

Article

Estimation of Oil Recovery Factor for Water Drive Sandy Reservoirs through Applications of Artificial Intelligence

Ahmed Abdulhamid Mahmoud , Salaheldin Elkatatny * , Weiqing Chen 
and Abdulazeez Abdulraheem 

College of Petroleum Engineering and Geosciences, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; eng.ahmedmahmoud06@gmail.com (A.A.M.); jerromy@kfupm.edu.sa (W.C.); toazeez@gmail.com (A.A.)

* Correspondence: elkatatny@kfupm.edu.sa; Tel.: +966-594-663-692

Received: 20 August 2019; Accepted: 23 September 2019; Published: 25 September 2019



Abstract: Hydrocarbon reserve evaluation is the major concern for all oil and gas operating companies. Nowadays, the estimation of oil recovery factor (RF) could be achieved through several techniques. The accuracy of these techniques depends on data availability, which is strongly dependent on the reservoir age. In this study, 10 parameters accessible in the early reservoir life are considered for RF estimation using four artificial intelligence (AI) techniques. These parameters are the net pay (effective reservoir thickness), stock-tank oil initially in place, original reservoir pressure, asset area (reservoir area), porosity, Lorenz coefficient, effective permeability, API gravity, oil viscosity, and initial water saturation. The AI techniques used are the artificial neural networks (ANNs), radial basis neuron networks, adaptive neuro-fuzzy inference system with subtractive clustering, and support vector machines. AI models were trained using data collected from 130 water drive sandstone reservoirs; then, an empirical correlation for RF estimation was developed based on the trained ANN model's weights and biases. Data collected from another 38 reservoirs were used to test the predictability of the suggested AI models and the ANNs-based correlation; then, performance of the ANNs-based correlation was compared with three of the currently available empirical equations for RF estimation. The developed ANNs-based equation outperformed the available equations in terms of all the measures of error evaluation considered in this study, and also has the highest coefficient of determination of 0.94 compared to only 0.55 obtained from Gulstad correlation, which is one of the most accurate correlations currently available.

Keywords: hydrocarbon reserve estimation; oil recovery factor; water drive sandy reservoirs; artificial intelligence

1. Introduction

The petroleum industry is characterized by the need to make critical investment decisions under several uncertainties. Different techniques are currently applied to diminish these uncertainties in key areas such as reserve estimation, data management, and/or reservoir characterization.

Oil recovery factor (RF) is the most significant parameter for all exploration and development (E&P) companies mainly during the early reservoir life, because several investment decisions are based on the amount of hydrocarbon, which could be obtained from the target asset with the available techniques and operational practices [1].

The fact that RF is affected by several engineering and geological aspects makes the estimation of the RF very complicated, since no clear approach that considers all these aspects is available. Understanding

all the technical and non-technical parameters associated with the reservoir nature, technologies in use, economic conditions, and other factors is necessary for the reserve evaluation process.

Currently, there are mainly six available techniques for oil reserve estimation. (1) Analogy is based on comparing the geological and petrophysical properties of poorly defined or newly discovered reservoir to old ones, and setting an oil recovery factor range for the new asset based on those of the similar assets [2]. (2) Volumetric calculations [3] calculate the stock-tank oil initially in place (STOIIP) first based on the asset dimensions, fluid properties, and rock parameters by assuming the reservoir is sealed; then, based on the recovery mechanism of the reservoir, the reserve could be estimated. (3) Material balance calculations [3–6] require oil, water, and gas production data, as well as data related to water encroachment from the reservoir. (4) The application of decline curve analysis [5,7] also requires production data. (5) Numerical reservoir simulation combines both material balance equations and fluid flow equations to estimate hydrocarbon reserve [3,4]. (6) Lastly, several empirical correlations are currently available, and the accuracy of these correlations depends mainly on data availability [8,9]. The first two techniques are applicable early in the reservoir life, but they are not accurate; the accuracy of the recovery factor prediction could be increased by including production data into calculations and applying one of the last three techniques.

Data availability, which is strongly dependent on the reservoir age, is significantly affecting the accuracy of the RF estimation techniques. The highly accurate techniques require huge production data which restricts their applicability to the late reservoir life. On the other hand, the techniques applied during early reservoir life are not highly accurate.

2. Background of Empirical Correlations Used for Recovery Factor Estimation

In 1945, The American Petroleum Institute (API) initiated a data collection process aiming to correlate the recovery factor with reservoir rock parameters and the properties of the produced fluid. Then, an investigation was conducted by a special study committee on well spacing. They examined data from 103 oil reservoirs, 25% of which are depletion-drive reservoirs, and the remaining are water-drive reservoirs from sandstone, limestone, and dolomite formations.

Craze and Buckley [10] listed 19 parameters that have a major effect on recovery efficiency for 103 reservoirs. They reported that the rock properties, fluid properties, mode of production, drive mechanism, and structural aspects are highly affecting the oil recovery factor. Ten years later, Guthrie and Greenberger [8] suggested an empirical correlation (Equation (1)) to calculate the oil RF from a water-drive reservoir using five factors that affect the oil recovery in sandstone reservoirs:

$$R_o = 0.114 + 0.272 \log(k) + 0.256(S_w) - 0.136 \log(\mu_o) - 1.538(\phi) - 0.00035(h) \quad (1)$$

where R_o is the oil recovery factor (fraction), k represents the reservoir permeability (mD), S_w is the water saturation (fraction), μ_o is the oil viscosity (cp), ϕ represents the reservoir porosity (fraction), and h denotes the reservoir thickness (ft).

Muskat and Taylor [11] studied the effect of the rock characteristics and the reservoir fluid on oil production from gas-drive reservoirs. They reported that the increase in oil viscosity significantly decreased the ultimate oil recovery, while Arps and Roberts [12] found that the ultimate recovery increases with oil gravity, except for the higher solution gas–oil ratios.

Between 1956–1984, API published many correlations for RF calculation based on real performance data from producing fields rather than on theoretical or laboratory data. Equation (2) was suggested by the API for evaluating the oil RF from water-drive reservoirs:

$$R_o = 4259 \left[\frac{\phi[1 - S_w]}{B_{oi}} \right]^{1.0422} \left[\frac{k \mu_{wi}}{\mu_{oi}} \right]^{0.0770} [S_w]^{-0.1903} \left[\frac{p_i}{p_a} \right]^{-0.2159} \quad (2)$$

where B_{oi} and B_{oa} represent the oil formation volume factor at the original and abandonment reservoir pressures (STB/bbl), respectively, μ_{oi} is the original oil viscosity (cp), μ_{wi} denotes the water viscosity

at reservoir pressure (cp), and p_i and p_a are the original and abandonment reservoir pressures (psi), respectively.

Gulstad [9] used multiple linear regression techniques to study the determination of the oil RF. Out of his work, he developed RF models for water-drive and solution gas-drive reservoirs in both sandstone and carbonate formations. He observed that the STOIP is strongly correlated to the RF in both water and solution gas-drive reservoirs. Although the author pointed out that the heterogeneity is an important factor to be considered when developing the RF model, he did not include the heterogeneity, claiming that there is no specific parameter that could be used to clarify it. The Gulstad [9] model for water-drive sandstone reservoirs is shown in Equation (3) below:

$$R_o = -274.94 + 0.44(STOIP) - 56.70 \ln(\mu_{ob}) - 119.45 \ln(S_w) + 0.04(p_{ep}) - 4.73(\mu_{oi}) + 4.38(\mu_{oa}) + 0.24(STOIP_{calc}) - 0.88(T) \quad (3)$$

where $STOIP$ is the stock-tank oil initially in place at the original reservoir pressure as reported by the operator (STB/NAF), μ_{ob} is the oil viscosity at the bubble point pressure (cp), P_{ep} denotes the pressure at the end of the primary recovery (psig), μ_{oi} and μ_{oa} denote the viscosity of the oil at the original and abandonment reservoir pressures (cp), T represents the reservoir temperature (°F), and $STOIP_{calc}$ is the calculated value of the STOIP at original reservoir pressure, assuming a volumetric reservoir (STB/NAF).

3. Applications of Artificial Intelligence in the Petroleum Industry

Since the early 1990s, artificial intelligence (AI) techniques have had many applications in several scientific and engineering fields, including the petroleum industry. Currently, AI has been being used by petroleum engineers and geologists to solve problems related to unconventional resources evaluation [13,14], predicting the bubble point pressure [15], real-time estimation of the drilling fluids rheological parameters [16,17], estimating rock mechanical parameters [18,19], reservoir characterization [20–22], optimizing the rate of penetration [23], evaluating the wellbore casing integrity [24,25], drilling hydraulic optimization [26], pore pressure and fracture pressure estimation [27,28], and others.

Adrian and Chukwueke [29] applied the artificial neural networks (ANNs) to predict the oil RF for water-drive Niger Delta reservoirs. The authors used data from 94 reservoirs in the Niger Delta in this study. They divided the data into three groups: 60% of the data was used to train the ANNs model, 20% was used to validate the model, and the remaining 20% was used to test the trained model. They used the backpropagation network to build the model, and the porosity, permeability, reservoir original and abandonment pressures, oil formation volume factor, oil viscosity, connate water saturation, and connate water viscosity as input parameters to predict the oil RF. Although this model was able to estimate the RF more accurately compared to the available correlations, the authors were not able to extract an empirical equation out of it, which restricts the use of this model by others.

Noureldeen and El-Banbi [1] generated two ANN models for RF estimation. The first model (simple model) utilizes readily available data of net pay, STOIP, the original reservoir pressure, asset area, porosity, Lorenz coefficient, effective permeability, API gravity, oil viscosity, and initial water saturation. This simple model predicted the RF with an absolute average percentage error (AAPE) of 9.5%. The second model (sophisticated model) utilized additional operational and technological parameters. This model has a prediction accuracy of 8.0% for the testing dataset, but since it requires the availability of operational and technological parameters, its application early in the reservoir life is restricted.

Nolemhemhen et al. [30] came up with three models to predict the oil RF for water drive, solution gas drive, and secondary recovery with water injection in the Niger delta. The authors used the data from 136 reservoirs to develop their models. They pointed out that no correlation existed between the porosity, permeability, reservoir thickness, water viscosity, initial water saturation, temperature,

and the RF. Hence, the authors did not include any information related to the reservoir formation and formation-water properties, which are believed to affect the accuracy of these models drastically when applied on different environments.

In this study, four AI techniques—the ANNs, radial basis neuron networks (RNNs), adaptive neuro-fuzzy inference system with subtractive clustering (ANFIS-SC), and support vector machines (SVM) were used to estimate oil RF based on 10 parameters (net pay, STOIP, the original reservoir pressure, asset area, porosity, Lorenz coefficient, effective permeability, API gravity, oil viscosity, and initial water saturation), which are readily available for all assets at their early stages; the use of these parameters was suggested recently by Noureldien and El-Banbi [1] for RF prediction using ANNs. Noureldien and El-Banbi [1] did not extract an empirical correlation from their model, while in this study, the extracted weights and biases of the optimized ANNs were used to develop an empirical equation that could be easily programmed and used for RF estimation. The predictability of the developed empirical equation will be compared with three available empirical correlations from the works of literature.

4. Materials and Methods

A dataset of 173 lessons was collected from literature for this study [1,9]. The datasets were analyzed statistically, and outliers were removed based on the standard deviation (SD) where any data point out of the range of ± 0.3 SD was considered as an outlier. Five lessons were removed from the data based on the SD criteria. Then, the remaining datasets (from 168 lessons) were used to develop the AI models. These models were trained using 77% of the data, and the remaining (23%) were used to test the trained models. The parameters used to generate the AI models (10 parameters) are explained in Table 1.

Table 1. The parameters used to generate the artificial intelligence (AI) models. API: American Petroleum Institute.

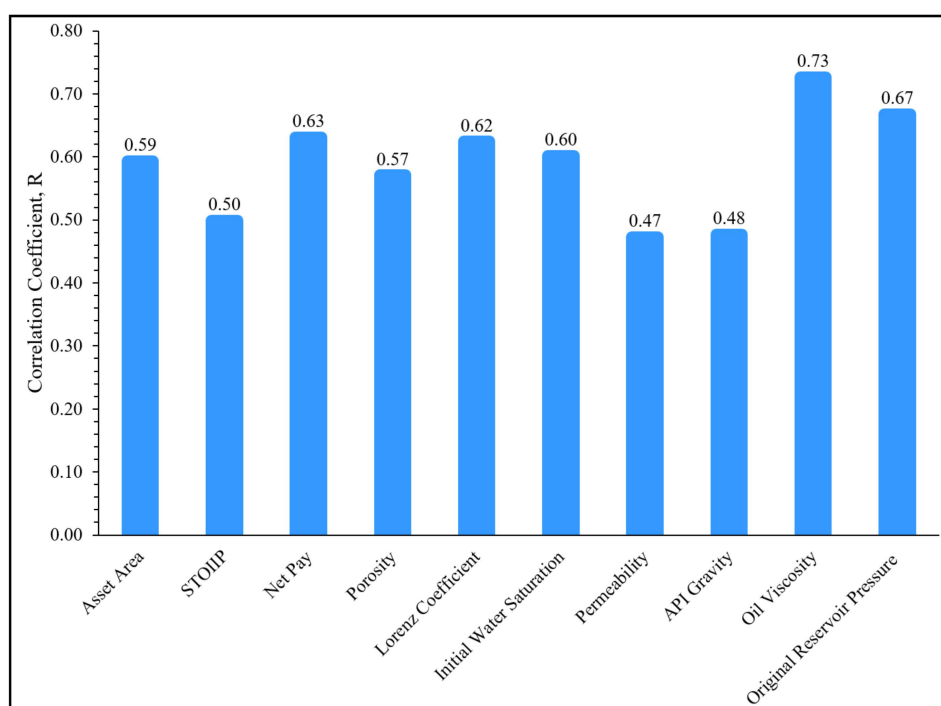
| Group | Parameter | Definition |
|------------------|---------------------------------------|--|
| Asset Size | Asset area STOIP | Asset size in terms of its areal extent and reservoir size. The estimated value of the stock-tank oil initially in place. |
| Rock Parameters | Net pay thick (h) | The net thickness of oil-saturated sand within the entire reservoir. |
| | Porosity (ϕ) | The pore volume relative to the total bulk rock volume of the rock. |
| | Lorenz coeff. | Represents the vertical heterogeneity of the reservoir. |
| | Initial water saturation (S_{wi}) | Value of initial water saturation. |
| Fluid Properties | Permeability (k) | Absolute permeability from core analysis. |
| | API | API gravity from PVT. |
| | Oil viscosity (μ_o) | Measured or calculated oil viscosity. |
| Reservoir Energy | Reservoir pressure (p) | Reservoir pressure, referenced at 10,000 ft TVD. |

These parameters could be divided into four groups (asset size, rock parameters, fluid properties, and reservoir energy). Table 2 summarizes the statistical description of the data (130 reservoirs) used to train the AI models. It shows the ranges of training data: asset area from 446 to 15,515 acres, STOIP from 5.0 to 1072.5 MMSTB, porosity from 0.12 to 0.32, connate water saturation from 0.16 to 0.31, permeability from 15 to 1270 mD, API gravity from 23.0 to 42.2° API, oil viscosity at reservoir conditions from 0.16 to 2.59 cp. These ranges represent the applicable ranges for the developed models. Later on, the testing data must fall within the same ranges as the training data to predict the RF with acceptable accuracy. Figure 1 compares the relative importance of the parameters used in this study to train the AI models developed to estimate the RF; as shown in this figure, all parameters have a moderate to high correlation coefficient with the oil RF.

Table 2. The statistical description of the data used to train the AI models developed in this study.

| Input Parameters | Min | Max | Mean | Mode | Range | SD | Coef. of Variation | R ² |
|-------------------------------|------|--------|--------|------|--------|--------|--------------------|----------------|
| Asset Area (acres) | 446 | 15515 | 4787.5 | 3970 | 15069 | 3094 | 0.646 | 0.352 |
| STOIIP (MMSTB) | 5.00 | 1072.5 | 326.6 | 88 | 1067.5 | 279.1 | 0.854 | 0.249 |
| Net Pay "h" (ft) | 12.9 | 471.45 | 152.9 | 25.7 | 458.6 | 107.8 | 0.705 | 0.399 |
| Porosity (fraction) | 0.12 | 0.32 | 0.234 | 0.23 | 0.2 | 0.04 | 0.173 | 0.326 |
| Lorenz Coefficient (fraction) | 0.15 | 0.77 | 0.442 | 0.24 | 0.62 | 0.118 | 0.268 | −0.390 |
| S_{wi} (fraction) | 0.16 | 0.31 | 0.225 | 0.2 | 0.15 | 0.036 | 0.16 | −0.362 |
| Permeability "k" (md) | 15.0 | 1270 | 450 | 1000 | 1255 | 318.3 | 0.707 | −0.224 |
| API (degree) | 23.0 | 42.20 | 34.50 | 33.0 | 19.2 | 4.446 | 0.129 | 0.228 |
| Oil Viscosity (cp) | 0.16 | 2.59 | 0.874 | 0.6 | 2.43 | 0.548 | 0.627 | 0.528 |
| P "@ 10,000 TVD" (psi) | 1672 | 11470 | 5833.7 | 7000 | 9798 | 2286.4 | 0.392 | 0.446 |

R² is the coefficient of determination.

**Figure 1.** The relative importance of the parameters used to train the AI models developed in this study to estimate the recovery factor (RF).

The first technique used in this study is the backpropagation ANNs. The suggested model is selected based on the lowest AAPE and highest R² after testing different combinations of the ANNs model design parameters, such as the number of hidden layers, the number of neurons per layer, the training functions, the transferring functions, and the number of iterations. Inserted for loops were constructed to test the predictability of the ANNs model using different combinations of these design parameters, where every design factor was represented by one loop. The number of the hidden layers was optimized in the range from one to three layers. The effect of the number of neurons in each hidden layer was tested in the range of 5 to 25 neurons. The effect of different training functions (trainlm, traingdm, traingcf, trainbr, and traingda) and two transfer functions—namely, the tan-sigmoid function and pure line function—were also evaluated. An empirical correlation will be developed based on the extracted weights and biases of the ANN optimized model.

The second model used in this study is the RNNs model, which is optimized on the design parameters of the mean squared error goal (Goal), the maximum number of neurons (MN), the spread of radial basis functions (Spread), and the number of neurons to add between displays (DF). The optimization process was conducted in the same way as that followed earlier to optimize

the ANNs model. ANFIS-SC was also used to obtain the recovery factor, which is optimized for the radius of the cluster. The last technique considered in this study is the SVM, which is optimized for the kernel type, kernel option, epsilon value, lambda, and C.

5. Results and Discussion

The use of one hidden layer with five neurons, the trainlm (Levenberg–Marquardt) function to train the model, and one output layer with a tan-sigmoid transfer function was found to give the optimum predictability of the suggested model with R^2 and AAPE values of 0.95 and 5.80%, respectively, based on the training dataset. Then, the trained model was used to develop the empirical correlation in Equation (4), which predicted the oil RF for the testing data with R^2 and AAPE values of 0.94 and 7.92%, respectively. Figure 2 is a cross-plot that compares the actual and predicted RF for training and testing datasets using ANNs. For the testing set, the predictability of the ANNs-based equation over-performed all other AI models in term of AAPE and R^2 of the tested data. Equation (4) was developed on the same base as that followed by Mahmoud et al. [13].

$$RF = \left[\sum_{i=1}^N w_{2-i} \tanh \left(\sum_{j=1}^J w_{1-i,j} Y_j + b_{1i} \right) \right] + b_2 \quad (4)$$

where RF is the recovery factor (dimensionless), N represents the total neurons in the hidden layer (dimensionless), J is the total number of input parameters (dimensionless, in this case, 10 inputs are used as summarized in Table 1), w_1 is the hidden layer weights (dimensionless), b_1 is the hidden layer bias (dimensionless), w_2 denotes the output layer weights (dimensionless), b_2 represents the output layer weights (dimensionless, the extracted weights and biases of the hidden and output layers are summarized in Table 3), Y represents the normalized input parameters (dimensionless). The use of an equation form (such as that in Equation (4)) for determining the desired output (the RF in this case) is explained before by many authors e.g., Mahmoud et al. [13].

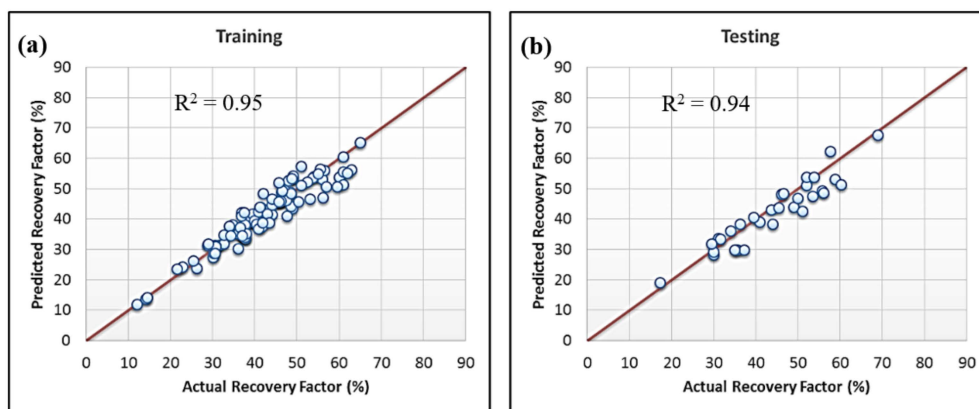


Figure 2. Predicted vs. actual RF based on the artificial neural network (ANN) model (a) training dataset, (b) testing dataset.

The second model used in this study is the RNN model. As a result, the use of a goal of zero, spread of 3.0, MN of 16, and DF of 4.0 was found to give the lowest AAPE of 6.86% and the highest R^2 0.95 based on the same training data. The R^2 of 0.95 as per RNNs is higher than that of the ANN model. However, when the RNN model is applied to the testing dataset, it gave the lowest R^2 among all the techniques used ($R^2 = 0.88$). The AAPE for testing data is 8.78%, which is also higher than that obtained by the ANN model. The cross-plots of the actual and predicted RF using the RNN model on both the training and testing datasets are shown in Figure 3.

Table 3. The proposed ANN-based weights and biases for RF calculations with Equation (4).

| | | Input Layer | | | | | | | | | | Output Layer | | |
|-------------------|---------|-------------------|---------|---------|---------|---------|---------|---------|---------|---------|----------|------------------|-------------------|----------------|
| | | Weights (w_1) | | | | | | | | | | Biases (b_1) | Weights (w_2) | Bias (b_2) |
| | | $j = 1$ | $j = 2$ | $j = 3$ | $j = 4$ | $j = 5$ | $j = 6$ | $j = 7$ | $j = 8$ | $j = 9$ | $j = 10$ | | | |
| No. of Neurons | $i = 1$ | 42.15 | 3.82 | 11.27 | −57.13 | −8.42 | −23.59 | −10.34 | 27.35 | 50.09 | 40.05 | −2.06 | 0.07 | |
| | $i = 2$ | 40.32 | 4.05 | −7.60 | 8.91 | −5.01 | 9.13 | 11.68 | −14.11 | 16.53 | −21.10 | −1.48 | 0.18 | |
| | $i = 3$ | −11.13 | 10.73 | −10.68 | −23.31 | −1.18 | 23.48 | 32.23 | 6.22 | −41.03 | −19.40 | −2.27 | 0.29 | −0.49 |
| | $i = 4$ | −0.29 | 1.14 | −0.13 | 0.07 | −0.59 | −1.05 | 0.07 | 0.11 | 1.34 | 0.68 | 0.39 | 0.59 | |
| | $i = 5$ | 2.34 | 0.54 | 0.83 | 0.14 | 0.92 | −4.49 | 0.12 | −9.08 | 4.76 | 0.55 | 2.60 | −0.38 | |

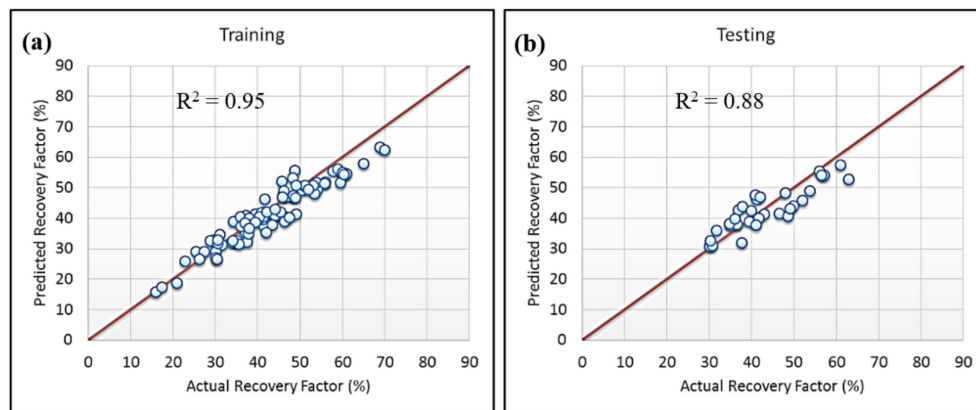


Figure 3. Predicted vs. actual recovery factor based on the radial basis neuron network (RNN) model (a) training dataset, (b) testing dataset.

The suggested ANFIS-SC model was optimized based on the cluster radius value. The results show that the ANFIS-SC model with the cluster radius of 0.7 has the lowest AAPE for a training dataset of 4.83% and a very high R^2 of 0.98. The tested dataset with this model also exhibits a good R^2 with actual oil recovery ($R^2 = 0.91$) and relatively low AAPE of 8.53%, as shown in Figure 4.

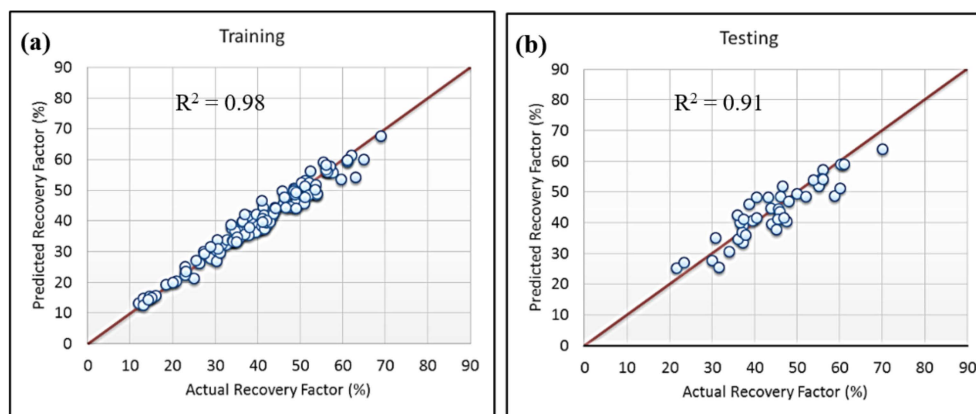


Figure 4. Predicted vs. actual recovery factor based on adaptive neuro-fuzzy inference system with subtractive clustering (ANFIS-SC) model (a) training dataset, (b) testing dataset.

The highest R^2 for the training dataset was achieved by the SVM model ($R^2 = 0.99$), as shown in Figure 5, with a relatively low AAPE of 5.11% for the optimum SVM, which has the following design parameters (lambda is 10^{-8} , epsilon is 2.0, the kernel is Gaussian with the kernel option of 1.7, verbose of zero, and the C is 2500). This design parameters combination was selected using the same way followed earlier to optimize the ANN model. Applying the optimized SVM model on the testing dataset showed the highest AAPE value of 10.44% and relatively low R^2 value of 0.90. Figure 5 is a cross-plot that compares the actual and predicted RF for the training and testing datasets using the optimized SVM.

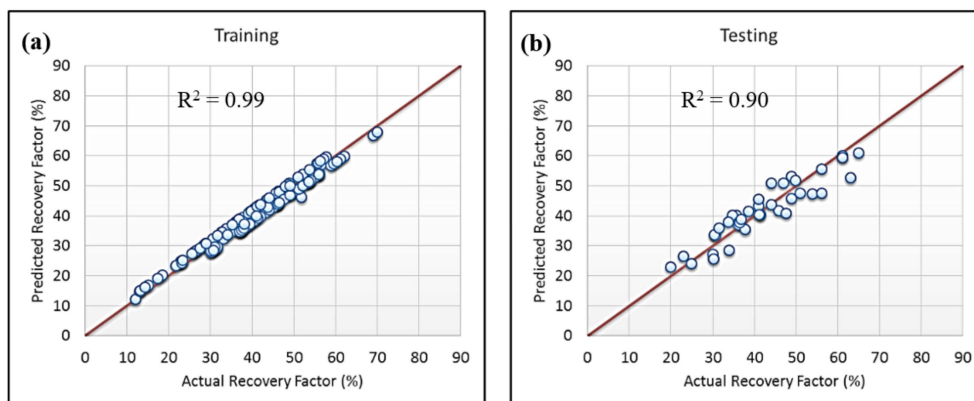


Figure 5. Predicted vs. actual recovery factor based on the support vector machine (SVM) model (a) training dataset, (b) testing dataset.

Table 4 summarizes the values of R^2 and AAPE for the training and testing datasets between the actual and predicted recovery factors estimated through the four AI techniques used in this work. This table indicates that the ANN-based correlation predicted the RF for the testing dataset with the highest R^2 and lowest AAPE compared to all the other AI models.

Table 4. Coefficient of determinations between actual and predicted RF and the absolute average percentage error (AAPE) in predicting the RF by different AI techniques.

| | Training | | Testing | |
|----------|----------|----------|---------|----------|
| | R^2 | AAPE (%) | R^2 | AAPE (%) |
| ANNs | 0.95 | 5.80 | 0.94 | 7.92 |
| RNNs | 0.95 | 6.86 | 0.88 | 8.78 |
| ANFIS-SC | 0.98 | 4.83 | 0.91 | 8.53 |
| SVM | 0.99 | 5.11 | 0.90 | 10.44 |

The performance of three available empirical correlations for RF estimation from water-drive sandstone reservoirs was compared with the prediction capability of the suggested ANN-based correlation, as shown in Figure 6. The ANN-based correlation over-performed all the other empirical equations in terms of the coefficient of determination with R^2 values of 0.95, 0.40, -0.18 , and 0.55 between the actual and estimated RF predicted using the ANN-based correlation, Guthrie and Greenberger correlation, API correlation, and Gulstad correlation, respectively. Figure 6 also compares the results of the deviation in RF values estimated using the ANNs correlation and the other three empirical correlations considered in this work from actual ones for the testing dataset. It indicates that the estimated RF by ANNs model has a lower deviation compared to all the other correlations studied, with a deviation of between -10% to $+20\%$ maximum. The low deviation of the RF estimated using the ANN-based correlation is attributed to the high accuracy of this equation, which is trained using 130 data points.

Figure 7 compares the error in the estimated RF from the ANNs-based equation with those of empirical correlations through different error measures (AAPE, RMSE, and R^2). All the measures used to quantify the errors indicate that the ANNs-based correlation has the lowest error and the highest correlation with the real RF. Appendix A summarizes the relationships used to calculate the errors.

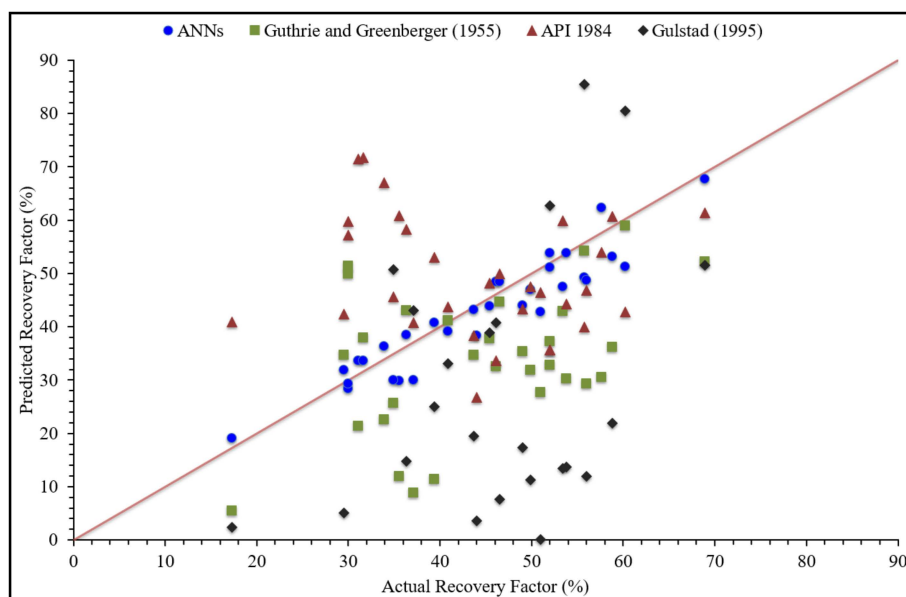


Figure 6. Cross-plot of the predicted and actual RF estimated by ANN-based, Guthrie and Greenberger, API, and Gulstad correlations.

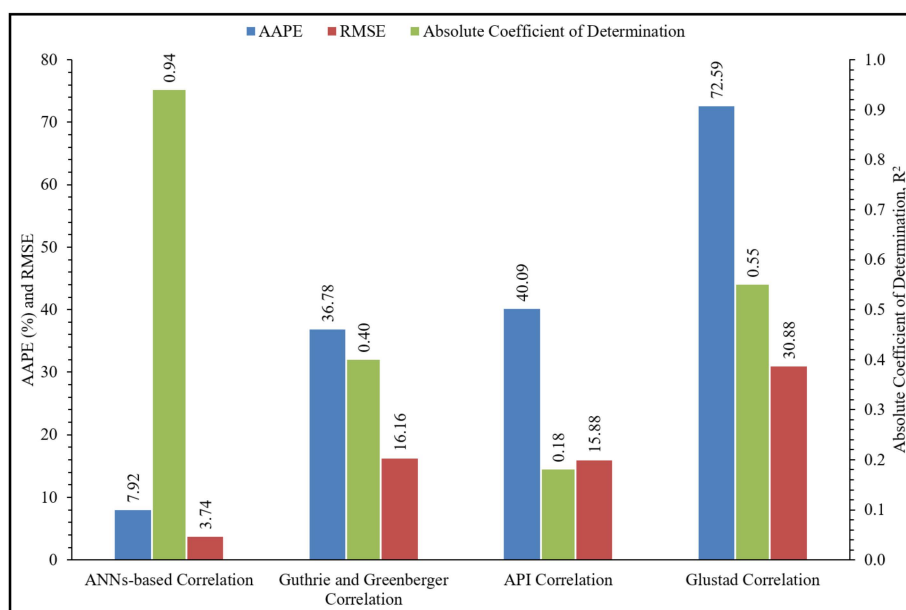


Figure 7. Comparison of AAPE, root mean square error (RMSE), and R^2 of ANNs-based, Guthrie and Greenberger, API, and Gulstad correlations for the testing data.

6. Conclusions

Recovery factor (RF) estimation is a very complicated problem. In this paper, four artificial intelligence (AI) models of the artificial neural networks (ANNs), radial basis neuron networks, adaptive neuro-fuzzy inference system with subtractive clustering, and support vector machines were optimized to predict RF by using 10 parameters regarding the reservoir rock and fluid properties readily available early in the reservoir's life. The ANNs is the best AI model to predict the RF because of its lowest AAPE of 7.92% and the highest R^2 of 0.94 for the predicted RF for the testing dataset (38 reservoirs). For the first time, an empirical correlation for RF prediction was developed based on the ANNs model, which could be modeled and used easily to predict the RF. The developed correlation outperformed the published correlations in terms of all measures of error evaluation considered in

this study; as well, it also has the highest R^2 of 0.94 compared to only 0.55 obtained from Gulstad correlation, which is one of the most accurate correlations currently available.

Author Contributions: Conceptualization, A.A.M., A.A.; methodology, A.A.M. and S.E.; software, A.A.M.; validation, A.A.M. and S.E.; formal analysis, A.A.M.; data curation, A.A.M.; writing—original draft preparation, A.A.M.; writing—review and editing, S.E. and W.C.; visualization, A.A.M., A.A. and S.E.; supervision, A.A., and S.E.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

| | |
|----------|---|
| AAPE | Absolute Average Percentage Error |
| AI | Artificial Intelligence |
| ANFIS-SC | Adaptive Neuro-Fuzzy Inference System with Subtractive Clustering |
| ANNs | Artificial Neural Networks |
| API | American Petroleum Institute |
| STOIIP | Stock-Tank Oil Initially in Place |
| R^2 | Coefficient of Determination |
| RF | Recovery Factor |
| RF_a | Actual Correlation Factor |
| RF_m | Estimated Correlation Factor |
| RNNs | Radial Basis Neural Networks |
| RMSE | Root Mean Square Error |
| SD | Standard Deviation |
| SVM | Support Vector Machines |

Appendix A

Absolute Average Percent Error (AAPE)

$$AAPE = \frac{1}{N} \sum_{i=1}^N \left(\left| \frac{RF_a|_i - RF_m|_i}{RF_a|_i} \right| \times 100 \right) \quad (A1)$$

The Coefficient of Determination (R^2)

$$R^2 = \frac{\sum_{i=1}^N \left[\left([RF_a]_i - \overline{RF_a} \right) - \left([RF_m]_i - \overline{RF_m} \right) \right]^2}{\sqrt{\sum_{i=1}^N \left([RF_a]_i - \overline{RF_a} \right)^2 \sum_{i=1}^N \left([RF_m]_i - \overline{RF_m} \right)^2}} \quad (A2)$$

Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N-2} \sum_{i=1}^N \left([RF_a]_i - [RF_m]_i \right)^2} \quad (A3)$$

where in all previous equations a and m denote actual and estimated, respectively.

References

1. Noureldien, D.M.; El-Banbi, A.H. Using Artificial Intelligence in Estimating Oil Recovery Factor. In Proceedings of the SPE North Africa Technology Conference and Exhibition, Cairo, Egypt, 14–16 September 2015. [\[CrossRef\]](#)
2. Ahmed, A.A.; Elkatatny, S.; Abdulraheem, A.; Mahmoud, M. Application of Artificial Intelligence Techniques in Estimating Oil Recovery Factor for Water Derive Sandy Reservoirs. In Proceedings of the SPE Kuwait Oil & Gas Show and Conference, Kuwait City, Kuwait, 15–18 October 2017. [\[CrossRef\]](#)

3. Craft, B.C.; Hawkins, M. *Applied Petroleum Reservoir Engineering*, 3rd ed.; Pearson Education, Inc.: London, UK, 2015.
4. Ahmed, T. *Reservoir Engineering Handbook*, 4th ed.; Gulf Professional Publishing: Houston, TX, USA, 2010.
5. Lee, J.; Wattenbarger, R.A. *Gas Reservoir Engineering*; Society of Petroleum Engineers: Houston, TX, USA, 1996.
6. Dake, L.P. Fundamentals of Reservoir Engineering. In *Chapter 4, Reservoir Engineering*; Elsevier Science: Amsterdam, The Netherlands, 1983.
7. Agarwal, R.G.; Gardner, D.C.; Kleinstieber, S.W.; Fussell, D.D. Analyzing Well Production Data Using Combined Type Curve and Decline Curve Analysis Concepts. In *Proceedings of the SPE Annual Technical Conference and Exhibition*, New Orleans, LA, USA, 27–30 September 1998. [\[CrossRef\]](#)
8. Guthrie, R.K.; Greenberger, M.H. The Use of Multiple-correlation Analyses for Interpreting Petroleum-engineering Data. In *Proceedings of the Drilling and Production Practice*; American Petroleum Institute: Washington, DC, USA, 1955; pp. 130–137.
9. Gulstad, R.L. 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, 1995.
10. Craze, R.C.; Buckley, S.E. A factual analysis of the effect of well spacing on oil recovery. In *Proceedings of the Drilling and Production Practice*; American Petroleum Institute: Washington, DC, USA, 1945; pp. 144–159.
11. Muskat, M.; Taylor, M.O. Effect of Reservoir Fluid and Rock Characteristics on Production Histories of Gas-Drive Reservoirs. *Trans. AIME* **1946**, *165*, 78–93. [\[CrossRef\]](#)
12. Arps, J.J.; Roberts, T.G. The effect of the relative permeability ratio, the oil gravity, and the solution gas-oil ratio on the primary recovery from a depletion type reservoir. *Pet. Trans.* **1955**, *204*, 120–127.
13. Mahmoud, A.A.A.; Elkatatny, S.; Mahmoud, M.; Abouelresh, M.; Abdulraheem, A.; Ali, A. Determination of the total organic carbon (TOC) based on conventional well logs using artificial neural network. *Int. J. Coal Geol.* **2017**, *179*, 72–80. [\[CrossRef\]](#)
14. Mahmoud, A.A.; Elkatatny, S.; Abdulraheem, A.; Mahmoud, M.; Ibrahim, M.O.; Ali, A. New Technique to Determine the Total Organic Carbon Based on Well Logs Using Artificial Neural Network (White Box). In *Proceedings of the SPE Kingdom Saudi Arabia Annual Technical Symposium and Exhibition*, Dammam, Saudi Arab, 24–27 April 2017. [\[CrossRef\]](#)
15. Elkatatny, S.A.; Mahmoud, M.A. Development of a New Correlation for Bubble Point Pressure in Oil Reservoirs Using Artificial Intelligent Technique. *Arab. J. Sci. Eng.* **2018**, *43*, 2491–2500. [\[CrossRef\]](#)
16. Elkatatny, S.M. Real Time Prediction of Rheological Parameters of KCl Water-Based Drilling Fluid Using Artificial Neural Networks. *Arab. J. Sci. Eng.* **2017**, *42*, 1655–1665. [\[CrossRef\]](#)
17. Abdelgawad, K.; Elkatatny, S.; Moussa, T.; Mahmoud, M.; Patil, S. Real Time Determination of Rheological Properties of Spud Drilling Fluids Using a Hybrid Artificial Intelligence Technique. *J. Energy Resour. Technol.* **2018**, *141*, 032908. [\[CrossRef\]](#)
18. Elkatatny, S. Application of Artificial Intelligence Techniques to Estimate the Static Poisson's Ratio Based on Wireline Log Data. *ASME. J. Energy Resour. Technol.* **2018**, *140*, 072905. [\[CrossRef\]](#)
19. Mahmoud, A.A.; Elkatatny, S.; Ali, A. Estimation of Static Young's Modulus for Sandstone Formation Using Artificial Neural Networks. *Energies* **2019**, *12*, 2125. [\[CrossRef\]](#)
20. Mohaghegh, S.; Arefi, R.; Ameri, S.; Hefner, M.H. A Methodological Approach for Reservoir Heterogeneity Characterization Using Artificial Neural Networks. In *Proceedings of the SPE Annual Technical Conference and Exhibition*, New Orleans, LA, USA, 25–28 September 1994. [\[CrossRef\]](#)
21. Barman, I.; Ouenes, A.; Wang, M. Fractured Reservoir Characterization Using Streamline-Based Inverse Modeling and Artificial Intelligence Tools. In *Proceedings of the SPE Annual Technical Conference and Exhibition*, Dallas, TX, USA, 1–4 October 2000. [\[CrossRef\]](#)
22. Moussa, T.; Elkatatny, S.; Mahmoud, M.; Abdulraheem, A. Development of New Permeability Formulation From Well Log Data Using Artificial Intelligence Approaches. *ASME. J. Energy Resour. Technol.* **2018**, *140*, 072903. [\[CrossRef\]](#)
23. Al-AbdulJabbar, A.; Elkatatny, S.M.; Mahmoud, M.; Abdelgawad, K.; Abdulaziz, A. A Robust Rate of Penetration Model for Carbonate Formation. *J. Energy Resour. Technol.* **2018**, *141*, 042903. [\[CrossRef\]](#)
24. Al-Shehri, D.A. Oil and Gas Wells: Enhanced Wellbore Casing Integrity Management through Corrosion Rate Prediction Using an Augmented Intelligent Approach. *Sustainability* **2019**, *11*, 818. [\[CrossRef\]](#)

25. Salehi, S.; Hareland, G.; Dehkordi, K.K.; Ganji, M.; Abdollahi, M. Casing collapse risk assessment and depth prediction with a neural network system approach. *J. Pet. Sci. Eng.* **2009**, *69*, 156–162. [[CrossRef](#)]
26. Wang, Y.; Salehi, S. Application of real-time field data to optimize drilling hydraulics using neural network approach. *J. Energy Resour. Technol.* **2015**, *137*, 062903. [[CrossRef](#)]
27. Ahmed, A.S.; Ahmed, A.A.; Elkatatny, S. Fracture Pressure Prediction Using Radial Basis Function. In Proceedings of the 2019 AADE National Technical Conference and Exhibition, Denver, CO, USA, 9–10 April 2019.
28. Ahmed, A.S.; Ahmed, A.A.; Elkatatny, S.; Mahmoud, M.; Abdulraheem, A. Prediction of Pore and Fracture Pressures Using Support Vector Machine. In Proceedings of the 2019 International Petroleum Technology Conference, Beijing, China, 26–28 March 2019. [[CrossRef](#)]
29. Adrian, O.; Chukwueke, N. Application of Neural Networks in Developing an Empirical Oil Recovery Factor Equation for Water Drive Niger Delta Reservoirs. In Proceedings of the SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, 5–7 August 2014. [[CrossRef](#)]
30. Onolemhemen, R.U.; Isehunwa, S.O.; Salufu, S.O. Development of Recovery Factor Model for Water Drive and Depletion Drive Reservoirs in The Niger Delta. In Proceedings of the SPE Nigeria Annual International Conference and Exhibition, Lagos, Nigeria, 2–4 August 2016. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).