



New proxy models for predicting oil recovery factor in waterflooded heterogeneous reservoirs

Mohamed Al-Jifri¹ · Hazim Al-Attar¹ · Fathi Boukadi¹

Received: 8 September 2020 / Accepted: 16 January 2021 / Published online: 16 February 2021
© The Author(s) 2021

Abstract

To predict the recovery factor (RF) in waterflooded layered oil reservoirs, two empirical relationships were derived. Both correlations use four independent variables. These are reservoir heterogeneity (characterized by permeability variation coefficient), permeability anisotropy (ratio of vertical to horizontal permeability), viscosity of the injected water, and water injection rate. One of the correlations estimates RF at water breakthrough time (RF_{BT}) and the other evaluates RF at the end of project (RF_{EOP}). Each correlation comes in an expanded form with more parameters and a reduced form with fewer parameters. Both models are based on the global linear model. Eclipse black-oil simulation was used to determine RF for generic reservoirs with different combinations of permeability variation, permeability anisotropy, injected water viscosities, and water injection rates. A total of 192 data sets have been generated. Out of these, 144 data sets (about 75% of the generated sets) were used for model development and 48 data sets (about 25% of the generated sets) were used for model testing and validation. The expanded forms of the new developed correlations gave reliable estimates of RF_{BT} and RF_{EOP} with absolute average percent difference (AAPCD) of 6.9 and 1.02, respectively. The reduced forms yielded slightly higher AAPCDs of 8.30 and 1.04, respectively. When tested against 48 simulation-generated data sets, the expanded forms yielded excellent fits for RF_{BT} and RF_{EOP} with AAPCDs of 14 and 6.5, respectively. The reduced forms showed comparable fit with AAPCDs of 16.9 and 6.70, respectively. The highest RF_{EOP} of 50.6% was achieved for a generic reservoir with a permeability variation in $V=0.1$ and a permeability anisotropy of $k_z/k_x=1.0$. This particular reservoir needs to be waterflooded using a water viscosity of $\mu_w=1.0$ cp and a water injection rate of $q_i=10,000$ bpd. Finally, when tested against the Guthrie–Greenberger and the API statistical study, using a single field data set, the proposed correlations gave higher absolute percent difference of 22.9 and 22.7 compared to 0.758 and 19.2 for Guthrie–Greenberger and the API statistical study, respectively.

Keywords Empirical correlations · Recovery factor · Heterogeneous reservoirs · Permeability anisotropy · Waterflooding

Abbreviations

AAPCD	Absolute average percent difference
AI	Artificial intelligence
AIC	Akaike information criterion
APCD	Average percent difference
BIC	Bayesian information criterion
BT	Breakthrough time
EOP	End of project
GLM	Global linear model
MLM	Machine learning methods
MS	Mean square

RF	Recovery factor
SS	Sum of squares

Introduction

Solving oil recovery problems requires the use of reservoir models. The computing time involved increases with increasing model complexity. One way of handling the increased computing time is to develop parallel algorithms that takes care of the heterogeneity of the computing system. However, every time the initial data used for predicting oil recovery is changed the calculations must be changed which, could significantly increase the computing time and resources. Another way of doing things is to use machine learning methods (MLM), which can be used for big data processing. The application of MLM in the oil industry has

✉ Hazim Al-Attar
hazim.alattar@uaeu.ac.ae

¹ Chemical and Petroleum Engineering Department, UAE University, Al Ain, UAE

been expanding (Li et al. 2020). There were many attempts to apply MLM to oil recovery estimation in waterflooding systems. Guthrie and Greenberger (1955) related oil recovery by water drive empirically to reservoir rock and fluid properties. The authors considered 73 sandstone reservoirs. Some had a water drive, others were under a combination drive. Oil recovery was expressed as a function of permeability, porosity, oil viscosity, formation thickness, connate water saturation, depth, oil reservoir volume factor, drainage area, and well spacing. The developed correlation (Eq. 1) produced an excellent fit. It predicted oil recovery within 6.2% of the reported value in 50% of the time and 9.0% of the reported value in 75% of the time.

$$RF = 0.2719 \log k + 0.25569S_{wi} - 0.1355 \log \mu_o - 1.538\phi - 0.0003488h + 0.114403 \quad (1)$$

In Guthrie–Greenberger’s empirical model, RF is the recovery factor, k is the absolute permeability in md, S_{wi} is the initial water saturation, ϕ is the porosity, h is the formation thickness in ft, and μ_o is the oil viscosity in cp. This equation infers that water drive recovery efficiency is lower in reservoirs of higher porosity.

An API statistical study used a database from 312 reservoirs (Arps et al. 1967) undergoing water drive to develop an empirical equation to predict recovery efficiency. The correlation (Eq. 2) yielded a coefficient of determination, R^2 , of 0.958, indicating an impeccable data fit.

$$RF = 54.898 \left[\phi(1 - S_{wi})/B_{oi} \right]^{0.0422} [k\mu_w/\mu_o]^{0.077} (S_{wi})^{-0.1903} (p_i/p_a)^{-0.2159} \quad (2)$$

In the API empirical equation, RF is the recovery factor, ϕ is the porosity, μ_w is the water viscosity in cp, μ_o is the oil viscosity in cp, k is the absolute permeability in md, S_{wi} is the initial water saturation, p_i is the initial pressure in psia, p_a is the abandonment pressure in psia, and B_{oi} is the oil formation volume factor at initial reservoir pressure in rb/stb.

Another secondary recovery model was developed to predict waterflood performance for different reservoir properties and under various design conditions (Khan 1971). Multiple regression equations used eight input variables to predict injection rate, ultimate secondary reserves, response time, peak oil rate, peak production year, and the production profile as a function of time. Historical data from 12 waterfloods located in the San Jorge Basin were used to develop the regional predictive model. To generate five output variables, six geometrical factors and two reservoir quality parameters were used. Input variables are reservoir depth, net sand thickness, pore volume, number of sand layers, number of injectors and producers, porosity, and primary recovery factor. Output variables are injectivity, secondary recovery factor, response time, project life, and recovered reserves

after injecting 28% of the required number of pore volumes (R28).

Recently, a new correlation for estimating oil recovery factor under waterflooding in core samples at constant water injection rate was proposed (Balhasan and Jumaa 2017). The coefficients of correlation and determination were calculated using nonlinear regression analysis. The empirical model (Eq. 3) correlated RF to reservoir-to-surface temperature ratio, oil-to-water viscosity ratio, and oil specific gravity.

$$RF = \left[0.165 \ln (T_r/T_s) \right]^{0.88} + 0.0066 \ln (\mu_o/\mu_w) + \left[0.280 \ln (1/\gamma_o) \right]^{1.55} + 0.26 \quad (3)$$

where RF is the recovery factor, %, T_r is the reservoir temperature, °F, T_s is the surface temperature, °F, μ_o is the oil viscosity, cp, μ_w is the water viscosity, cp, γ_o is the oil specific gravity. The authors noted that oil recovery factor increased to 48.8% at 194 °F, compared to 38% at 95 °F when one pore volume was injected. Furthermore, the authors indicated that the developed correlation was also applicable for three sandstone reservoirs in Libya and one sandstone reservoir in Kuwait.

Artificial intelligence (AI) was also used in the estimation of the oil recovery factor in waterflooded reservoirs (Mahmoud et al. 2019). In this study, the authors collected a data set of 173 cases and analyzed it statistically. Data outliers within ± 0.3 standard deviation (SD) were removed. Five cases were therefore discarded, and the remaining 168 cases were used to develop the AI models. These models were trained using 77% of the data. The remaining 23% were used for model testing and validation. The parameters used are divided into four groups (asset size, rock parameters, fluid properties, and reservoir energy). The authors noted that their developed model outperformed all the available empirical equations in the literature. The model yielded a coefficient of determination of 0.94 compared to only 0.55 obtained from Gulstad’s correlation (Gulstad 1995). Gulstad’s is considered as the industry standard. Sharma et al. (2010) used Tertiary Oil Recovery Information Systems (TORIS) for oil reservoirs and Gas Information System (GASIS) for gas reservoirs to match multivariant linear regression. These authors highlighted the accuracy of their linear model. Using descriptions of 34 low-permeability reservoirs, Han and Bian (2018) illustrated the application of a model based on support vector machine in combination with the particle swarm optimization (PSO-SVM) technique for oil recovery. Aliyuda and Howell (2019) demonstrated their successful application of support vector machine using a data set of 93 reservoirs descriptions from the Norwegian continental shelf. Using a huge data set of more than 2000 reservoirs, Makhotin et al. (2020) applied advanced MLMs to estimate oil recovery factor. They concluded that their

proposed data-driven approach overcomes several limitations of the traditional methods and that, it is suitable for reliable and rapid estimation of oil recovery factor.

In this work, new models were developed for predicting oil recovery factors at water breakthrough time and at end of project in waterflooded multilayered reservoirs. The new models incorporate four key parameters believed to impact the performance of such reservoirs. The statistical database was generated using ECLIPSE black-oil simulation, and the new empirical models were developed through multiple regression analysis using Global Linear Model.

Proposed design methodology

Data preparation for the simulator and assumptions

The Dykstra–Parsons method (Ahmed 2002) was applied to generate permeability distribution across reservoir thickness for different reservoir heterogeneities. Each reservoir heterogeneity corresponds to a specific value of Dykstra–Parsons permeability variation coefficient (V). By setting an average permeability of a 10-layer reservoir (k_{50}) equal to 68 md and selecting several values of V , the corresponding permeability distributions in a descending order are determined using Eqs. 4 and 5. For each selected value of V , Eq. 4 is solved for $k_{84.1}$ (at one standard deviation from the average value) as depicted in Eq. 5. Results of these calculations are presented in Table 1. Using a log-probability graph paper with permeability on the log scale and percent cumulative frequency on the probability scale and assuming log-normal permeability distribution, Fig. 1 is generated. The permeability distribution for each selected value of V was then deduced from Fig. 1 and the results are listed in Table 2.

Table 1 Values of k_{50} and $k_{84.1}$

V	k_{50} (md)	$k_{84.1}$ (md)
0.1	68	61.2
0.3	68	47.6
0.5	68	34
0.7	68	20.4

$$V = (k_{50} - k_{84.1}) / k_{50} \tag{4}$$

$$k_{84.1} = k_{50}(1 - V) \tag{5}$$

Fig. 1 Dykstra–Parsons permeability variation plot for V values between 0.1 and 0.7

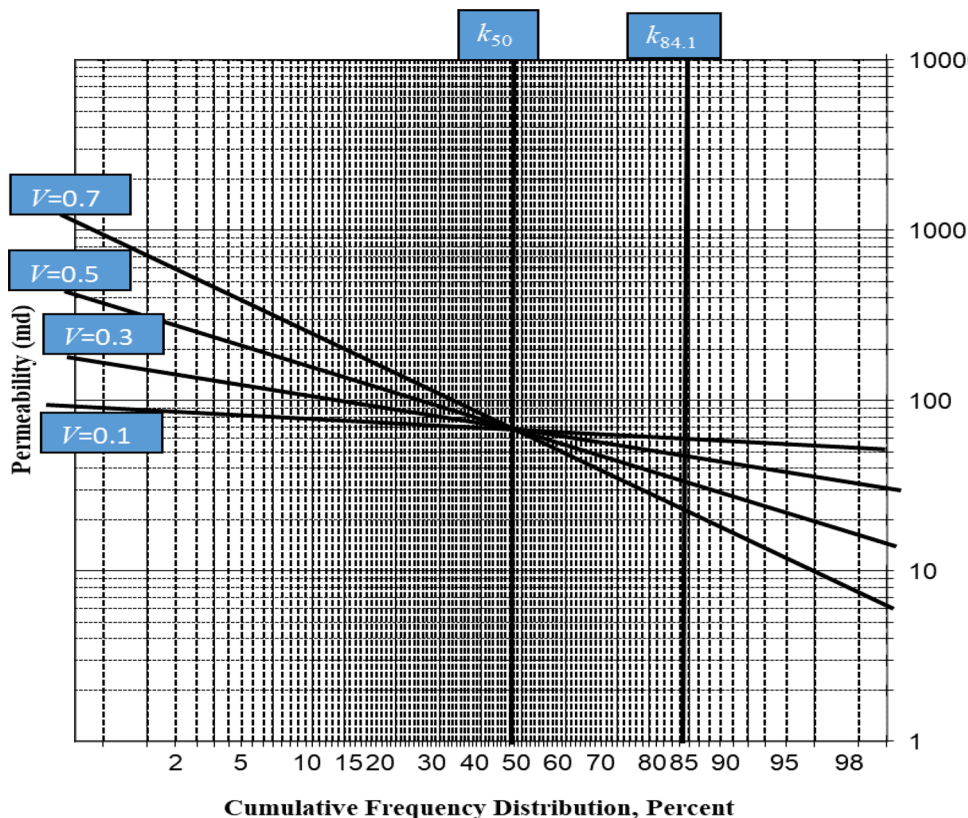


Table 2 Permeability distributions

<i>k</i> , md			
<i>V</i> =0.1	<i>V</i> =0.3	<i>V</i> =0.5	<i>V</i> =0.7
78	130	190	400
76	110	160	300
74	97	130	220
73	93	120	170
71	87	110	140
70	84	95	130
69	78	87	105
68	75	78	87
67	72	72	75
66	68	65	65

Corey's correlations (Ahmed 2002) were used to generate the relative permeability curves as presented in Eqs. (6) and (7), and the results of calculations are shown in Fig. 2.

$$k_{ro} = [(1 - S_w)/(1 - S_{wi})]^4 \quad (6)$$

$$k_{rw} = [(S_w - S_{wi})/(1 - S_{wi})]^4 \quad (7)$$

Additional input data for the simulator were necessary to generate the various recovery performances for various scenarios. The assumed data are listed in Table 3.

The basic assumptions made in this work are as follows: (1) water-wet reservoir, (2) no free gas saturation at all times during the flooding process, (3) waterflooding affects one quadrant of a five spot pattern with one injector and one producer, (4) capillary pressure effect is ignored, (5) constant

porosity, thickness and initial water saturation for all layers, (6) low volatility black oil, homogeneous individual layers, and (7) a lateral sweep efficiency of 100%.

Combination scenarios of various key parameters

Eclipse black-oil simulation was used to generate production performance profiles for various combination scenarios. In this work, four key parameters with a dominant impact on waterflooding projects were considered. These parameters are water injection rate, water viscosity, reservoir anisotropy, and reservoir heterogeneity. Table 4 presents values of key parameters used in the combination scenarios and thus, a total of 192 combination scenarios were obtained.

Generation of simulation data for statistical analysis

In addition to other performance profiles, the main output from eclipse after running the above combination scenarios includes water cut versus time and oil recovery factor versus time. A complete set of results of simulator calculations can be found in Ref. Al-Jifri (2020).

Application of minitab to develop the new empirical correlation

The simulator-generated data were then used as input data for the Minitab software for further statistical analysis. The objective of using the GLM is to generate the proposed empirical correlation for oil recovery factor in terms of the four key parameters listed in Table 4. The Global Linear Model (GLM) was implemented because it predicts values for new observations, identify the combination of predictor values that jointly optimize one or more fitted values,

Fig. 2 Relative permeability curves generated by Corey's correlations

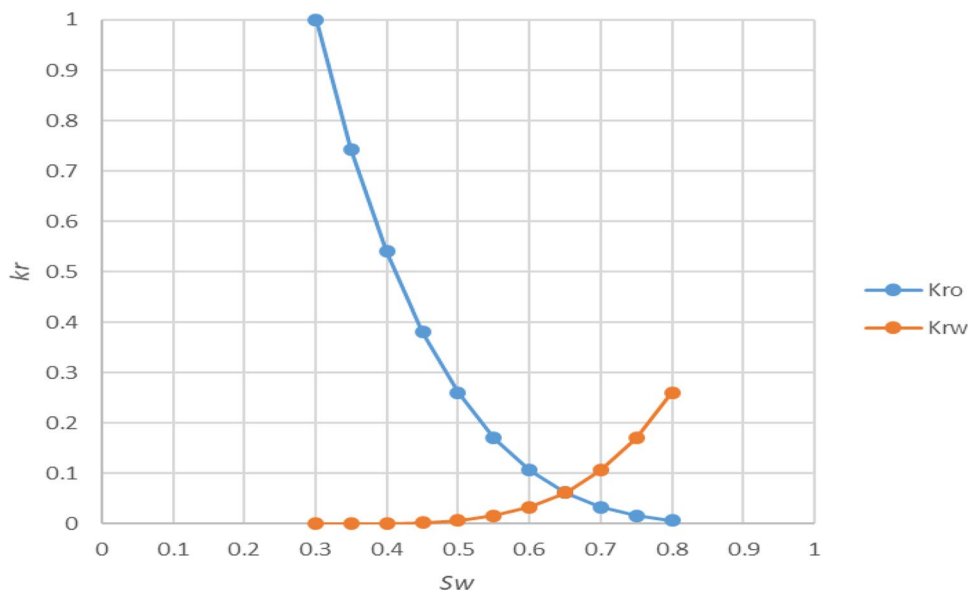


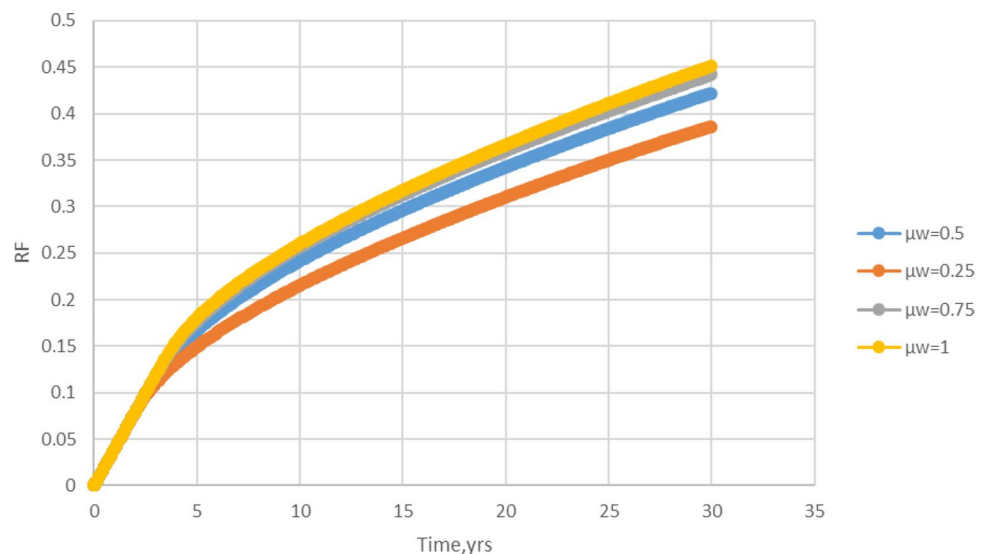
Table 3 Eclipse input data

Parameter	Value
Number of cells in <i>x</i> -direction	10
Number of cells in <i>y</i> -direction	10
Number of layers	10
Depth	8000 ft
Reservoir pressure	4500 psia
Reservoir temperature	240 °F
Reservoir thickness	50 ft
Area	72 acres
Porosity	0.20
Water formation volume factor	1.02 rb/stb
Oil viscosity at reservoir pressure	2.0 cp
Water compressibility	3E-06 psi ⁻¹
Water density	49 lbs/cf
Oil density	63 lbs/cf
Gas density	0.01 lbs/cf
Pore compaction	4E-06 psi ⁻¹
Water oil contact	8050 ft

Table 4 Values of key parameters used in the combination scenarios

Variables	Scenarios
Reservoir heterogeneity	$V=0.1; 0.3; 0.5$ and 0.7
Water injection rate	$q_i=2000; 5000$ and $10,000$ stb/day
Permeability anisotropy	$k_z/k_x=0.1; 0.3; 0.5$ and 1
Water viscosity	$\mu_w=0.25; 0.5; 0.75$ and 1 cp

and create surface plots, contour plots, and factorial plots. Also, it can identify the key parameters of highest and lowest impact on the predicted oil recovery, water cut, cumulative oil production, and reservoir pressure.

Fig. 3 Oil recovery factor performance

Validation of the new empirical correlation

The empirical correlation (s) thus developed by the GLM using the simulator-generated data were then validated using simulation-generated data not considered in their development. Further validation was accomplished using field data.

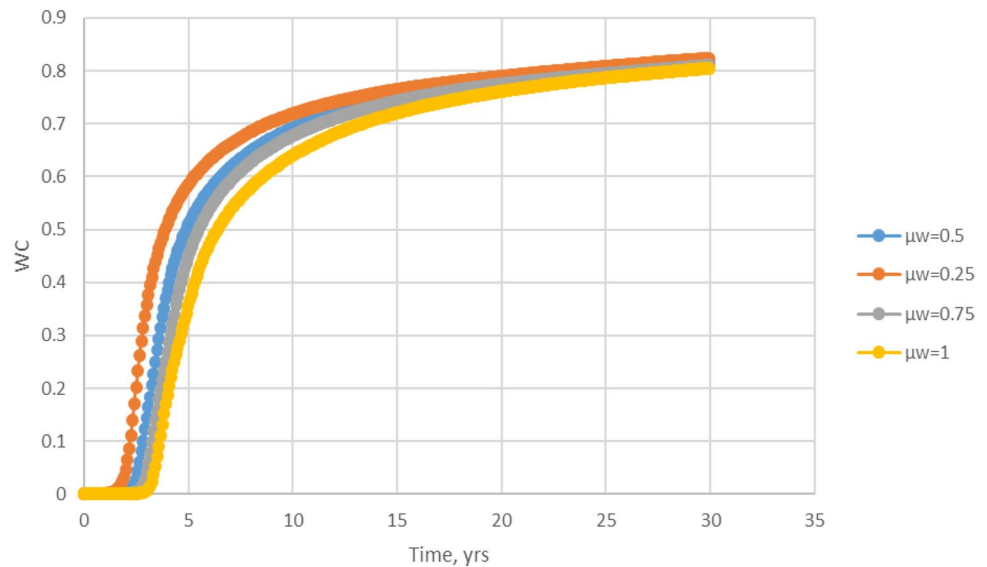
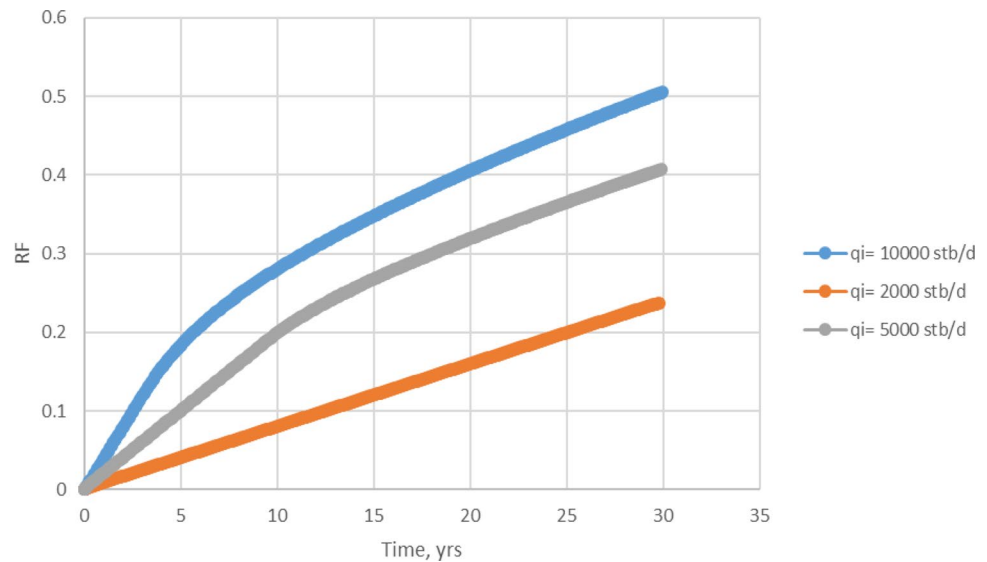
Simulated reservoir performance during waterflooding

Plots of specific performances for selected combination scenarios are shown in Figs. 3, 4, 5, 6, 7, 8, 9, and 10. These plots show the simulated performances of oil recovery factor and water cut during the waterflooding project.

Effect of water viscosity

Selected combination scenario: $V=0.5$, $k_z/k_x=1$, $q_i=10,000$ stb/day, and $\mu_w=0.25-1.0$ cp.

This selected scenario investigates the effect of changing water viscosity on the waterflood performance. In this scenario, all parameters were held constant and only water viscosity was changed and the results of the predicted performances are shown in Figs. 3 and 4. As can be observed from these plots, the most effective case is when the water viscosity approaches the oil viscosity which leads to favorable mobility ratio of one or less than one. Under such conditions, the highest RF of 45% is achieved for $\mu_w=1.0$ cp as shown in Fig. 3 at end of project. In addition, a water viscosity of 1 cp has been found to yield later breakthrough time and lower water cut than those predicted for the lower water viscosities as depicted in Fig. 4. Thus, favorable mobility ratio can be achieved as the water viscosity approaches the oil viscosity resulting in improved overall waterflooding performance.

Fig. 4 Water cut performance**Fig. 5** Oil recovery factor performance

Effect of water injection rate

Selected combination scenario: $V=0.7$, $k_z/k_x=1$, $\mu_w=1$ cp, and $q_i=2000$ – $10,000$ stb/day.

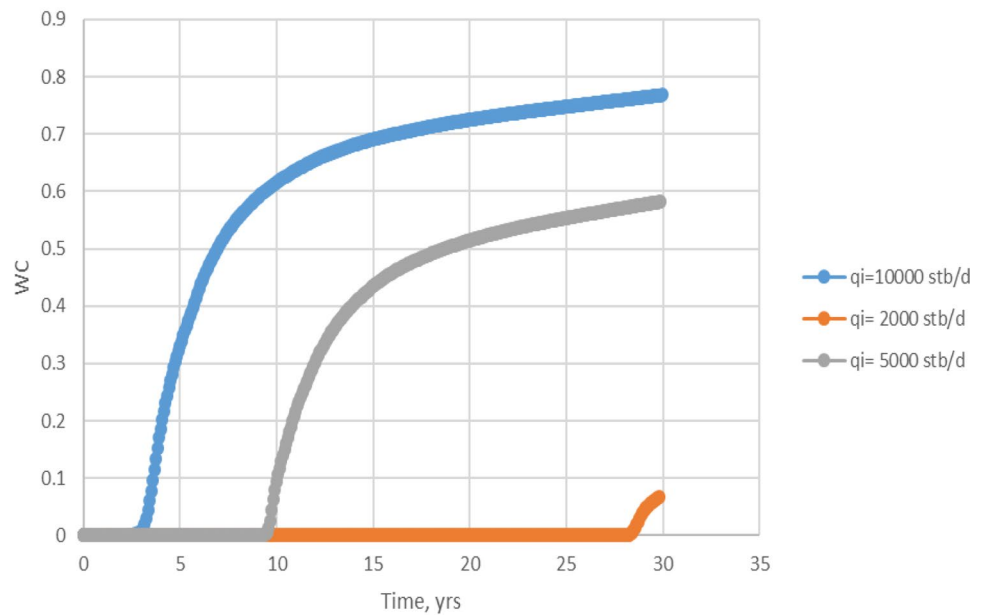
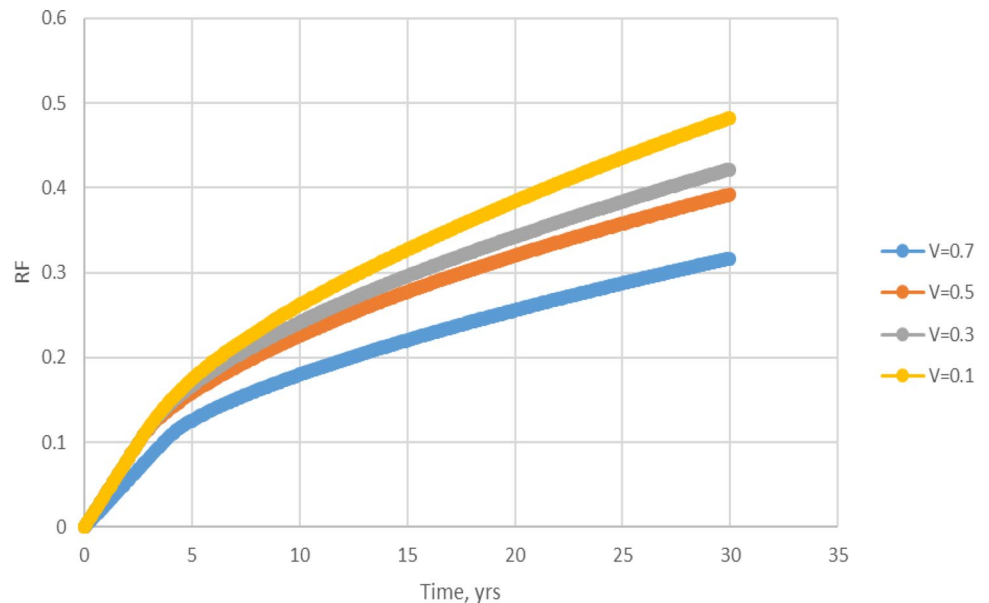
In this scenario, the effect of water injection rate on oil recovery factor and water cut was investigated as illustrated in Figs. 5 and 6. In these plots, the water injection rate was changed between 2000 and 10,000 stb/d. From Fig. 5, it can be observed that the oil recovery factor is increased from 23 to 51% as the water injection increased from 2000 to 10,000 stb/day. The negative aspect of injecting water at high rate, however, is revealed in Fig. 6, as it leads to earlier water breakthrough at the producing end and significant increase in the produced water cut. This negative aspect of high water injection rate of 10,000 stb/d may well be counterbalanced

by the significant increase in cumulative oil production of around 40 million stb at the end of the project (not shown). Therefore, the feasibility of any waterflooding project should be assessed based on similar performances as described above, and the final decision would be a compromise of the above effects.

Effect of reservoir heterogeneity

Selected combination scenario: $k_z/k_x=1$, $\mu_w=0.5$ cp, $q_i=10,000$ stb/day, and $V=0.1$ – 0.7

In this scenario, the effect of changing permeability variation coefficient on waterflooding performance was investigated and the results are plotted in Figs. 7 and 8. In these plots, V varied between 0.1 and 0.7 and other parameters

Fig. 6 Water cut performance**Fig. 7** Oil recovery factor performance

were held constants to observe the significance of reservoir heterogeneity on waterflooding performances. From Fig. 7, it can be observed that the highest RF of 46% is achieved in a homogeneous reservoir with $V = 0.1$ and that the lowest RF of 33% is realized in a heterogeneous reservoir with $V = 0.7$. The effect of reservoir heterogeneity on water cut performance is not as pronounced as shown in Fig. 8. The selected combination scenario addresses conditions of full cross-flow between layers and high water injection rate. In addition, the descending order of permeability across reservoir leads to what is known as coarsening upward situation in layered reservoirs which would normally lead to the development of piston-like displacement across the macroscopic section

(Dake 2001). The high injection rate of water, however, reduces the effect of gravity and thus the displacement front will not become so perfect which may explain the lower RF and later BT for $V = 0.7$ as shown in Fig. 8.

Effect of permeability anisotropy ratio

Selected combination scenario: $V = 0.7$, $\mu_w = 0.25$ cp, $q_i = 10,000$ stb/day, and $k_z/k_x = 0.1-1$.

In this scenario, the effect of changing permeability anisotropy on the performance of waterflood projects was studied. It can be observed in Figs. 9 and 10 that changing k_z/k_x from 0.1 (limited cross-flow between layers) to one (full

Fig. 8 Water cut performance

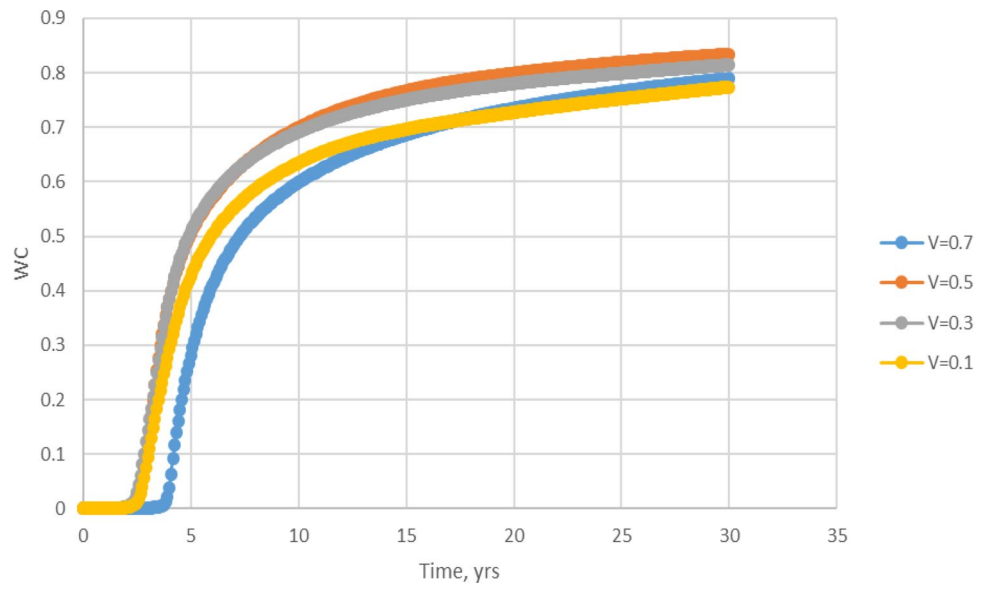


Fig. 9 Oil recovery factor performance

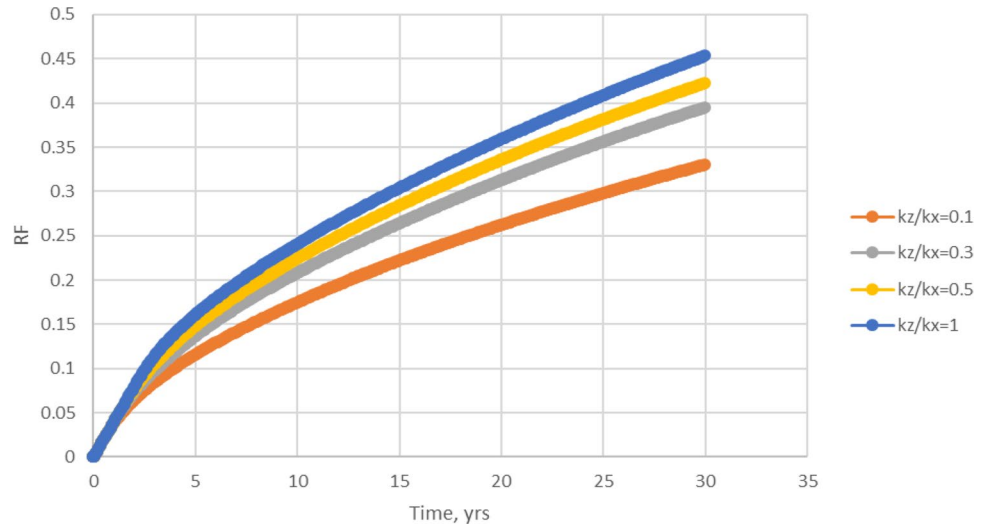
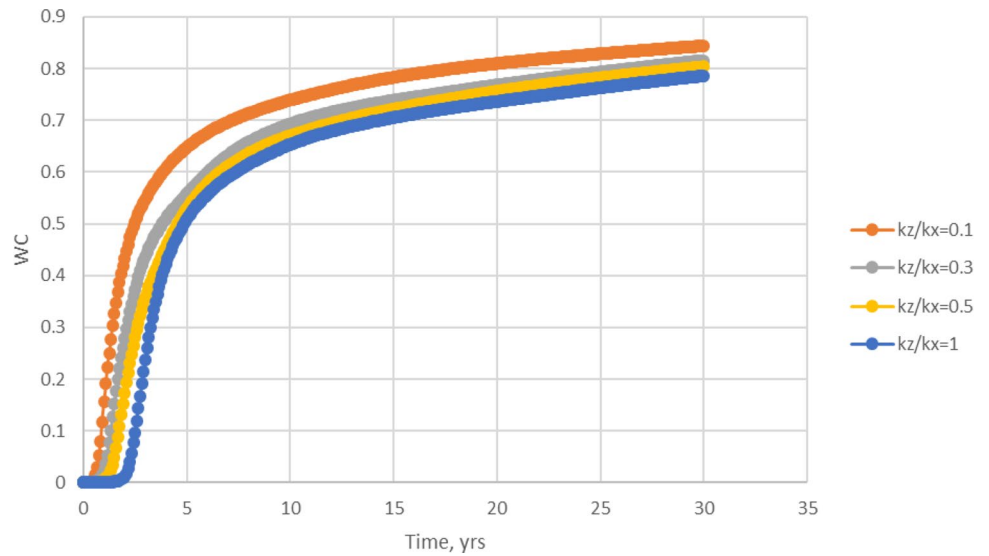


Fig. 10 Water cut performance



cross-flow between layers) can have significant impacts on the various performances considered in this work. For example, in Fig. 9, the RF is increased from 35 to 45% for k_z/k_x values of 0.1 and 1, respectively. In addition, higher water cuts are observed limited cross-flow which is indicative of unfavorable waterflood performances as shown in Fig. 10. The coarsening upward of permeability and the water injection rate-gravity relationship dictates the favorable displacement in layered reservoirs with cross-flow.

Predicting RF at water breakthrough time (RF_{BT})

A total of 144 simulation data sets have been generated, 96 data sets (about 75% of the generated sets) were used to develop the correlations and 48 data sets (about 25% of the generated sets) were used to test for the accuracy of the derived empirical models. The first correlation (Eq. 8), the *expanded form*, uses the four key parameters considered in this work. The second correlation (Eq. 9), the *reduced form*, which is based on the GLM analysis, still uses the same independent variables but considers fewer parameters.

$$\begin{aligned}
 RF_{BT} = & 3.23 + 44.44(k_z/k_x) + 17.03(\mu_w) - 3.9(V) - 2.467(q_i) - 72.16(k_z/k_x)^2 - 13.1(\mu_w)^2 - 0.7(V)^2 \\
 & + 0.1986(q_i)^2 + 21.4(k_z/k_x * \mu_w) - 12(k_z/k_x * V) - 0.388(k_z/k_x * q_i) - 0.3(\mu_w * V) - 1.171(\mu_w * q_i) \\
 & + 1.2(V * q_i) + 31.96(k_z/k_x)^3 + 3.73(\mu_w)^3 - 6.14(k_z/k_x^2 * \mu_w) + 4.47(k_z/k_x^2 * V) + 2.23(k_z/k_x^2 * q_i) \\
 & - 2.81(k_z/k_x * \mu_w^2) - 9.62(k_z/k_x * \mu_w * V) - 0.855(k_z/k_x * \mu_w * q_i) - 0.24(k_z/k_x * V^2) + 0.46(k_z/k_x * V * q_i) \\
 & - 0.1779(k_z/k_x * q_i^2) - 0.25(\mu_w^2 * V) + 0.361(\mu_w^2 * q_i) + 5(\mu_w * V^2) + 0.097(\mu_w * V * q_i) + 0.0497(\mu_w * q_i^2) \\
 & + 0.72(V^2 * q_i) - 0.1165(V * q_i^2) + 1.16(k_z/k_x * \mu_w * q_i * V)
 \end{aligned} \tag{8}$$

In Eqs. 8 and 9, RF_{BT} is the oil recovery factor at breakthrough, k_z/k_x is the anisotropy, μ_w is the water viscosity in cp, V is the permeability variation coefficient, and q_i is the water injection rate in Mstb/d.

$$\begin{aligned}
 RF_{BT} = & 3.77 + 42.28(k_z/k_x) + 11.14(\mu_w) - 2.22(V) - 2.202(q_i) - 69.3(k_z/k_x)^2 - 5.51(\mu_w)^2 \\
 & + 0.1372(q_i)^2 + 11.52(k_z/k_x * \mu_w) - 6.4(k_z/k_x * V) + 3.57(\mu_w * V) - 0.1894(\mu_w * q_i) + 31.96(k_z/k_x)^3 \\
 & - 6.14(k_z/k_x^2 * \mu_w) + 2.126(k_z/k_x^2 * q_i) - 0.1945(k_z/k_x * q_i^2) + 0.0843(V * q_i^2)
 \end{aligned} \tag{9}$$

Development of new empirical correlations

Generating the new empirical correlation(s)

To develop new correlations for the oil recovery factor, simulation-generated data were used in Minitab.

The results of Minitab software in terms of significance of parameters and model summary for RF_{BT} are illustrated in Table 8 and Fig. 19 of "Appendix".

Predicting the RF at end of project (RF_{EOP}).

Similar developments for oil recovery factor at the end of project were attempted and the resulting correlations include an expanded form (Eq. 10) and a reduced form (Eq. 11).

$$\begin{aligned}
 RF_{EOP} = & 14.09 + 14.16(k_z/k_x) - 1.41(\mu_w) - 9(V) + 1.624(q_i) - 29.39(k_z/k_x)^2 - 9.96(\mu_w)^2 \\
 & + 4.93(V)^2 - 0.0848(q_i)^2 + 11.22(k_z/k_x * \mu_w) - 17.61(k_z/k_x * V) + 2.968(k_z/k_x * q_i) + 16.99(\mu_w * V) \\
 & + 3.527(\mu_w * q_i) + 3.382(V * q_i) + 12.7(k_z/k_x)^3 + 4.74(\mu_w)^3 - 0.29(k_z/k_x^2 * \mu_w) + 9.4(k_z/k_x^2 * V) \\
 & - 0.294(k_z/k_x^2 * q_i) - 5.82(k_z/k_x * \mu_w^2) - 0.37(k_z/k_x * \mu_w * V) - 0.122(k_z/k_x * \mu_w * q_i) + 4.89(k_z/k_x * V^2) \\
 & - 0.846(k_z/k_x * V * q_i) - 0.1109(k_z/k_x * q_i^2) + 1.92(\mu_w^2 * V) - 0.453(\mu_w^2 * q_i) - 8.63(\mu_w * V^2) \\
 & - 2.728(\mu_w * V * q_i) - 0.136(\mu_w * q_i^2) - 2.659(V^2 * q_i) - 0.0763(V * q_i^2) + 0.075(k_z/k_x * \mu_w * q_i * V)
 \end{aligned} \tag{10}$$

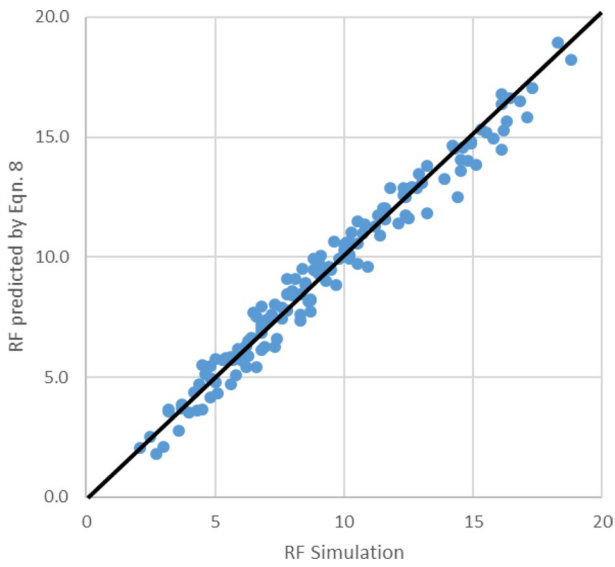


Fig. 11 Comparison between predicted and simulated values of RF_{BT}

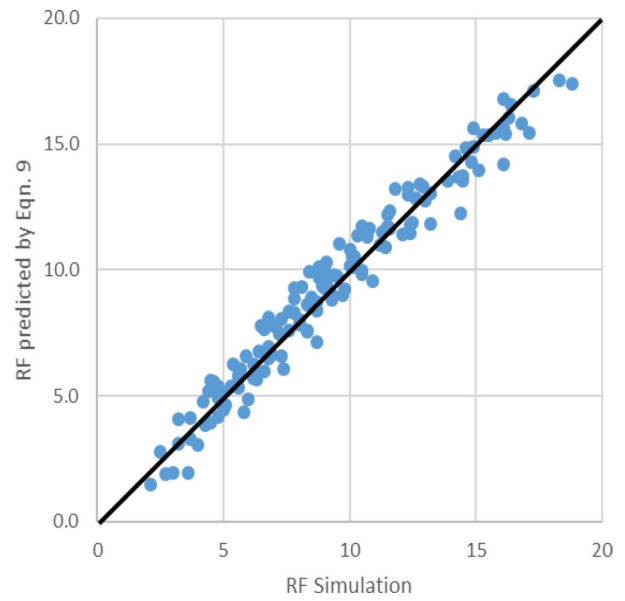


Fig. 13 Comparison between predicted and simulated values of RF_{BT}

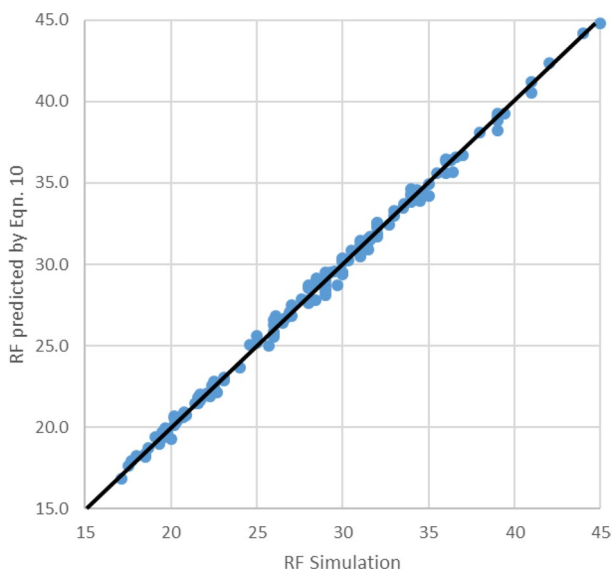


Fig. 12 Comparison between predicted and simulated values of RF_{EOP}

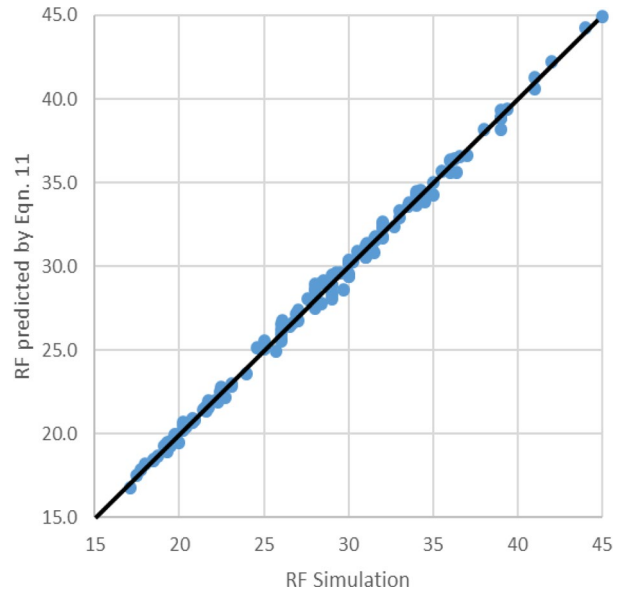


Fig. 14 Comparison between predicted and simulated values of RF_{EOP}

$$\begin{aligned}
 RF_{EOP} = & 13.862 + 13.66(k_z/k_x) - 5.399(\mu_w) - 4.56(V) + 1.713(q_i) - 29.57(k_z/k_x)^2 + 0.0848(q_i)^2 \\
 & + 10.32(k_z/k_x * \mu_w) - 12.95(k_z/k_x * V) + 2.892(k_z/k_x * q_i) + 10.58(\mu_w * V) + 3.482(\mu_w * q_i) \\
 & + 3.114(V * q_i) + 12.7(k_z/k_x)^3 + 9.4(k_z/k_x^2 * V) - 0.294(k_z/k_x^2 * q_i) - 5.9(k_z/k_x * \mu_w^2) - 0.8(k_z/k_x * V * q_i) \\
 & - 0.1109(k_z/k_x * q_i^2) - 0.463(\mu_w^2 * q_i) - 2.692(\mu_w * V * q_i) - 0.136(\mu_w * q_i^2) - 2.413(V^2 * q_i) - 0.0763(V * q_i^2)
 \end{aligned} \tag{11}$$

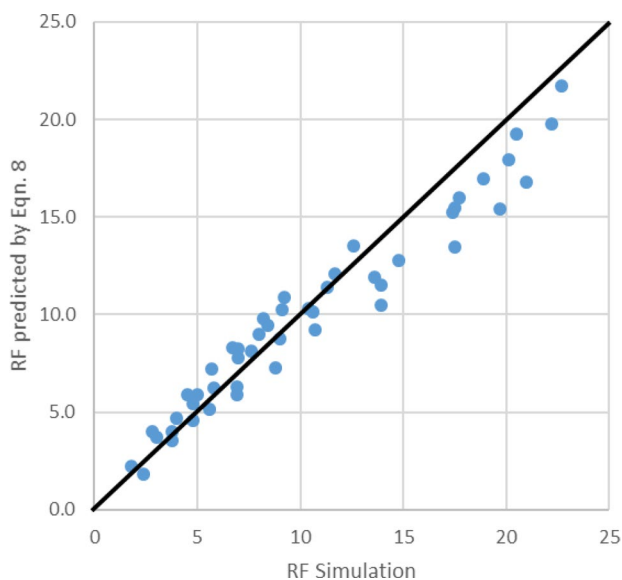


Fig. 15 Comparison between predicted and simulated values of RF_{BT}

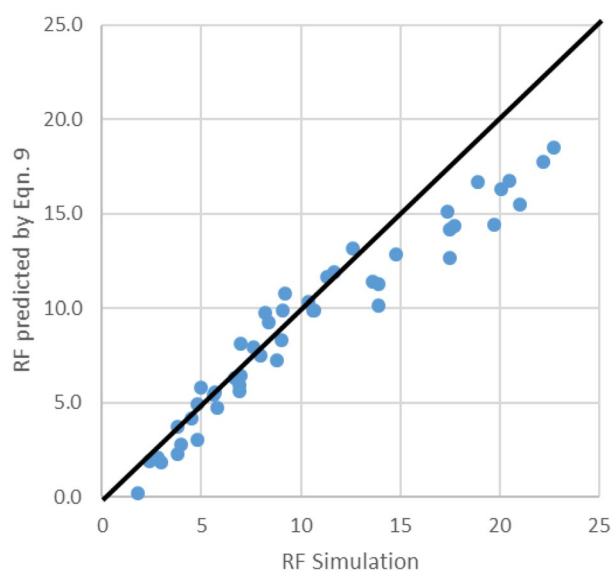


Fig. 17 Comparison between predicted and simulated values of RF_{BT}

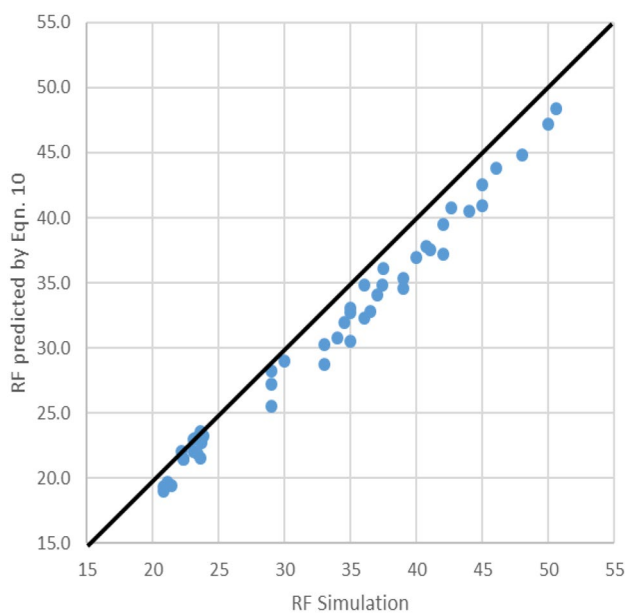


Fig. 16 Comparison between predicted and simulated values of RF_{EOP}

The results of Minitab software in terms of significance of parameters and model summary for RF_{EOP} are illustrated in Table 9 and Fig. 20 of "Appendix".

Validation of the new correlations

Validation of the expanded forms using 144 data points

The accuracy of the proposed correlations developed in the previous section was tested by comparing the values of RF

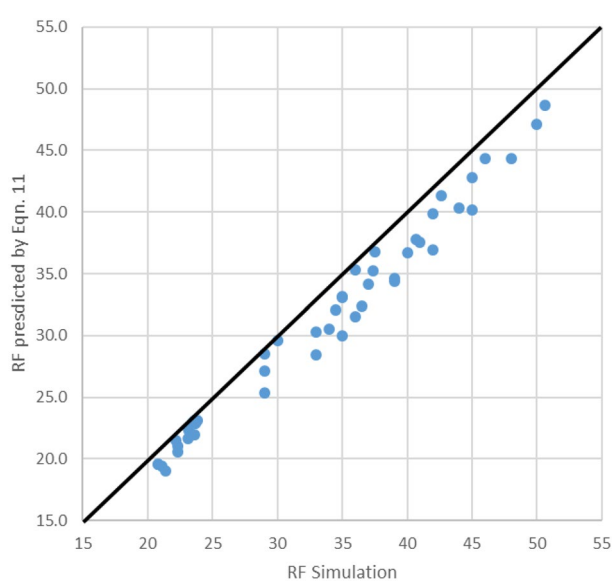


Fig. 18 Comparison between predicted and simulated values of RF_{EOP}

generated by the simulator with those predicted by the new correlations. The results of these comparisons are illustrated in Figs. 11 and 12.

Validation of the reduced forms using 144 data points

Similar comparisons were performed for testing the *reduced forms* expressed in Eqs. 9 and 11 and the results are shown in Figs. 13 and 14. The 45° line in Figs. 11, 12, 13, 14, 15,

Table 5 Results of calculations of AAPCD for various cases investigated in this work

Case	RF _{BT} by Eq. 8	RF _{BT} by Eq. 9	RF _{EOP} by Eq. 10	RF _{EOP} by Eq. 11
144 data points	6.90	8.30	1.02	1.04
48 data points	14.00	16.90	6.50	6.70

Table 6 Field case data (Espinel 2010)

Reservoir name	Field A
Oil viscosity, cp	1.2
Water viscosity, cp	0.9
Corey exponent for oil (n_o)	3.017
Corey exponent for water (n_w)	1.8045
End point-relative permeability to oil (k_{ro_e})	0.96865
End point-relative permeability to water (k_{rwe})	0.551
Residual oil saturation (S_{or})	0.23
Connate water saturation (S_{wc})	0.38
Dykstra–Parson coefficient (V)	0.8
Wettability index: water wet = 1 and oil wet = 2	1.0
Estimated max operational WOR	26.3
Permeability anisotropy ratio (k_z/k_x)	1.0
Injection rate, stb/day	8000

Table 7 Predicted values of RF_{EOP} versus field value

Correlation	Field case	Guthrie–Greenberger method (Eq. 1)	API statistical study (Eq. 2)	New correlation (Eq. 10)	New correlation (Eq. 11)
RF _{EOP}	0.396	0.399	0.472	0.308	0.309
APCD	–	0.758	19.2	22.9	22.7

16, 17, and 18 represents the perfect match location between the simulated and predicted RF values.

Validation of the expanded forms using the remaining 48 data points

The second part of the validation was accomplished by considering the remaining 48 data of the total 192 data points. These 48 data points represent data which were not used in the development of the new correlations. Equation 8 is applied to calculate the RF_{BT} for the 48 data points and the results were compared with the simulation-generated results. The results of this comparison are shown in Fig. 15.

A similar comparison between RF_{EOP} predicted by Eq. 10 and by simulation is illustrated in Fig. 16.

In addition, the reduced forms, Eqs. 9 and 11, have been applied for the 48 data points and the results of the comparison are shown in Figs. 17 and 18.

A summary of the results of calculations of average absolute percent difference (AAPCD) for the individual combination scenarios of RF_{BT} and RF_{EOP} is listed in Table 5.

Validation of the new correlations using field data

The new correlations were also validated using Field A data listed in Table 6 (Espinel 2010). The recovery factor at end of project, RF_{EOP}, for this field has been determined by Visual Basic for Applications (VBA) program and found equal to 0.396. *The value of RF_{BT} is not available for this field and thus, the only possible comparison was between the predicted values of RF_{EOP} by various methods and the above field value.*

The value of RF_{EOP} was estimated by three empirical correlations, namely Guthrie–Greenberger correlation, Eq. 1, API Statistical Study, Eq. 2, and the proposed new correlations, Eqs. 10 and 11. The predicted values of RF_{EOP} were then compared with the field observation and the absolute percent difference for each method was calculated. The results of these calculations are illustrated in Table 7.

Discussion of results

Eclipse simulation results

The relative permeability curves generated by Corey's correlations (Fig. 2) intersect at water saturation of 0.65 which is indicative of a water-wet system. The reservoir is assumed to consist of ten layers which have different permeabilities and that there is a significant variation in the values of permeability of these layers across the reservoir thickness. This variation is illustrated in Table 2 which shows that as the permeability variation coefficient (V) increases from 0.1 to 0.7, the reservoir becomes more heterogeneous. For all selected values of V , the permeabilities of the ten layers were arranged in a descending order which represents a situation known as *coarsening upward* across the reservoir (Dake 2001). Such permeability arrangement scenario would promote gravity effects during the process of waterflooding provided that cross-flow exists between layers. The net effect of gravity and rate of water injection would significantly impact the shape of the advancing displacement front during the process of waterflooding. To a lesser importance is the water–oil end point mobility ratio impact on the development of the shape of the displacement front in flooding heterogeneous reservoirs. These factors, namely rate of water

injection, gravity, layering, cross-flow, and end point mobility ratio, are the key parameters in waterflooding projects. The success of a waterflood project significantly relies on the right combination of these parameters. Accordingly, these key parameters were considered in this work to develop the new correlations and to interpret the simulator results presented in Figs. 3, 4, 5, 6, 7, 8, 9, and 10. For the hypothetical layered reservoir considered in this study, the lateral sweep efficiency has been assumed equal to 1.0.

Validity of proposed correlations (GLM results)

The GLM was used to generate the new correlations for predicting RF_{BT} and RF_{EOP} . The program can also identify the relative importance of the four key parameters which were included in the developed correlations. Using Global Linear Model (GLM), the water injection rate has been found as the most effective parameter and that water viscosity as the least effective parameter as far as the recovery factor is concerned.

The accuracy of the expanded forms of the new proposed correlations (Eq. 8 for predicting RF_{BT} , and Eq. 10 for predicting RF_{EOP}), which include all four key parameters, was tested against: (1) the 144 simulation-generated data points used in their development, (2) against the remaining 48 simulation-generated data points not included in the development of the new correlations, and (3) against real field case data. Similarly, the reduced forms of the new proposed correlations (Eq. 9 for predicting RF_{BT} and Eq. 11 for predicting RF_{EOP}) were tested as described above. The results of comparisons of (1) and (2) are shown in Figs. 11, 12, 15, and 16. From Figs. 11 and 12, one can observe that the nearly perfect match between the predicted values and simulator-generated values of RF_{BT} and RF_{EOP} , respectively. From Figs. 15 and 16, however, the new correlation Eqs. 8 and 10 seem to generally under predict the values of RF_{BT} and RF_{EOP} , respectively. Similar results were observed for Eqs. 9 and 11 as shown in Figs. 13, 14, 17, and 18. A summary of the AAPCD of comparisons of (1) and (2) is shown in Table 5. For both validity cases considered in this study, and as expected, it can be observed that the proposed expanded forms yield more accurate results than the proposed reduced forms. For the 144 data points case, an AAPCD as low as 1.02 has been obtained for predicting RF_{EOP} with Eq. 10.

The reliability of the proposed correlations was further tested using published data of a waterflood project with data listed in Table 6 (Espinel 2010). The value of field RF_{EOP} was compared with RF_{EOP} predicted by three methods, namely Guthrie–Greenberger method, API Statistical Study, and the proposed correlations. The results of these comparisons are presented in Table 7. These results are not indicative of the superiority of any of the methods considered in this study simply because a single field data point has been used. Nevertheless, the high APCD values of the proposed

correlations are comparable to those of the API method. The fact that the permeability variation coefficient for this field data ($V=0.8$) is higher than the maximum value considered in the development of the proposed correlations ($V=0.7$) may explain the relatively high APCD values in Table 7. Another reason for the relatively high APCD obtained by the new proposed correlations can be realized by observing Figs. 16 and 18. In these two figures, the proposed new correlation under predicts the RF_{EOP} and noticeably at higher values of this parameter. Additional field data are necessary to examine the proposed correlations.

Limitations of the proposed correlations

The new empirical correlations developed in this study are based on four key parameters believed to impact the overall performance of waterflood operations. Therefore, the proposed correlations will depend very much on the availability of these key parameters, which puts a limitation on their application. Moreover, the selected four parameters may not be enough to evaluate the effectiveness of the waterflood performance in terms of oil recovery and water cut profiles. Other reservoir characteristics, such as wettability preference, initial free gas saturation, dip angle, and rock consolidation, could very much affect the accuracy of the proposed correlations. The proposed empirical correlations were developed for specific ranges of key parameters as shown in Table 4. Therefore, the application of Eqs. 8 through 11 outside these ranges, and specifically for V and q_i , may not yield accurate RF_{BT} and RF_{EOP} .

Conclusion and recommendation

Conclusion

Based on the results of this study, the following conclusions are drawn:

- Two sets of new empirical correlations have been developed to predict the performance of a five-spot waterflood in a stratified reservoir. These correlations encompass four key parameters believed to significantly affect the oil recovery factors in waterflood operations. These key parameters include water injection rate, water viscosity, permeability anisotropy, and reservoir heterogeneity.
- When tested against 144 simulation-generated data points, the expanded forms of the new correlations have been found to give reliable estimates of RF_{BT} and RF_{EOP} with AAPCD of 6.9 and 1.02, respectively. The reduced forms were found to yield slightly higher AAPCDs for the same data set.

- When tested against 48 simulation-generated data points representing ranges of key parameters outside the ones used in their development, the expanded forms of the new correlations gave fairly good estimates of RF_{BT} and RF_{EOP} with AAPCDs of 14 and 6.5, respectively.
- The new correlations gave more accurate estimates of RF_{EOP} than for RF_{BT} . The highest RF_{EOP} of 50.6% was achieved for a combination scenario with $q_i = 10,000$ bpd, $\mu_w = 1.0$ cp, $k_z/k_x = 1.0$, and $V = 0.1$.
- When tested against two published empirical correlations, using a single field data point, the proposed correlations were found to produce a comparable fit with the API statistical study.
- The new developed correlations can be used to get quick and reliable estimates of RF_{BT} and RF_{EOP} .
- The water injection rate has been found as the most effective parameter in predicting the oil recovery factor, followed by the Dykstra–Parson coefficient, followed by the k -anisotropy. The water viscosity has been found to be the least effective parameter as far as the recovery factor is concerned.

Recommended measures to improve the accuracy of the proposed correlations

To improve the accuracy and reliability of the proposed correlations, it is recommended to:

- Include the effects of free gas saturation, angle of dip, and wettability preference indicator.
- Benchmark with other analytical methods and simulation results using more field data.
- Increase the total number of simulation-generated data points and consider additional combination scenarios.

Appendix

Results of GLM for RF_{BT}

See Table 8 and Fig. 19.

Table 8 Significance of parameters

Source	DF	Seq. SS	% Contribution	Adj. SS	Adj. MS	F-value	P value
<i>Analysis of variance</i>							
k_z/k_x	1	873.89	22.27	93.033	93.033	103.32	0.000
μ_w	1	953.41	24.30	31.514	31.514	35.00	0.000
V	1	18.56	0.470	56.650	56.650	62.91	0.000
q_i	1	1002.93	25.56	234.966	234.966	260.95	0.000
$(k_z/k_x)^2$	1	209.58	5.340	50.857	50.857	56.48	0.000
$(\mu_w)^2$	1	21.13	0.540	21.134	21.134	23.47	0.000
$(V)^2$	1	26.93	0.690	18.829	18.829	20.91	0.000
$(q_i)^2$	1	111.67	2.850	151.254	151.254	167.98	0.000
$(k_z/k_x)(\mu_w)$	1	38.96	0.990	17.161	17.161	19.06	0.000
$(k_z/k_x)(V)$	1	79.40	2.020	79.405	79.405	88.19	0.000
$(\mu_w)(V)$	1	10.73	0.27	10.726	10.726	11.91	0.001
$(V)(q_i)$	1	205.46	5.24	38.299	38.299	42.53	0.000
$(k_z/k_x)^3$	1	26.17	0.67	26.175	26.175	29.07	0.000
$(k_z/k_x)^2(\mu_w)$	1	4.38	0.11	4.379	4.379	4.86	0.029
$(k_z/k_x)^2(q_i)$	1	1.25	0.03	156.338	156.338	173.63	0.000
$(k_z/k_x)(q_i)(\mu_w)$	1	40.63	1.04	14.626	14.626	16.24	0.000
$(k_z/k_x)(q_i)^2$	1	124.82	3.18	124.816	124.816	138.62	0.000
$(V)^2(q_i)$	1	4.00	0.10	3.999	3.999	4.44	0.037
$(V)(q_i)^2$	1	15.01	0.38	15.011	15.011	16.67	0.000
Error	172	154.87	3.95	154.874	0.900		
Total	191	3923.78	100				
S	R-sq	R-sq(adj)	PRESS	R-sq(pred)	AIC	BIC	
Model summary	0.948911	96.05%	95.62%	193.604	95.07%	551.05	614.02

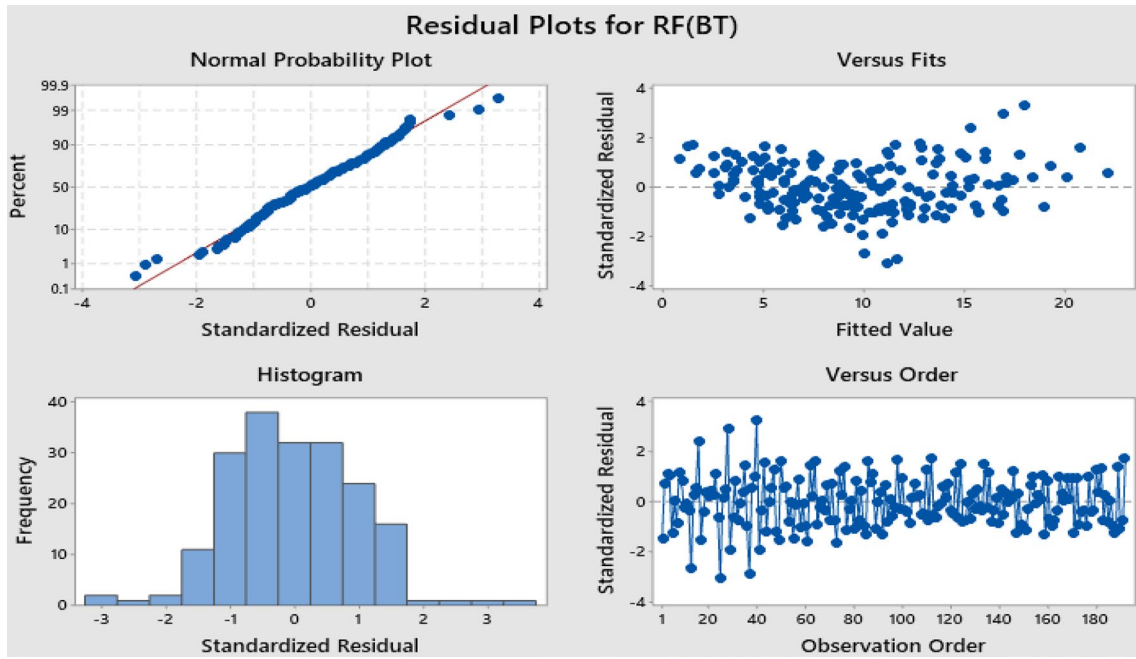


Fig. 19 Residual plots for RF_{BT}

Results of GLM for RF_{EOP}

See Table 9 and Fig. 20.

Table 9 Significance of parameters

Source	DF	Seq. SS	% Contribution	Adj. SS	Adj. MS	F-value	P value
<i>Analysis of variance</i>							
k_z/k_x	1	685.40	6.29	10.355	10.3545	39.19	0.000
μ_w	1	246.30	2.26	18.657	18.6574	70.61	0.000
V	1	1530.9	14.05	23.845	23.8451	90.24	0.000
q_i	1	6682.7	61.33	62.016	62.014	234.69	0.000
$(k_z/k_x)^2$	1	99.80	0.92	11.539	11.5392	43.67	0.000
$(V)^2$	1	0.20	0.00	17.623	17.6227	66.69	0.000
$(q_i)^2$	1	704.90	6.47	26.775	26.7753	101.33	0.000
$(k_z/k_x)(\mu_w)$	1	15.70	0.14	8.453	8.4533	31.99	0.000
$(k_z/k_x)(V)$	1	72.80	0.67	12.221	12.2206	46.25	0.000
$(k_z/k_x)(q_i)$	1	205.90	1.89	57.042	57.0419	215.87	0.000
$(\mu_w)(V)$	1	0.30	0.00	21.216	21.2161	80.29	0.000
$(\mu_w)(q_i)$	1	0.50	0.00	22.096	22.0959	83.62	0.000
$(k_z/k_x)^3$	1	5.90	0.05	5.891	5.891	22.29	0.000
$(V)^3$	1	11.90	0.11	11.948	11.9483	45.22	0.000
$(k_z/k_x)^2(V)$	1	12.40	0.11	12.421	12.4214	47.01	0.000
$(k_z/k_x)^2(q_i)$	1	3.80	0.04	3.847	3.8472	14.56	0.000
$(k_z/k_x)(\mu_w)^2$	1	20.70	0.19	4.437	4.4372	16.79	0.000
$(k_z/k_x)(V)^2$	1	1.70	0.02	1.705	1.7054	6.45	0.012
$(k_z/k_x)(V)(q_i)$	1	387.60	3.56	16.846	16.8457	63.75	0.000
$(k_z/k_x)(q_i)^2$	1	21.60	0.20	21.571	21.5708	81.63	0.000
$(\mu_w)^2(q_i)$	1	2.50	0.02	2.472	2.4724	9.36	0.003
$(\mu_w)(V)^2$	1	8.20	0.08	8.196	8.1955	31.01	0.000
$(\mu_w)(V)(q_i)$	1	113.8	1.04	37.884	37.884	143.37	0.000
$(\mu_w)(q_i)^2$	1	8.40	0.08	8.428	8.4279	31.89	0.000
$(V)^2(q_i)$	1	3.70	0.03	7.837	7.8372	29.66	0.000
$(V)(q_i)^2$	1	4.20	0.04	4.165	4.1649	15.76	0.000
Error	165	43.60	0.40	43.601	0.2642		
Total	191	10,895.6	100				
S	R -sq	R -sq(adj)	PRESS	R -sq(pred)	AIC	BIC	
<i>Model summary</i>							
0.514049	99.60%	99.54%	58.6762	99.46%	326.21	407.46	

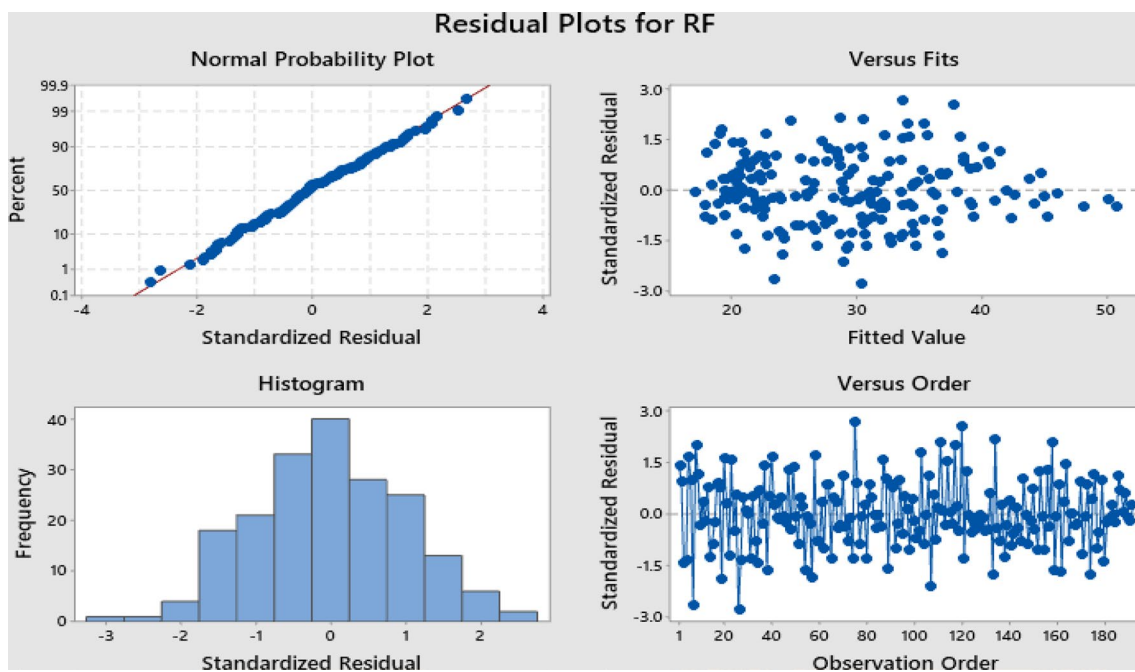


Fig. 20 Residual plots for RF_{EOP}

Funding No funding.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahmed T (2002) Reservoir engineering handbook, 2nd edn. Gulf Professional Publishing, Houston, TX
- Al-Jifri MK (2020) Developing a new formula for predicting oil recovery factor in water flooded-heterogeneous reservoirs. Master's thesis, United Arab Emirates University, Department of Chemical and Petroleum Engineering, Al-Ain, UAE
- Aliyuda K, Howell J (2019) Machine-learning algorithm for estimating oil recovery factor using a combination of engineering and stratigraphic dependent parameters. Interpretation 7:SE151–SE159. <https://doi.org/10.1190/INT-2018-0211.1>
- Arps JJ, Brons F, van Everdingen AF, Buchwald RW, Smith AE (1967) A statistical study of recovery efficiency. API Bull 14D:1–37
- Balhasan S, Jumaa M (2017) Development of a correlation to predict water-flooding performance of sandstone reservoirs based on reservoir fluid properties. Int J Appl Eng Res 12(10):2586–2597 (ISSN 0973-4562)
- Dake LP (2001) The practice of reservoir engineering, Revised. Elsevier, Amsterdam
- Espinel AL (2010) Generalized correlation to estimate oil recovery and pore volumes injected in waterflooding projects. Ph.D. dissertation, Texas A&M University, College Station, Texas
- Gulstad RL (1995) The determination of hydrocarbon reservoir recovery factors by using modern multiple linear regression techniques. Master's thesis, Texas A&M University, College Station, TX, USA
- Guthrie RK, Greenberger MH (1955) The use of multiple-correlation analyses for interpreting petroleum engineering data. Drill Prod Pract API 130–137:1955
- Han B, Bian X (2018) A hybrid pso-svm-based model for determination of oil recovery factor in the low-permeability reservoir. Petroleum 4:43–49
- Khan A (1971) An empirical approach to waterflood predictions. J Pet Technol 23(05):565–573. <https://doi.org/10.2118/2931-pa>
- Li H, Yu H, Cao N, Tian H, Cheng S (2020) Applications of artificial intelligence in oil and gas development. Arch Comput Methods Eng. <https://doi.org/10.1007/s11831-020-09402-8>
- Mahmoud AA, Elkatatny S, Chen W, Abdurraheem A (2019) Estimation of oil recovery factor for water drive sandy reservoirs through applications of artificial intelligence. Energies 12(19):3671. <https://doi.org/10.3390/en12193671>
- Makhotin I, Orlov D, Koroteev D, Burnaev E, Karapetyan A, Antonenko D (2020) Machine learning for recovery factor estimation of an oil reservoir: a tool for de-risking at a hydrocarbon asset evaluation. J Pet Sci Eng (Submitted)
- Sharma A, Srinivasan S, Lake L, et al (2010) Classification of oil and gas reservoirs based on recovery factor: a data-mining approach. In: SPE annual technical conference and exhibition, society of petroleum engineers. <https://doi.org/https://doi.org/10.2118/130257-MS>

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