



Water alternating gas incremental recovery factor prediction and WAG pilot lessons learned

Lazreg Belazreg¹ · Syed Mohammad Mahmood¹

Received: 2 November 2018 / Accepted: 13 May 2019 / Published online: 28 May 2019
© The Author(s) 2019

Abstract

Water alternating gas (WAG) injection process is a proven enhanced oil recovery (EOR) technology with many successful field applications around the world. WAG pilot projects demonstrate that WAG incremental recovery factor typically ranges from 5 to 10% of original oil in place, though up to 20% has been observed in some fields. Despite its proven success, WAG application growth has been very slow. One of the reasons is the unavailability of robust analytical predictive tools that could estimate WAG incremental recovery factor, which is required for preliminary economic analysis before committing to expensive and time-consuming detailed technical studies and field pilot test that often requires a lot of input data. A semi-numerical model for WAG incremental recovery factor prediction was developed based on data mining of published WAG pilots to fill this gap. An extensive review of published WAG pilot projects was carried out, and consequently, 33 projects from 28 fields around the world were selected for this research study. Field WAG incremental recovery factor and parameters with total of one hundred and seventy-seven (177) observations were inputted to the predictive model. A predictive model was developed using both regression and group method of data handling (GMDH) techniques; 70% of the 33 WAG pilot projects data were used as validation, whereas remaining 30% of data set were used for validation. The predictive model results achieved with coefficient of determination (R^2) from the regression method were ranging from 0.892 to 0.946 and 0.854 to 0.917 for training and validation sets, respectively. However, the prediction model coefficient of determination (R^2) using GMDH method was ranging from 0.964 to 0.981 and 0.934 to 0.974 for training and validation, respectively. The developed predictive model can predict WAG incremental recovery factor versus multiple input parameters that include rock type, WAG process type, hydrocarbon pore volume of injected gas, reservoir permeability, oil gravity, oil viscosity, reservoir pressure, and reservoir temperature. The results of the study demonstrated that few input parameters have a significant impact on WAG incremental recovery factor as reservoir permeability and hydrocarbon pore volume of injected gas. This research study uses a novel approach in pre-defining the expected incremental WAG recovery factor before committing resources for building complex numerical reservoir simulation models and running WAG pilot tests, which are very time-consuming and costly and require extensive data input.

Keywords Water alternating gas injection · Recovery factor prediction · Data mining · Lessons learned

List of symbols

API	Oil gravity	$f(x_i)$	Calculated incremental recovery factor for observation i
D	Depth (ft)	GOR	Gas–oil ratio
EOR	Enhanced oil recovery	h_p	Reservoir thickness (ft)
FVF	Formation volume factor (Rbbl/stb)	HCPVI	Hydrocarbon pore volume injected (fraction)
		HCPVI _{i}	Hydrocarbon pore volume injected for observation i (fraction)
		K	Reservoir permeability (md)
		K_h	Reservoir horizontal permeability (md)
		K_v	Reservoir vertical permeability (md)
		MMP	Minimum miscibility pressure
		n	Total number of observations
		OOIP	Original oil-in-place (MMstb)
		P	Reservoir pressure (psi)

✉ Syed Mohammad Mahmood
mohammad.mahmood@utp.edu.my
Lazreg Belazreg
Lbelazreg@gmail.com

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

RF	Recovery factor (%)
T	Reservoir temperature (°C)
WAG	Water alternating gas
WOR	Water–oil ratio
y_i	Observed incremental recovery factor for observation i
μ_o	Oil viscosity (cp)

Prediction model input vectors

$P1$	Reservoir horizontal permeability
$P2$	API
$P3$	Oil viscosity
$P4$	Reservoir pressure at WAG start-up
$P5$	Reservoir temperature
$P6$	Hydrocarbon pore volume of injected gas (fraction)

Introduction

Due to increasing demand for energy during the last decades, adding substantial oil reserves is a necessity. Besides exploration and unconventional field development, EOR technology has demonstrated very promising results in improving oil recovery factor.

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

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 in many fields around the world (Christensen et al. 2001). However, the application of this technology is still very limited due to the complexity of the WAG process and difficulty in quantifying the expected performance prior building full field complex simulation model and running a pilot test.

During the last decades, many research papers were published focusing on WAG performance (Giordano et al. 1998), WAG simulation (Kohata et al. 2017), WAG pilot results analysis (Christensen et al. 2001), WAG process mechanisms (Afzali et al. 2018), WAG hysteresis (Lazreg et al. 2017), and WAG management (Sanchez 1999). However, WAG incremental recovery factor over secondary recovery is still unknown.

Furthermore, field-scale WAG project approval required detailed technical study based on laboratory experiments and WAG modeling followed by WAG pilot test for WAG performance calibration which are very time and costly.

Figure 1 demonstrates the typical WAG project approval workflow.

The objective of this study was to develop a semi-numerical model that can predict WAG incremental recovery factor

based on data mining of published WAG pilots worldwide. This research study started by literature review on WAG process and data mining techniques, followed by building a database of WAG pilot test results and input, and then building a predictive model from the WAG pilots using regression and GMDH.

This technical research paper will share the background of WAG process and mechanisms, summarizes WAG pilot lessons learned, and demonstrates a new approach for predicting WAG incremental recovery factor using semi-numerical model developed based on WAG data mining.

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 (Ahmed 2010):

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

where RF = the overall recovery factor (0 to 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 fluid mobilities, injection patterns, areal and vertical heterogeneities, the degree of gravity segregation, and total pore volume injected (Ahmed 2010).

WAG incremental recovery factor is a result of the increase in both displacement and volumetric sweep efficiencies due to the reduction in the residual oil saturation and improvement of both areal and vertical sweep efficiencies. WAG injection is one of the promising and effective EOR technologies in enhancing oil recovery factor over secondary recovery. However, WAG recovery mechanisms are complex process that require in-depth understating of factors controlling its performance.

Multiple research papers were published during the last decades on WAG recovery mechanisms which include 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 laboratory 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 waterflood and gas injection.

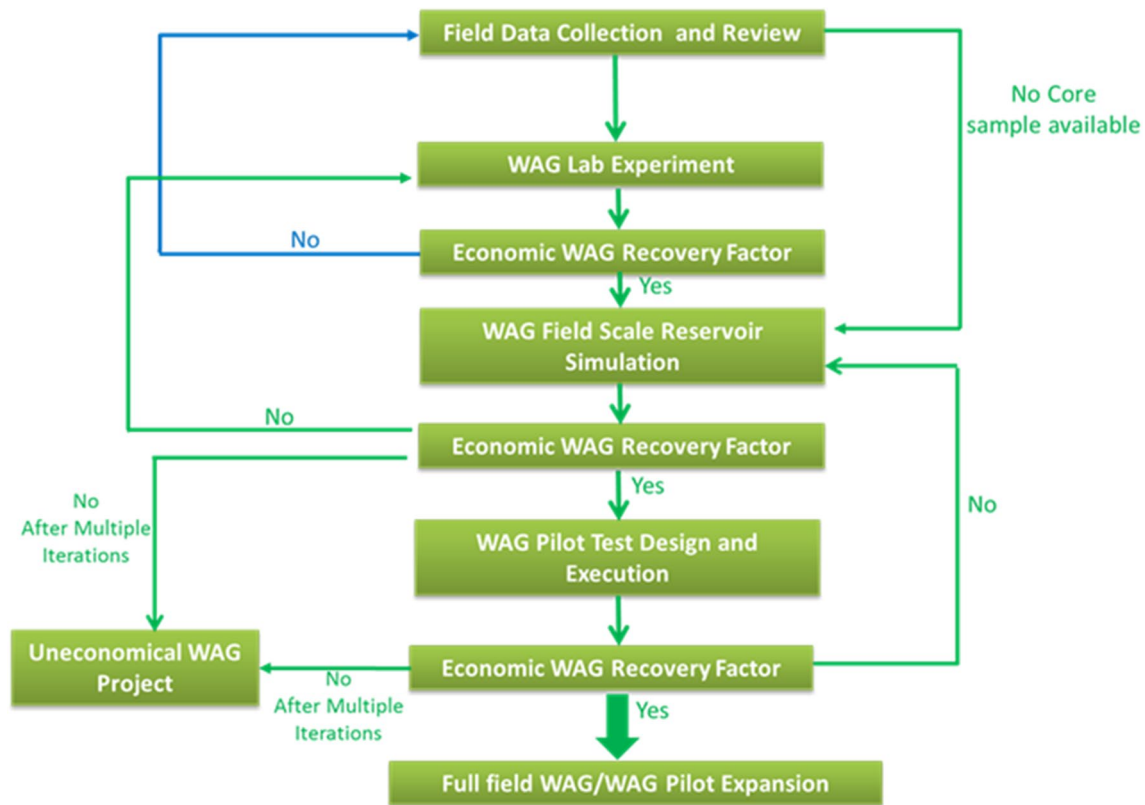


Fig. 1 Typical workflow for field-scale WAG project approval

Mobility ratio is an important factor that controls the gas injection process volumetric sweep efficiency with a favorable mobility of less than one (< 1). Reduction in 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 are the components of the WAG incremental recovery factor. The improvement of oil recovery during gas EOR includes 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; Chen et al. 2010). 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.

Water alternating gas injection technology application

Water alternating gas (WAG) injection technology is a method which may improve oil recovery efficiency by combining the effects from two traditional techniques—water and gas flooding. Both microscopic oil displacement and sweep efficiency can be enhanced by WAG implementation (Surguchev et al. 1992). WAG injection consists of injecting water and gas from the same injection wells which contribute to the improvement of overall sweep efficiency by:

- Reducing oil viscosity and increasing reservoir oil volume due to oil swelling (Moffitt and Zornes 1992).
- Vaporizing of light oil components and striping of higher hydrocarbons (Jakobsson and Christian 1994).
- Reducing residual oil saturation under gas injection as compared to waterflooding (Moreno et al. 2010).
- Increasing oil recovery factor through improved water-flood conformance (Champion and Shelden 1989).
- WAG hysteresis which involves reduction in water and gas mobilities, and reduction in residual oil saturation (Lazreg et al. 2017).

Table 1 Fields used during this research study (Christensen et al. 2001; Ramachandran et al. 2010; Sanchez 1999; Rosman et al. 2011)

	Field name	Location		Field name	Location
1	Kelly Snyder	Texas, USA	15	Daging	China
2	Lick Creek	Arkansas, USA	16	Dollarhide	Texas, USA
3	Lost Soldier Field	Wyoming, USA	17	Easi Vacuum	New Mexico, USA
4	Maljamar	New Mexico, USA	18	Fenn Big Valley	Alberta, Canada
5	Mitsue	Alberta, Canada	19	Garber	Oklahoma, USA
6	N. Ward Estes	Texas, USA	20	Unknown	India
7	Turdy Springer NE	Oklahoma, USA	21	Judy Creek	Alberta, Canada
8	Quatantine Bay	Louisiana, USA	22	Kapaiuk	Alaska, USA
9	Kaiigcly Weber	Colorado, USA	23	Kaybob North	Alberta, Canada
10	San Andres Means	Texas, USA	24	Staffjord	North Sea, Norway
11	Slaughter Estate	Texas, USA	25	Sundown Slaughter	Texas, USA
12	Wilmington	California, USA	26	Garber	Oklahoma, USA
n	Swan Hills	Alberta, Canada	71	Maracaiho	Venezuela
14	Ekovisk	North Sea, Norway	28	Unknown	Malaysia

- Increasing attic oil production through gas injection (Owens et al. 2018; Christensen et al. 2001).
- Maintaining reservoir pressure (Hermansen et al. 1997).

The improvements in gas handling and in situ lifting are some other benefits of WAG injection.

WAG injection process is divided into miscible and immiscible processes depending on the occurrence of miscibility between injection gas and in situ reservoir oil under reservoir conditions. If gas injection happens at a pressure higher than minimum miscibility pressure (MMP), then the WAG process is miscible otherwise the opposite. Field application has demonstrated that miscible WAG injection performance and recovery factor are generally higher than immiscible WAG injection (IWAG) (Christensen et al. 2001).

Injected gas during the WAG process can be hydrocarbon or non-hydrocarbon. Hydrocarbon gases are paraffin of lower molecular weight like methane and ethane, and non-hydrocarbon gases are carbon dioxide (CO₂) and nitrogen (N₂).

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

Most of the published successful WAG pilot projects are from fields located in USA and Canada (Christensen et al. 2001). Few published successful projects are in North Sea, China, Venezuela, North Africa, Middle East, Malaysia, and Croatia.

With the progress in computer hardware during the last decades and development of multiple reservoir modeling software, evaluating large-scale WAG project performance is now possible. However, modeling the WAG process in full scale is expensive, is very time-consuming, and depends on the modeling quality and input data uncertainty.

Pilot WAG data summary

Twenty-eight oil fields with published pilot WAG data were used in this research study. Field location includes USA, Canada, Malaysia, China, Norway, Venezuela, and India. Table 1 summarizes selected fields location. Many other reviewed WAG pilot papers did not reveal the observed WAG incremental recovery factor which is required in this research work.

Figure 2 shows the WAG incremental recovery factor versus hydrocarbon pore volume of injected gas from the selected fields. Figure 3 shows each WAG projects parameter which includes WAG process type, rock and fluid properties, and WAG injection scheme. Few projects shared WAG recovery factor at specific hydrocarbon pore volume of injected gas and/or ultimate recovery factor from the WAG project only, while others shared full trend of the pilot WAG recovery factor.

WAG pilot data analysis

One method to qualify dependence between random elements is to rank them according to how close one is a function of another. Several measures based on this idea have been proposed. Many researchers have published different approaches for defining the dependency between random variables, including those based on copulas associated with Markov operators (Boonmee and Tasena 2016).

Understanding the dependency of input vectors is very important in reducing the number of predictors that simplify the study and cut the required time of the analysis. Linear regression is a widely useful statistical learning method for predicting a quantitative response. Moreover, it serves as a good jumping-off point for newer approaches as many fancy

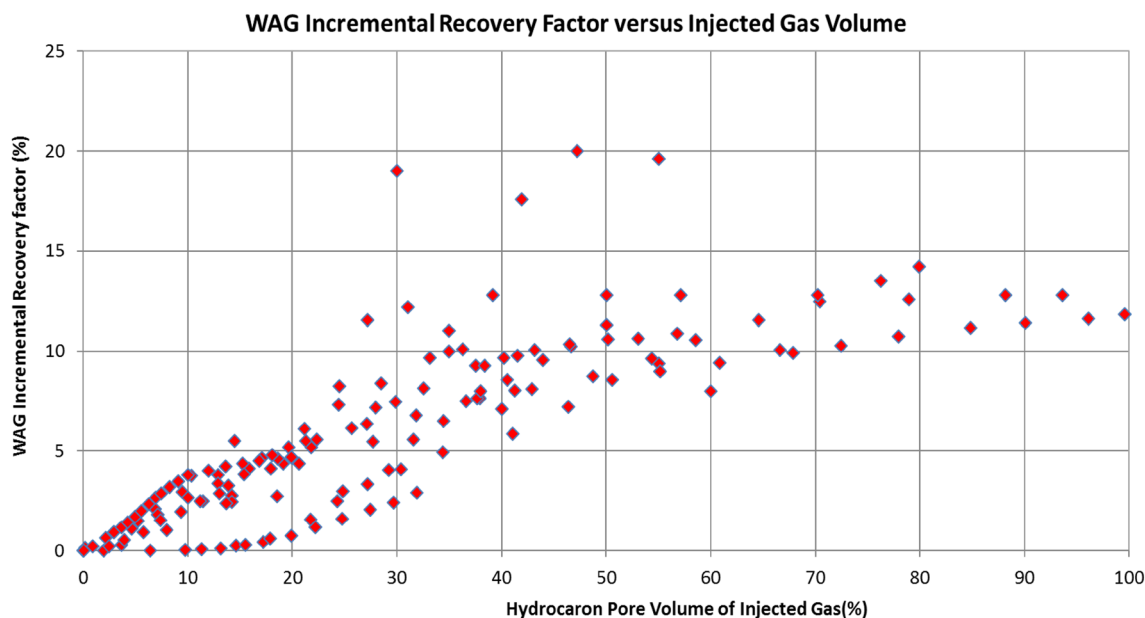


Fig. 2 Summary of WAG incremental recovery factor from WAG pilots

statistical learning approaches can be seen as generalizations or extension of linear regression (James et al. 2015).

Pearson correlation coefficient (PCC) is a statistical metric that measures the strength and direction of a linear relationship between two random variables. It has been applied to various indices in statistics, such as data analysis and classification, data analysis, clustering, decision making, finance analysis, and biological research (Zhou et al. 2016).

The Pearson correlation coefficient of two variables X and Y is formally defined as the covariance of the two variables divided by the product of their standard deviations (which acts as a normalization factor), and it can be equivalently defined by (Zhou et al. 2016):

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \sum (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}} \quad (2)$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^N x_i$ denotes the mean of x and $\bar{y} = \frac{1}{n} \sum_{i=1}^N y_i$ denotes the mean of y .

The coefficient r_{xy} indicates the strength of the linear relationship between the two variables and it ranges between 1 and -1 . The r_{xy} is positive if the two variables are directly related and negative if they are inversely related. The closer the $|r_{xy}|$ to 1, the stronger the linear relationship between the two variables. r_{xy} of 0 indicates uncorrelated variables.

The results of the input vector dependency analysis demonstrated low correlation coefficients which reflect the independence of the input vectors. Table 2 shows the results of the input vector dependency.

Data mining

Data mining 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. 2015).

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, data mining (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. 4) (Han et al. 2012).

Statistics studies the collection, analysis, interpretation or 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).

Field	Formation	Process Type	K (md)	hp (ft)	D (ft)	API	μ_o (cp)	FVF (bbl/Stb)	P (psi)	T (F)	Well Spacing (ft)	WAG Ratio	Cycle (Month)	HCPV (%)	Incr. RF (%)	Gas Injected
Wilming ton	Sand	Immiscible	500	141.00	7546	14	180.00	1.05	1000	120		1.0	0.50		9.60	CO2
Kaybob Beaverhill	Carbonate	Miscible		56.00			0.19	1.61	4550		951	1.0	2.00			
Mitsue	Sand	Miscible	190	15.70		41	0.60		2500	143				4.90	1.30	
Kuparuk	Sand	Miscible	150	33.00		24			3100		1476	1.0		14.50	5.50	
Lick Creek	Sand	Immiscible	1200	8.86	2549	17	160.00	1.05	1200	118					13.20	CO2
Judy Creek	Sand	Miscible									1476	1.0	1.00		8.90	
Lick Creek	Sand	Immiscible	1200	8.86		17		1.05		118				242.00	11.10	CO2
Quarantine Bay 4RC	Sand	Miscible	100	15.00	8179	32	0.94	1.23	3534	183		1.5	1.00	18.90	16.90	CO2
Swan Hills	Carbonate	Miscible				42	0.38	1.41	2098		1476	0.7		47.20	20.00	NGL
Ekovish	Chalk			400.00	10400	36			7135						2.50	
Fenn Big Valley		Miscible		80.00			1.40		1650	136		1.3		55.00	19.60	
Ekovish	Chalk			500.00		36										
Statfjord		Immiscible							4785						2.20	
Daging		Immiscible	1628	30.18	3742				1777	120	1476	3.0		20.00	4.67	CO2
Kuparuk	Sand	Immiscible				24	2.00		3100	158		0.5				CO2
Purdy Springer	Sand	Miscible	44	40.00	7500	35	1.45	1.19	2400	148				36.59	7.50	CO2
Purdy Springer	Sand	Miscible	44	40.00	7500	35	1.41	1.19	2000	148	1050	3.0				CO2
Sundown Slaughter	Carbonate	Miscible	5	87.00	5000	33	1.00		1750	107	656			31.91	2.92	CO2
Slaughter Estate	Dolomite	Miscible	5	79.00	4985	32			1750	107		0.5		50.03	11.30	CO2
Ward Estate	Sand	Miscible	37	108.00	8530	37	1.40	1.20	1500	83		1.0		50.00	8.00	CO2
India	Sand	Immiscible	100			47		1.73	3530	130		1.0	4.00	110.48	13.83	
Kelly Synder	Limestone		19		6700	41	0.35	1.50	3122	130				19.92	4.67	CO2
Mamcaibo					12500	32			1500	236					11.00	CO2
Rangly Weber		Miscible	10	250.00	5300	34	1.70		3350	160	525	1.0		55.17	9.00	CO2
Lost Soldier	Sand	Miscible	31	210.00	5000	35	1.38	1.12	2520	158	492	1.0		67.89	9.90	CO2
Dollarhide	Mixture	Miscible	9	76.00	7800	40	0.40	1.35	3200	120	755	1.0		30.00	19.00	CO2
Pennsylvanian	Sand	Miscible	57	21.00	1950	47	2.10	1.20	1250	95				35.00	14.00	CO2
Maljamar	Carbonate		18	27.00	3665	37	0.80	1.25		90				41.90	17.60	CO2
Maljamar	Carbonate		18	27.00	4000	36	0.80	1.25		92				36.30	10.10	CO2
Carber														35.00	10.00	
San Andres														40.00	7.10	
East Vaccum														10.00	3.80	
Malaysia	Sand	Immiscible													7.00	HC

Fig. 3 WAG pilot data used in this research study

Table 2 WAG pilot input vectors Pearson’s correlation coefficients

	K	H _p	Depth	API	Viscosity	FVF	Pressure	Temperature	Well spacing	WAG ratio	Gas HCPVI
K	1.00	0.20	0.10	-0.11	0.33	-0.02	0.18	0.24	0.22	0.21	0.35
H _p	-0.20	1.00	0.04	-0.05	0.04	0.13	0.17	-0.06	-0.48	-0.18	-0.02
Depth	0.10	0.04	1.00	-0.27	-0.05	0.08	0.57	0.70	0.08	0.29	0.09
API	-0.11	-0.05	-0.27	1.00	-0.30	0.41	-0.01	-0.39	-0.27	0.03	-0.05
Viscosity	0.33	0.04	-0.05	-0.30	1.00	-0.20	-0.18	-0.07	0.04	-0.05	0.03
FVF	-0.02	0.13	0.08	0.41	-0.20	1.00	0.45	-0.06	0.03	-0.34	0.24
Pressure	0.18	0.17	0.57	-0.01	-0.18	0.45	1.00	0.60	0.16	0.01	0.28
Temperature	0.24	-0.06	0.70	-0.39	-0.07	-0.06	0.60	1.00	0.25	0.30	0.15
Well spacing	0.22	-0.48	0.08	-0.27	0.04	0.03	0.16	0.25	1.00	-0.02	0.16
WAG ratio	0.21	-0.18	0.29	0.03	-0.05	-0.34	0.01	0.30	-0.02	1.00	-0.14
Gas HCPVI	0.35	-0.02	0.09	-0.05	0.03	0.24	0.28	0.15	0.16	-0.14	1.00

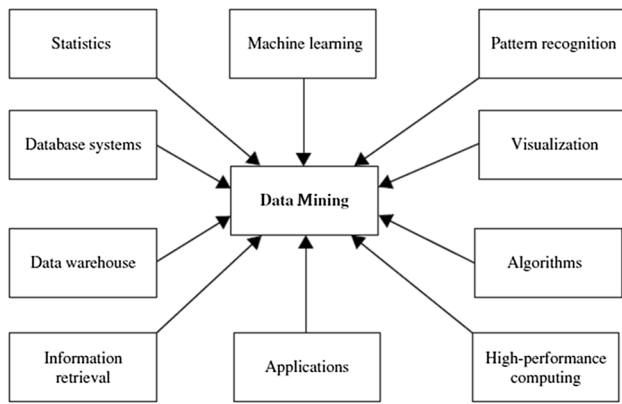


Fig. 4 Data mining adopts techniques from many domains (Han et al. 2012)

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 (Xiuyi and Yuxia 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 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 facilitates the program ability to get an accurate conclusion with new data.
- Unsupervised machine learning where the program is given a 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 (Michael 2011). Artificial neural networks (ANNs) are 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 5 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

Fig. 5 Neural network typical structure

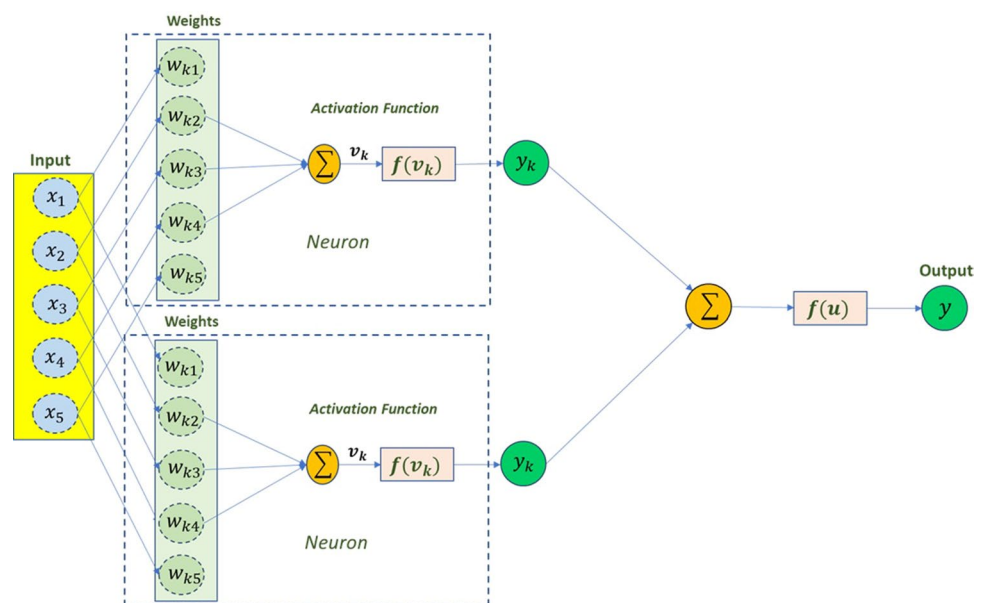
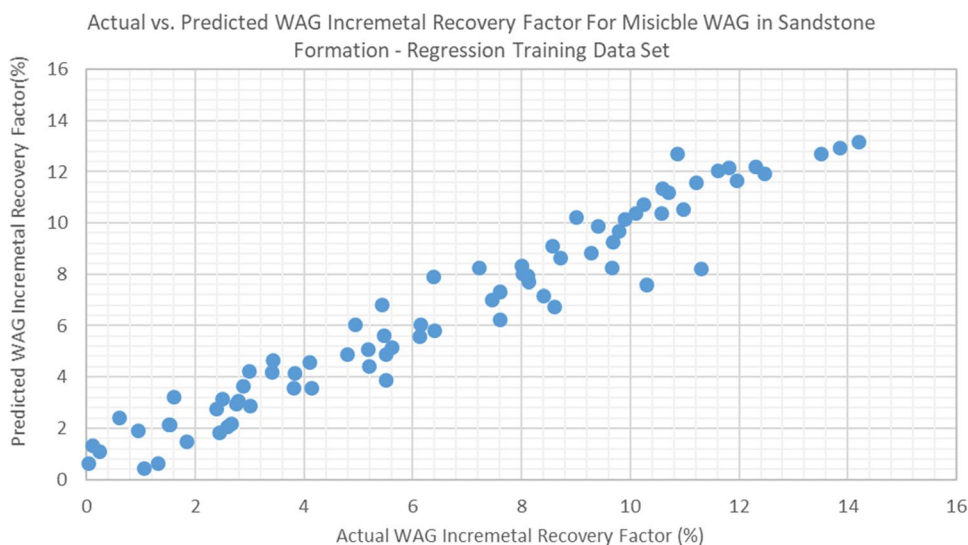


Fig. 6 Regression WAG incremental recovery factor prediction model training results for miscible WAG in sandstone formation



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).

WAG Incremental Recovery Factor Prediction

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.

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 to w_n implies the coefficients, y_1 to y_n implies the independent variables, and ϵ implies the random error (Bini and Mathew 2016).

Regression model and validation

The WAG data were divided into three categories: miscible WAG in sandstone formation, immiscible WAG in sandstone formation, and miscible WAG in carbonate formation.

For the development of WAG incremental recovery factor predictive model for each category, 70% of the WAG data was used for training the model and 30% was used for model validation.

Figures 6, 7, 8, 9, 10 and 11, and Table 3 summarize the results of WAG incremental recovery factor for training and validation.

The prediction model details are shared under [Appendix](#).

Group Method of Data Handling

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 (Tsai and Yen 2017).

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 is not only time-consuming but also produces unreliable prediction models.

GMDH was developed to produce a model by looking only at input data and the desired output (Semenov et al. 2010).

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:

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

Fig. 7 Regression WAG incremental recovery factor prediction model training results for immiscible WAG in sandstone formation

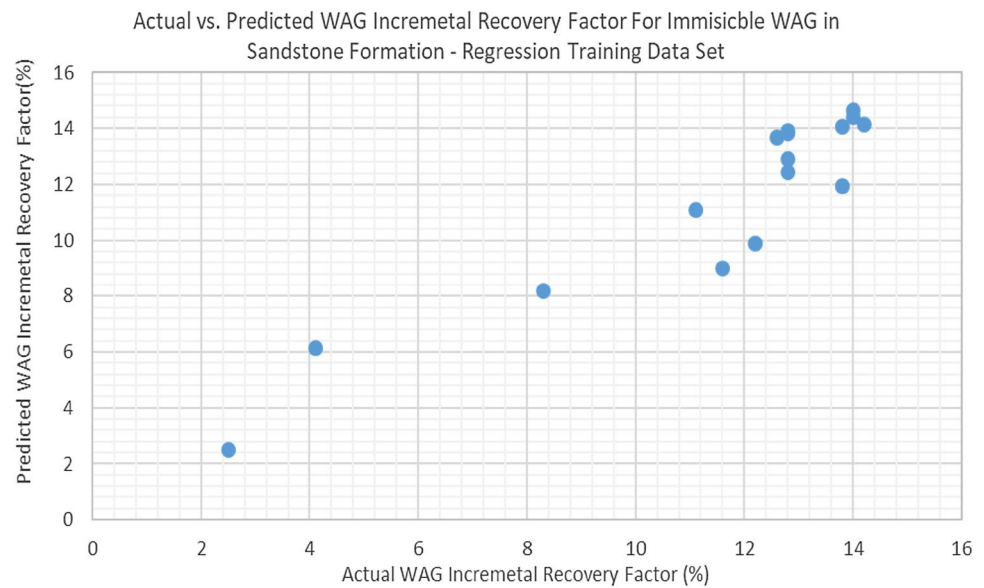
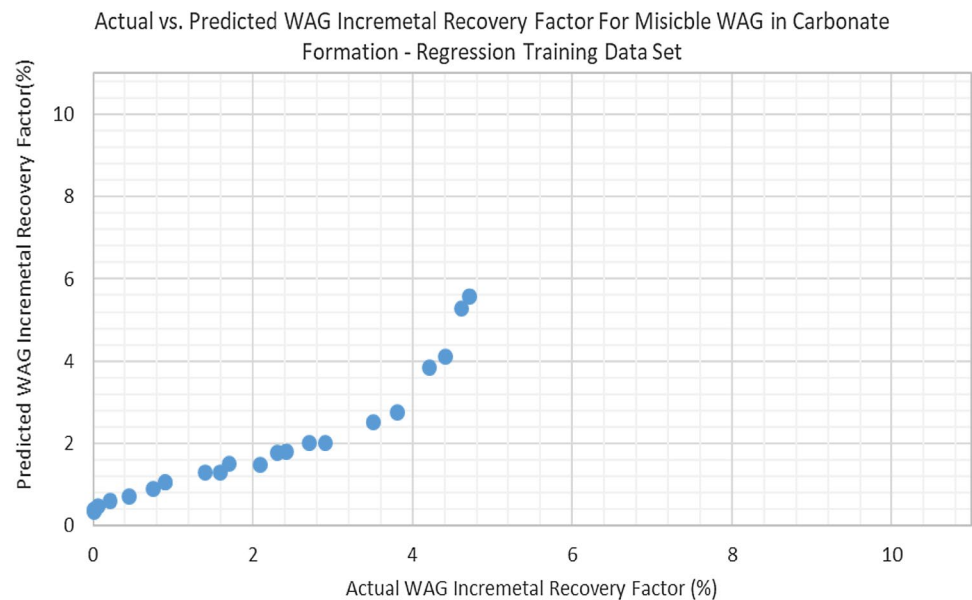


Fig. 8 Regression WAG incremental recovery factor prediction model training results for miscible WAG in carbonate formation



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.

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

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 to 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 the input to the second layer, and so on. Less effective variables will be eliminated using either regulatory criterion or root-mean-squared error.

Training and iteration will continue to produce new variables that describe the solutions better than the previous

Fig. 9 Regression WAG incremental recovery factor prediction model validation results for miscible WAG in sandstone formation

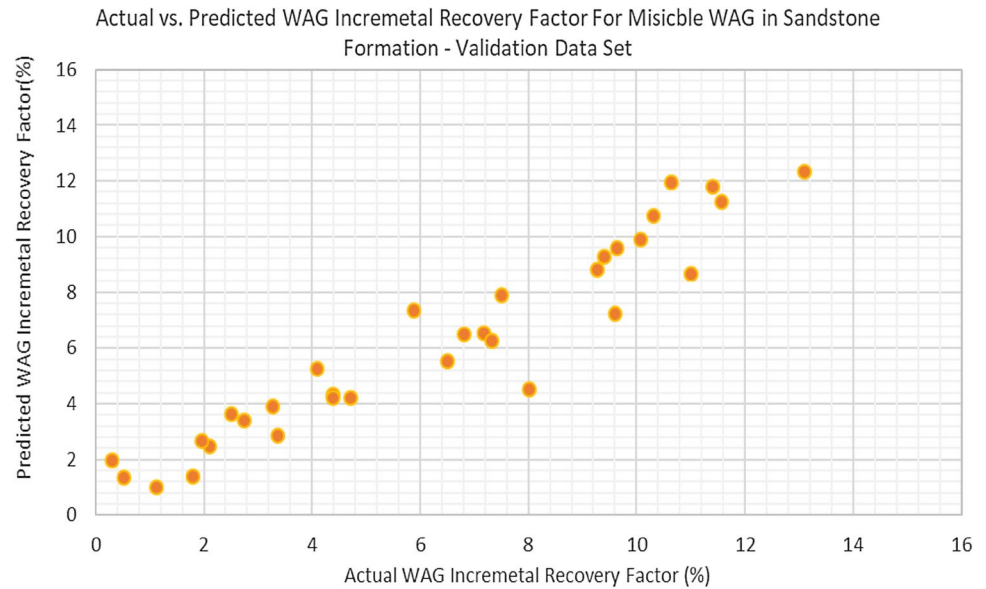
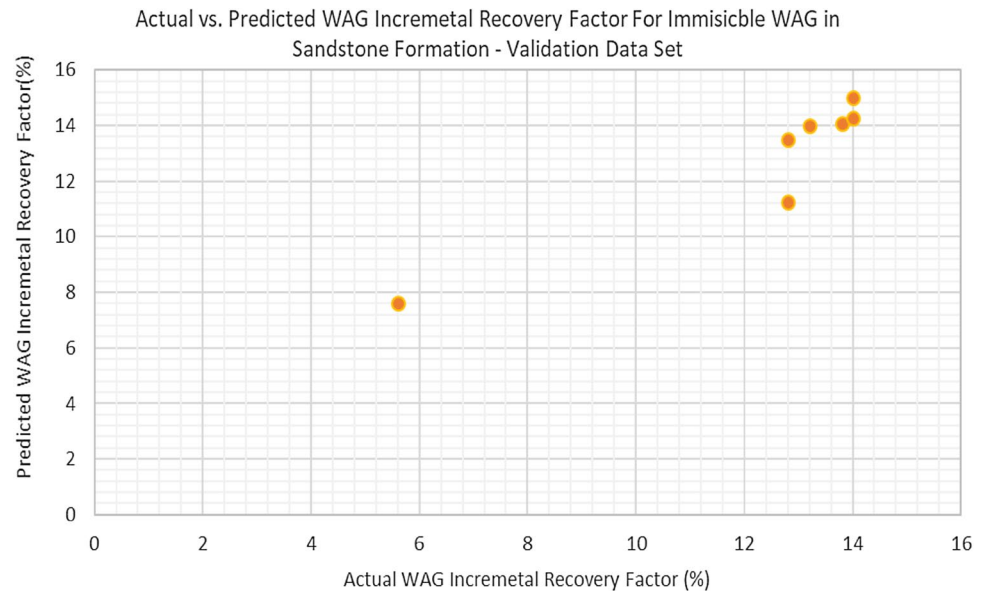


Fig. 10 Regression WAG incremental recovery factor prediction model validation results for immiscible WAG in sandstone formation



ones until the minimum error value of the current layer is greater than the previous one. Figures 12 and 13 show a typical GMDH method layout and GMDH algorithm.

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

GMDH external criterion

It is a 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 used as an 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 are then used to validate the model and determine the complete set of characteristics.

Fig. 11 Regression WAG incremental recovery factor prediction model validation results for miscible WAG in carbonate formation

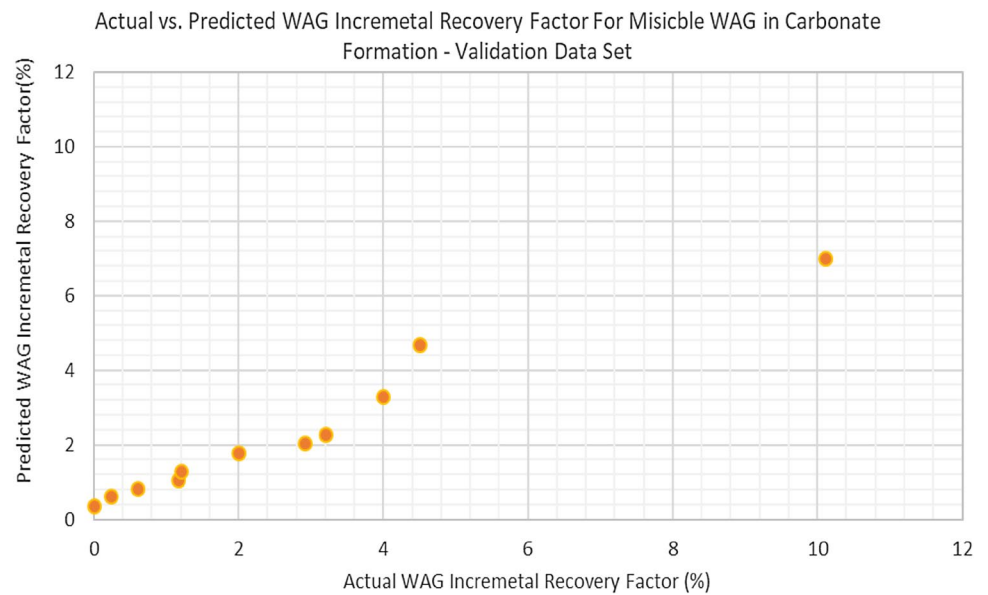


Table 3 Regression WAG incremental recovery factor prediction model output

Regression training and validation model results						
	Training			Validation		
	Miscible WAG in sandstone formation	Immiscible WAG in sandstone formation	Miscible WAG in carbonate formation	Miscible WAG in sandstone formation	Immiscible WAG in sandstone formation	Miscible WAG in carbonate formation
Samples (%)	82	17	25	35	7	11
Mean absolute error (MAE)	0.728	0.830	0.461	0.850	0.935	0.653
Root-mean-square error (RMSE)	0.931	1.184	0.530	1.113	1.110	1.05
Correlation coefficient	0.972	0.961	0.945	0.960	0.937	0.977
Coefficient of determination (R^2)	0.946	0.923	0.892	0.917	0.840	0.854

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. RMSE criterion in this case is penalized by the difference between training and testing sample RMSE.

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 fall below a certain objective default value;

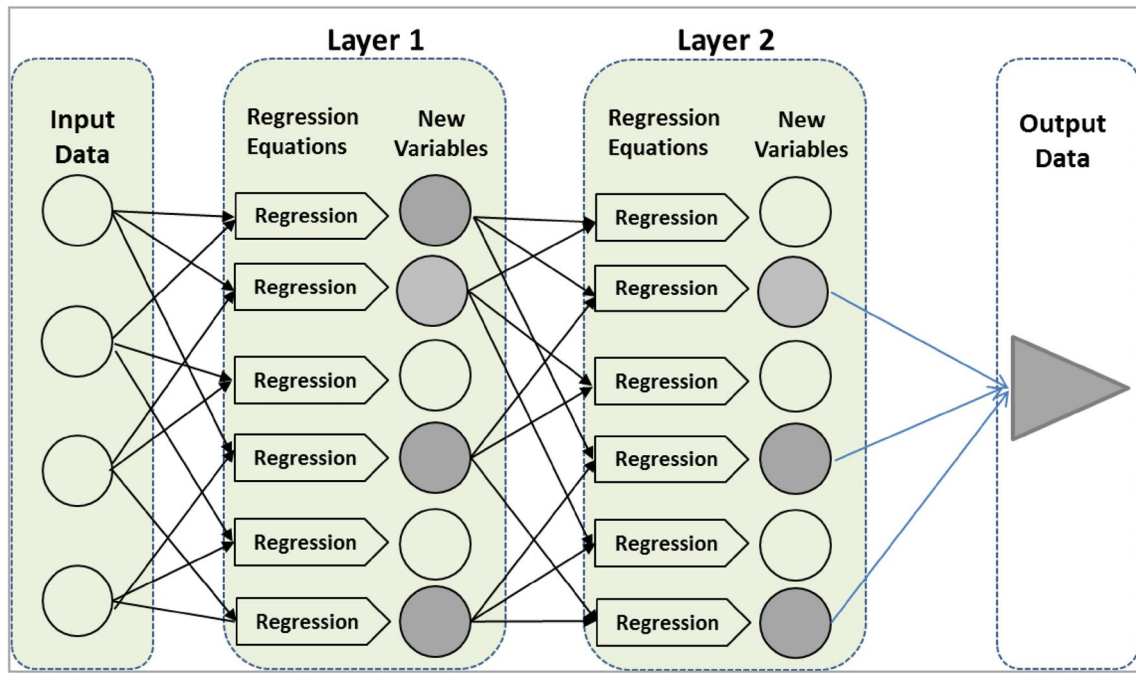


Fig. 12 GMDH method layout

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

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

Fig. 13 GMDH method algorithms

otherwise, steps 2 to 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 the features of both NN and self-organizing modeling in a variety of categories as shown in Table 4.

Group method of data handling results

The WAG data were divided into three categories: miscible WAG in sandstone formation, immiscible WAG in sandstone formation, and miscible WAG in carbonate formation.

For the development of the prediction model for WAG incremental recovery factor, 70% of the WAG data were used for training the model, while 30% was used for model validation.

Figures 14, 15, 16, 17, 18, 19, and Table 5 summarize the results of WAG incremental recovery factor for training and validation.

The prediction model details are shared under [Appendix](#).

Results summary and discussions

The research study results can be divided into two parts: the WAG incremental recovery factor prediction model results and WAG pilots lessons learned.

Table 4 Comparison between artificial neural network and group method of data handling

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

WAG incremental recovery factor predictive model

The WAG data obtained from the WAG pilots include rock properties, fluid properties, and injection scheme; however, complete data set was not available for all fields. The preliminary analysis of the of WAG parameters as per Table 2 demonstrated that selected parameters are independent.

WAG pilot reservoir parameters and WAG incremental recovery factor with one hundred and seventy-seven (177) observations were prepared and inputted to the two widely used data mining techniques which are regression and GMDH. Input data include rock type, WAG process (miscible, immiscible), reservoir permeability, reservoir pressure and temperature, oil gravity and viscosity, and hydrocarbon pore volume of injected gas.

All prediction models were trained utilizing 70% of the observations and validated using the remaining 30% for three input classes that are miscible WAG in sandstone formation, immiscible WAG in sandstone formation, and miscible WAG in carbonate formation. Sensitivity on data splitting (i.e., 80/20, 60/40, 70/30) demonstrated no significant improvement of prediction models accuracy. After multiple iterations, both regression and GMDH models were optimized. For regression method, linear regression model was tested first leading to low coefficient of determination (R^2) (i.e., $R^2 < 65\%$ for validation); hence, complexity was added to the regression model till reaching minimum mismatch error between observed and predicted data.

For the GMDH method, multiple iterations were taken adjusting maximum number of layers and neurons till optimum model achieved with minimum mismatch error. RMSE balance was used as validation criterion. The final WAG incremental recovery factors models were achieved with high coefficient of determination (R^2) as per Tables 6 and 7.

Tables 6 and 7 summarize the results of WAG incremental recovery factor prediction model parameters achieved for both regression and GMDH methods. GMDH algorithm results demonstrate its strength and ability in selecting the effective predictors in the prediction model, optimizing the network structure, and achieving predictive model with high accuracy as compared to regression technique.

Figure 20 shows the comparison between the regression and GMDH models' prediction results. Both predictive models showed good predictivity of the WAG recovery factor; however, regression model showed significant mismatch for few of the observed data.

As compared to the traditional way of estimating WAG incremental recovery factor from time-consuming reservoir simulation studies which depend on reservoir models' history matching quality, technical team expertise, and input data uncertainty, these WAG prediction models are an easy tool for quick WAG evaluation.

Fig. 14 GMDH WAG incremental recovery factor prediction model training results for miscible WAG in sandstone formation

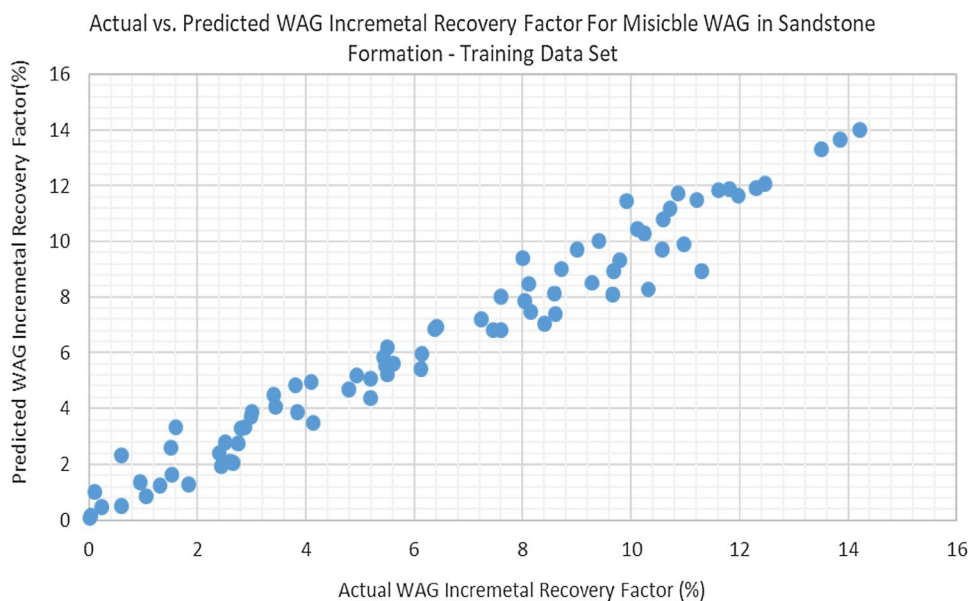
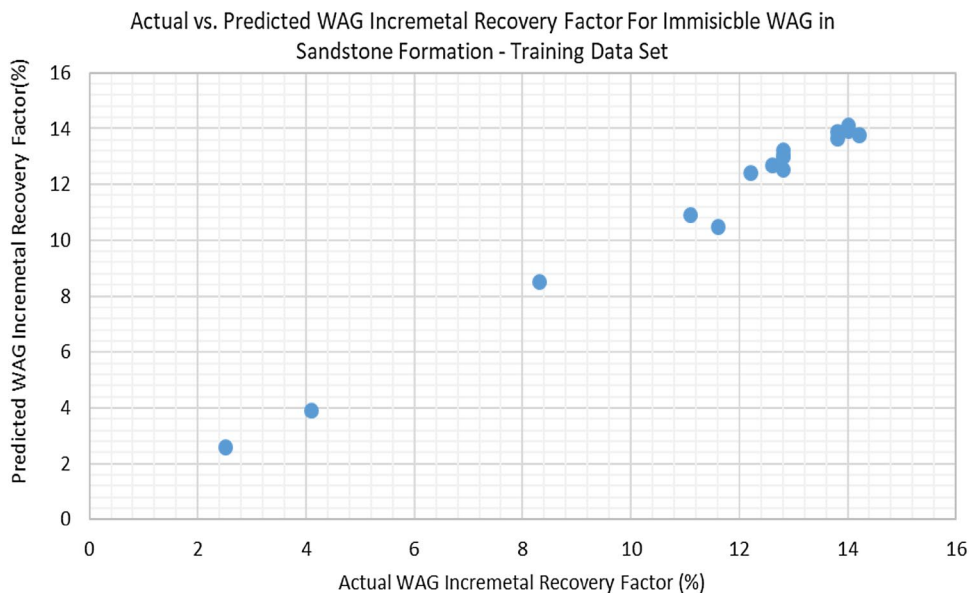


Fig. 15 GMDH WAG incremental recovery factor prediction model training results for immiscible WAG in sandstone formation



These WAG prediction models are expected to support technical people to estimate field WAG recovery factor prior launching very detailed technical study which is very expensive and costly and assess the risks associated with the WAG project.

These WAG predictive models apply to CO₂ WAG mainly since most of the WAG projects used in this study are CO₂-WAG as per Fig. 4.

WAG lesson learnt

WAG success criteria:

- Higher production rate and reduced water production (Ma and Youngren 1994).

- Immediate increase in oil production and an increase in the ultimate recovery (Hinderaker et al. 1996).
- Oil viscosity was decreased, and light components of oil were increased (Jingcun et al. 1997).
- Laboratory coreflood indicated that residual oil to waterflood, s_{orw} , is reduced by trapped gas. To reduce this residual “immobile” oil saturation, the oil mobility must have increased (Champion and Shelden 1989).
- WAG injection tends to increase waterflood efficiency and reduce water–oil ratio (WOR) (Champion and Shelden 1989).
- A gently increase in well GOR and flattening on water cut and oil production rate (Choudhary et al. 2011).

Fig. 16 GMDH WAG incremental recovery factor prediction model training results for miscible WAG in carbonate formation

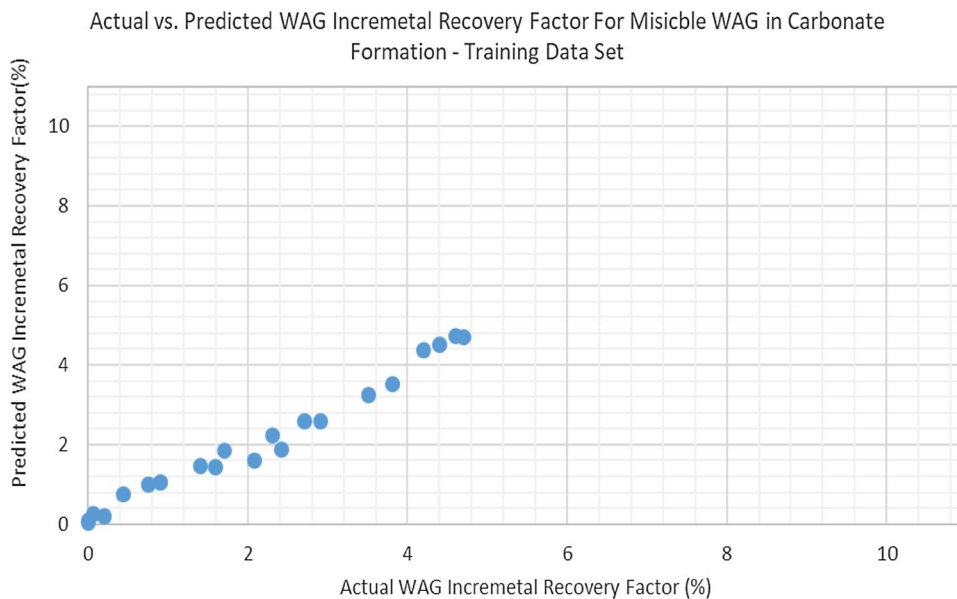
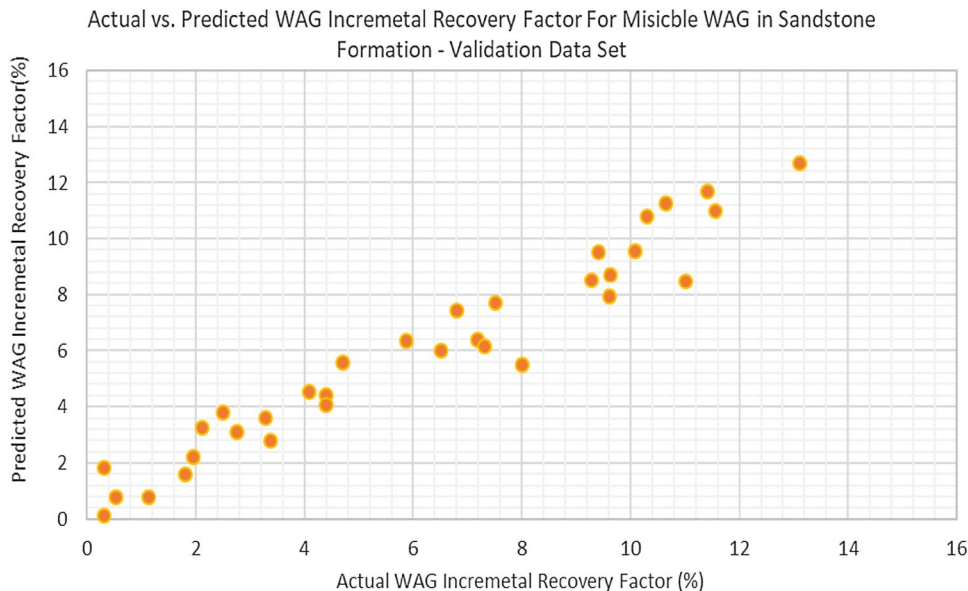


Fig. 17 GMDH WAG incremental recovery factor prediction model validation results for miscible WAG in sandstone formation



- The decline rate prior to WAG was about 55% and after a year since the commencement of WAG the decline rate dropped to below 25% (Choudhary et al. 2011).
- Laboratory displacement study indicated an improvement in displacement efficiency by 14.5% over water flood in the target sand (Ramachandran et al. 2010).
- Oil production from pilot wells was improved after initialization of WAG with a decline in water cut (Ramachandran et al. 2010).

WAG recovery mechanisms and benefits:

- Reduction in oil viscosity and an increase in reservoir oil volume due to oil swelling (Moffitt and Zornes 1992).
- Pressure maintenance, gravity drainage, compositional effects such as vaporization and viscous displacement of oil by gas (Hermansen et al. 1997).
- Pressure support, swelling of the oil and vaporization, and stripping of higher hydrocarbons (Jakobsson and Christian 1994).
- Viscous displacement, molecular diffusion, gravity drainage, vaporization/stripping—below bubble point, swelling of oil—above bubble point (Jakobsson and Christian 1994).
- Gas injection mobility control, thereby reducing gas-handling requirements (Champion and Shelden 1989).
- Oil recovery factor increases through improved water-flood conformance (Champion and Shelden 1989).

Fig. 18 GMDH WAG incremental recovery factor prediction model validation results for immiscible WAG in sandstone formation

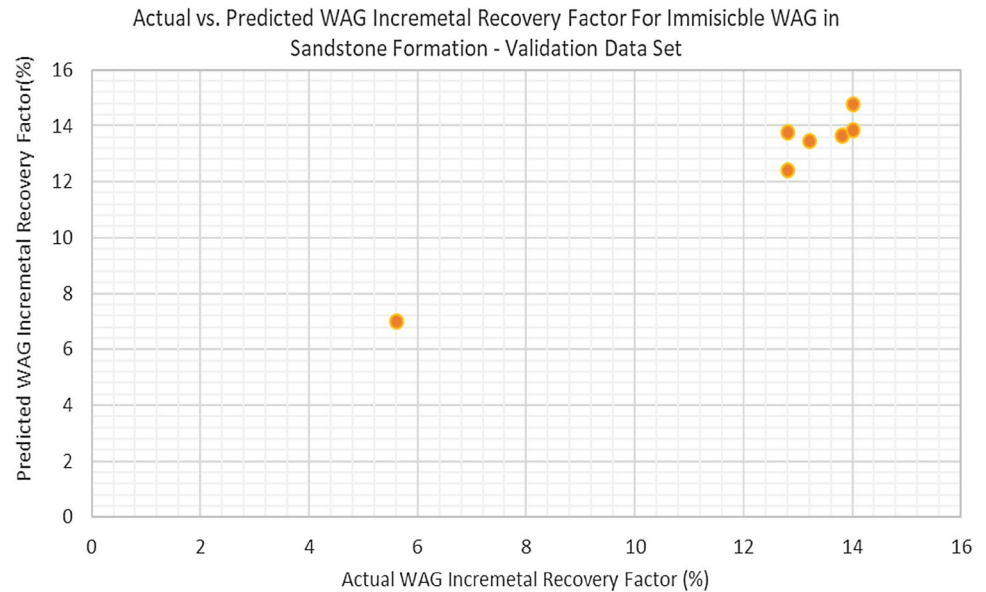
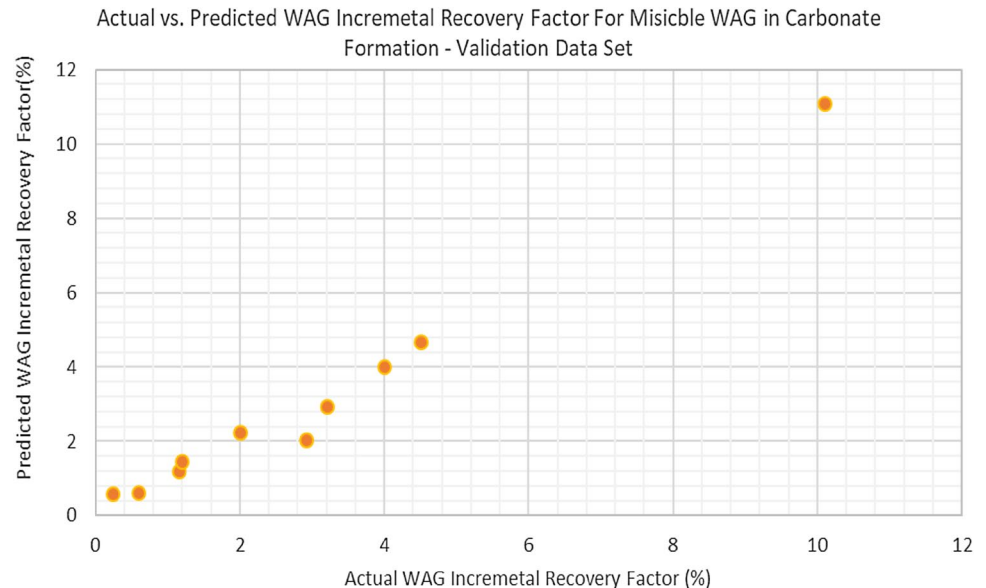


Fig. 19 GMDH WAG incremental recovery factor prediction model validation results for miscible WAG in carbonate formation



- Oil swelling and viscosity reduction contribute in the mobilization of the oil (Hsie and Moore 1998).
- Reservoir heterogeneity, dispersion, and diffusion were believed to promote spreading of CO₂ and mixing between CO₂ and reservoir Oil, thereby increasing volumetric sweep (Hsie and Moore 1998).
- Better reservoir management (improved gas handling, in situ lifting) (Ma and Youngren 1994).
- Erosion caused by sand production accelerated downhole corrosion (Hsie and Moore 1998).
- Downdip injection is dominated by gas segregation, as shown in gas injection (Leif Hinderaker et al. 1996).
- Reservoir complexity/thickness of the formation, communication between main field and east flank, production of well coverage, the distance between the WAG injector and the production wells, extensive gas migration upwards in the formation, WAG performance was underestimated in simulation (Crogh et al. 2002).
- Due to grid size and vertical reservoir permeability, representation of the WAG process in a full field simulation model is not satisfactory (Crogh et al. 2002).
- Gas override during WAG process (Pritchard et al. 1990).

Pilot WAG challenges and issues:

- Formation of in situ carbonic acid is believed to cause dissolution of carbonate cementing materials in the reservoir (Hsie and Moore 1998).

Table 5 GMDH WAG incremental recovery factor prediction model output

	Training			Validation		
	Miscible WAG in sandstone formation	Immiscible WAG in sandstone formation	Miscible WAG in carbonate formation	Miscible WAG in sandstone formation	Immiscible WAG in sandstone formation	Miscible WAG in carbonate formation
<i>GMDH training and validation model results</i>						
Samples (%)	82	17	25	35	7	11
Mean absolute error (MAE)	0.586	0.244	0.180	0.757	0.74	0.303
Root-mean-square error (RMSE)	0.764	0.340	0.220	1.004	0.74	0.440
Correlation coefficient	0.982	0.997	0.991	0.968	0.981	0.992
Coefficient of determination (R^2)	0.964	0.994	0.981	0.934	0.930	0.974

Table 6 Summary of the two prediction model results: training set

	Regression			GMDH		
	Miscible WAG in sandstone formation	Immiscible WAG in sandstone formation	Miscible WAG in carbonate formation	Miscible WAG in sandstone formation	Immiscible WAG in sandstone formation	Miscible WAG in carbonate formation
<i>Regression versus GMDH training model results</i>						
Samples (%)	82	17	25	82	17	25
Mean absolute error (MAE)	0.728	0.830	0.461	0.586	0.244	0.180
Root-mean-square error (RMSE)	0.931	1.184	0.530	0.764	0.340	0.180
Correlation coefficient	0.972	0.961	0.945	0.982	0.997	0.991
Coefficient of determination (R^2)	0.946	0.923	0.892	0.964	0.994	0.981

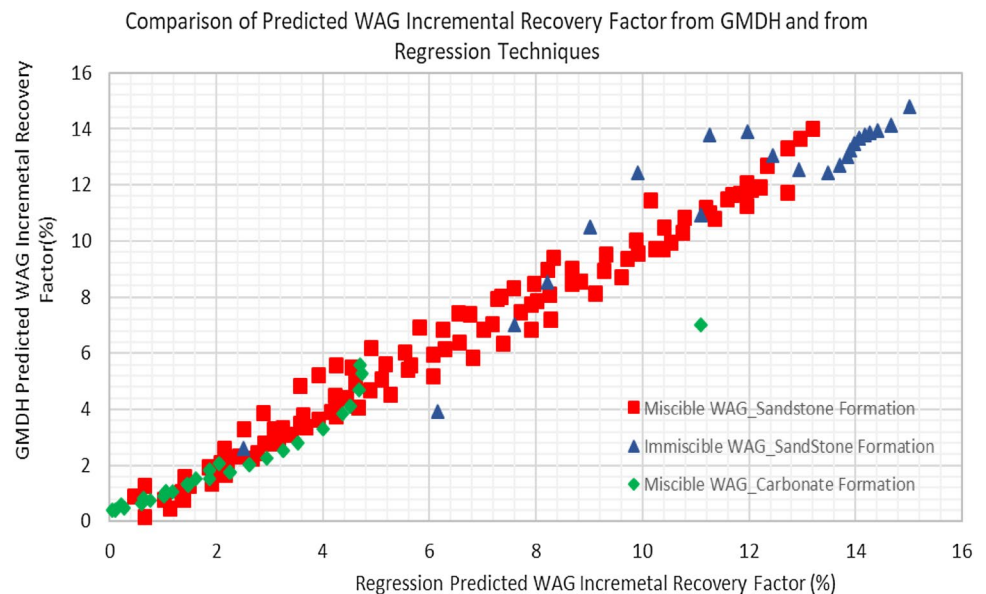
- Injected CO₂ moved mainly along the high permeable path (Jingcun et al. 1997).
- The injection pressure of water injector was increased (Jingcun et al. 1997).
- Produced liquid was slightly acidic when more calcium, magnesium, chloride, and salinity were increased (Jingcun et al. 1997).
- Wellhead pressure increased with the increase in gas–oil ratio (Jingcun et al. 1997).
- Well temperature decreased, and liquid yield of producers declined due to mobility decrease in the reservoir (Jingcun et al. 1997).
- Controlling the injection between upper and lower zones was a key factor in deciding the success of the development (Al Shamsi et al. 2012).

Conclusions

- Recovery factor from the WAG pilots was mostly ranging from 5 to 10%, with few WAG pilots showing an incremental recovery factor of up to 20%.
- Three WAG incremental recovery factor predictive models were developed for sandstone formation with immiscible gas injection, sandstone formation with miscible gas injection, and carbonate formation with miscible gas injection.
- A predictive model that estimates WAG incremental recovery factor was developed based on the input from WAG pilots with one hundred and seventy-seven (177) observations.
- Seventy percent of the data were used for training the model and 30% were used in validating the model. Split-

Table 7 Summary of the two prediction model results: training set

	Regression			GMDH		
	Miscible WAG in sandstone formation	Immiscible WAG in sandstone formation	Miscible WAG in carbonate formation	Miscible WAG in sandstone formation	Immiscible WAG in sandstone formation	Miscible WAG in carbonate formation
<i>Regression versus GMDH validation model results</i>						
Samples (%)	35	7	11	35	7	11
Mean absolute error (MAE)	0.850	0.935	0.653	0.757	0.74	0.303
Root-mean-square error (RMSE)	1.113	1.110	1.05	1.004	0.74	0.440
Correlation coefficient	0.960	0.937	0.977	0.968	0.981	0.992
Coefficient of determination (R^2)	0.917	0.840	0.854	0.934	0.930	0.974

Fig. 20 Comparison between GMDH and regression model prediction results

ting data into 50/50, 60/40, 80/20 for model training and validation, respectively, changes the coefficient of determination by 0 to 0.05.

- Coefficient of determination (R^2) for GMDH WAG predictive models was ranging from 0.964 to 0.981 and 0.934 to 0.974 for training and validation sets, respectively.
- Coefficient of determination (R^2) for regression predictive models was ranging from 0.892 to 0.946 and 0.854 to 0.917 for training and validation sets, respectively.
- GMDH results demonstrate its ability in developing predictive model, with high accuracy based, which is easy to use by reservoir engineers.
- WAG incremental recovery factor is estimated as a function of rock type, WAG process type, reservoir horizontal permeability, API, oil viscosity, reservoir pressure and temperature, and hydrocarbon pore volume of injected gas.
- This WAG incremental recovery factor prediction model is expected to help reservoir engineers generate WAG incremental production profiles and assess the WAG project risk prior launching detailed technical studies.
- The developed WAG incremental recovery factor models apply mainly to CO₂ WAG process.

Acknowledgements The authors would like to thank the University of 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

WAG incremental recovery factor prediction models for immiscible WAG process in sandstone formation

Regression model

WAG incremental recovery factor (WAG Incr. RF) from regression method for immiscible WAG process in sandstone formation is shown in Eq. 3:

$$\begin{aligned} \text{WAG Incr. RF} = & -25.6071 + 15.6822P_2^{\frac{1}{2}}P_3^{\frac{1}{3}} \\ & + 0.00264226P_4P_6 + 16.0902P_6^2 \quad (3) \\ & + 0.00390735P_1P_3^{1/3} - 44.5674P_6^{4/3} \end{aligned}$$

Group Method of Data Handling (GMDH) WAG incremental recovery factor prediction model

WAG incremental recovery factor (WAG Incr. RF) from GMDH method for immiscible WAG process in sandstone formation is shown in Eq. 4.

$$\begin{aligned} \text{WAG Incr. RF} = & 0.3873 + 0.0765779FA - 0.0454176F^2 \\ & + 0.924003A - 0.0274141A^2 \quad (4) \end{aligned}$$

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

WAG incremental recovery factor prediction models for miscible WAG process in sandstone formation

Regression model

WAG incremental recovery factor (WAG Incr. RF) from regression method for miscible WAG process in sandstone formation is shown in Eq. 5:

$$\begin{aligned} \text{WAG Incr. RF} = & -11.2068 + 0.285319P_2P_6 - (2.60876e - 05)P_1P_4 \\ & + 0.130359P_3P_5 - 8.40194P_6^2 + 18.0267P_6 \\ & + 0.00275274P_1P_5 - 2.20219P_3P_6 - 0.00100856P_5^2 \\ & - 0.133367P_1P_3 - 0.000756092P_1^2 \quad (5) \end{aligned}$$

Group Method of Data Handling (GMDH) WAG incremental recovery factor prediction model

WAG incremental recovery factor (WAG Incr. RF) from GMDH method for miscible WAG process in sandstone formation is shown in Eq. 6.

$$\begin{aligned} \text{WAG Incr. RF} = & -0.947853 + 0.0191723P_1 - 0.00100608P_1A \\ & - (5.99956e - 05)P_1^2 + 1.09467A \quad (6) \end{aligned}$$

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

Parameter	Expressions
<i>A</i>	$-0.350505 - 0.585691D - 0.348318DB + 0.212468D^2 + 1.6464B + 0.133252B^2$
<i>B</i>	$-2.60129 - 1.65265CD + 0.822269C^2 + 1.23574D + 0.826308D^2$
<i>C</i>	$2.2009 + 19.5991P_6 + 1.80286P_6P_3^{1/3} - 10.1732P_6^2 - 0.485981P_3^{1/3}$
<i>D</i>	$8.3847 - 30.4875P_6^{1/3} + 2.64131P_6^{1/3}E + 2.11164E - 0.15743E^2$
<i>E</i>	$-2.89678 + 93.3058P_6 - 90.6825P_6^{4/3} + 14.1629P_6^2$
<i>F</i>	$7.22846 + 0.138751P_1^{1/3}G - 0.130646P_1^{2/3}$
<i>G</i>	$71.1926 - 6.27601P_4^{1/3} + 4.70902P_4^{1/3}P_6^{1/3} - 33.6107P_6^{2/3}$

Parameter	Expressions
A	$0.359427 + 0.137386F + 1.23268FB - 0.676988F^2 + 0.881179B - 0.559896B^2$
B	$-10.5624 + 0.142121P_5 - 0.00233359P_5C - 0.000459685P_5^2 + 1.35854C$
C	$-0.106753 - 0.810565ED + 0.443619E^2 + 0.930795D + 0.372548D^2$
D	$3.16551 - 0.303684P_2 + 0.00529144P_2^2 + 27.0673P_6 - 13.5092P_6^2$
E	$1.34765 - 3.8589P_3 - 3.76902P_3P_6 + 1.63146P_3^2 + 28.4982P_6 - 9.06872P_6^2$
F	$-1.32253 + 0.0122785P_1 - 0.0145529P_1P_6 + 26.1293P_6 - 12.0378P_6^2$

WAG incremental recovery factor prediction models for miscible WAG process in carbonate formation

Regression model

WAG incremental recovery factor (WAG Incr. RF) from regression method for miscible WAG process in carbonate formation is shown in Eq. 7:

$$\begin{aligned} \text{WAG Incr. RF} = & 0.388736 + 0.841917P_1P_6 - 11.6735P_3P_6 \\ & + 21.0854P_6^2 + 0.00233251P_4P_6 \\ & - 0.021164P_6 + 0.0555556P_2P_6 \end{aligned} \quad (7)$$

Group Method of Data Handling (GMDH) WAG incremental recovery factor prediction model

WAG incremental recovery factor (WAG Incr. RF) from GMDH method for miscible WAG process in carbonate formation is shown in Eq. 8.

$$\begin{aligned} \text{WAG Incr. RF} = & 0.160245 - 0.420475P_6^{1/3} \\ & + 0.354172P_6^{1/3}A + 0.831651A \end{aligned} \quad (8)$$

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

Parameter	Expressions
A	$-0.0795559 - 0.0461346FB + 1.19553B$
B	$0.401215 - 1.18075P_6^{1/3} + 0.396649P_6^{1/3}C + 0.853362C$
C	$-0.840964 + 1.13819E - 0.202916ED + 0.97546D$
D	$0.181846 + 0.0162205P_4P_6 - 23.971P_6$
E	$0.181846 + 1.54545P_1P_6 - 3.31248P_6$
F	$-1.16046 + 6.17696P_1^{1/3}P_6^{1/3} - 6.90452P_6^{1/3}$

References

- Afzali S, Rezaei N, Zendehboudi S (2018) A comprehensive review on enhanced oil recovery by water alternating gas (WAG) injection. *Fuel* 227:218–246. <https://doi.org/10.1016/j.fuel.2018.04.015>
- Ahmed TH (2010) Reservoir engineering handbook, 4th edn. Gulf Professional Publishing, Houston. <https://doi.org/10.1016/B978-1-85617-803-7.50001-8> (ISBN 978-1-85617-803-7)

Ahmed AM, Rizaner A, Ulusoy AH (2016) Using data mining to predict instructor performance. *Proc Comput Sci* 102:137–142. <https://doi.org/10.1016/j.procs.2016.09.380>

Al Shamsi HA, Al-Katheeri AB, Al Ameri AF, Abdulrahman AS, Sajeel K, Al Yaaqoobi A (2012) Immiscible WAG injection pilots performance and lessons learnt in carbonate reservoir. In: Presented at Abu Dhabi international petroleum conference and exhibition, 11–14 November, Abu Dhabi, UAE. SPE-162165-MS. <https://doi-org.proxy1.athensams.net/10.2118/162165-MS>

Anastasakis L, Mort N (2001) The development of self-organization techniques in modelling: a review of the Group Method of Data Handling (GMDH). ACSE research report No 813, University of Sheffield, UK

Bini BS, Mathew T (2016) Clustering and regression techniques for stock prediction. *Proc 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 Manage* 34(9):1197–1204

Boonmee T, Tasena S (2016) Measure of complete dependence of random vectors. *J Math Anal Appl* 443(1):585–595. <https://doi.org/10.1016/j.jmaa.2016.05.051> (ISSN 0022-247X)

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

Champion JH, Shelden JB (1989) An immiscible WAG injection project in the Kuparuk river unit. Society of Petroleum Engineers. *J Pet Technol* 41:533–540. <https://doi.org/10.2118/16719-PA> (ISSN 0149-2136)

Chen S, Li H, Yang D, Tontiwachwuthikul P (2010) Optimal parametric design for water-alternating-gas (WAG) process in a CO₂-miscible flooding reservoir. *J Can Pet Technol* 49(10):75–82

Chordia M, Trivedi JJ (2010) Diffusion in naturally fractured reservoirs—a review. Society of Petroleum Engineers, Brisbane

Choudhary MK, Parekh B, Dezabala E, Solis HA, Pujijono P, Zambrano De Narvaez JZ (2011) Design, implementation and performance of a down-dip WAG pilot. In: Presented at international petroleum technology conference, 15–17 November, Bangkok, Thailand. IPTC-14571-MS. <https://doi-org.proxy1.athensams.net/10.2523/IPTC-14571-MS>

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> (ISSN 1094-6470)

Crogh NA, Eide K, Morterud SE (2002) WAG injection at the Statfjord Field, a success story. In: Presented at European petroleum conference, 29–31 October, Aberdeen, UK. SPE-78348-MS. <https://doi-org.proxy1.athensams.net/10.2118/78348-MS>

Farlow S (1981) The GMDH algorithm of Ivakhnenko. *Am Stat* 35:210–215. <https://doi.org/10.1080/00031305.1981.10479358>

Giordano MR, Redman RS, Bratvedt F (1998) A new approach to forecasting miscible WAG performance at the field scale. *SPE Reserv Eval Eng* 1:192–200. <https://doi.org/10.2118/36712-PA>

Han J et al (2012) Data mining concepts and techniques, 3rd edn. Elsevier Inc. <https://doi.org/10.1016/C2009-0-61819-5>

- Hsie JC, Moore JS (1998) The Quarantine Bay 4RC CO₂ WAG pilot project: a postflood evaluation. *SPE Reserv Eng* 3:809–814. <https://doi.org/10.2118/15498-PA> (ISSN 0885-9248)
- Jakobsson NM, Christian TM (1994) Historical performance of gas injection of Ekofisk. In: Presented at SPE annual technical conference and exhibition, 25–28 September, New Orleans, Louisiana. SPE-28933-MS. <https://doi-org.proxy1.athensams.net/10.2118/28933-MS>
- James G, Witten D, Hastie T, Tibshirani R (2015) An introduction to statistical learning: with applications in R. Springer, New York
- Jingcun Z, Shangxian X, Peihui H, Jiabao C, Zhenghua Z, Encang G (1997) Field test of immiscible CO₂ drive in Daqing oil field. *SPE advanced technology series*, vol 5, pp 49–55, SPE-30847-PA. <https://doi-org.proxy1.athensams.net/10.2118/30847-PA> (ISSN 1076-0148)
- Kim KKK, Patrón ER, Braatz RD (2018) Standard representation and unified stability analysis for dynamic artificial neural network models. *Neural Netw* 98:251–262
- Kohata A, Willingham T, Yunus Khan M, Al Sowaidi A (2017) Extensive miscible water alternating gas WAG simulation study for a giant offshore oil field. *Soc Pet Eng*. <https://doi.org/10.2118/188292-MS>
- Labastie A (2011) En route: increasing recovery factors: a necessity. *J Pet Technol* 63:8. <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, p 11
- Ma TD, Youngren GK (1994). Performance of immiscible water-alternating-gas (IWAG) injection at Kuparuk river unit, North Slope, Alaska. In: Presented at SPE annual technical conference and exhibition, 25–28 September, New Orleans, Louisiana. SPE-28602-MS. <https://doi-org.proxy1.athensams.net/10.2118/28602-MS>
- Medvedev V, Kurasova O, Bernatavičienė J, Treigys P, Marcinkevičius V, Dzemyda G (2017) A new web-based solution for modelling data mining processes. *Sim Model Pract Theory* 76:34–46. <https://doi.org/10.1016/j.simpat.2017.03.001> (ISSN 1569-190X)
- Michael N (2011) Artificial intelligence a guide to intelligence systems, 3rd edn. Pearson education, London
- Moffitt PD, Zornes DR (1992) Postmortem analysis: lick creek meakin sand unit immiscible CO₂ waterflood project. In: Presented at SPE annual technical conference and exhibition, 4–7 October, Washington, D.C. SPE-24933-MS. <https://doi-org.proxy1.athensams.net/10.2118/24933-MS>
- 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>
- Ramachandran KP, Gyani ON, Sur S (2010) Immiscible hydrocarbon WAG: laboratory to field. In: Presented at SPE oil and gas India conference and exhibition, 20–22 January, Mumbai, India. SPE-128848-MS. <https://doi-org.proxy1.athensams.net/10.2118/128848-MS>
- Rosman A, Riyadi S, Kifli AM (2011) Oil recovery optimization by immiscible WAG in offshore mature field: Dulang case study. Society of Petroleum Engineers. <https://doi.org/10.2118/144531-MS>
- Sanchez NL (1999) Management of water alternating gas (WAG) injection projects. 10.2118/53714-MS
- Semenov AA, Oshmarin RA, Driller A, Boutakova A (2010) Application of Group Method of Data Handling for geological modeling of Vankor field. In: Presented at SPE North Africa technical conference and exhibition, Cairo, Egypt, 14–17 February. 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
- Surguchev LM, Korbol R, Krakstad OS (1992) Optimum water alternate gas injection schemes for stratified reservoirs. In: Presented at SPE annual technical conference and exhibition, 4–7 October, Washington, D.C. SPE-24646-MS. <https://doi-org.proxy1.athensams.net/10.2118/24646-MS>
- Tsai T-M, Yen P-H (2017) GMDH algorithms applied to turbidity forecasting. *Appl Water Sci* 7(3):1151. <https://doi.org/10.1007/s12205-019-1298-1>
- 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
- Xiuyi T, Yuxia G (2018) Research on application of machine learning in data mining. In: IOP conference series, materials science and engineering, vol 392, p 5, Intelligent Manufacturing and Information Technology. <https://doi.org/10.1088/1757-899X/392/6/062202>
- 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. <https://doi-org.proxy1.athensams.net/10.2118/118226-MS>
- Zhou H, Deng Z, Xia Y, Fu M (2016) A new sampling method in particle filter based on Pearson correlation coefficient. *Neurocomputing* 216:208–215. <https://doi.org/10.1016/j.neucom.2016.07.036> (ISSN 0925-2312)

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