Petroleum 2 (2016) 40–48

Contents lists available at ScienceDirect

Petroleum

journal homepage: www.keaipublishing.com/en/journals/petlm

Original article

The most accurate heuristic-based algorithms for estimating the oil formation volume factor

Mohammad Reza Mahdiani*, Ghazal Kooti

Faculty of Petroleum Engineering, Amirkabir University of Technology (Tehran Polytechnic), Hafez Avenue, Tehran, Iran

A R T I C L E I N F O

Article history:
Received 6 November 2015
Received in revised form
30 December 2015
Accepted 30 December 2015

Keywords:
Genetic programming
Neural network
Modeling

A B S T R A C T

There are various types of oils in distinct situations, and it is essential to discover a model for estimating their oil formation volume factors which are necessary for studying and simulating the reservoirs. There are different correlations for estimating this, but most of them have large errors (at least in some points) and cannot be tuned for a specific oil. In this paper, using a wide range of experimental data points, an artificial neural network model (ANN) has been created. In which its internal parameters (number of hidden layers, number of neurons of each layer and forward or backward propagation) are optimized by a genetic algorithm to improve the accuracy of the model. In addition, four genetic programming (GP)-based models have been represented to predict the oil formation volume factor In these models, the accuracy and the simplicity of each equation are surveyed. As well as, the effect of modifying of the internal parameters of the genetic programming (by using some other values for its nodes or changing the tree depth) on the created model. Finally, the ANN and GP models are compared with fifteen other models of the most common previously introduced ones. Results show that the optimized artificial neural network is the most accurate and genetic programming is the most flexible model, which lets the user set its accuracy and simplicity. Results also recommend not adding another operator to the basic operators of the genetic programming.

1. Introduction

Oil formation volume factor (Bo) is one of the PVT properties of oil, which is the ratio of a specific petroleum volume at the reservoir condition to its corresponding volume at standard condition. Bo's value is essential in calculating various parameters such as the depletion rate, oil in place, predicting the future of the reservoir, optimizing the rate of production and some other simulation and optimization problems. Distinctive correlations are used whenever the experimental value for Bo of a specific oil at particular pressure and temperature is not available. Thus, an accurate correlation for different ranges of parameters is necessary. There are numerous correlations for estimating Bo in the literature, but most of them have poor prediction for some points even in their applicable range. In addition to that they can't be tuned for a special oil, and their accuracy and speed can't be changed base on the need of the problem. They are not flexible to be used in different problems and also there is not a comprehensive study to help users to select one for their studies. Here, one artificial neural network (ANN), one hybrid of ANN and genetic algorithm, four genetic programming (GP) methods, and fifteen previously introduced correlations will be surveyed, and their strengths and weaknesses will be discussed. Considering the pressure, reservoirs can be classified into three categories: below the bubble point pressure, at bubble point pressure and above the bubble point pressure. Usually, there is some free gas in the reservoir at pressures below the bubble point. Above that, Bo changes because of the oil and gas compressibility. Knowing the value of Bo at bubble point condition, Bo at higher pressures can be easily

* Corresponding author.
E-mail address: mahdiani@aut.ac.ir (M.R. Mahdiani).
Peer review under responsibility of Southwest Petroleum University.

Production and Hosting by Elsevier on behalf of KeAi

http://dx.doi.org/10.1016/j.petlm.2015.12.001

calculated using compressibility factor. Thus, most of the correlations estimate Bo at bubble point pressure [1–3]. The models of this paper also estimate the B_o at bubble point pressure.

Different correlations have been represented for predicting the oil formation volume factor. In 1942, Katz [4] using 117 measurement data developed a correlation to predict the oil formation volume factor using gas gravity, oil gravity, gas oil ratio, and reservoir pressure and reservoir temperature. In 1947, Standing [5–8] created a correlation which was the first correlation that needs only four parameters (gas gravity, oil gravity, and solution gas oil ratio and reservoir temperature) to estimate the oil formation volume factor. Vasquez and Begg [9](1980) developed a correlation which was highly dependent on gas gravity. In 1980, Glaso [10] introduced a correlation for a methane contained black oil based on gas–oil equilibrium. Al-Marhoun [11,12] (1988) used the linear regression for Middle East oil. After that a lot of researchers tried to introduce coefficients for previous correlations. Among them are Abdul-Majeed, 1988 [13], Dokla (1992) [14], Al-Marhoun (1992) [15], Petrosky (1993) [17], Omar (1993) [18], Kartoamodj 1994 [19], Farshad 1996 [20]. In more recent years some researchers tried to introduce completely new correlations. Maybe by adding parameters or different techniques. Sutton [24] developed a correlation for density and using that estimated Bo. Sebakhy [25] (2009) used support vector machine modeling Nassar (2013) [26]considered separator pressure, temperature and its configuration (two or three stage) and introduced a correlation for Bo, In 2014 Sulaimon [27] used a group method of data handling to create a model.

As it can be seen, there are different correlations for the oil formation volume factor in the literature. Now, the question is which one of them is more accurate. There are a lot of studies that compare these correlations. In 1994, using 195 data points, Ghetto [28]compared different correlations and suggested the Vasquez and Begg model for estimating the oil formation volume factor. In 1997, Almehaideb [22] used the data of 13 oil fields in UAE to show the performance of common correlations. He suggested Al-Marhoun, Glaso, and Standing as the top three accurate correlations. In 2012, Godefroy [29] compared various correlations and suggested Glaso and Al-Marhoun to estimate Bo. Unfortunately none of the above studies have a comprehensive investigation on distinctive models, they are old and are not surveyed the modern models and the new methods for modeling such as neural network, genetic programming and simulated annealing programming [46,47] and the database of most of them is not wide enough. Here, using an extensive database, 15 models of the most common literature correlations in addition to modern models such as neural network and different genetic programming models are compared and their strengths and weaknesses are illustrated.

2. Methodology

As mentioned earlier, here using some artificial neural network and genetic programming some new models are created and compared with the most common previously introduced ones in literature. For this purpose there is need to have a dataset of the measured oil formation volume factors in various conditions. This data set, gained from different literature works and experiment, consists of 160 data points of the Middle East oil of Al-Marhoun [12] study, 195 points of Ghetto [28] from Persian gulf, middle east and Africa oils, 41 data point of Glaso [10] from north sea, 177 points of Katz [4] study 93 points of Malaysian oil of Omar's study as well as 200 other experimental data points which were measured of Iranian south oil fields. The number of all points was 866. From them 395 points that they were not used in building of any surveyed correlations of this paper put aside for comparing different correlations and 471 points used for building new models for oil formation volume factor using artificial neural network and genetic programming.

Table 1
Range of parameters, used in building the model.

	Gas gravity	Oil gravity	Temperature (°F)	Solution gas oil ratio (SCF/STB)
Minimum	0.612	0.766	100	104
Maximum	1.315	0.895	280	1439

Artificial neural network (ANN) is a modeling method that is based on biological neurons. This network is consisted of units called artificial neurons. These neurons receive inputs and sum them to produce the output. Each neuron has a weight and thus the summation is weighted. The sum is passed through a function called transfer function. And finally the model is created [30]. ANN has been used widely in the different courses of science [31–33].

To create the ANN model, the same data that were used in creating the genetic programming models is used. Their range is shown in Table 1 and the property of the neural network model is shown In Table 2. In this Table, 'Levenberg–Marquart' is the name of a method; for more information see Ref. [34]. As shown in this table, type of neural network is a forward propagation and all its details are listed in that. Here using artificial neural network a model for estimating the oil formation volume factor created. In next section the created ANN model statistical properties will be discussed and compared with other B_o correlations.

Artificial neural network is a good method but it is very random base and different runs of that, results to different models. As well as changing its internal parameters such as the number of hidden layers, the number of neurons of each layer and having forward or backward propagation can change the final model and its efficiency. Thus, here ANN is coupled with GA. GA optimizes the number of hidden layers and the neurons of each layer to improve the quality of the models in addition to that it selects the type of propagation (forward or backward). The basic properties of the artificial neural network are mentioned in Table 2. The properties of GA are listed in Table 3. For knowing more about the parameters of this table the interested reader can refer to [35]. As mentioned before. For that reason in each iteration for every set of inputs (number of hidden layers and number of neurons of each layer and the type of propagation) five times the ANN ran and the best model selected. The output of fitness function of genetic algorithm is average relative error that should be minimized. Five was selected because it is not so small that random structure of algorithm affects that and is not so large that causes the low speed of the algorithm. In this study,

Table 2
The parameters of the artificial neural network model used in this study.

Parameter	Value
Type of ANN	MLPs, forward
Training	Levenberg–Marquart
Performance	Mean Square Error
Stop tolerance	1.00E-05
Number of nodes of each layer	8
Number of input layers	4
Number of hidden layers	10
Number of output layers	1