



A Self-Adaptive Artificial Intelligence Technique to Predict Oil Pressure Volume Temperature Properties

Salaheldin Elkatatny[®], Tamer Moussa, Abdulazeez Abdulraheem *[®] and Mohamed Mahmoud

Department of Petroleum Engineering, King Fahd University of Petroleum & Minerals, Dhahran 31261, Saudi Arabia; elkatatny@kfupm.edu.sa (S.E.); g201105270@kfupm.edu.sa (T.M.); mmahmoud@kfupm.edu.sa (M.M.) * Correspondence: toazeez@gmail.com

Received: 24 October 2018; Accepted: 6 December 2018; Published: 14 December 2018



Abstract: Reservoir fluid properties such as bubble point pressure (*Pb*) and gas solubility (*Rs*) play a vital role in reservoir management and reservoir simulation. In addition, they affect the design of the production system. Pb and Rs can be obtained from laboratory experiments by taking a sample at the wellhead or from the reservoir under downhole conditions. However, this process is time-consuming and very costly. To overcome these challenges, empirical correlations and artificial intelligence (AI) models can be applied to obtain these properties. The objective of this paper is to introduce new empirical correlations to estimate Pb and Rs based on three input parameters-reservoir temperature and oil and gas gravities. 760 data points were collected from different sources to build new AI models for *Pb* and *Rs*. The new empirical correlations were developed by integrating artificial neural network (ANN) with a modified self-adaptive differential evolution algorithm to introduce a hybrid self-adaptive artificial neural network (SaDE-ANN) model. The results obtained confirmed the accuracy of the developed SaDE-ANN models to predict the *Pb* and *Rs* of crude oils. This is the first technique that can be used to predict Rs and Pb based on three input parameters only. The developed empirical correlation for Pb predicts the Pb with a correlation coefficient (CC) of 0.99 and an average absolute percentage error (AAPE) of 6%. The same results were obtained for Rs, where the new empirical correlation predicts the Rs with a coefficient of determination (R^2) of 0.99 and an AAPE of less than 6%. The developed technique will help reservoir and production engineers to better understand and manage reservoirs. No additional or special software is required to run the developed technique.

Keywords: self-adaptive differential evolution; artificial intelligence (AI); bubble point pressure correlation; gas solubility correlation; pressure volume temperature (PVT) properties prediction

1. Introduction

Reservoir fluid pressure volume temperature (PVT) properties such as bubble point pressure, gas solubility, and oil and gas formation volume factors and viscosities are critical in reservoir engineering management and computations. These PVT properties are required to obtain the initial hydrocarbons in place, optimum production schemes, ultimate hydrocarbon recovery, design of fluid handling equipment, and reservoir volumetric estimates. Bubble point pressure (*Pb*) and gas solubility (*Rs*) are two of the most critical quantities used to characterize an oil reservoir. Therefore, the accurate determination of these properties is one of the main challenges in reservoir development and management. There are also other factors that affect reservoir management, such as permeability. Jia et al. [1] illustrated that for shale reservoirs with a permeability of 0.01 mD, continuous gas injection is preferred, while for ultra-low permeability reservoirs, CO_2 huff-*n*-puff is recommended. For CO_2 huff-*n*-puff injection in oil shale reservoirs, the reservoir heterogeneity is not a favorable



function during the primary production period. Meanwhile, the fracture length plays a key role in oil production [2].

Conventionally, PVT properties are determined by laboratory measurements. However, these experiments are costly, time-consuming, and highly dependent on the quality and quantity of collected samples [3–5]. Therefore, several empirical correlations such as the equation of states (EOS) as well as linear, non-linear, and multiple regression correlations have been introduced to predict PVT properties [6–9]. However, the accuracy of these correlations is highly dependent on fluid types and the chosen equation [10–12].

Recently, artificial intelligence (AI) techniques have been extensively applied in the petroleum industry, especially in predicting well/field performance. Alajmi et al. predicted the choke performance using artificial neural network (ANN) [13]. Alarifi et al. [14] estimated the productivity index for oil horizontal wells using ANN, functional network, and fuzzy logic. Chen et al. [15] applied neural network and fuzzy logic to evaluate the performance of an inflow control device (ICD) in a horizontal well. Moussa et al. [16] used optimized an ANN model to predict average reservoir permeability using well-log data. Van and Chon [17] evaluated the performance of CO₂ flooding using ANN. Elkatatny et al. [18] applied ANN to estimate the rheological properties of drilling fluids based on real-time measurements.

Therefore, several AI approaches and data-driven models have been introduced to predict PVT properties and overcome the challenges associated with laboratory measurements and analytical correlations. Abedini et al. [19] used ANN and fuzzy logic approaches to predict the oil viscosity of undersaturated oil reservoirs. Two models were introduced and showed accurate prediction of oil viscosity compared to the measured values in the laboratory. The input parameters of their models were oil gravity, reservoir temperature, gas oil ratio (GOR), and bubble point pressure. Moghadasi et al. [20] used ANN to estimate the values of Pb for Iranian oil fields. The input parameters utilized in their model were reservoir temperature, GOR, and oil and gas gravities. They compared their prediction with previous models and showed that ANN yielded the highest accuracy. Al-Marhoun et al. used ANN to determine the *Pb* from the oil composition as well as the GOR, oil and gas gravities, and reservoir temperature [21]. They compared the developed ANN model with other equations of states (EOS) and other available models in the literature and they concluded that ANN yielded very accurate prediction compared to the previous methods. Tatar et al. [22] used ANN models to estimate the water density in oil and gas reservoirs. Water density is necessary in reservoir simulation and material balance calculations. Their model predicted the formation water density with a correlation coefficient (CC) close to unity and error close to zero. They used reservoir pressure, temperature, and sodium chloride concentration as inputs to predict the water density. Ahmadi and Bahadori [23] used AI tools such as fuzzy logic to evaluate the enhanced oil recovery (EOR) processes. They coupled the fuzzy approach with commercial reservoir simulators to enhance the accuracy of selecting and ranking the appropriate EOR method for the specified oil reservoirs. Choubineh et al. [24] used 693 data points to develop an ANN model to predict the natural gas density for different temperature and pressure ranges. Their model showed that the gas pressure and temperature have a great effect on the natural gas density. Their model can be used in a temperature range of 250 to 450 K and a pressure range of 15 to 65 MPa. The model accuracy was high compared with previous models; the regression coefficient was more than 0.99 and the average absolute error was less than 0.5%.

Although the data-driven models developed by different AI approaches have shown good accuracy compared to laboratory measurements and have outperformed analytical correlations, the input parameters of these models still require expensive laboratory experiments. For example, in order to estimate bubble point pressure (*Pb*), gas solubility (*Rs*) or oil composition is required as an input parameter. In other words, the abovementioned models did not eliminate the requirement of the expensive and time-consuming laboratory experiments. Therefore, the objective of this paper is to introduce two data-driven models to: (1) predict the *Pb* of crude oil samples based on three input

parameters—reservoir temperature and oil and gas gravities; and (2) predict the *Rs* using the three input parameters as well as the predicted *Pb* as the fourth input parameter.

The proposed methods require no expensive laboratory experiments. Hence, it is a step toward minimizing PVT laboratory experiments. The proposed data-driven models are developed using a modified self-adaptive differential evolution algorithm (MSaDE) [25] combined with ANN. In the subsequence sections of this paper, the proposed hybrid algorithm is referred to as SaDE-ANN.

2. Artificial Neural Network Modeling

Artificial neural network (ANN) is computational method derived from the biological neural network [26,27]. In the architecture of ANN, the input and the output are connected by specific neurons. A normal ANN contains an input layer, one or more hidden layers, and an output layer. Information is received by the input layer. In the hidden layer(s), a relationship between input(s) and output(s) is developed. Every neuron of one layer is linked to every neuron in the following layer and every connection has a related weight [28]. The relationship between the neuron and the source is controlled by weights and biases [29].

To avoid overfitting and underfitting, an optimization process is performed to determine the optimum number of neurons [28,30]. Training is the first step in the formation of the network. After training the network with training data, the testing output can be predicted using the weighted average of the outputs of training dataset, where the weights are calculated using the Euclidean distance between the training and testing data [31,32].

3. Methodology

ANN has several control parameters, such as the number of hidden layers, number of neurons at each layer, training and transferring functions, and ratio of testing over training datasets. Conventionally, the values of these control parameters are assigned by several sensitivity trials. In each single trial, different values of one parameter are assigned while keeping other parameters constant. Then the value that achieved the minimum error between the measured (real) and predicted output is selected. Similar processes are applied to the remaining parameters to find their best values. However, because of the interdependency of these control parameters, this "trial" approach does not ensure the accomplishment of optimum results.

Therefore, the methodology approached in this paper involves the simultaneous optimization of these parameters to achieve the minimum average absolute percentage error (AAPE) and the maximum CC. The definitions of AAPE and CC are shown in Appendix A. The stochastic optimization method used in this paper is modified by self-adaptive differential evolution (MSaDE) [25]. In MSaDE, the control variables of a differential evolution algorithm, such as scale factor, crossover, and mutation strategy, are self-adapted during each iteration. In this paper, MSaDE is integrated with ANN to optimize the control parameters of ANN.

The input parameters to the ANN are: reservoir temperature (T), oil gravity (American Petroleum Institute (API)), and gas specific gravity (GG). The outputs are bubble point pressure (Pb) and solution gas ratio (Rs). As mentioned earlier, ANN consists of two phases—training and testing. In the training phase, the optimization process of SaDE-ANN continues running until one of two conditions: (1) the AAPE is less than 5%, or (2) the maximum number of function evaluations (1000) is reached. Then the optimized SaDE-ANN model is validated on unseen testing datasets to predict the values of Pb and Rs using the input parameters T, API, and GG.

Data Analysis and Acquisition

The data points utilized in this paper were collected from the literature [7–9,33–35]. The data includes different oil sources with different concentrations. Data from the Middle East (Al-Marhoun) [8], data from Malaysian Crudes (Omar and Todd) [34], data from North Sea Glasø [7], data from fields all over the world (Vazquez and Beggs) [9], and data from the Mediterranean Basin,

Africa, the Persian Gulf, and the North Sea (De Ghetto) [35] were employed. Each data point contains input parameters (reservoir temperature (*T*), oil gravity (API), and GG) and output parameters (solution gas oil ratio (*Rs*) and bubble point pressure (*Pb*)). Table 1 shows the statistical parameters of the studied 460 datasets after outlier removal using mean-standard deviation method; in which the dataset (x_i) would be considered as outliers if the condition shown in Equation (1) is achieved.

$$\left|\mathbf{x}_{j} - \overline{x}_{j}\right| > 3 \sigma_{j} \tag{1}$$

where \mathbf{x}_j is the data vector for the j^{th} parameter, $\mathbf{x}_j = [x_{j,1}, x_{j,2}, x_{j,3}, \dots, x_{j,N}]$, $j = 1, 2, \dots, J$, J is the total number of input parameters (in this case, J = 3), \overline{x}_j is the mean of the j^{th} parameter, $\overline{x}_j = \frac{1}{N} \sum_{i=1}^{N} x_{j,i}$, N is the total number of datasets, and σ_j is the standard deviation of the j^{th} parameter.

Statistical Parameter	Gas Specific Gravity	Oil Gravity, °API	Reservoir Temperature, °F	Solution Gas Oil Ratio, SCF/STB	Bubble Point Pressure, psi
Maximum	1.367	56.200	294.000	1718.000	4735.000
Minimum	0.589	15.300	74.000	9.000	126.000
Arithmetic Mean	0.904	36.048	170.926	518.981	1754.048
Geometric Mean	0.891	35.273	163.511	356.563	1357.907
Harmonic Mean	0.879	34.449	155.754	189.980	907.683
Mode	0.802	33.300	100.000	61.000	500.000
Range	0.778	40.900	220.000	1709.000	4609.000
Mid-Range	0.201	10.400	74.000	554.000	1633.000
Variation	0.389	20.450	110.000	854.500	2304.500
Interquartile Range	0.025	53.085	2,394	145,930	1101,077
Standard Deviation	0.160	7.286	48.933	382.008	1049.322
Skewness	0.820	-0.043	0.082	0.823	0.357
Kurtosis	0.285	-0.385	-0.772	0.202	-0.624
Coefficient of Variation	0.176	0.202	0.286	0.736	0.598

Table 1. Statistical parameters for the studied data.

The CC of the input parameters (*T*, API, and GG) with output parameters (*Pb* and *Rs*) are shown in Figure 1. In this paper, a combined correlation coefficient (cCC) parameter is introduced to indicate the combined CC of T, API, and GG to *Pb* and *Rs*. cCC is the arithmetic mean of the CCs of the three input parameters calculated by Equation (2). cCC is estimated for *Pb* and *Rs* to determine which output should be estimated first.

$$cCC = \frac{|CC_{GG}| + |CC_{API}| + |CC_T|}{3}$$
 (2)

where CC_{GG} , CC_{API} , and CC_{T} are the correlation coefficients between the output parameter and GG, oil gravity, and reservoir temperature, respectively. Figure 1 shows that *Pb* has a higher cCC with the input parameters (0.37) compared to *Rs* (0.32). Therefore, it is more convenient to estimate *Pb* first, and then use the estimated *Pb* with the three input parameters to predict *Rs*.



Figure 1. Correlation coefficients of reservoir temperature (*T*), oil API, and gas gravity (GG) to bubble point pressure (*Pb*) and gas solubility (*Rs*).

4. Results and Discussion

4.1. Bubble Point Pressure Estimation

The SaDE-ANN model was built to correlate *Pb* with T, API, and GG. The optimum parameters of ANN to generate the best results in terms of the lowest AAPE and highest CC were found to be an ANN structure of 3-18-17-1; the input layer consisted of three neurons representing the input parameters—reservoir temperature, oil API, and gas gravity. 18 neurons made up the first hidden layer, 17 neurons made up the second hidden layer, and *Pb* was the only parameter in the output layer. Data were divided into three sets—training (65%), validation (11%), and testing (24%). The optimum training and transfer functions were Bayesian regularization backpropagation and symmetric sigmoid, respectively.

Figure 2 shows the cross plot of the predicted values of *Pb* using the SaDE-ANN model vs. the actual *Pb* values. Figure 2 shows that the AAPE was 5.18% and the CC was 0.994 for the training data, while for the testing data the AAPE was 6.37% and the CC was 0.993. These results confirm the stability and high accuracy of the SaDE-ANN model, which can be used to predict the *Pb* based on reservoir temperature, oil API gravity, and GG.



Figure 2. Cross plot between real *Pb* and predicted values using the self-adaptive differential evolution-artificial neural network (SaDE-ANN) model. (**a**) Training data; and (**b**) testing data.

An AAPE of 6.37% for bubble point pressure prediction is acceptable considering two important factors: (1) this approach is utilized when the solution gas oil ratio data are not available and depends only on reservoir temperature and oil and gas gravities; (2) most of the published bubble point pressure correlations and models, including this proposed model, use fluid properties data from oilfield service companies, which has a lot of concerns in terms of data quality. In addition, when comparing the outputs of the proposed SaDE-ANN model with other correlation and models, in which the solution gas oil ratio is considered as the fourth input with the three inputs used in the SaDE-ANN model, the results are superior compared to other models and correlations. Figure 3 shows the performance comparison of validation data between different models and correlations. The outputs from the proposed SaDE-ANN method has the highest R² compared to other models. Figure 4 shows the comparison between different models and correlations based on the AAPE and CC—as shown in the figure, SaDE-ANN has the lowest AAPE and highest CC.



Figure 3. Validation cross plot comparison of bubble point pressure prediction achieved using different models. The SaDE-ANN model achieved the best match compared to other correlations.



Figure 4. Comparison between different models and correlations based on (**a**) the average absolute percentage error (AAPE) and (**b**) the correlation coefficient (CC).

Mathematical Model for Bubble Point Pressure

The mathematical model to estimate Pb derived from the optimized ANN model using GG, API, and T as input parameters, where the limitations of each parameter are shown in Table 1, is shown as follows:

$$Pb_n = \sum_{j=1}^{N_2} w_{3_j} Y_j + b_3$$
(3)

where Pb_n is normalized Pb (psi) and Y_j is calculated as:

$$Y_j = \sum_{i=1}^{N_1} w_{2_{j,i}} \frac{2}{1 + e^{-2 X_i}} - 1 + b_{2_j}$$
(4)

and X_i is calculated as:

$$X_{i} = \frac{2}{1 + e^{-2 (w_{1_{i,1}} \gamma_{g_{n}} + w_{1_{i,2}} \operatorname{API}_{n} + w_{1_{i,3}} T_{n} + b_{1_{i}})} - 1$$
(5)

where:

 N_1, N_2 number of neurons in the first and second hidden layers, respectively;i, jneuron index in the first and second hidden layers, respectively, as shown in Tables 2 and 3; w_{1j}, b_{1i} weights and bias between the input and first hidden layers, respectively, as shown in Table 2; $w_{2j,i}, b_{2j}$ weights and bias between the first hidden and output layers, respectively; the values of $W_{2j,i}$ are shown in Table 3;

 w_{3_j} , b_3 weights and bias between the second hidden and output layers, respectively; $b_3 = -0.2626$ and the values of W_{3j} are shown in Table 3;

 γ_{g_n} normalized GG, as calculated by Equation (6); API_n normalized oil API gravity, as calculated by Equation (7);

 T_n normalized reservoir temperature (°F), as calculated by Equation (8).

$$\gamma_{g_n} = 2.5707 (\gamma_g - 0.589) - 1 \tag{6}$$

$$API_n = 0.0453(API - 15.3) - 1 \tag{7}$$

$$T_n = 0.0091(T - 74) - 1 \tag{8}$$

 Pb_n presented in Equation (3) is the normalized value of Pb. Pb is calculated as:

$$Pb = \frac{Pb_n + 1}{0.000286} + 126\tag{9}$$

Table 2. Weights and biases of the first hidden layer in the ANN structure of the bubble point pressure model.

i	$w_{1_{i,1}}$	$w_{1_{i,2}}$	$w_{1_{i,3}}$	b_{1_i}
1	0.0847	3.8463	0.5493	0.7595
2	3.91	-1.1935	4.9371	2.2716
3	-1.2509	2.3835	2.1648	-1.5377
4	0.3323	3.9022	0.3577	-1.27
5	-3.4482	-0.7113	-2.0318	0.3747
6	-4.184	0.6819	3.4924	-0.6621
7	-1.7434	4.3069	-2.8004	-1.7342
8	0.4419	-3.0066	4.0254	-0.8286
9	0.4158	4.5222	-2.392	1.6974
10	-3.9243	-1.6785	-1.0686	-1.4006
11	1.3103	0.1133	-1.5418	-1.4221
12	-2.0139	-0.2613	4.4812	-0.53
13	1.3382	-2.6467	-2.8971	-0.5727
14	3.0649	-3.291	-0.4142	0.0906
15	-2.6736	-3.06	1.1079	0.0082
16	-2.5812	5.0496	-0.5947	-1.4075
17	2.0543	1.3933	-1.2484	1.0663
18	-1.3636	-1.2778	-3.4035	1.2546

i	$w_{2_{j,i}}$														ha	7122				
)	i = 1	<i>i</i> = 2	<i>i</i> = 3	i = 4	<i>i</i> = 5	<i>i</i> = 6	i = 7	<i>i</i> = 8	<i>i</i> = 9	i = 10	i = 11	i = 12	i = 13	i = 14	i = 15	i = 16	i = 17	i = 18	0 _{2j}	
1	-0.410	0.184	0.134	-0.573	-0.403	-0.683	-0.406	-1.883	-2.377	0.335	0.963	0.693	-0.502	0.426	0.631	-0.165	1.085	-0.564	-0.140	2.45
2	1.839	0.385	-0.889	0.058	0.149	-1.337	0.649	0.125	1.038	-0.344	0.513	-1.343	-0.352	-2.096	0.143	-1.614	-0.257	-1.045	-0.043	2.33
3	-0.691	-0.169	-0.392	-0.738	0.966	-0.037	2.469	1.518	-0.477	0.860	0.060	-1.226	-0.331	0.086	0.986	0.814	-1.231	-1.200	0.658	-2.55
4	-0.066	-1.132	0.415	0.910	0.027	-0.677	-0.639	-0.017	-1.124	0.380	-0.161	0.106	1.061	-1.934	0.167	1.766	-1.055	0.291	0.475	3.03
5	-0.871	0.580	-0.018	1.068	0.210	-0.130	0.122	-0.208	0.603	1.078	-0.179	0.406	1.156	-0.458	0.470	-1.116	2.286	1.042	-0.308	-2.45
6	-0.398	0.099	0.129	-0.735	0.309	-0.698	1.262	0.292	0.307	-0.349	-0.255	-0.003	-0.508	-1.289	0.797	0.309	0.956	-2.374	0.030	-1.69
7	2.022	0.363	-0.565	0.045	1.093	0.642	1.615	-0.573	-0.532	-0.361	0.372	0.643	0.274	-0.125	0.750	-2.309	-1.358	1.216	-1.258	1.59
8	0.745	-0.001	-0.369	-0.999	0.947	0.635	-0.097	1.481	-1.005	0.766	0.153	0.541	1.084	-1.943	-0.291	0.116	0.892	-0.271	-0.179	2.88
9	-1.396	2.816	-1.460	-0.517	0.891	-1.169	-1.263	-1.146	0.977	-0.275	0.575	0.896	-0.382	1.573	-1.410	-0.920	0.249	-0.653	-0.376	2.16
10	0.178	0.040	0.676	-0.682	-0.926	0.075	1.437	-1.222	1.177	0.304	-0.626	1.318	0.117	0.112	0.620	1.398	0.775	-0.206	0.032	2.22
11	0.599	1.986	-0.004	1.060	-0.798	-2.162	-1.054	-2.858	-1.155	-0.851	-0.474	0.723	1.148	0.355	1.151	-0.598	0.802	-0.841	0.428	-2.01
12	1.621	1.076	-0.721	0.871	-0.406	-0.070	-0.468	0.738	-1.347	2.547	0.202	0.095	-0.247	-0.356	0.930	-0.132	-1.160	-1.164	-0.336	-2.26
13	0.754	0.041	0.499	-1.609	-0.795	-0.997	0.627	-0.995	0.422	0.316	0.195	-0.564	-0.831	-0.339	1.346	1.092	0.062	0.030	-0.425	-2.49
14	-0.106	2.295	-1.196	1.851	0.387	-1.079	0.522	-0.724	1.553	0.901	0.517	0.671	0.445	1.174	-0.451	-0.530	-0.803	0.275	-0.358	3.01
15	0.413	-0.554	0.719	-0.232	-0.347	-0.208	-0.588	-0.839	-1.993	-0.045	-1.883	0.534	0.749	0.201	-0.239	-0.480	0.826	-1.309	-0.795	-2.04
16	0.619	-0.603	1.597	-0.640	-0.289	-0.909	-0.746	-0.293	-1.093	0.489	-0.850	-1.021	-0.823	0.522	0.175	0.414	1.333	-1.094	0.347	3.26
17	-0.073	-0.903	1.402	-0.143	-1.191	0.156	0.652	-0.068	-1.299	0.760	-0.864	0.733	0.539	0.285	0.904	-1.034	-0.233	0.345	0.205	2.84

Table 3. Weights and biases of the second hidden layer in the ANN structure of the bubble point pressure model.

4.2. Gas Solubility Estimation

In this section, gas solubility (*Rs*) is estimated based on the value of *Pb* predicted by the proposed model as well as the three input parameters (*T*, API, and GG). The optimum values of ANN parameters to generate the best results in terms of lowest AAPE and highest CC are an ANN with a structure of 4-15-15-1; the input layer consists of four neurons (the input parameters of oil API reservoir temperature, gas gravity, and *Pb*), while both the first and second hidden layers consist of 15 neurons. *Rs* is the only output parameter in the output layer. Data were divided into three sets—training (67%), testing (21%) and validation (12%). The best training function was trainbr and the best transferring function was logsig.

Figure 5 shows the relative importance of input parameters (GG, oil API (API), reservoir temperature (T), and predicted bubble point pressure (from the previous step) with solution gas oil ratio (*Rs*)). As shown in Figure 5, the bubble point pressure had the highest relative importance with *Rs*, which is why it was very challenging to predict *Rs* without considering *Pb* as an input parameter. Therefore, GG, API, T, and predicted *Pb* were considered as the four inputs for SaDE-ANN to predict *Rs*.



Figure 5. Correlation coefficients of Rs with inputs, including the predicted Pb.

Training and testing cross plots of actual and predicted values of Rs, from SaDE-ANN, are shown in Figure 6. The figure shows that the R^2 was 0.99 for both the training and testing data, while the AAPEs of the training and testing data were 5.89% and 6.54%, respectively. These results confirm the capability of the SaDE-ANN model to predict Rs based on four parameters—Pb, T, API, and GG.



Figure 6. Cross plot between the real *Rs* and the values predicted using the SaDE-ANN model. (a) Training data; and (b) testing data.

Mathematical Model for Gas Solubility

The mathematical model extracted from the optimized SaDE-ANN model to estimate *Rs* using *Pb*, T, API, and GG, with the limitation of each value as shown in Table 1, is introduced by equations as follows:

$$Rs_n = \sum_{j=1}^{N_2} w_{3_j} Y_j + b_3 \tag{10}$$

where Rs_n is the normalized Rs (SCF/STB) and Y_j is calculated as follows:

$$Y_j = \sum_{i=1}^{N_1} w_{2_{j,i}} \frac{1}{1 + e^{-X_i}} + b_{2_j}$$
(11)

X_i is calculated as:

$$X_{i} = \frac{1}{1 + e^{-(w_{1_{i,1}} \gamma_{g_{n}} + w_{1_{i,2}} \operatorname{API}_{n} + w_{1_{i,3}} \operatorname{T}_{n} + w_{1_{i,4}} \operatorname{Pb}_{n} + b_{1_{i}})}$$
(12)

The definitions of N_1 , N_2 , i, j, b_{1i} , b_{2j} , b_3 , w_{1i} , $w_{2j,i}$, w_{3j} , γ_{gn} , API_n, and T_n are similar to those introduced in Section 4.1. The values of i, b_{1i} , and w_{1i} are listed in Table 4; the values of j, b_{2j} , $w_{2j,i}$, and w_{3j} are listed in Table 5; $b_3 = 1.2091$. Pb_n is the normalized bubble point pressure (psi) and is calculated as follows:

$$Pb_n = 0.00029(Pb - 161.96) - 1 \tag{13}$$

The value of Rs_n from Equation (10) is normalized and can be converted to Rs as follows:

$$Rs = \frac{Rs_n + 1}{0.000765} + 21\tag{14}$$

Table 4. Weight and biases of each neuron in the first hidden layer in the ANN structure of the gas solubility model.

i	$w_{1_{i,1}}$	$w_{1_{i,2}}$	$w_{1_{i,3}}$	$w_{1_{i,4}}$	\boldsymbol{b}_{1_i}
1	7.8528	-3.4231	2.4038	-2.2436	1.7801
2	-2.2089	-3.6357	-1.3463	4.094	0.1083
3	1.5981	-3.9961	4.4487	-0.4893	-1.9289
4	-0.3229	2.2407	5.5417	3.2776	0.0307
5	-5.2848	2.7897	3.7404	3.9076	1.3228
6	3.1948	6.484	3.2778	4.2005	2.2259
7	-1.6744	1.3062	2.0996	-4.7358	-1.6276
8	-1.0052	-0.6798	-1.2986	6.0228	3.8066
9	-7.2086	0.9991	3.2067	4.6079	-0.7049
10	-1.9889	3.5852	-1.5909	-4.3559	-1.9949
11	-1.7207	3.2091	-3.3657	6.883	-0.5073
12	-5.5474	0.9029	0.6965	0.6955	1.1485
13	-2.7482	-1.8183	-4.8786	-0.4023	1.0552
14	-6.9512	-4.748	3.7096	-2.0801	-0.5928
15	0.5709	-4.1858	-3.5728	3.9487	1.5681

;								$w_{2_{j,i}}$								ha	W2.
)	<i>i</i> =1	<i>i</i> = 2	<i>i</i> = 3	<i>i</i> = 4	<i>i</i> = 5	<i>i</i> = 6	<i>i</i> = 7	<i>i</i> = 8	<i>i</i> = 9	<i>i</i> = 10	i = 11	<i>i</i> = 12	<i>i</i> = 13	<i>i</i> = 14	<i>i</i> = 15	v_{2_j}	3 _j
1	1.554	2.800	1.085	-0.860	-1.485	-2.396	-0.242	0.840	1.776	1.773	-2.829	0.694	-0.441	-0.082	1.385	-0.922	-5.681
2	-2.932	-0.424	0.263	-0.633	-0.212	-0.844	0.465	4.355	-2.354	1.788	-1.613	1.671	0.636	-0.997	-2.740	-0.434	4.585
3	-0.938	-0.749	-1.130	-1.269	4.179	1.264	1.402	0.393	-2.974	-2.532	1.927	-1.845	0.234	2.833	-1.008	0.310	-3.292
4	-2.711	-1.708	2.747	1.537	-0.662	-1.630	2.046	-1.875	-1.191	1.346	-2.360	0.289	2.147	-0.430	-0.108	2.778	-4.724
5	3.450	2.797	-1.875	0.605	-2.980	-4.805	3.414	-1.952	2.863	-0.029	2.960	1.690	-2.389	-0.267	0.929	0.854	4.937
6	3.137	-2.165	1.345	1.902	-0.208	-1.544	-1.150	-0.256	-1.293	1.383	-4.041	1.017	-0.175	0.102	-1.149	0.683	-2.076
7	1.598	0.386	0.274	1.047	1.528	0.021	0.477	-1.698	-0.732	0.598	-1.437	0.031	0.247	0.940	-2.502	-0.063	4.667
8	3.038	1.812	0.852	-4.579	2.510	1.750	-0.412	-1.840	-0.295	-0.441	-2.573	1.692	-0.174	1.624	1.293	-1.129	3.736
9	1.148	1.240	0.769	2.059	1.937	-1.253	0.274	-2.670	1.178	-0.013	-2.148	-3.262	1.213	1.012	1.594	-0.535	4.452
10	0.166	2.730	-3.771	1.449	-0.412	-0.376	-2.410	0.841	2.582	0.553	2.596	0.003	0.336	-1.986	-2.667	1.420	-4.039
11	-2.558	-2.640	0.543	-2.271	0.328	2.071	-0.171	1.974	-3.259	-1.002	2.292	-0.392	-0.611	1.001	1.955	-0.067	3.891
12	1.815	0.977	0.500	-0.962	-1.080	-1.207	0.502	-3.151	-0.801	-0.441	-0.533	1.825	-0.441	2.414	-1.809	1.262	-4.627
13	-2.361	-3.992	-3.165	-2.139	0.709	1.626	1.510	1.054	0.260	1.670	-0.604	-0.122	1.297	-1.834	-1.540	-1.014	2.243
14	-2.109	0.314	-0.084	-0.826	0.154	0.630	1.856	-0.959	-0.841	0.530	-2.510	0.546	2.185	-1.291	-0.992	-0.712	-3.847
15	0.563	-0.695	0.811	-0.402	1.825	1.529	-2.245	-1.279	-0.432	-0.295	4.792	-1.066	-1.730	-0.036	-0.070	0.922	-2.925

Table 5. Weight and biases of each neuron in the second hidden layer of the ANN structure of the bubble point pressure model.

5. Conclusions

Bubble point pressure (*Pb*) and gas solubility (*Rs*) have a significant effect on the accuracy of modeling fluid flow in porous media. This paper introduced two data-driven correlations to predict *Pb* and *Rs* using reservoir temperature and oil and gas gravities. These empirical correlations were developed using a self-adaptive artificial neural network (SaDE-ANN). SaDE-ANN is a hybrid ANN integrated with a modified self-adaptive differential evolution (MSaDE) algorithm. The proposed correlations by SaDE-ANN were validated using previous experimental data reported in the literature (760 data points).

The developed empirical correlation for *Pb* predicted the *Pb* with a CC of 0.99 and an average absolute error (AAPE) of 6%. The same results were obtained for *Rs*, where the new empirical correlation predicted the *Rs* with a coefficient of determination (R^2) of 0.99 and an AAPE of less than 6%.

The proposed correlations showed the highest prediction accuracy when compared to different empirical correlations. The proposed method outperformed other previously reported methods, as it obtained the highest CC of 0.992 and lowest AAPE of 5.42% between measured and predicted values. The correlations introduced in this paper used reservoir temperature and oil and gas gravities as input parameters to predict *Pb* and *Rs*. Hence, this represents a breakthrough that minimizes the need for the expensive and time-consuming PVT laboratory experiments commonly used to determine *Pb* and *Rs*.

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

Funding: This research received no external funding.

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

Appendix A

The AAPE is the measure of the relative deviation from the experimental data and is defined by:

$$AAPE = \frac{1}{N} \sum_{i=1}^{N} |E_i|$$

where E_i is the relative deviation of an estimated value (Y_{est}) from an experimental value (Y_{exp});

$$E_i = \left[\frac{Y_{\text{exp}} - Y_{\text{est}}}{Y_{\text{exp}}}\right] \times 100 \quad i = 1, 2, \dots, N$$

The CC represents the degree of success in reducing the standard deviation by regression analysis, defined by:

$$CC = \sqrt{1 - \frac{\sum_{i=1}^{N} [Y_{exp} - Y_{est}]_{i}^{2}}{\sum_{i=1}^{N} [Y_{exp} - \overline{Y}]}}$$

where:

$$\overline{Y} = \frac{1}{N} \sum_{i=1}^{N} [Y_{\exp}]_{i}$$

References

- Jia, B.; Tsau, J.; Barati, R. A Review of the Current Progress of CO₂ Injection EOR and Carbon Storage in Shale Oil Reservoirs. *Fuel* 2019, 236, 404–427. [CrossRef]
- Jia, B.; Tsau, J.; Barati, R. Role of Molecular Diffusion in Heterogeneous, Naturally Fractured Shale Reservoirs during CO₂ Huff-*n*-Puff. J. Pet. Sci. Eng. 2018, 164, 31–42. [CrossRef]
- Baker, R.O.; Regier, C.; Sinclair, R. PVT Error Analysis for Material Balance Calculations. In Proceedings of the Canadian International Petroleum Conference, Calgary, AB, Canada, 10–12 June 2003.

- 4. Dong, C.; O'Keefe, M.D.; Elshahawi, H.; Hashem, M.; Williams, S.M.; Stensland, D.; Hegeman, P.S.; Vasques, R.R.; Terabayashi, T.; Mullins, O.C.; et al. New Downhole-Fluid-Analysis Tool for Improved Reservoir Characterization. *SPE Reserv. Eval. Eng.* **2008**, *11*, 1107–1116. [CrossRef]
- Nnochiri, M.O.; Lawal, K.A. How Variable Fluid PVT Model Affects the Performance of an Integrated Production System. In Proceedings of the SPE EUROPEC/EAGE Annual Conference and Exhibition, Barcelona, Spain, 14–17 June 2010.
- 6. Standing, M.B. A Pressure-Volume-Temperature Correlation for Mixtures of California Oils and Gases. In *Drilling and Production Practice;* American Petroleum Institute: New York, NY, USA, 1947.
- 7. Glaso, O. Generalized Pressure-Volume-Temperature Correlations. J. Pet. Technol. 1980, 32, 785–795. [CrossRef]
- 8. Al-Marhoun, M.A. PVT Correlations for Middle East Crude Oils. J. Pet. Technol. 1988, 40, 650–667. [CrossRef]
- 9. Vazquez, M.; Beggs, H.D. Correlations for Fluid Physical Property Prediction. J. Pet. Technol. 1980, 32, 968–970. [CrossRef]
- 10. Wang, L.-S.; Gmehling, J. Improvement of the SRK equation of state for representing volumetric properties of petroleum fluids using Dortmund Data Bank. *Chem. Eng. Sci.* **1999**, *54*, 3885–3892. [CrossRef]
- Peng, W.; Pope, G.A. Proper Use of Equations of State for Compositional Reservoir Simulation. *J. Pet. Technol.* 2001, 53, 74–81. [CrossRef]
- 12. Guo, X.Q.; Sun, C.Y.; Rong, S.X.; Chen, G.J.; Guo, T.M. Equation of state analog correlations for the viscosity and thermal conductivity of hydrocarbons and reservoir fluids. *J. Pet. Sci. Eng.* **2001**, *30*, 15–27. [CrossRef]
- AlAjmi, M.D.; Alarifi, S.A.; Mahsoon, A.H. Improving Multiphase Choke Performance Prediction and Well Production Test Validation Using Artificial Intelligence: A New Milestone. In Proceedings of the SPE Digital Energy Conference and Exhibition, The Woodlands, TX, USA, 3–5 March 2015.
- 14. Alarifi, S.; AlNuaim, S.; Abdulraheem, A. Productivity Index Prediction for Oil Horizontal Wells Using different Artificial Intelligence Techniques. In Proceedings of the SPE Middle East Oil & Gas Show and Conference, Manama, Bahrain, 8–11 March 2015.
- 15. Feifei, C.; Yonggang, D.; Junbin, Z.; Weifeng, W. Application of neural network and fuzzy mathematic theory in evaluating the adaptability of inflow control device in horizontal well. *J. Pet. Sci. Eng.* **2015**, *134*, 131–142. [CrossRef]
- Moussa, T.; Elkatatny, S.; Mahmoud, M.; Abdulraheem, A. Development of New Permeability Formulation From Well Log Data Using Artificial Intelligence Approaches. J. Energy Resour. Technol. 2018, 140, 072903. [CrossRef]
- Van, S.L.; Chon, B.H. Effective Prediction and Management of a CO₂ Flooding Process for Enhancing Oil Recovery Using Artificial Neural Networks. *J. Energy Resour. Technol.* 2017, 140, 032906. [CrossRef]
- Elkatatny, S.; Tariq, Z.; Mahmoud, M. Real time prediction of drilling fluid rheological properties using Artificial Neural Networks visible mathematical model (white box). *J. Pet. Sci. Eng.* 2016, 146, 1202–1210. [CrossRef]
- 19. Abedini, R.; Esfandyari, M.; Nezhadmoghadam, A.; Rahmanian, B. The Prediction of Undersaturated Crude Oil Viscosity: An Artificial Neural Network and Fuzzy Model Approach. *Pet. Sci. Technol.* **2012**, *30*, 2008–2021. [CrossRef]
- 20. Moghadasi, J.; Kazemi, K.; Moradi, S. The Application of Artificial Neural Networks in Determination of Bubble Point Pressure for Iranian Crude Oils. *Pet. Sci. Technol.* **2013**, *31*, 2475–2482. [CrossRef]
- 21. Al-Marhoun, M.A.; Ali, S.S.; Abdulraheem, A.; Nizamuddin, S.; Muhammadain, A. Prediction of Bubble Point Pressure from Composition of Black Oils Using Artificial Neural Network. *Pet. Sci. Technol.* **2014**, *32*, 1720–1728. [CrossRef]
- 22. Tatar, A.; Naseri, S.; Bahadori, M.; Rozyn, J.; Lee, M.; Kashiwao, T.; Bahadori, A. Evaluation of Different Artificial Intelligent Models to Predict Reservoir Formation Water Density. *Pet. Sci. Technol.* **2015**, *33*, 1749–1756. [CrossRef]
- 23. Ahmadi, M.A.; Bahadori, A. A simple approach for screening enhanced oil recovery methods: Application of artificial intelligence. *Pet. Sci. Technol.* **2016**, *34*, 1887–1893. [CrossRef]
- 24. Choubineh, A.; Khalafi, E.; Kharrat, R.; Bahreini, A.; Hosseini, A.H. Forecasting gas density using artificial intelligence. *Pet. Sci. Technol.* **2017**, *35*, 903–909. [CrossRef]
- 25. Moussa, T.M.; Awotunde, A.A. Self-adaptive differential evolution with a novel adaptation technique and its application to optimize ES-SAGD recovery process. *Comput. Chem. Eng.* **2018**, *118*, 64–76. [CrossRef]

- 26. Arabjamaloei, R.; Shadizadeh, S. Modeling and Optimizing Rate of Penetration Using Intelligent Systems in an Iranian Southern Oil Field (Ahwaz Oil Field). *Pet. Sci. Technol.* **2011**, *29*, 1637–1648. [CrossRef]
- 27. Schalkoff, R.J. Artificial Neural Networks; McGraw-Hill: New York, NY, USA, 1997.
- 28. Haykin, S. Neural Networks: A Comprehensive Foundation; Prentice Hall: Upper Saddle River, NJ, USA, 1998.
- 29. Ali, J.K. Neural Networks: A New Tool for the Petroleum Industry? In Proceedings of the European Petroleum Computer Conference, Aberdeen, UK, 15–17 March 1994.
- Van der Aalst, W.M.P.; Rubin, V.; Verbeek HM, W.; van Dongen, B.F.; Kindler, E.; Günther, C.W. Process mining: A two-step approach to balance between underfitting and overfitting. *Softw. Syst. Model.* 2010, *9*, 87. [CrossRef]
- 31. Lippmann, R. An introduction to computing with neural nets. *IEEE ASSP Mag.* **1987**, *4*, 4–22. [CrossRef]
- 32. Jain, A.K.; Jianchang, M.; Mohiuddin, K.M. Artificial neural networks: A tutorial. *Computer* **1996**, *29*, 31–44. [CrossRef]
- 33. Katz, D.L. Prediction of The Shrinkage of Crude Oils. In *Drilling and Production Practice;* American Petroleum Institute: New York, NY, USA, 1942.
- 34. Omar, M.I.; Todd, A.C. Development of New Modified Black Oil Correlations for Malaysian Crudes. In Proceedings of the SPE Asia Pacific Oil and Gas Conference, Singapore, 8–10 February 1993.
- 35. De Ghetto, G.; Villa, M. Reliability Analysis on PVT Correlations. In Proceedings of the European Petroleum Conference, London, UK, 25–27 October 1994.



© 2018 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/).