories as shown in Table 6.

WAG prediction models construction

Reservoir simulation study outcomes which consist of WAG incremental recovery factor and the thirteen parameters were input to the regression and GMDH models. A total of four thousand two hundred ninety (4290) observations were used in this study: 70% of the data used for training the model and 30% used for model validation.

Table 6 Comparison between artificial neural network and group method of data handling

	Neural network	GMDH
Data analysis	Universal approximator	Structure identificatory
Architecture	Indirect approximation	Direct
Network synthesis	Preselected unbounded network structure, Experimental selection of adequate architecture demands time and experience	Sounded network structure
Threshold	Globally optimized fixed network structure	Structure evolved during the estimation process
Self-organization	Threshold transfer functions	Adaptive synthesized structure
Parameter estimation	Deductive, given number of layers and number of nodes	Threshold objective functions
Optimization	In a recursive way demands long samples	Inductive, number of layers and of nodes estimated by minimum of external criterion
On/off line	Global search of a highly multimodal surface, result depends on initial solutions, slow and tedious, requiring the user to set various algorithmic parameters by trail and error, time-consuming techniques	Estimation in batch by means of maximum likelihood techniques using all the observational data, extremely short samples
Regularization	Observation is available transiently in a real-time Environment	Group method of data handling, not-time-consuming technique adaptively synthesised networks are more parsimonious, parts of the network which are inappropriate are automatically not included
A priori information knowledge	Without, only internal information	Data are usually stores and repeatedly accessible
	Without transformation in the world of neural networks not	Estimation on training set, selection on testing set
		Can be used directly to select the reference functions and criteria

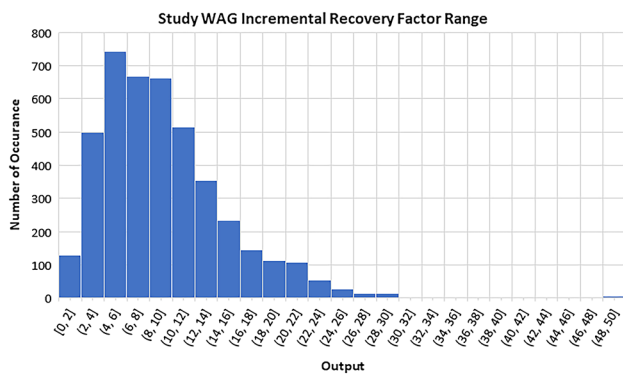


Fig. 12 WAG incremental recovery factor distribution

The reservoir simulation model WAG incremental recovery factor distribution is illustrated in Fig. 12. The study showed WAG incremental recovery factor on top of waterflooding is generally ranging from 5 to 15% with few reservoir models showing less than 5% or higher than 15% values.

Regression model training and validation

The reservoir simulation data were split into training and validation sets and input to the regression model, and a prediction model was achieved after multiple iterations with a correlation coefficient of 0.766 and 0.761 for training and validation. Figures 13 and 14, and Table 7 summarize the results of WAG incremental recovery factor model from regression method.

The WAG incremental recovery factor prediction model details are shared under “Appendix B” section.

Group method of data handling results

The reservoir simulation data were split into training and validation sets and input to the GMDH model, and a prediction model was achieved after multiple iterations with a correlation coefficient of 0.853 and 0.848 for training and validation. Figures 15 and 16, and Table 8 summarize the results of WAG incremental recovery factor model from GMDH method.

The WAG incremental recovery factor prediction model details are shared under “Appendix B” section.

Results and discussion

The research study results can be summarized in three main results related to model sensitivity parameters selection, reservoir modeling of WAG and waterflooding, and data mining prediction models.

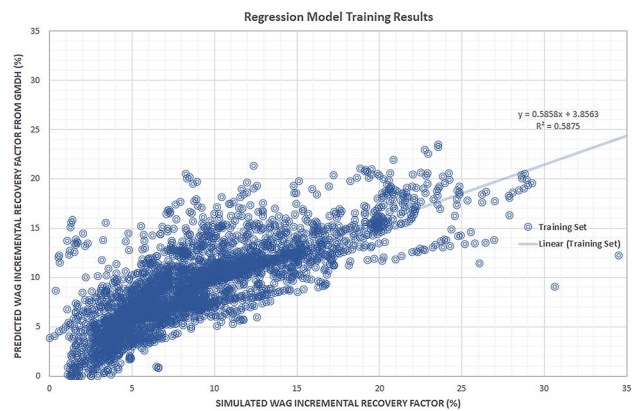


Fig. 13 Regression WAG incremental recovery factor prediction model training results

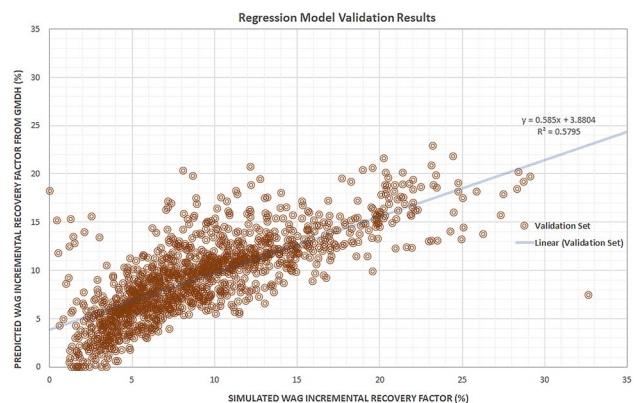


Fig. 14 Regression WAG incremental recovery factor prediction model validation results

Table 7 Regression WAG incremental recovery factor prediction model output

Regression method output	Training	Validation
Number of observations	3003	1287
Mean absolute error (MAE)	2.45989	2.44422
Root mean square error (RMSE)	3.57118	3.56289
Correlation coefficient	0.766466	0.761228
Coefficient of determination (R^2)	0.587464	0.57932

First of all, one factor at time (OFAT) study and WAG literature reviews demonstrated that few parameters have an impact on the recovery factor trend and ultimate value as it was case for horizontal permeability and injected gas volume. On the other hand, few parameters showed an impact on the shape of the WAG recovery factor but less effect on the ultimate recovery factor as it was case for WAG ratio.

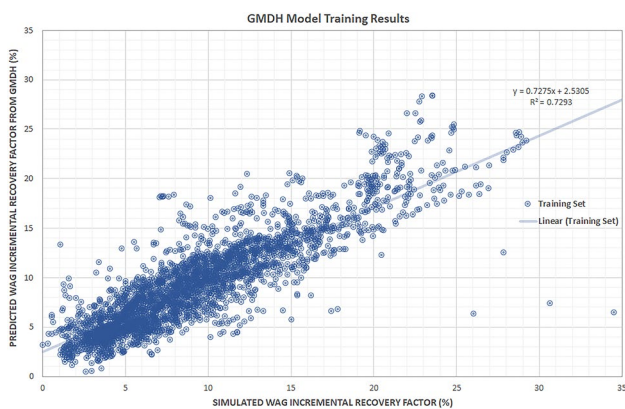


Fig. 15 GMDH WAG incremental recovery factor prediction model training results

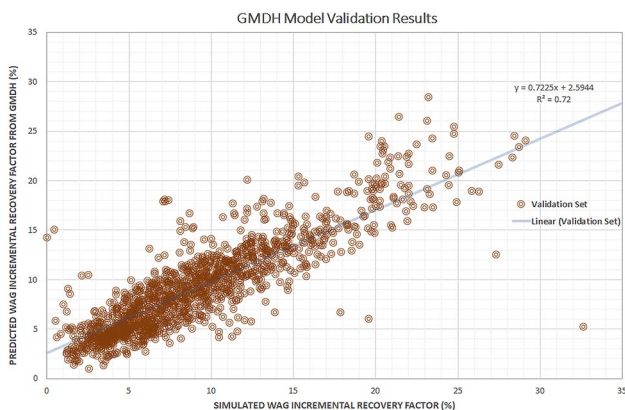


Fig. 16 GMDH WAG incremental recovery factor prediction model validation results

Table 8 GMDH WAG incremental recovery factor prediction model output

Group method of data handling (GMDH) output	GMDH output	
	Training	Validation
Number of observations	3003	1287
Mean absolute error (MAE)	1.86852	1.86476
Root mean square error (RMSE)	2.89307	1.86476
Correlation coefficient	0.853969	0.848507
Coefficient of determination (R^2)	0.729258	0.719913

Few reservoir models with high WAG ratio showed high initial oil buildup but lower ultimate recovery factor compared to other cases with lower WAG ratio. High WAG ratio may lead to early gas breakthrough and production well shut-in; hence, WAG optimization is critical for the success of the WAG project.

Secondly, WAG simulation main results can be summarized in the following points:

- Typical WAG incremental recovery factor is mostly ranging from 5 to 15%.
- WAG gave similar recovery factor as waterflooding under few reservoir conditions where the presence of gas did not improve the overall recovery factor (volumetric sweep efficiency, displacement efficiency). This is mainly due to low gas injection performance caused by combination of multiple factors (i.e., gas override, low trapped gas saturation, low amount of injected gas, long WAG cycle plus high WAG ratio).
- Desired WAG ratio should be updated periodically based on wells performance (i.e., GOR and pressure).
- Reservoir voidage ratio control required for optimum WAG project.
- Water injectivity was reduced under few reservoir conditions due to the increase of gas saturation in the vicinity of the injection well.
- Hysteresis is an important factor that control WAG performance.

Thirdly, data mining study based on regression and GMDH techniques utilizing the reservoir modeling outcomes can be summarized in the following points:

- WAG incremental recovery factor prediction models were developed based on both regression and GMDH with correlation coefficient of 0.766 and 0.853, respectively.
- Predictive models validation correlation coefficients were 0.761 and 0.848, respectively.
- GMDH has shown strength and ability in selecting the effective input parameter, optimizing the network structure, and achieving predictive model with high accuracy.

Table 9 summarizes the prediction model parameters for both regression and GMDH. The two prediction models are applicable to hydrocarbon WAG process only with input parameters ranges defined earlier.

The developed WAG incremental recovery factor predictive models are expected to help reservoir engineers:

- Run quick evaluation of the WAG injection process based on a field data and WAG injection scheme,
- Run preliminary economical study based on the incremental WAG production profiles,
- Perform WAG optimization varying few of the input parameters (i.e., WAG ratio, WAG cycle, reservoir pressure, WAG startup timing, and so on),
- Assess the risk of the WAG project by understating the low and high expected additional reserves from WAG injection.

Table 9 Summary of the two prediction models results

Data mining method	Regression		GMDH	
	Training	Validation	Training	Validation
Number of observations	3003	1287	3003	1287
Mean absolute error (MAE)	2.45989	2.44422	1.86852	1.86476
Root mean square error (RMSE)	3.57118	3.56289	2.89307	1.86476
Correlation coefficient	0.766466	0.761228	0.853969	0.848507
Coefficient of determination (R^2)	0.587464	0.57932	0.729258	0.719913

Based on the results from the WAG predictive model study, reservoir modeling team will be able to decide on launching detailed reservoir simulation study or running WAG pilot to support the findings prior going for full field WAG project. The developed WAG incremental recovery factor predictive models are mathematical equations which are easy for reservoir engineers to implement on full field scale using average properties (i.e., horizontal permeability) or apply on selective reservoir or region.

Conclusions

- One factor at time reservoir study demonstrated that few parameters have high impact on WAG incremental recovery factor as it is a case of reservoir permeability, and injected gas volume. Few other parameters showed an impact on WAG incremental recovery factor shape but low impact on the ultimate WAG recovery as it is case for WAG ratio and WAG startup timing.
- Vertical permeability sensitivity on WAG performance has shown that gravity segregation may have a positive effect on WAG performance at low reservoir permeability.
- WAG incremental recovery factor over waterflooding from reservoir simulation study generally ranging from 5 to 15%; however, there are few reservoir models that showed a WAG incremental recovery factor up to 30%.
- Two incremental recovery factor predictive models were developed based on regression and GMDH method.
- WAG incremental recovery factor predictive model's correlation coefficients of 0.766 and 0.853 were achieved for regression and GMDH, respectively.
- GMDH results demonstrated its robustness and capabilities in building more accurate predictive models, optimizing network structure, and selecting effective input parameters
- WAG incremental recovery factor is predicted as a function of reservoir horizontal and vertical permeabilities, oil and gas gravity, solution gas–oil ratio, water viscosity, reservoir pressure, residual oil saturation to gas, trapped gas saturation, WAG ratio, WAG cycle length, WF recov-

ery factor prior WAG startup, and hydrocarbon pore volume of injected gas. This WAG incremental recovery factor prediction models are applicable to immiscible hydrocarbon WAG process only.

Acknowledgements The authors would like to thank Universiti Teknologi PETRONAS for their support and permission to publish this paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix A: PVT correlation used in this research study

Tarek (2010)

Bubble point pressure

$$p_b = \left(\frac{R_s}{C_1 \gamma_g e \left(C_3 \left(\frac{\gamma_{API}}{T+459.67} \right) \right) \right)^{\frac{1}{C_2}}$$

Solution gas–oil ratio

$$R_s = C_1 \gamma_g P^{C_2} \cdot e \left(C_3 \left(\frac{\gamma_{API}}{T+459.67} \right) \right)$$

Oil FVF—saturated

$$B_o = 1 + A_1 R_s + A_2 (T - 60) \left(\frac{\gamma_{API}}{\gamma_g} \right) + A_3 R_s (T - 60) \left(\frac{\gamma_{API}}{\gamma_g} \right)$$

Oil FVF—undersaturated

$$B_o = B_{ob} e(c_o (p_b - p))$$

Compressibility—saturated

$$c_o = \frac{-1433 + 5R_{sb} + 17.2T - 1180\gamma_g + 12.61\gamma_{API}}{10^5 p} + \frac{B_g \cdot \frac{dB_g}{B_o} \cdot \frac{dR_s}{dp}}{5.6145835}$$

Compressibility—undersaturated

$$c_o = \frac{-1433 + 5R_{sb} + 17.2T - 1180\gamma_g + 12.61\gamma_{API}}{10^5 p}$$

Coefficient	$\gamma_o \leq 30^\circ \text{ API}$	$\gamma_o > 30^\circ \text{ API}$
A_1	0.0362	0.0178
A_2	1.0937	1.1870
A_3	25.7240	23.9310
C_1	4.677E-04	4.670E-04
C_2	1.751E-05	1.100E-05
C_3	-1.811E-08	1.377E-09

Beal's viscosity correlation (Tarek 2010)

$$\mu_{od} = \left(0.32 + \frac{1.8(10^7)}{\text{API}^{4.53}} \right) \left(\frac{360}{T - 260} \right)^a$$

with

$$a = 10^{(0.43 + \frac{8.33}{\text{API}})}$$

where T is temperature, $^\circ\text{R}$, and μ_{od} is the viscosity of the dead oil as measured at 14.7 psia and reservoir temperature, cp.

Appendix B: WAG incremental recovery factor prediction models

Regression model

WAG incremental recovery factor (WAG Incr. RF) from regression method is shown in Eq. 3:

$$\begin{aligned} \text{WAG Incr. RF} = & -21.0688 - 0.097691P1^{\frac{1}{3}} + 0.891202P2^{\frac{1}{3}} \\ & + 1.39881P3^{\frac{1}{3}} + 27.5687P4^{\frac{1}{3}} - 9.59638P5^{\frac{1}{3}} \\ & - 5.46256P6^{\frac{1}{3}} - 0.507397P7^{\frac{1}{3}} - 0.0433339P8^{\frac{1}{3}} \\ & + 0.390731P9^{\frac{1}{3}} + 2.09834P10^{\frac{1}{3}} - 5.42296P11^{\frac{1}{3}} \\ & + 0.763709P12^{\frac{1}{3}} + 9.43669P13^{\frac{1}{3}} \end{aligned} \quad (3)$$

Group method of data handling wag incremental recovery factor prediction Model

WAG incremental recovery factor (WAG Incr. RF) from GMDH method is shown in Eq. 4

$$\text{WAG Incr. RF} = -1.06647 + 0.725772C + 0.0820085CS - 0.0864785C^2 + 0.635375S - 0.0175908S^2 \quad (4)$$

where $A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S$ are calculated below.

Parameter	Expressions
A	$3.64503 - 0.417105P3^{2/3} + 1.74196P9^{1/3}$
B	$15.935 - 7.48585P5^{1/3} - 4.39881P6^{2/3}$
C	$-7.68316 + 0.472008A - 0.00907984AB + 0.0338545A^2 + 1.10521B$
D	$8.74928 + 3.93524P2^{1/3} - 7.02062P2^{1/3}P6^{1/3} + 0.55184P6^{1/3}$
E	$-8.66825 + 17.7079P4^{1/3}P13^{1/3} + 9.08092P4^{2/3} - 3.95216P13^{2/3}$
F	$19.4033 - 20.4009P^{1/3}P11^{1/3} + 4.3348P5^{2/3} - 19.6509P11^{1/3} + 21.0261P11^{2/3}$
G	$-0.968145 - 0.403418E + 0.113997EF + 0.0200969E^2 + 0.529535F - 0.0283807F^2$
H	$16.5784 - 4.27868P12^{\frac{1}{3}} + 0.158907P12^{\frac{1}{3}}G + 0.249929P12^{\frac{2}{3}} + 0.0030124G^2$
I	$-7.70913 - 0.22216P3^{\frac{1}{3}}H + 0.620351P3^{\frac{2}{3}} + 1.84436H - 0.00334797H^2$
J	$-10.7234 - 4.62122P3^{1/3}P12^{1/3} + 3.35336P3^{2/3} + 9.31215P12^{1/3} + 0.502607P12^{2/3}$
K	$10.4426 + 0.390282P1^{1/3} - 0.0599042P1^{\frac{2}{3}} - 16.225P10^{\frac{1}{3}} + 12.9965P10^{2/3}$
L	$12.9493 + 0.189344JK - 0.0410685J^2 - 2.84699K + 0.115597K^2$
M	$14.6834 - 57.87P11^{\frac{1}{3}} - 8.6405P11^{\frac{1}{3}}P13^{\frac{1}{3}} + 48.4343P11^{\frac{2}{3}} + 20.5669P13^{\frac{1}{3}} - 3.3457P13^{2/3}$
N	$11.5993 + 24.8253P5^{\frac{1}{3}} - 22.0144P5^{\frac{2}{3}} - 4.78272P9^{1/3} + 0.575425P9^{2/3}$
O	$-5.27228 + 0.713909M + 0.0585581MN - 0.018546M^2 + 0.691731N - 0.0191964N^2$
P	$1.94024 + 0.275679EO - 0.135212E^2 + 0.553732O - 0.10671O^2$
Q	$3.73295 - 0.789248L + 0.043444LP + 0.0352881L^2 + 0.703719P - 0.00868122P^2$
R	$0.0796747 + 0.204468I + 0.783717IQ - 0.363835I^2 + 0.835446Q - 0.420041Q^2$
S	$49.9659 - 11.1501D + 0.0434839DR + 0.620772D^2 + 0.541157R + 0.00264495R^2$

References

- Afzali S, Rezaei N, Zendehboudi S (2018) A comprehensive review on enhanced oil recovery by water alternating gas (WAG) injection. *Fuel* 227:10. <https://doi.org/10.1016/j.fuel.2018.04.015>
- Ahmed AM, Rizaner A, Ulusoy AH (2016) Using data mining to predict instructor performance. *Procedia Comput Sci* 102:137–142. <https://doi.org/10.1016/j.procs.2016.09.380>
- Anastasakis L, Mort N (2001) The development of self-organization techniques in modeling: a review of the group method of data handling (GMDH). ACSE research report no. 813, University of Sheffield, UK
- Behrouz T, Kharrat R, Ghazanfari MH (2007) Experimental study of factors affecting heavy oil recovery in solvent floods. *Pet Soc Can.* <https://doi.org/10.2118/2007-006>
- Bini BS, Mathew T (2016) Clustering and regression techniques for stock prediction. *Procedia Technol* 24:1248–1255. <https://doi.org/10.1016/j.protcy.2016.05.104>
- Blunt M, Fayers FJ, Orr FM (1993) Carbon dioxide in enhanced: oil recovery. *Energy Convers Manag* 34(9):1197–1204
- Cao M, Gu Y (2013) Physicochemical characterization of produced oils and gases in immiscible and miscible CO₂ flooding processes. *Energy Fuels* 27(1):440–453
- Chen S, Li H, Yang D, Tontiwachwuthikul P (2010) Optimal parametric design for water-alternating-gas (WAG) process in a CO₂-miscible flooding reservoir. *Soc Petrol Eng.* <https://doi.org/10.2118/141650-pa>
- Chordia M, Trivedi JJ (2010) Diffusion in naturally fractured reservoirs—a review. Society of Petroleum Engineers, Brisbane
- Christensen JR, Stenby EH, Skauge A (2001) Review of WAG field experience. *SPE Reserv Eval Eng* 4:97–106. <https://doi.org/10.2118/71203-PA>
- Denney D (2012) Fluid- and rock-property effects on reserves estimation. *Soc Petrol Eng.* <https://doi.org/10.2118/1212-0095-jpt>
- Farlow S (1981) The GMDH algorithm of Ivakhnenko. *Am Stat* 35:210–215. <https://doi.org/10.1080/00031305.1981.10479358>
- Farshid T, Benyamin YJ, Ostap Z, Brett AP, Nevin JR, Ryan RW (2010) Effect of oil viscosity, permeability and injection rate on performance of waterflooding, CO₂ flooding and WAG processes on recovery of heavy oils. In: Canadian unconventional resources and international petroleum conference. 2010, Calgary, Alberta, Canada: Society of Petroleum Engineers, 138188-MS
- Feigl A (2011) Effect of trapped gas saturation on oil recovery during the application of secondary recovery methods in exploitation of petroleum reservoirs. *Nafta Explor Prod Process Petrochem* 62:5–6
- Han J et al (2012) Data mining concepts and techniques, 3rd edn. Elsevier Inc. <https://doi.org/10.1016/C2009-0-61819-5>
- Jaber AK, Awang MB, Lenn CP (2017) Box–Behnken design for assessment proxy model of miscible CO₂-WAG in heterogeneous clastic reservoir. *J Nat Gas Sci Eng* 40(2017):236–248. <https://doi.org/10.1016/j.jngse.2017.02.020>
- Jackson DD, Andrew GL, Claridge EL (1985) Optimum WAG Ratio vs. rock wettability in CO₂ flooding. In: SPE annual technical conference and exhibition, Las Vegas, Nevada, September 22, 1985
- James G, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning with R application. Book printed in USA
- Kim KKK, Patrón ER, Braatz RD (2018) Standard representation and unified stability analysis for dynamic artificial neural network models. *Neural Netw* 98:251–262
- Labastie A (2011) En route: increasing recovery factors: a necessity. *J Petrol Technol.* <https://doi.org/10.2118/0811-0012-JPT>
- Lazreg B, Raub MRA, Hanifah MAB, Ghadami N (2017) WAG cycle dependent hysteresis modelling through an integrated approach from laboratory to field scale, Malaysia oil fields. In: Presented at SPE/IATMI Asia Pacific Oil & Gas conference and exhibition, 17–19 October, Jakarta, Indonesia. SPE-186379-MS. <https://doi-org.proxy1.athensams.net/10.2118/186379-MS>
- Lemke F, Motzev M (2016) Self-organizing data mining techniques in model based simulation games for business training and education. *Vanguard Scientific Instruments in Management*, vol 11.
- Ling K, Shen Z (2011) Effects of fluid and rock properties on reserves estimation. *Soc Petrol Eng.* <https://doi.org/10.2118/148717-ms>
- Mahesh PE, Britt MH (2015) A study of the effect of relative permeability and residual oil saturation on oil recovery, pp 339–346. <https://doi.org/10.3384/ecp15119339>. <https://doi-org.proxy1.athensams.net/10.2118/118226-MS>
- Medvedev V, Kurasova O, Bernatavičienė J, Treigys P, Marcinkevičius V, Dzemyda G (2017) A new web-based solution for modelling data mining processes. *Simul Model Pract Theory* 76:34–46. <https://doi.org/10.1016/j.simpat.2017.03.001>
- Moreno R, Gonçalves R, Okabe C, Schiozer D, Trevisan O, Bonet JE, Iatchuk S (2011) Comparison of residual oil saturation for water and supercritical CO₂ flooding in a long core, with live oil at reservoir conditions. *J Porous Media* 14:699–708. <https://doi.org/10.1615/JPorMedia.v14.i8.40>
- Negnevitsky M (2011) Artificial intelligence: a guide to intelligence systems. Pearson Education Limited. Printed in Great Britain, England
- Panjalizadeh H, Alizadeh A, Ghazanfari MH, Alizadeh N (2015) Optimization of the WAG injection process. *Pet Sci Technol* 33:294–301. <https://doi.org/10.1080/10916466.2014.956897>
- Pritchard DWL, Georgi DT, Hemingson P, Okazawa T (1990) Reservoir surveillance impacts management, of the Judy creek hydrocarbon miscible flood. In: Presented at SPE/DOE enhanced oil recovery symposium, 22–25 April, Tulsa, Oklahoma. SPE-20228-MS. <https://doi-org.proxy1.athensams.net/10.2118/20228-MS>
- Rogers JD, Grigg RB (2000) A literature analysis of the WAG injectivity abnormalities in the CO₂ process. In: Presented at SPE/DOE improved oil recovery symposium, 3–5 April, Tulsa, Oklahoma. SPE-59329-MS. <https://doi.org/10.2118/59329-MS>
- Satter A, Iqbal GM (2016) Reservoir fluid properties, pp 81–105. <https://doi.org/10.1016/B978-0-12-800219-3.00004-8>
- Eclipse Simulator Reference Manual (2014), Schlumberger
- Semenov AA, Oshmarin RA, Driller A, Butakova A (2010) Application of group method of data handling for geological modeling of vankor field. In: Presented at North Africa technical conference and exhibition, 14–17 February, Cairo, Egypt. SPE-128517-MS. <https://doi-org.proxy1.athensams.net/10.2118/128517-MS>
- Simon A, Mahima SD, Venkatesan S, Babu Ramesh DR (2016) An overview of machine learning and its applications. *Int J Electr Sci Eng* 1:22–24
- Skauge A, Larsen JA (1994) Three-phase relative permeabilities and trapped gas measurements related to WAG processes. In: International symposium of the society of core analysts, SCA 9421
- Tarek AH (2010) Reservoir engineering handbook, 4th edn. ISBN: 978-1-85617-803-7. <https://doi.org/10.1016/B978-1-85617-803-7.50001-8>
- Teng X, Gong Y (2018) Research on application of machine learning in data mining. In: IOP conference series: materials science and engineering, vol 392, p 062202. <https://doi.org/10.1088/1757-899x/392/6/062202>
- Tham B, Raif BD, Saaid IB, Abillah E (2011) The effects of kv/kh on gas assisted gravity drainage process. *Int J Eng Technol* 11(3):153–185
- Tsung-Min T, Pei-Hwa Y (2017) GMDH algorithms applied to turbidity forecasting. *Appl Water Sci* 7(3):1151. <https://doi.org/10.1007/s13201-016-0458-4>

- Tunio SQ, Tunio AH, Ghirano NA, El Adawy ZM (2011) Comparison of different enhanced oil recovery techniques for better oil productivity. *Int J Appl Sci Technol* 5(1):143–153
- Wu X, Ogbe DO, Zhu T, Khataniar S (2004) Critical design factors and evaluation of recovery performance of miscible displacement and WAG process. Petroleum Society of Canada. <https://doi.org/10.2118/2004-192>
- Yang B, Jiang H, Chen M, Fang Y (2008) Experimental and numerical comparison of flooding schemes to enhance recovery of light/medium heavy oil in an offshore oilfield. In: Presented at Abu Dhabi international petroleum exhibition and conference, 3–6 November, Abu Dhabi, UAE. SPE-118226-MS
- Yavuz O, Kemal GN, Eray S, Hüseyin B (2019) Factors affecting global passenger flow and a model proposal for forecasting. *Am J Sci Technol* 6:1–13
- Yu B, Zhang X, Du M, Ju Y (2017) Evaluation and selection of CO₂-water alternating flooding favorable area based on Flow-units analysis. *Energy Procedia* 114:4557–4563. <https://doi.org/10.1016/j.egypro.2017.03.1574>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.