Evaluation of Different Artificial Intelligent Models to Predict Reservoir Formation Water Density

A. Tatar, S. Naseri, M. Bahadori, J. Rozyn, M. Lee, T. Kashiwao & A. Bahadori


To link to this article: http://dx.doi.org/10.1080/10916466.2015.1091840

Published online: 17 Dec 2015.
Evaluation of Different Artificial Intelligent Models to Predict Reservoir Formation Water Density

A. Tatar,1 S. Naseri,2 M. Bahadori,3 J. Rozyn,4 M. Lee,5 T. Kashiwao,6 and A. Bahadori 4

1 Young Researchers and Elite Club, North Tehran Branch, Islamic Azad University, Tehran, Iran
2 Department of Petroleum Engineering, Ahwaz Faculty of Petroleum Engineering Petroleum University of Technology (PUT), Ahwaz, Iran
3 University of Tehran, School of Soil Science and Engineering, Tehran, Iran
4 Southern Cross University, School of Environment, Science and Engineering, Lismore, NSW, Australia
5 School of Chemical Engineering, Yeungnam University, Gyeongsan, Republic of Korea
6 Department of Electronics and Control Engineering, Niihama College, Niihama, Japan

Nearly all hydrocarbon reservoirs are bounded by water-saturated rocks, namely aquifers. In addition to natural water drive, there is an artificial water drive mechanism in which water is injected into formation to intensify the reservoir pressure. This method, employed to induce the hydrocarbon production, is called water flooding. Several laboratory researches have shown that oil recovery can be heightened by making some alterations to injected brine salinity through water flooding. Accordingly, acquiring exact information about the PVT characteristics of brine is necessary. Density is a property of great importance as it is employed in various physical, chemical, geothermal, and geochemical aspects. The authors aimed to develop a dependable intelligent method to accurately predict the brine density at elevated temperatures and pressures. MLP and GA-RBF models were utilized in this study. The results showed that the proposed model is capable of accurately predicting the brine density at elevated pressures and temperatures for different concentrations of brine. The correlation factor of 1.0000 and root mean squared error of 3.27E-05 demonstrate the accuracy of the proposed model.

Keywords: brine, density, intelligent method, simulated annealing, support vector machine

1. INTRODUCTION

Nearly all hydrocarbon reservoirs are bounded by water-saturated rocks, namely, aquifers. In many cases, a strategy called natural water drive helps to produce oil and gas from hydrocarbon reservoirs. When hydrocarbon is being produced from the reservoir, the reservoir pressure drops, causing water influx from the aquifer in response to the pressure drop (Ahmed, 2006). Increasing the amount of pressure drop that occurs throughout the production life increases the reservoir brine (formation water) that would be produced (Bailey et al., 2000). In addition to natural water drive, there is an artificial water drive mechanism in which water is injected into formation to intensify the reservoir pressure. This method, employed to induce the hydrocarbon production, is called water flooding.
In both mechanisms water coning causes the upward motion of brine, and hence brine would be produced along with hydrocarbons (Zeidani and Bahadori, 2006). Even if excellent field operation approaches are employed, it is probable to produce brine to a degree that constitutes more than 90% of the total volume of produced liquid (Bahadori and Vuthaluru, 2010).

Such brine production can significantly influence hydrocarbon production and impact the extent of depletion and generally recovery efficiency of the reservoirs (Zeidani and Bahadori, 2006). Wet crude, produced from the reservoir, sometimes does not possess a good quality, and in some cases wells should be closed in due to the lack of acceptable treatment facilities (Zeidani and Bahadori, 2006). Furthermore, several laboratory researches have shown that oil recovery can be heightened by making some alterations to injected brine salinity through water flooding (Ogunberu and Ayub, 2007). Accordingly, acquiring exact information about the PVT characteristics of brine is of great importance. Additionally, many of the computations in the petroleum industry require the PVT characteristics of brine in either a direct or indirect way. Thus, providing accurate PVT characteristics of brine is necessary in preventing errors in other estimations of petroleum properties (Yildiz and Morrow, 1996; Morrow et al., 1998; Bagci et al., 2001; Sharma and Filoco, 2000; Spivey et al., 2004).

Among viscosity, thermal conductivity, enthalpy, density, solubility, vapor pressure, and other brine properties, density is considered an important factor. Density is employed in various physical, chemical, geothermal, and geochemical areas, which include high extents of temperatures and pressures. Fluid inclusion researches, simulating fluid flow, and enhanced oil recovery are the areas that require precise knowledge of brine density (Al Ghafri et al., 2013). In addition, brine density is also needed to construct dependable equations of states and computations of other characteristics of brine such as interfacial tension, isothermal compressibility, conversion from molarity to molality, and dynamic viscosity (Al Ghafri et al., 2012a). Among diverse types of solutes, NaCl is one of the most common solutes in natural brines. Therefore precise density information for brines, especially for common ones such as NaCl (aq), are a basic requirement in order to simulate the fluid flow in various areas (Driesner, 2007).

Due to the significance of water formation properties, several studies have been conducted. To find the brine properties, two different approaches have been utilized: conducting laboratory studies, and presenting estimative models. In the experimental field, several researchers have reported the density of brine in various ranges of temperature, pressure, and concentration. Gates and Wood (1985) measured the density of NaCl (aq) and other salts from 0.1013 to 40 MPa and at molality in the range of (0.05–5.0 mol/kg) and at temperature of 298.15 K. Crovetto et al. (1993) reported the vapor pressure and density of NaCl (aq) at a molality of 0.25, 0.5, 1, and 3 mol/kg and temperature of 623 K. Al Ghafri et al. (2012b) measured the density of NaCl (aq) at temperatures between 283 and 472 K; pressures up to 68.5 MPa; and molality of 1.06, 3.16, and 6 mol/kg. Other experimental investigations can be found in the literature (Kumar, 1986; Sharygin and Wood, 1998). Despite a lot of efforts in laboratory methods, experimental approaches are costly and time-consuming (Arabloo et al., 2013). As a result, the latter method is employed in the absence of experimental equipment. In this regard, several studies have investigated the PVT characteristics of brine. Ghafri et al. (Al Ghafri et al., 2013) proposed an empirical correlation using experimental data of various brine samples to estimate their density, apparent molar volume, and isothermal compressibility through the whole spectrum of molality, pressure, and temperature of experimental data. Another density formula of brine is presented by Phillips et al. (1981). In this formula, NaCl (aq) density can be prognosticated through a range of (10–350°C) for temperature, (0.25–5 mol/kg) for molality, and for pressures up to 50 MPa. This model can estimate the experimental data with a maximum deviation of ±2%. Haas (1970) employed the empirical Masson’s Rule to develop a model to predict the density of vapor-saturated NaCl-H2O solutions. This model can do the estimation from 75 to 325°C and up to a saturation of 7.3 molal. Rogers and Pitzer (1982) presented very detailed computations for NaCl
(aq). They provide extensive reports for values of expansivity, specific volume, and compressibility versus temperature, pressure, and molality. They developed a semiempirical equation that has the capability to describe thermal characteristics of NaCl (aq) for a range of (0.1–5 molal) and was utilized to estimate the volumetric data for the temperature of (0–300°C) and pressure of (1–1000 bars). Bahadori et al. (2009) developed an Arrhenius type function to prognosticate the formation water characteristics for a specific range of temperature and salt content. Using the Arrhenius type function, this model avoids some complicated mathematical methods. Besides these models to predict the brine properties, soft computing techniques such as support vector machines (SVMs), fuzzy logic (FL), genetic algorithms (GAs), and artificial neural networks (ANNs) have gained great attention in various researches concerning the petroleum industry due to their great capability to analyze and model vague and complicated subjects. Regarding estimation of the formation water properties, Arabloo et al. (2015) proposed a model employing least squares SVM strategy to prognosticate liquid saturation, vapor pressure, density, and enthalpy of formation water. They indicated that the results of this model are in good agreement with experimental data. Although aforementioned methods are valuable, more studies are required to simplify modeling and increase the precision of models. Consequently, ANN, a type of soft computing technique that has the capacity to solve many complex problems, was used in this study. We aimed to develop a dependable intelligent method to accurately predict the brine density at elevated temperatures and pressures. Due to its great flexibility and power in pattern recognition as well as handling large databases, MLP and GA-RBF intelligent methods are utilized in this study. In order to determine the optimum values of tuning parameters, coupled simulated annealing is utilized.

2. DETAILS OF INTELLIGENT MODEL

In the stage of modeling, it is desired to develop a mathematical model based on statistical learning theory to relate the input variables to the output. This model learns and recognizes the governing pattern among the input variables and the output, and is capable of generalizing the results to the unseen data by the network. In the recent decade, several machine-learning methods are evolved and they are used to solve numerous complex problems in science and engineering.

2.1 Multilayer perceptron networks

Multilayer perceptron (MLP) neural networks consist of three different layer types called input, hidden, and output layers. Each MLP neural network has one input and one output layer and one or more hidden layers. Each layer consists of some neurons. The number of neurons in the input and output layers is corresponded to the number of input and output variables, respectively. The number of neurons in the hidden layer(s) is optional and can be determined either by trial and error or incorporating an optimization method. In such networks, the performance is determined by a parameter called mean squared error (MSE). The error is back-propagated through the network and weights and biases are adjusted through some iteration called epochs. The number of iterations should be such that the network neither overtrain nor undertrain. In the former the network memorizes instead of learning and in the latter the network does not have enough time to learn the pattern.

2.2 Radial basis function networks

The structure of radial basis function (RBF) neural networks is somewhat different from MLP neural networks. The main advantage of RBF neural networks is their easier design. MLP neural networks
TABLE 1
The statistical parameters of the input and out parameters of the network

<table>
<thead>
<tr>
<th>Input parameter</th>
<th>Status</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Average</th>
<th>Standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure (bar)</td>
<td>input</td>
<td>100</td>
<td>1000</td>
<td>466.6667</td>
<td>309.1732</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>input</td>
<td>5</td>
<td>350</td>
<td>177.5</td>
<td>101.0432</td>
</tr>
<tr>
<td>NaCl concentration (Molality)</td>
<td>input</td>
<td>0.25</td>
<td>5.00</td>
<td>2.25</td>
<td>1.685305</td>
</tr>
<tr>
<td>Density g/cm³</td>
<td>output</td>
<td>0.608087</td>
<td>1.250468</td>
<td>0.979551</td>
<td>0.109701</td>
</tr>
</tbody>
</table>

have the fixed structure of three layers of input, output, and just one hidden layer. RBF neural networks benefit from high power of generalization, high tolerance of noises, and online learning (Santos et al., 2013). Regarding the power of generalization, RBF neural networks respond very well to the data unseen by the network during the training process (Yu et al., 2011). Localized basis functions and iterative function approximation form the basis of the RBF neural networks (Broomhead and Lowe, 1988; Dayhoff, 1990; Huang and Zhang, 1994; Zurada, 1992). This kind of network is a type of feed-forward neural network and uses a supervised training scheme. Due to a simpler structure of fixed three layers, the training process is much faster in RBF networks compared to MLP networks. The classification technique in MLP and RBF networks is quite different. The former uses hyper surfaces to separate the clusters while the latter utilizes hyper spheres.

3. RESULT AND DISCUSSION

3.1 Data acquisition

In order to develop a dependable model it is necessary to use valid data that covers a wide range of variables (Shokrollahi et al., 2015; Tatar et al., 2013; Tatar et al., 2015). The data used in this study was gathered from open literature (S. L. Phillips et al., 1981). The aim of this study is to determine the NaCl brine density over a range of pressures (100–1000 bar) and temperatures (5–350 °C) in different NaCl concentrations. The brine density is a function of temperature, pressure, and salt concentration. In this study the input variables are pressure (bar), temperature (°C), and NaCl concentration. A total of 2,940 data points were used to develop and evaluate the proposed model. The statistical values of the input and output variables are listed in Table 1.

3.2 Model development

First, the dataset was divided into two sets of train (2,352 data points) and test (588 data points) such that there is no local concentration of train or test data. Then, the MLP and GA-RBF were developed. The details of development of each model are presented in following. Matlab 2014a (The MathWorks, Natick, MA) is utilized for implementing MLP code. Cybenko (1989) stated that it is mathematically proved that every function can be estimated adequately by an ANN with only one hidden layer. Therefore, only networks with one hidden layer are investigated in this study. The number of neurons in the input and output layer are fixed and equal to the number of input and output parameters respectively. Different neuron numbers of 4–25 were tested for the hidden layer and the network performance was monitored. It was concluded that 16 neurons in the hidden layer results in the best performance of the MLP network. In Matlab 2014a the RBF network training code has some adjusting parameters among which spread and maximum number of neurons (MNN) are of
great importance. The aforementioned parameters have the most influence on the performance of the GA-RBF network. The optimal values of spread and MNN result in the best performance of the GA-RBF neural network. In this study the values of the aforementioned parameters were determined by trial and error to be 0.640 and 1077 for spread and MNN respectively.

3.3 Accuracy of the proposed model and validation

Both statistical and graphical attempts were utilized to show the accuracy of the developed methods. The scatter plot of the predicted versus the experimental values is depicted in Figure 1. In this figure, the vertical and horizontal axes denote the predicted and experimental values, respectively. Closer predicted values to the experimental ones results in accumulation of data points on the line. As it is evident, both the train and test data points are concentrated in close vicinity of the line for the trio of developed models. The relative error deviation is depicted in Figure 2. As it is shown, the maximum absolute relative error for the trio models is 0.5%. As it is depicted, the GA-RBF method has the least error deviation. Figure 3 shows the simultaneous depiction of experimental and predicted data.
TABLE 2
The Statistical Parameters of the Developed CSA-LSSVM Model

<table>
<thead>
<tr>
<th>Method</th>
<th>Data set</th>
<th>$R^2$</th>
<th>AARD</th>
<th>STD</th>
<th>RMSE</th>
<th>$N^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP</td>
<td>Train Data</td>
<td>1.0000</td>
<td>0.002102</td>
<td>0.000106</td>
<td>8.16E $-05$</td>
<td>2352</td>
</tr>
<tr>
<td></td>
<td>Test Data</td>
<td>1.0000</td>
<td>0.002108</td>
<td>9.66E $-05$</td>
<td>6.35E $-05$</td>
<td>588</td>
</tr>
<tr>
<td></td>
<td>All Data</td>
<td>1.0000</td>
<td>0.002103</td>
<td>0.000104</td>
<td>7.83E $-05$</td>
<td>2940</td>
</tr>
<tr>
<td>RBF</td>
<td>Train Data</td>
<td>1.0000</td>
<td>0.000664</td>
<td>3.17E $-05$</td>
<td>3.37E $-05$</td>
<td>2352</td>
</tr>
<tr>
<td></td>
<td>Test Data</td>
<td>1.0000</td>
<td>0.000577</td>
<td>1.22E $-05$</td>
<td>1.28E $-05$</td>
<td>588</td>
</tr>
<tr>
<td></td>
<td>All Data</td>
<td>1.0000</td>
<td>0.000627</td>
<td>2.88E $-05$</td>
<td>3.07E $-05$</td>
<td>2940</td>
</tr>
</tbody>
</table>

*Number of data points

for total data sets. As can be seen, the predicted values are highly consistent with the experimental data.

The utilized statistical parameters in this study are correlation factor, absolute average relative deviation (AARD), root mean squared error (RMSE), and standard deviation (STD). The formulation of the mentioned parameters is shown in Eqs. (1) to (4).

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (\rho_{\text{Pr pred}}(i) - \rho_{\text{Exp}}(i))^2}{\sum_{i=1}^{N} (\rho_{\text{Pr pred}}(i) - \bar{\rho}_{\text{Exp}})^2} \tag{1}
\]

\[
\text{%AARD} = \frac{100}{N} \sum_{i=1}^{N} \frac{(\rho_{\text{Pr pred}}(i) - \rho_{\text{Exp}}(i))}{\rho_{\text{Exp}}(i)} \tag{2}
\]

\[
RMSE = \left( \frac{\sum_{i=1}^{N} (\rho_{\text{Pr pred}}(i) - \rho_{\text{Exp}}(i))^2}{N} \right)^{0.5} \tag{3}
\]

\[
STD = \left( \frac{\sum_{i=1}^{N} (\rho_{\text{Pr pred}}(i) - \bar{\rho}_{\text{Exp}}(i))^2}{N} \right)^{0.5} \tag{4}
\]

The mentioned parameters for the train, test, and total data sets for MLP and GA-RBF are listed in Table 2. All the presented models has yielded correlation factor ($R^2$) equal to 1.0000. The very low value of RMSE indicates the agreement between the predicted values and the experimental ones.

FIGURE 3  Simultaneous projection of predicted values for the brine density.
4. CONCLUSION

In this study, MLP and GA-RBF models were applied successfully to predict the brine density in a wide temperature and pressure range in different NaCl concentrations. Data samples in this study were gathered from open literature (Phillips et al., 1981). A total number of 2,940 experimental data points were utilized, among which 2,352 data points were allocated for training and the rest (588 data points) were used to validate the proposed model. Pressure, temperature, and NaCl concentration were considered as the input of the model and the brine density was the output. The intelligent models were developed using the train data set and was successfully validated by the data unseen to the network during training. Coupled simulated annealing was used as the optimization algorithm to find the optimum values of the tuning parameters. Based on the results, it was found that the developed MLP and GA-RBF models predict the NaCl brine density with an average absolute error deviation of 0.0005%. Thus, the proposed model provided accurate estimation of NaCl brine densities without performing experimental activities and complex formulations.

REFERENCES

Al Ghafri, S., Maitland, G. C., and Trusler, J. M. (2012). Densities of Aqueous MgCl2 (aq), CaCl2 (aq), KI (aq), NaCl (aq), KC1 (aq), AlCl3 (aq), and (0.964 NaCl + 0.136 KCl)(aq) at Temperatures Between (283 and 472) K, Pressures up to 68.5 MPa, and Molalities up to 6 mol·kg−1. J. Chem. Eng. Data 57:1288–1304.


