

Development of New Models for Predicting Crude Oil Bubble Point Pressure, Oil Formation Volume Factor, and Solution Gas-Oil Ratio Using Genetic Algorithm

Abdelwahab, M. H.¹, Soliman, A. A.^{1,2,*}, Attia, A. M.^{1,2}

¹Petroleum Engineering Department, Faculty of Engineering, British University in Egypt (BUE), Elshorouk city, Cairo, Egypt

²Faculty of Energy and Environmental Engineering, British University in Egypt (BUE), Elshorouk city, Cairo, Egypt

*Corresponding author e-mail: ahmed.abdelhafez@bue.edu.eg

Article Info

Received 6 June, 2020

Revised 12 Aug, 2020

Accepted 1 Sep, 2020

Abstract

Bubble point pressure (P_b), oil formation volume factor (B_o), and solution gas-oil ratio (R_s) are considered the key parameters required to describe and characterize the crude oil. Accurate determination for crude oil properties are necessary for multi-operation in reservoir evaluation, such as reserve estimation, enhanced oil recovery (EOR), oil reservoir performance prediction, designing pipelines and production equipment, and reservoir simulation. Traditional techniques used to calculate PVT data are usually expensive or unavailable, so there are a huge number of empirical correlations developed to estimate PVT properties as a function of production data. But when we used these correlations to predict crude oil properties, big errors are attained. The main target of this study is to find a better and accurate approach for predicting the properties of crude oil. This paper developed new empirical correlations for predicting the properties of reservoir oil as a function of PVT properties such as (P , T , B_o , R_s) using genetic algorithm technique (GA). The simulation model is built using MATLAB software which contains the optimization tool that includes a genetic algorithm tool in it. To validate these correlations, 130 data sets of different crude oils were used. The results obtained showed that the developed empirical correlations from the genetic algorithm model (GA) appeared excellent accuracy of predicting crude oil properties compared to their relevant published correlations. The average absolute error for all correlations that the genetic algorithm applied to them is decreased. This technique can be applied to predict crude oil properties with a high level of accuracy.

Keywords

PVT properties; Bubble point pressure; oil formation volume factor; solution gas-oil ratio; empirical correlations; Genetic algorithm.

Introduction

Dake [1] reported that the evaluation of the oil reserves, prediction of future production performance, planning for Enhanced oil recovery projects, and designing the facilities of production require the PVT characteristics of the oil reservoir. According to Dake, the importance of PVT analysis was concentrated on the relation between the volume of gas that is produced at the surface to the corresponding underground formation withdrawal. Accurate determination of reservoir rock properties and reservoir fluid properties plays a crucial role in estimating original oil in place and original gas in place [2,3].

The properties of the oil reservoir involve oil formation volume factor, solution gas/oil ratio, and oil bubble point pressure. The oil formation volume factor (B_o) is described as the ratio between the oil reservoir volume (in addition to the dissolved gas in solution) at pressure and temperature of reservoir to the volume of oil at the standard conditions. All the Calculations of the material

balance require oil formation volume factor to establish the oil reservoir volume after the process of reservoir depletion. The bubble point pressure (P_b) is defined as the pressure at which the first bubble of the gas will release from the liquid oil solution. Bubble point pressure is very important for the analysis of the future performance of wells, calculations of production engineering, and reservoir simulation. Solution gas/oil ratio (R_s) determines the amount of gas dissolved in liquid oil at any pressure and temperature. Besides, the Solution gas/oil ratio and oil formation volume factor is a strong function with the reservoir pressure [4].

Commonly, PVT properties are calculated in an experimental laboratory by using bottom hole samples or using surface samples at reservoir condition which are temperature and pressure. If there is no data available for use, so an empirical correlation will be used to compute PVT properties, however these correlation successes or not, the predicted data depends mainly on the range of data, not on the correction of these data [5]. There are

several empirical equation correlations were developed to predict PVT properties.

Standing [6] developed a correlation about bubble point pressure, formation volume factor for oil, and formation volume factor for gas plus liquid phase as an empirical function of pressure, temperature, solution gas-oil ratio (R_s), gas gravity (γ_g), and oil gravity (γ_o), using a graphical solution, which depends on 105 samples. The following correlations mainly established on California crude oils and gases:

$$P_b = 18.2 \left[\left(\frac{R_{sb}}{\gamma_g} \right)^{0.83} * 10^{(0.00091T - 0.0125\gamma_{API}) - 1.4} \right] \quad (1)$$

$$R_s = \gamma_g \left[\left(\frac{P_b}{18.2} + 1.4 \right) \times 10^{(-0.00091T + 0.0125\gamma_{API})} \right]^{1.2048} \quad (2)$$

$$B_{ob} = 0.972 + 1.47 \times 10^{-4} \left[R_s \left(\frac{\gamma_g}{\gamma_o} \right)^{0.5} + 1.25 T \right]^{1.175} \quad (3)$$

Lasater [7] developed empirical correlation for the bubble point pressure where the standard physical-chemical equation is used to develop this correlation. The correlation is predicated on 158 experimentally measured bubble point pressures. This correlation is based totally on the same parameter established by Standing, but the data used in that system produced in Canada. The final form for bubble point pressure correlation is:

$$P_b = \frac{(Pf)(t+459.6)}{\gamma_g} \quad (4)$$

Vazquez et al. [8] reported that the development of an empirical correlation to estimate oil properties like oil formation volume factor and solution gas-oil ratio, it was needed laboratory result, so they use more than 600 samples of crude oil to develop this empirical correlation. They have a large database that contains around 6000 data points where these data points measured over a large scale of several pressure, temperature, oil gravity, and gas gravity. The solution gas-oil ratio correlation that presented was originally to expect the value of solution gas, but this correlation can be arranged to estimate bubble point pressure. There are other correlations of this type, but the limitation of this correlation to work for pressure below bubble point pressure.

$$P_b = \left[\frac{C_1 R_{sb}}{\gamma_{gs}} 10^{\frac{-C_3 \gamma_{API}}{(T+459.67)}} \right]^{1/C_2} \quad (5)$$

$$R_{sb} = \left[\frac{\gamma_g P_b^{C_2}}{C_1} 10^{\frac{C_3 \gamma_{API}}{(T+459.67)}} \right] \quad (6)$$

$$B_{ob} = 1 + C_1 R_{sb} + C_2 (T - 60) \left(\frac{\gamma_{API}}{\gamma_{gs}} \right) + C_3 R_{sb} (T - 60) \left(\frac{\gamma_{API}}{\gamma_{gs}} \right) \quad (7)$$

Glaso [9] stated that pressure, volume, and temperature correlations are important tools for the development of reservoir technology. The uses of these measurements are estimating the amount of hydrocarbon in the reservoir, production capacity, and the change in gas/oil ratio during production time. PVT correlations are also needed to calculate the recovery efficiency of the reservoir. Especially in the exploration period, the PVT correlations are so important, at the time that only fluids that produced are available from flowing tests, so we are

heading to empirically derived PVT correlations. According to Glaso, there are two different reasonable factors in standing original correlation, the first one is the hydrocarbon from different region which have different properties, and the second one is the surface gases from different reservoir which have a different amount of CO_2 , H_2S , and N_2 .

$$\log P_b = 1.7669 + 1.7447 \cdot \log P_b^* - 0.30218 (\log P_b^*)^2 \quad (8)$$

$$P_b = 10^{\log(P_b)} \quad (9)$$

Al-Marhoun [10] compared his correlations that resulted from non-linear regression with the correlations of Standing and Glaso which develop their correlation by graphic estimation, curve-fitting, and linear regression respectively. Al-Marhoun confirmed that the data used to predict these correlations exclusively for the middle east sample, so these correlations must be valid for any samples falling in the range of properties of data used for this study. Al-Marhoun used around 69 bottom hole sample to make development to these correlations, these correlations presented as a function of the stock tank oil gravity, Solution gas-oil ratio, gas gravity, and finally the temperature of reservoir. All these previous functions must be at bubble point pressure. In contrast to Standing and Glaso, Al-Marhoun used non-linear regression method to make development to these correlations:

$$P_b = 5.38088 \times 10^{-3} R_{sb}^{0.715082} \gamma_g^{-1.87784} \gamma_o^{3.1437} (T+459.67)^{1.32657} \quad (10)$$

$$R_{sb} = \left[\frac{P_b \gamma_g^{1.87784}}{0.00538088 \gamma_o^{3.1437} (T+459.67)^{1.32657}} \right]^{1/0.71508} \quad (11)$$

$$B_{ob} = 0.497069 + 8.26963 \times 10^{-4} (T + 459.67) + 1.82594 \times 10^{-3} F + 3.18099 \times 10^{-6} F^2 \quad (12)$$

$$F = R_{sb}^{0.74239} \gamma_g^{0.323294} \gamma_o^{-1.202040} \quad (13)$$

Kartoatmodjo et al. [11] mentioned the importance of study black oil properties and what should be considered in this study. Kartoatmodjo & Schmidt make the development of a new generation of empirical correlation which based on a huge number of these data that used for developing these empirical correlations, these data came from reservoir all over the world. Kartoatmodjo & Schmidt used two different databases, which used to develop these correlations. The first database that mainly used to develop correlations, on the other hand, the second database which used mainly as verification for their correlations. The first data set consist of 740 different crude oil samples (5392 points), but the second data set consist of 998 data point. These new empirical PVT correlations developed due to gas specific gravity to a source separator pressure, oil formation volume factor, solution gas-oil ratio, bubble point pressure, dead oil viscosity, live oil viscosity, under-saturated oil viscosity, isothermal compressibility, and a conversion factor of gas liberation.

The new correlations established by Kartoatmodjo & Schmidt are:

$$P_b = \left[\frac{R_{sb}}{0.05958 \gamma_g^{0.7972} 10^{13.1405 \gamma_{API} / (T+459.67)}} \right]^{0.9986} \quad (\text{API} \leq 30) \quad (14)$$

$$P_b = \left[\frac{R_{sb}}{0.03150 \gamma_g^{0.7587} 10^{11.2895 \gamma_{API} / (T+459.67)}} \right]^{0.9143} \quad (\text{API} \geq 30) \quad (15)$$

$$R_{sb} = 0.05958 \gamma_g^{0.7972} p^{1.0014} 10^{13.1405 \gamma_{API} / (T+459.67)} \quad (\text{API} \leq 30) \quad (16)$$

$$R_{sb} = 0.03150 \gamma_g^{0.7587} p^{1.0937} 10^{11.2895 \gamma_{API} / (T+459.67)} \quad (\text{API} \geq 30) \quad (17)$$

$$B_o = 0.98496 + 0.0001 \times F^{1.50} \quad (18)$$

$$F = R_{sb}^{0.755} \gamma_g^{0.25} \gamma_o^{-1.5} + 0.45T \quad (19)$$

Gharbi et al. [12] mentioned the important role that pressure, volume, and temperature (PVT) properties played in the petroleum field. To estimate these properties there are two types of methods the first one by using the equation of state and the second one by using PVT correlations. The main target for Gharbi & Elsharkawy is to develop a new empirical correlation using a neural network. According to Gharbi & Elsharkawy, the neural network can predict bubble point pressure and oil formation volume factor, these predictions have done as a function in those properties' solution gas-oil ratio, gas gravity, and oil gravity.

Dindoruk et al. [13] established a new empirical correlation for developed correlations of Mexico crude oil gulf by using a function of a field data. These correlations are established for: Solution gas-oil ratio, Bubble point pressure, Under-saturated isothermal oil compressibility, and Oil formation volume factor at bubble point pressure. Dindoruk & Christman mentioned that there are many numbers of data used for developing a new correlation for Gulf of Mexico crude oil, there are over one hundred PVT data, Dindoruk & Christman also reported that Standing and Farshad correlations used to test Gulf of Mexico crude oil data, they mentioned that their correlation is more accurate for Gulf of Mexico crude oil data than the past correlation. They reported that to use this correlation of their study, they build a program in the form of simulation that can be generated PVT data that used to develop their correlation. The most important parameter in correlation parameters is gas gravity, this gas gravity which comes from that simulation program, but unfortunately this parameter on of the variable that measured with the highest degree of uncertainty. Pressure and temperature which is the gravity of evolved gas depend on, but it wasn't available.

Hemmati et al. [14] mentioned that there is no specific correlation for Iran crude oil established before, in this paper they used to around 300 samples of data to develop a new correlation for Iran crude oil. The correlations that developed are bubble point pressure correlation, solution gas-oil ratio correlation, and oil formation volume factor correlation. Hemmati & Kharrat also mentioned that there are many numbers of data used for developing a new correlation for Iranian crude oil, there are over one hundred PVT data. The most important parameter in correlation parameters is gas gravity, this gas gravity which comes from that simulation program, but unfortunately this parameter on of the variable that measured with the highest degree of uncertainty.

Moreover, Hemmati & Kharrat developed correlations that contain equations to estimate solution gas-oil ratio, oil formation volume factor and bubble point pressure. In the solution gas-oil ratio equation, there are two types of equations, the first one at bubble point, which is like every correlation before, and the second one below bubble point pressure.

Mazandarani et al. [15] developed empirical correlations depending on the correlations of Al-Marhoun to estimate the oil formation volume factor at bubble point pressure, solution gas-oil ratio, and bubble point pressure. The technique was used to develop these correlations that was Multiple regression analysis. According to Mazandarani & Asghari, the evaluation was conducted through using 55 of data sets of samples of reservoir fluid came from different locations in Iran. The following developed correlations can be used to predict (P_b , R_s , B_o) with high accuracy:

$$P_b = 1.09373 * 10^{-4} R_s^{0.5502} \gamma_g^{-1.71956} \gamma_o^{2.5486} (T + 460)^{2.0967} \quad (20)$$

$$R_s = 994.3718 \gamma_g^{2.113367} P_b^{1.45558} \gamma_o^{-5.48944} (T + 460)^{-1.90488} \quad (21)$$

$$B_{ob} = 0.99117 + 0.00021 R_s - 2.32 * 10^{-6} R_s \left(\frac{\gamma_g}{\gamma_o} \right) - 4.30 * 10^{-7} R_s (T - 60)(1 - \gamma_o) + 0.00071 (T - 60) \quad (22)$$

Ikiensikimama et al. [16] stated that the best important factor for them is knowledge, to establish any correlation the knowledge of bubble point pressure must be known. The bubble point pressure resulted from the material balance equation. Ikiensikimama & Ogboja mentioned that there are many numbers of data used for developing a new correlation for Nigerian crude oil, there are over one hundred PVT data. The most important parameter in correlation parameters is gas gravity, this gas gravity which comes from that simulation program, but unfortunately this parameter on of the variable that measured with the highest degree of uncertainty. The important role that pressure, volume, and temperature (PVT) properties played in the petroleum field. To estimate these properties there are two types of methods the first one by using the equation of state and the second one by using PVT correlations. The main target for Ikiensikimama & Ogboja is to develop a new empirical correlation using MATLAB.

Sadiq et al. [17] presented a new correlation utilizing nonlinear regression technique for calculating the oil formation volume factor as a function of reservoir pressure, reservoir temperature, gas specific gravity, and oil gravity at and below the bubble point pressure. Sadiq & Hassn used more than thirty reports of PVT data which taken from many Iraqi fields, and also used cross plot and statistical techniques to check the correlation performance. The tests showed that the following new correlation gave a good fitting with Iraqi oils experimental data:

$$B_o = a_1 P^{a_2} T^{a_3} API^{a_4} \gamma_g^{a_5} + a_6 \quad (23)$$

Hassan [18] formulated a new correlation using multiple linear regression technique to calculate the solution gas-oil ratio below the bubble point pressure. The building of new correlation is based on thirty-seven reports of PVT data which collected together from different Iraqi fields. Hassan used Statistical and graphical tools to be able to check the new correlation performance. The results showed that the following new correlation to predict the solution gas-oil ratio with five variables (reservoir pressure, reservoir temperature, the oil gravity, bubble point pressure, and relative gas density) has high accuracy than the original laboratory data:

$$R_s = A_0 P_b^{A_1} \gamma_g^{A_2} T^{A_3} API^{A_4} R_{sb}^{A_5} P^{(A_6 P_b^{A_7} \gamma_g^{A_8} T^{A_9} API^{A_{10}} R_{sb}^{A_{11}})} \quad (24)$$

El-Hoshoudy et al. [19] used three laboratory tests which are differential vaporization test, primary study, and constant mass depletion to understand and characterize the phase behavior of hydrocarbon fluids (Black Oils). The study of PVT properties is very important for petroleum engineers especially reservoir engineers to detect the hydrocarbon fluids phase behavior and initiate the material balance equation.

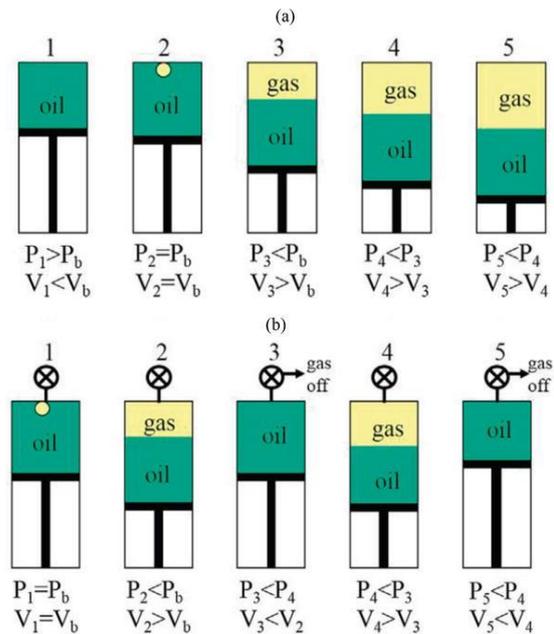
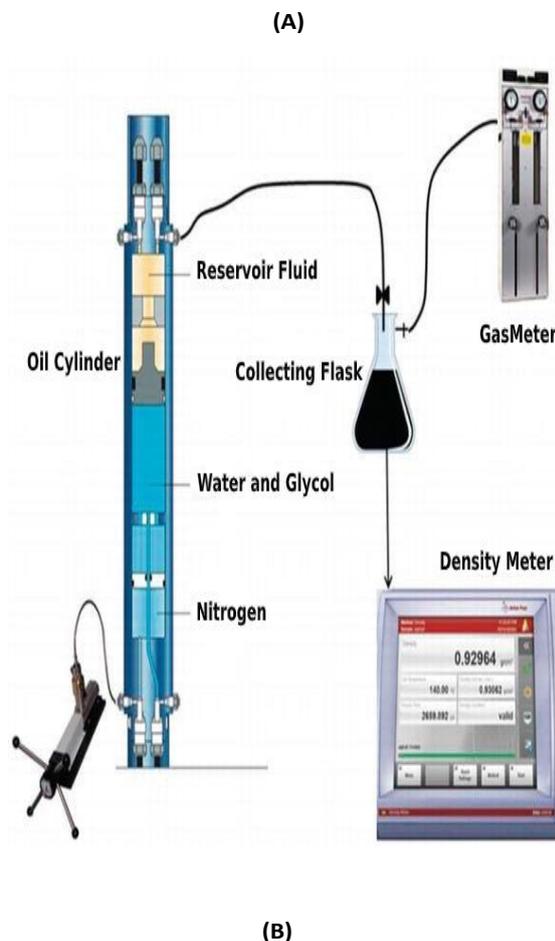


Figure 1 (A): Diagram of Primary study Steps; (B) Schematic illustration of (a) constant mass depletion, (b) differential liberation.

The objective of this research is to develop a new model to estimate the Bubble point pressure, oil formation volume factor, and solution gas-oil ratio based on the pressure and temperature of the reservoir, gas specific gravity, and stock tank oil API gravity. A new model using the genetic algorithm by Matlab was used to develop the (P_b , B_o , R_s) models. The obtained results of this model will be compared based on the highest correlation coefficient and the lowest average absolute error percent.

Genetic Algorithm Techniques

A genetic algorithm is one of the most essential computing algorithms and optimization techniques. These algorithms can be used to find out the minimum and maximum of each function used for. Genetic algorithm is one of many optimization algorithms that used to find the optimum solution for computable problems, also it maximizes and minimizes the function used by genetic algorithm. The genetic algorithm understands the decision variables of solving a problem into strings which are finite-length of alphabetic elements in a group. Those strings are the only solution to solve these problems and are expressed in chromosomes. The alphabets are called as genes, but the value of those genes is called alleles, the genetic algorithm technique work different rather than any other solution methods, it's depended on the coding of parameters and it does not depend on parameter themselves.

JH Holland [20] reported that a genetic algorithm is based on some principles which are search methods, these principles such as (selection and genetics). The Holland theory still the theory that existed in most of the recent theories, which make his came firstly for the model introduced by Holland. The recent theory that used genetic modeling also applies canonical genetic algorithm.

Pelikan et al. [21] stated that the genetic algorithm techniques depend on the notion of population which is considered as the main theory in genetic algorithm techniques, which differ from any other traditional methods, genetic algorithm depending on population for expected solution. The population size, that is usually considered as the most significant factor that affects the result and performance of the genetic algorithm. This population size is specified as a user parameter.

Genetic Algorithm Applications

The genetic algorithm is a strategy for comprehending both compelled and unconstrained streamlining issues that depend on common determination, the procedure that drives natural development. The hereditary calculation over and over again alters a populace of individual arrangements. The genetic algorithm technique started initially for computing and calculating humans to simulate a biological process, so since this period there are many terminologies taken from biology to this system.

Selection depends on putting on pressure relative to population size, weak performance individuals it was illustrated and switched to strong and better performance. Individuals are good for genetic algorithms technique since it has a high chance of promoting the information which will continue to the next generations.

Mutation While recombination works on at least two parental chromosomes, change locally however arbitrarily adjusts an answer. Once more, there are numerous varieties of transformation, however, it, as a rule, includes at least one change being made to a person's attribute or qualities. At the end of the day, transformation plays out an arbitrary stroll in the region of an applicant arrangement.

Cross over administrator is practically equivalent to multiplication and natural hybrid. In this, more than one parent is chosen and at least one off-springs are delivered utilizing the hereditary material of the guardians. Hybrid is normally connected in a GA with a high probability.

Coley [22] reported that beside all functions above such as (mutation, crossover, and selection) which applied and developed to the initial population, the generated new population and the generational number counter will be increased, this cycle of calculation will continue for selecting, mutation, and crossover until reach to the fixed number that is wanted or some convergence form will be shown a criterion wanted to be met.

Here we are thinking about the utilization of hereditary calculations to discover answers for different advancement issues like Backpack, Travelling Salesman Problems (TSP), work minimization and expansion and so forth. These issues have been in presence for long and numerous endeavors have been made to discover

productive systems to take care of these issues. Hereditary calculations are among the best methods utilized for such issues.

GA is utilized in various parts of a building-related to the issue of gas pipeline control [23]. Davis et al. [24] used the genetic algorithm in the system plan. Also, Grammatical Evolution (GE) and Raspberry PI make the possibility of using the GA in-stream motor turbine structure. Moreover, John Holland [25] used GA in displaying environments, airship configuration, and planning, emblematic math. The issue of untimely joins in hereditary calculations streamlining was examined. Recreated toughening was utilized to keep up a decent variety of the populace. An examination of the altered hereditary calculation approach with the straightforward hereditary calculation is completed taking a backpack issue.

To locate the most limited way, hereditary calculations can be utilized to encode away in the chart into a chromosome. The proposed methodology has been tried by Gen et al. [26] with various sizes from 6 nodes to 70 nodes and from 10 edges to 211 edges. The empowering results utilizing hereditary calculations can locate the ideal in all respects quickly and with high likelihood.

The value of heuristic calculations as the scan strategy for different enhancement issues is inspected by Chun et al. [27]. Hereditary calculations, transformative calculations were looked at on assorted enhancement issues and the outcomes uncover the outperformance of hereditary calculations.

The issue of finding hearty or adaptable answers for booking issues for true application was proposed by Jensen [28]. Tentatively, it is demonstrated that utilizing a hereditary calculation, it is conceivable to discover powerful and adaptable timetables.

In view of Takahashi [29], there are two sorts of hybrid administrators for settling the Travelling salesman problem (TSP). Traditional encoding of the TSP which is an exhibit portrayal of chromosomes where each component of this cluster is a quality that in the TSP demonstrates a city. The primary sort of hybrid administrator relates to this chromosome structure. In this administrator, two guardians are chosen and with the trading of certain parts in guardians, the youngsters are imitated. In the second technique, it is attempted to hold helpful data about connections of parent's edges which prompts combination. Nguyen et al. [30] proposed a half breed GA to discover excellent answers for the TSP.

Methodology

Data Acquisition

Our methodology is divided into two steps. The first step is the acquisition of the data which collected together from various sources. Where there are 130 of Data sets came from the Mediterranean Basins, Africa, the Persian Gulf and the North Sea for dead, gas-saturated and under-saturated oils as illustrated in Table 1 [31]. The data includes reservoir temperature T ($^{\circ}\text{F}$), Bubble point pressure P_b (Psia), oil formation volume factor B_o (bbl/STB), solution gas/oil ratio R_s (SCF/STB), the gas specific gravity γ_g , the stock tank oil API, and the oil specific gravity γ_o . Figure 2 shows the relation between bubble point pressure, temperature, and specific gravity for oil and gas.

Table 1 Collected Data.

T, $^{\circ}\text{F}$	P_b , Psia	B_o	R_s	γ_g	API	γ_o
248	1680	1.46 8	55 7	1.195 5	37. 2	0.8387 7
248	1415	1.43 2	48 6	1.246 8	37. 2	0.8387 7
248	1215	1.40 4	43 3	1.295 5	37. 2	0.8387 7
248	1015	1.37 8	38 1	1.353 9	37. 2	0.8387 7
248	815	1.35 2	32 8	1.427 2	37. 2	0.8387 7
248	615	1.32 2	27 3	1.526 4	37. 2	0.8387 7
248	415	1.29 2	21 5	1.661 1	37. 2	0.8387 7
.
.
.
244	515	1.31 3	26 1	1.787 1	37. 5	0.8372 8
244	315	1.27 7	19 9	1.993 7	37. 5	0.8372 8
244	187	1.24 1	14 7	2.209	37. 5	0.8372 8

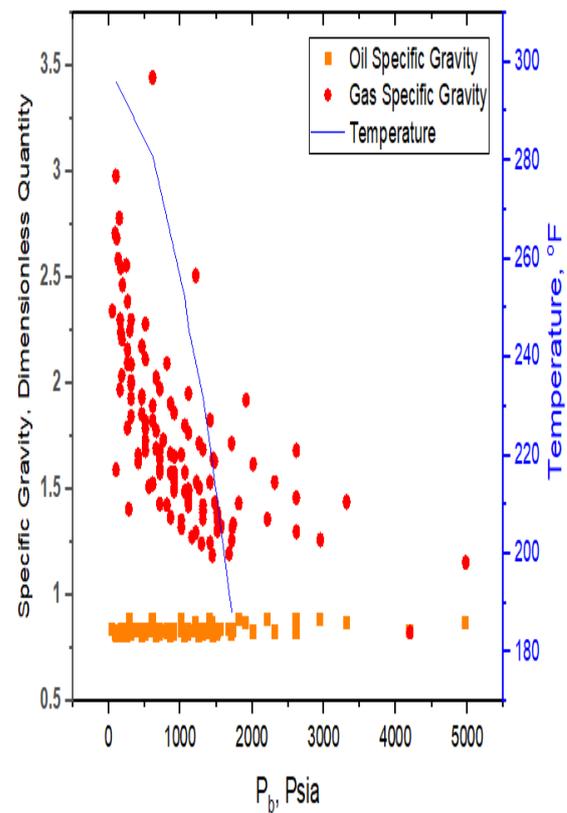


Figure 2 Relationship between bubble point pressure, temperature, and specific gravity for oil and gas.

Statistical Error

When starting to build a new model using set of a new data, it needs to make an analysis for this data, there are more than one type of errors in this data, which are; The data procedures collection, gross data capture or encoding error, and there is a big error came from the approach of data selecting or analysis.

These conditions refer to random errors, also it's considered as a non-systematic error, which observed during the measurements. There are many types of errors. One of them is known as Average Absolute error which is the difference between the true value and calculated value which the true value is X_t , and the calculated is X_c .

$$E = (X_c - X_t) / X_t * 100 \quad (25)$$

The comparison between published correlations of bubble point pressure, solution gas/oil ratio, and oil formation volume factor according to Standing, Lasater, Vazquez & Beggs, Glaso, Al-Marhoun, and Kartoatmodjo & Schmidt based on the lower value of average absolute error taken place as shown in Figures (3, 4, and 5). This selection depends on the comparison of data of presented correlations and the collected data point.

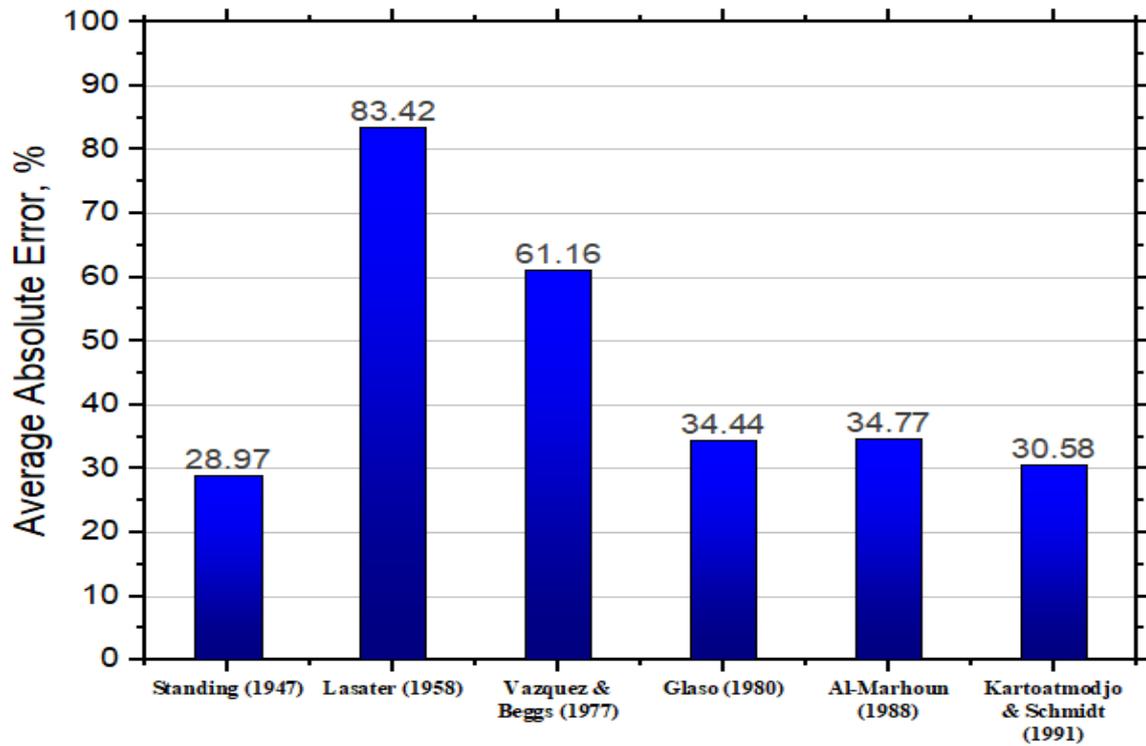


Figure 3 Average Absolute error for Bubble Point Pressure.

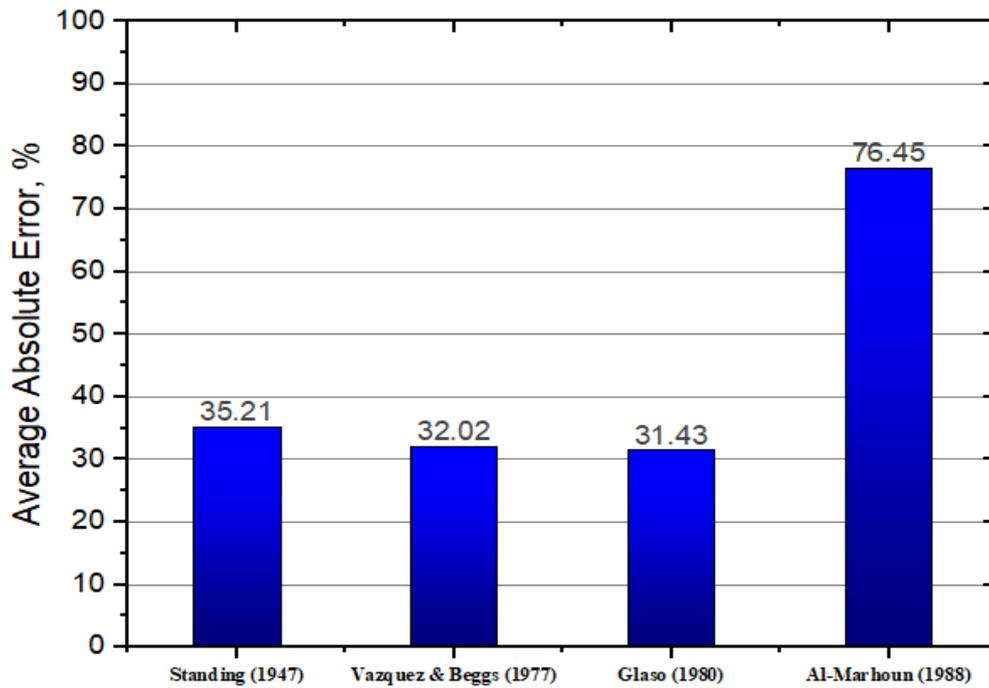


Figure 4 Average Absolute error for Solution Gas-Oil Ratio.

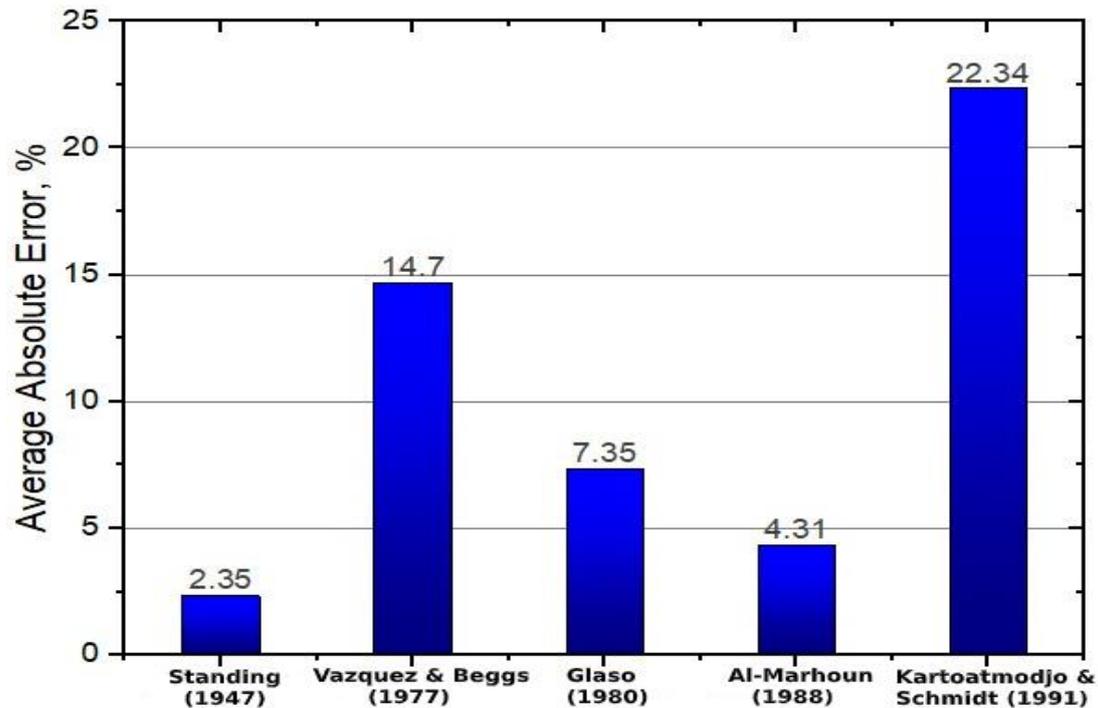


Figure 5 Average Absolute error for Oil Formation Volume Factor.

Genetic Algorithm Technique

The second step in our methodology is Genetic algorithm which is a family of computational models propelled by evolution, these calculations a potential arrangement to a specific issue on a straight forward chromosome like information structure and apply recombination administrators to these structures, so as to protect basic information. Genetic algorithms are frequently seen as work optimizers in spite of the fact that the run of issues to which hereditary calculations have been connected is quite a broad.

An implementation of the genetic algorithm starts with a populace of typically random chromosomes. One at that point assesses these structures as shown in Figure (6) and apportions regenerative openings in such a way that those chromosomes which offering better solution to the target problem will give more chance than poorer solution to solve the problem. The goodness of an arrangement is regularly ended with regard to the current population.

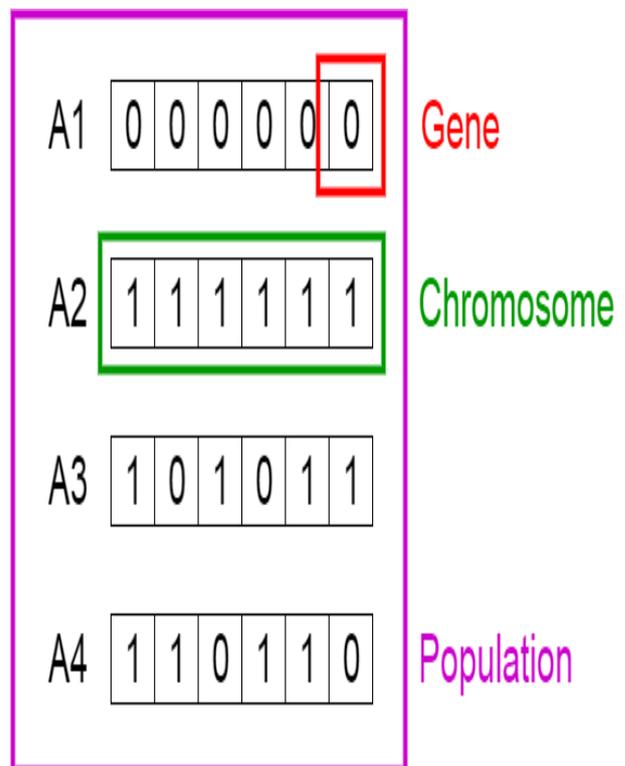


Figure 6 A chromosome in genetic algorithm.

Nelson et al. [32] stated that genetic algorithms are a global optimization algorithm that are able to solve combinatorial problems. They are a very effective way to traverse through huge datasets in particular when the space of search is enlarged due to excessive dimensionality. The benefits of Genetic Algorithms had been implemented in several fields, such as biology, for example, identify cancer cells from microarray records, predict the secondary shape in RNA, protein-primarily

based mass spectrometry as well as regression models to map quantitative trait. Primary steps in the genetic algorithm, loci (QTL) in genomic information, consisting of epistatic fashions and environmental factors. Large genomic datasets had been investigated with genetic algorithms to a confined extent. It's far, but, important to pick the ideal version to optimize. A number of systems gaining knowledge and techniques of feature selection have been used in genomic records and keep the promise to be a great place to begin while growing techniques incorporating EAs. All genetic algorithms have the simple define shown in Figure (7). Following is a short description of the steps required:

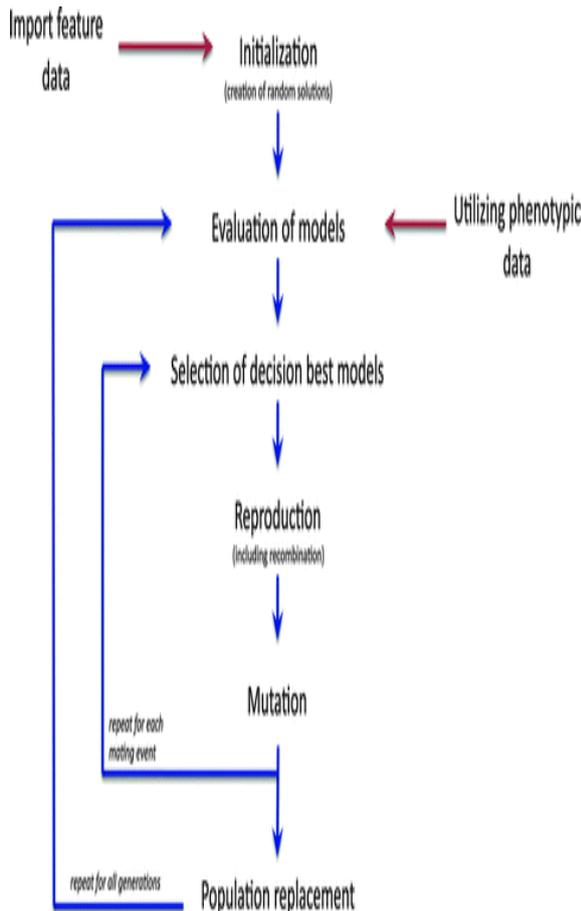


Figure 7 Basic Steps in genetic algorithm.

- Initialization: A random population answers is produced. Every solution is programmed in a typical way (the subcomponents are regularly known as chromosomes). Solutions may be characterized as a set of bits, mathematical features, subprograms, or tree-like structures.
- Evaluation: each solution is evaluated using a fitness feature. The fitness function serves because of the surroundings to which the solutions need to adapt.
- Selection: The solutions that better adapt to the surroundings are chosen for reproduction. Numerous selection tactics can be used. However, selection technique

has turned out to be better that allows less-suitable solutions are sometimes chosen. This decrease the chance of the solutions to grow to be stuck on local optima.

- Reproduction: the chosen solutions can be reproduced with every different. The latter entails that some parts of two solutions recombine to create a unique solution.
- Mutation: a number of the new, or cloned, solutions are decided on to go through mutation to form a novel variant at a number of the modules.

Building a new model using Genetic Algorithm by MATLAB

There are steps for building a model which include firstly, Create Fitness Function. Secondly, Running optimization tool from command window. Thirdly, Limit The boundaries matrix for the fitness function. Fourthly, Choose the Mutation function. Fifthly, Choose the Crossover function. Sixthly, Calculate Variable from GA as following. Seventhly, Calculate the P_b , R_s , or B_o value from GA. Figure (8) shows the simulation steps for bubble point pressure.

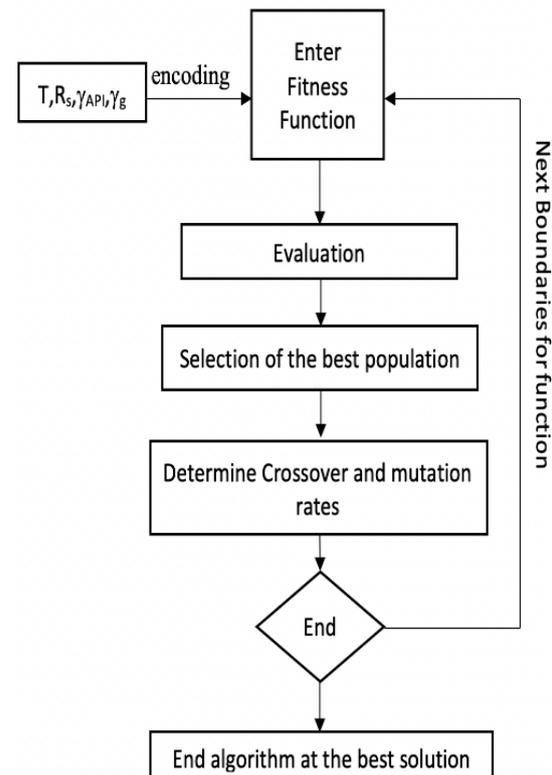


Figure 8 Simulation steps for P_b .

Model Initiation

Evaluation means that each solution or individual in the population is evaluated using fitness function. Where Fitness function is evaluation function, that determines what solutions are better than others and how well an individual is adapted to the environment. Higher fitness is more likely to be selected and contributed to the next

generation. The following fitness function which used as shown in figures (9, 10, and 13):

Standing function: Function $y = \text{Standing}(x)$

$$\text{Fit}(i) = (\sum_{j=1}^{nn} PFit(i, j)) / nn \quad (26)$$

The data used in this paper have multiple variables (Five Variables). Hence,

- Very big population size usually does not improve the performance of genetic algorithm.
- Good population size is about (20-30), however sometimes sizes 50-100 are reported as the best for problems with large number of variables. So, the population size used in this paper as a result of large number of variables is 50 as shown in figures (14, and 16).

Selection Deals with the fitness function with selects the best individuals from population for reproduction. The idea of selection phase is to select the fittest individuals and let them pass their genes to the next generation. Individuals with high fitness have more chance to be selected for reproduction. There are few possible ways to implement selection:

- 1- Only the strongest survive
 - Choose the individuals with the highest fitness for the next generation.
- 2- Some weak solutions survive
 - Assign a probability that a particular individual will be selected for the next generation.
 - More diversity.
 - Some bad solutions might have good parts.

There are many methods of selection:

- Fitness proportional selection (also known as Roulette Wheel Selection): chromosome's probability of being selected is directly proportional to its fitness. Two individuals are then chosen randomly based on these probabilities and produce offspring.
- Stochastic selection: select each chromosome the number of times equals to its expectation of being selected under the fitness proportional method. This method had been chosen to use in this paper as shown in figure (18) because most functions are stochastic and designed so that a small proportion of fewer fit solutions are

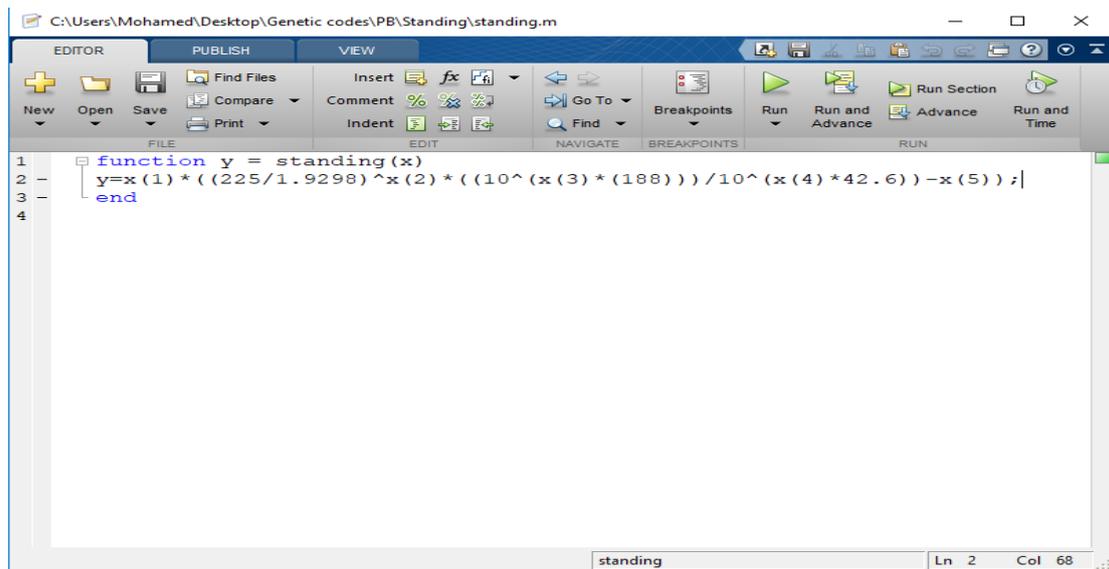
selected. This helps keep the diversity of the population large, preventing premature convergence on poor solutions.

- Tournament selection: first selects two chromosomes with uniform probability, then choose the one with the highest fitness.
- Truncation selection: selects at random from the population having first eliminated a fixed number of the least fit chromosomes.

Tables (2,3, and 4) illustrate the new genetic algorithm correlations for bubble point pressure, solution gas-oil ratio, and oil formation volume factor.

MATLAB Model Steps

Step 1: Create Fitness Function.

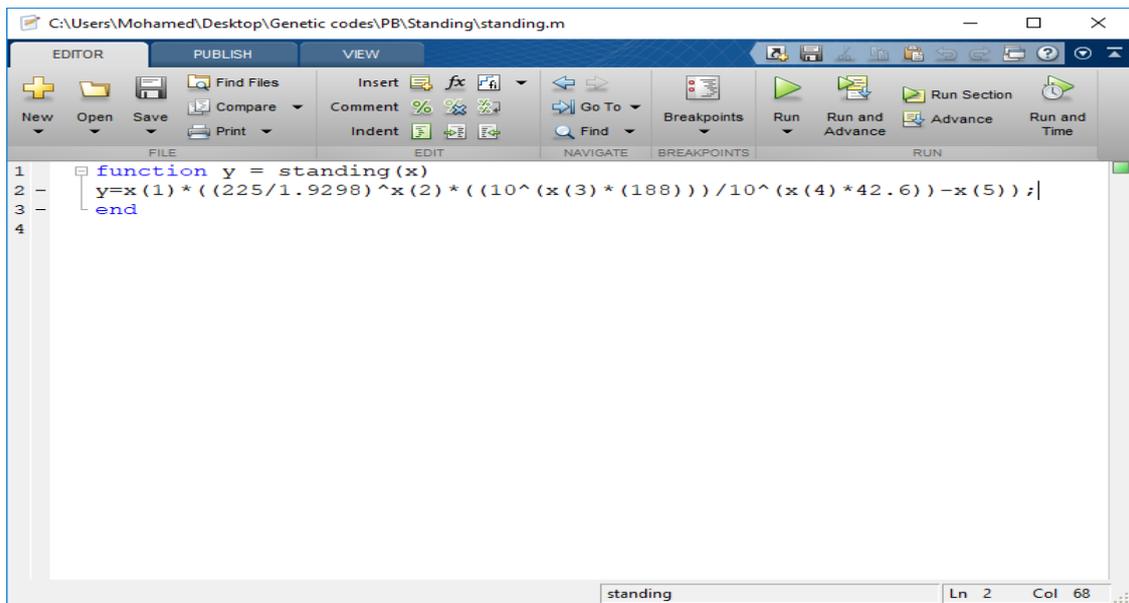


The screenshot shows the MATLAB editor interface with the following code in the 'standing.m' file:

```
1 function y = standing(x)
2     y=x(1) * ((225/1.9298)^x(2) * ((10^(x(3) * (188))) / 10^(x(4) * 42.6)) - x(5)) ;|
3     end
4
```

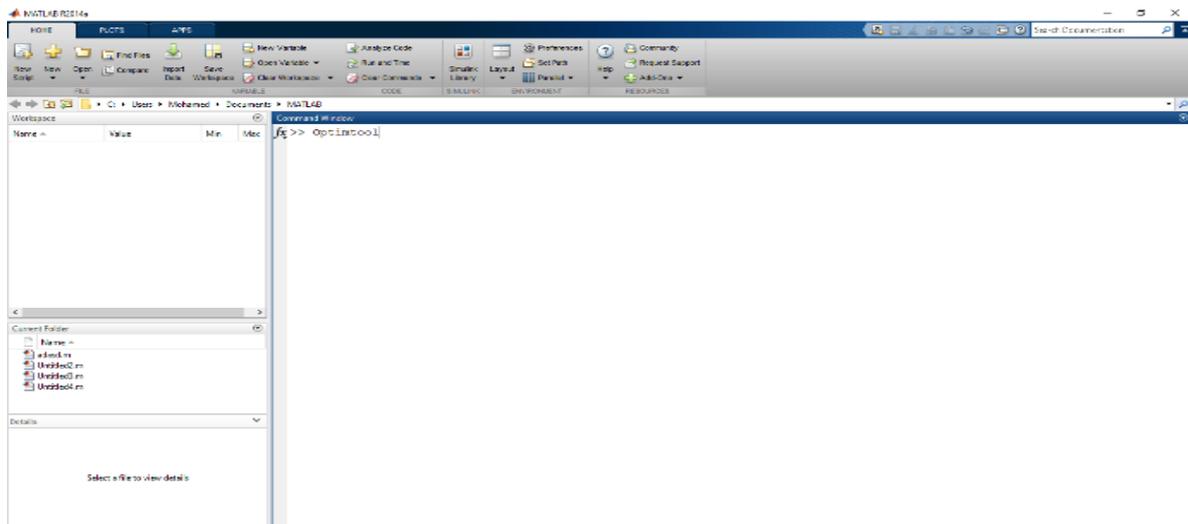
Figure 9 Create Fitness Function.

Step 2: Run Fitness Function File.



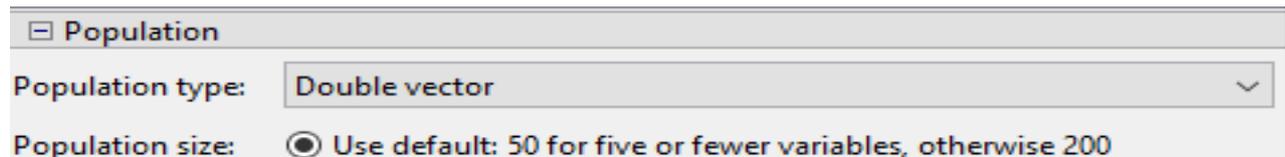
The screenshot shows the same MATLAB editor interface as Figure 9, with the code for the 'standing.m' file. The 'Run' button in the toolbar is highlighted, indicating the next step is to execute the function.

Figure 10 Run Fitness Function File.

Step 3: Running optimization tool from command window.**Figure 11** Running optimization tool from command window.**Step 4:** Choose the solver type from Optimization tool.**Figure 12** Choose the solver type from Optimization tool.**Step 5:** Write the name of fitness function with @.**Figure 13** Write the name of fitness function with @.**Step 6:** Entering the number of Variables.**Figure 14** Entering the number of Variables.

Step 7: Limit The boundaries matrix for fitness function

Bounds: Lower: [15 0.83 0.00091 0.0125 1.4] Upper: [18.2 0.831 0.000910 0.01251 1.0]

Figure 15 Limit The boundaries matrix for fitness function.**Step 8:** Determine the population size and its initial range.

Population

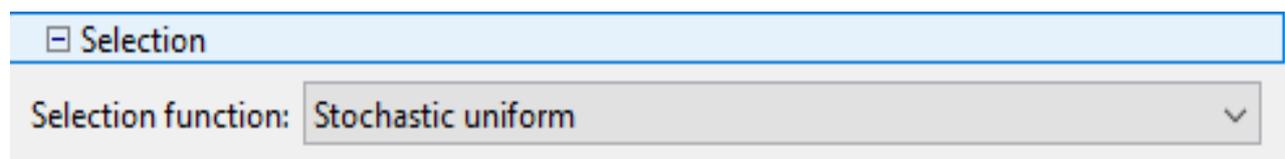
Population type: Double vector

Population size: Use default: 50 for five or fewer variables, otherwise 200

Figure 16 Population Size.

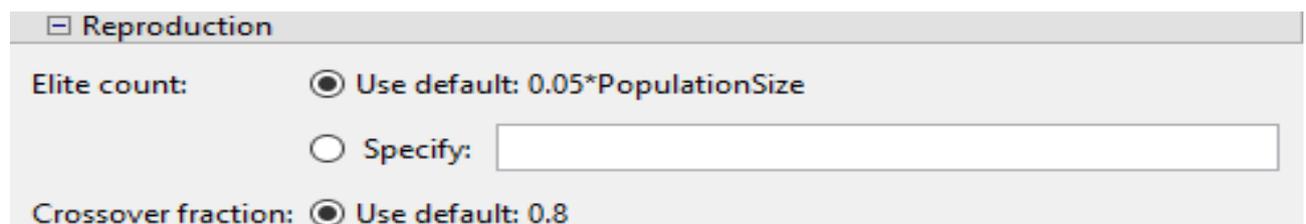
Initial range: Use default: [-10;10]

Specify:

Figure 17 Population initial Range.**Step 9:** Choosing selection function.

Selection

Selection function: Stochastic uniform

Figure 18 Selection function.**Step 10:** Choose Elite count and Crossover Fraction.

Reproduction

Elite count: Use default: 0.05*PopulationSize

Specify:

Crossover fraction: Use default: 0.8

Figure 19 Elite count and Crossover Fraction.

Step 11: Choose Mutation function.

Mutation function:

Figure 20 Maturation Function.

Step 12: Choose Crossover function.

Crossover function:

Figure 21 Crossover Function.

Step 13: Calculate Variable from GA as following.

Pareto front - function values and decision variables						
Index	x1	x2	x3	x4	x5	
1	15	0.83	0.001	0.013	1.598	

Figure 22 Variable Values.

Step 14: Calculate the P_b value from GA.

Optimization running,

Objective function value: 314.5704731726788

Optimization terminated: average change in the spread of Pareto solutions less than options, TolFun.

Figure 23 P_b Calculated from GA Model.

Step 15: check the value of P_b from model with laboratory P_b if agreed finish.

Step 16: If the value not agreed the steps from 7 to 15 will be repeated until the value of P_b agreed with Laboratory value.

Table 2 Initiation model for bubble point pressure correlation

PVT Properties	Correlations	Original Correlation	GA Model Initiation	New GA Correlation
Bubble point pressure (P _b)	Standing	$P_b = 18.2 \left[\left(\frac{R_{sb}}{\gamma_g} \right)^{0.83} * 10^{(0.00091T - 0.0125\gamma_{API}) - 1.4} \right]$	$P_b = x(1) \left[\left(\frac{R_{sb}}{\gamma_g} \right)^{x(2)} * 10^{(0.00091x(3) - 0.0125x(4)) - x(5)} \right]$	$P_b = 15 \left[\left(\frac{R_{sb}}{\gamma_g} \right)^{0.83} * 10^{(0.00091T - 0.0125\gamma_{API}) - 1.598} \right]$
	Lasater	$P_b = \frac{(Pf)(t+459.6)}{\gamma_g}$ P _f = 6.7 according to gas mole fraction	$P_b = \frac{(x(1)(t+x(2)))}{\gamma_g}$	$P_b = \frac{(2.9)(t+380)}{\gamma_g}$
	Vasquez & Beggs	$P_b = \left[\frac{C_1 R_{sb}}{\gamma_{gs}} 10^{\frac{-C_3 \gamma_{API}}{(T+459.67)}} \right]^{1/C_2}$ API ≥ 30 so C ₁ = 56.060 C ₂ = 1.187 C ₃ = 10.393	$P_b = \left[\frac{x(1) R_{sb}}{\gamma_{gs}} 10^{\frac{-x(2) \gamma_{API}}{(T)}} \right]^{1/x(3)}$	$P_b = \left[\frac{35 * R_{sb}}{\gamma_{gs}} 10^{\frac{-8.25 * \gamma_{API}}{(T)}} \right]^{1/1.158}$
	Glaso	$P_b^* = \left[\left(\frac{R}{\gamma_g} \right)^{0.816} \frac{T^{0.172}}{\gamma_{API}^{0.989}} \right]$ $\log P_b = 1.7669 + 1.7447 \cdot \log P_b^* - 0.30218 (\log P_b^*)^2$ $P_b = 10^{\log(P_b)}$	$P_b^* = \left[\left(\frac{R}{\gamma_g} \right)^{x(1)} \frac{T^{x(2)}}{\gamma_{API}^{x(3)}} \right]$ $\log P_b = X(4) + x(5) \cdot \log P_b^* - x(6) (\log P_b^*)^2$ $P_b = 10^{\log(P_b)}$	$P_b^* = \left[\left(\frac{R}{\gamma_g} \right)^{0.808} \frac{T^{0.168}}{\gamma_{API}^{0.978}} \right]$ $\log P_b = 1.7324 + 1.7113 \cdot \log P_b^* - 0.3323 (\log P_b^*)^2$ $P_b = 10^{\log(P_b)}$
	Al-Marhoun	$P_b = 5.38088 \times 10^{-3} R_{sb}^{0.715082} \gamma_g^{-1.87784} \gamma_o^{3.1437} (T+459.67)^{1.32657}$	$P_b = x(1) \times 10^{-3} R_{sb}^{x(2)} \gamma_g^{-x(3)} \gamma_o^{x(4)} (T+459.67)^{x(5)}$	$P_b = (5.88088 * 10^{-3}) * (R_{sb}^{0.615082}) * (\gamma_g^{-1.67784}) * (\gamma_o^{3.2437}) * (T)^{1.72998}$
	Kartoatmodjo & Schmidt	$P_b = \left[\frac{R_{sb}}{0.03150 \gamma_g^{0.7587} 10^{11.28955 \gamma_{API} / (T+459.67)}} \right]^{0.9143}$	$P_b = \left[\frac{R_{sb}}{X(1) \gamma_g^{X(2)} 10^{X(3) \gamma_{API} / (T)}} \right]^{X(4)}$	$P_b = \left[\frac{R_{sb}}{0.0298 * \gamma_g^{0.3578} 10^{4.7597 \gamma_{API} / (T)}} \right]^{0.9143}$

Table 3 Initiation model for solution gas-oil ratio correlations

PVT Properties	Correlations	Original Correlation	GA Model Initiation	New GA Correlation
Solution gas-oil ratio (R _s)	Standing	$R_s = \gamma_g \left[\left(\frac{P_b}{18.2} + 1.4 \right) \times 10^{\left(\frac{0.00091T + 0.0125\gamma_{API}}{1.2048} \right)} \right]$	$R_s = \gamma_g \left[\left(\frac{P_b}{x(1)} + x(2) \right) \times 10^{\left(\frac{x(3)T + x(4)\gamma_{API}}{x(5)} \right)} \right]$	$R_s = \gamma_g \left[\left(\frac{P_b}{15} + 1.598 \right) \times 10^{\left(\frac{0.00091T + 0.013\gamma_{API}}{1.1548} \right)} \right]$
	Vasquez & Beggs	$R_{sb} = \left[\frac{\gamma_g P_b^{c_2}}{C_1} 10^{\frac{C_3 \gamma_{API}}{T + 459.67}} \right]$	$R_{sb} = \left[\frac{\gamma_g P_b^{X(2)}}{X(1)} 10^{\frac{X(3)\gamma_{API}}{T}} \right]$	$R_{sb} = \left[\frac{\gamma_g P_b^{1.168}}{35} 10^{\frac{3.25 \gamma_{API}}{T}} \right]$
	Glaso	$N_{pb} = 10^{[2.8869 - [14.1811 - 3.3093 \log(P_b)]^{0.5}]}$ $R_{sb} = \gamma_g \left[N_{pb} \frac{\gamma_{API}^{0.989}}{T^{0.172}} \right]^{1.2255}$	$N_{pb} = 10^{[x(1) - [x(2) - x(3) \log(P_b)]^{0.5}]}$ $R_{sb} = \gamma_g \left[N_{pb} \frac{\gamma_{API}^{x(4)}}{T^{x(5)}} \right]^{x(6)}$	$N_{pb} = 10^{[2.8869 - [14.1811 - 3.3093 \log(P_b)]^{0.5}]}$ $R_{sb} = \gamma_g \left[N_{pb} \frac{\gamma_{API}^{0.978}}{T^{0.168}} \right]^{1.2412}$
	Al-Marhoun	$R_{sb} = \left[\frac{P_b \gamma_g^{1.87784}}{0.00538088 \gamma_o^{3.1437} (T + 459.67)^{1.32657}} \right]^{1/0.715082}$	$R_{sb} = \left[\frac{P_b \gamma_g^{X(1)}}{0.00538088 \gamma_o^{X(2)} (T)^{X(3)}} \right]^{1/X(4)}$	$R_{sb} = \left[\frac{P_b \gamma_g^{1.98784}}{0.007899088 \gamma_o^{3.3437} (T)^{1.69857}} \right]^{1/0.615082}$

Table 4 Initiation model for oil formation volume factor correlations

PVT Properties	Correlations	Original Correlation	GA Model Initiation	New GA Correlation
Oil formation volume factor (Bo)	Standing	$B_{ob} = 0.972 + 1.47 \times 10^{-4} [R_s (\frac{\gamma_g}{\gamma_o})^{0.5} + 1.25 T]^{1.175}$	$B_{ob} = x(1) + x(2) \times 10^{-4} [R_s (\frac{\gamma_g}{\gamma_o})^{0.5} + x(3) T]^{x(4)}$	$B_{ob} = 0.989 + 1.532 \times 10^{-4} [R_s (\frac{\gamma_g}{\gamma_o})^{0.5} + 1.26 T]^{1.169}$
	Vasquez & Beggs	$B_{ob} = 1 + C_1 R_{sb} + C_2 (T - 60) (\frac{\gamma_{API}}{\gamma_{gs}}) + C_3 R_{sb} (T - 60) (\frac{\gamma_{API}}{\gamma_{gs}})$	$B_{ob} = 1 + x(1) R_{sb} + x(2) (T - 60) (\frac{\gamma_{API}}{\gamma_{gs}}) + x(3) R_{sb} (T - 60) (\frac{\gamma_{API}}{\gamma_{gs}})$	$B_{ob} = 1 + 6.3 \times 10^{-4} R_{sb} + 2.12 \times 10^{-5} (T - 60) (\frac{\gamma_{API}}{\gamma_{gs}}) + 2.897 \times 10^{-9} R_{sb} (T - 60) (\frac{\gamma_{API}}{\gamma_{gs}})$
	Glaso	$B_{ob}^* = R_{sb} (\frac{\gamma_g}{\gamma_o})^{0.526} + 0.986 T$ $B_{ob} = 1 + 10^{[-6.58511 + 2.91329 \log(B_{ob}^*) - 0.27683 [(\log(B_{ob}^*))^2]}$	$B_{ob}^* = R_{sb} (\frac{\gamma_g}{\gamma_o})^{x(1)} + x(2) T$ $B_{ob} = 1 + 10^{[-X(3) + X(4) \log(B_{ob}^*) - X(5) [(\log(B_{ob}^*))^2]}$	$B_{ob}^* = R_{sb} (\frac{\gamma_g}{\gamma_o})^{0.538} + 0.989 T$ $B_{ob} = 1 + 10^{[-6.58511 + 3.06241 \log(B_{ob}^*) - 0.32097 [(\log(B_{ob}^*))^2]}$
	Al-Marhoun	$B_{ob} = 0.497069 + 8.26963 \times 10^{-4} (T + 459.67) + 1.82594 \times 10^{-3} F + 3.18099 \times 10^{-6} F^2$ $F = R_{sb}^{0.74239} \gamma_g^{0.323294} \gamma_o^{-1.202040}$	$B_{ob} = x(1) + x(2) \times 10^{-4} (T + 459.67) + x(3) \times 10^{-3} F + x(4) \times 10^{-6} F^2$ $F = R_{sb}^{x(5)} \gamma_g^{x(6)} \gamma_o^{x(7)}$	$B_{ob} = 0.50214 + 8.36963 \times 10^{-4} (T + 459.67) + 1.84964 \times 10^{-3} F + 2.98099 \times 10^{-6} F^2$ $F = R_{sb}^{0.75968} \gamma_g^{0.398647} \gamma_o^{-1.093015}$
	Kartoatmodjo & Schmidt	$B_o = 0.98496 + 0.0001 \times F^{1.50}$ $F = R_{sb}^{0.755} \gamma_g^{0.25} \gamma_o^{-1.5} + 0.45 T$	$B_o = x(1) + x(2) \times F^{x(3)}$ $F = R_{sb}^{x(4)} \gamma_g^{x(5)} \gamma_o^{-x(6)} + x(7) T$	$B_o = 0.9999 + 0.0001 \times F^{1.49}$ $F = R_{sb}^{0.7789} \gamma_g^{0.5} \gamma_o^{-1.3} + 0.49 T$

Results and Discussions

This section illustrates the validation and the results of genetic algorithm correlation and comparing the genetic algorithm correlations results with the results of the original correlations which related to Standing, Lasater, Vazquez & Beggs, Glaso, Al-Marhoun, and Kartoatmodjo & Schmidt. This comparing includes statistical error for original correlations and genetic algorithm correlations. There are around 70% of the data bank used to build a model, that used to generate the crude oil GA model. The other 30% of the data was used to assessment and validate the GA correlations.

Bubble Point Pressure Correlations

The genetic algorithm model is used to calculate P_b which validated by using the other 30% of the data. The average absolute error for bubble point correlations has been decreased. For Standing: the average absolute error decreased from 28.97 % to 18.98 %, for Lasater: the average absolute error decreased from 83.42 % to 14.36 %, for Vazquez & Beggs: the average absolute error decreased from 61.16 % to 23.34 %, for Glaso: the average absolute error decreased from 34.44 % to 13.23 %, for Al-Marhoun: the average absolute error decreased from 34.77 % to 12.43 %, and for Kartoatmodjo & Schmidt: the average absolute error decreased from 30.58 % to 17.28 %. Table 5 illustrates the least average absolute error for these correlations. Figure (24) shows the comparison between the least average absolute error of these correlations and after using the genetic algorithm on them.

Table 5 Validation of genetic algorithm bubble point pressure correlation.

Correlations	Average absolute error	
	Before using GA	After using GA
Standing	28.97 %	18.98 %
Lasater	83.42 %	14.36 %
Vazquez & Beggs	61.16 %	23.34 %
Glaso	34.44 %	13.23 %
Al-Marhoun	34.77 %	12.43 %
Kartoatmodjo & Schmidt	30.58 %	17.28 %

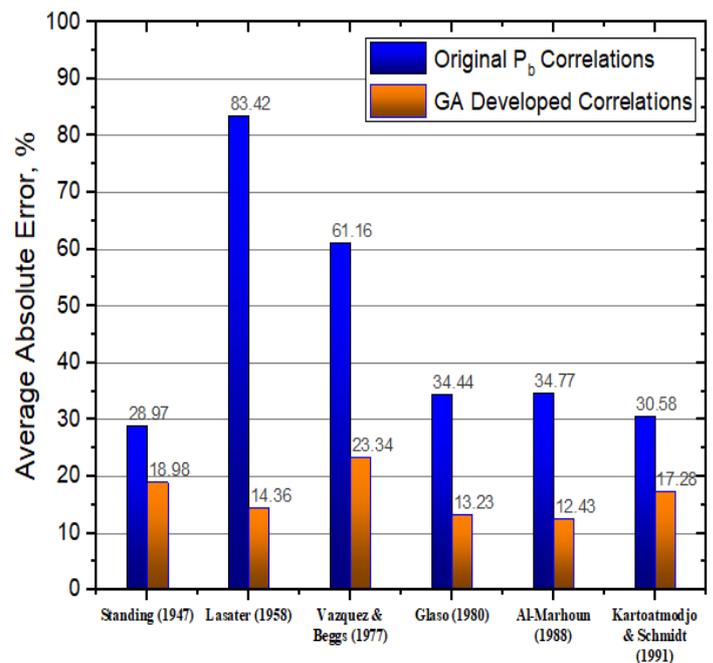


Figure 24 Results of applying the genetic algorithm on published correlation for Bubble point pressure.

Solution Gas-Oil Ratio

The genetic algorithm model is used to calculate R_s which validated by using the other 30% of the data. The average absolute error for the solution gas-oil ratio has been decreased. For Standing: the average absolute error decreased from 35.21 % to 14.43 %, for Vazquez & Beggs: the average absolute error decreased from 32.02 % to 22.93 %, for Glaso: the average absolute error decreased from 31.43 % to 13.42 %, and for Al-Marhoun: the average absolute error decreased from 76.45 % to 18.97 %. Table 6 illustrates the least average absolute error for these correlations. Figure (25) shows the comparison between the least average absolute error of these correlations and after using the genetic algorithm on them.

Table 6 Validation of genetic algorithm solution gas-oil ratio correlation.

Correlations	Average absolute error	
	Before using GA	After using GA
Standing	35.21 %	14.43 %
Vazquez & Beggs	32.02 %	22.93 %
Glaso	31.43 %	13.42 %
Al-Marhoun	76.45 %	18.97 %

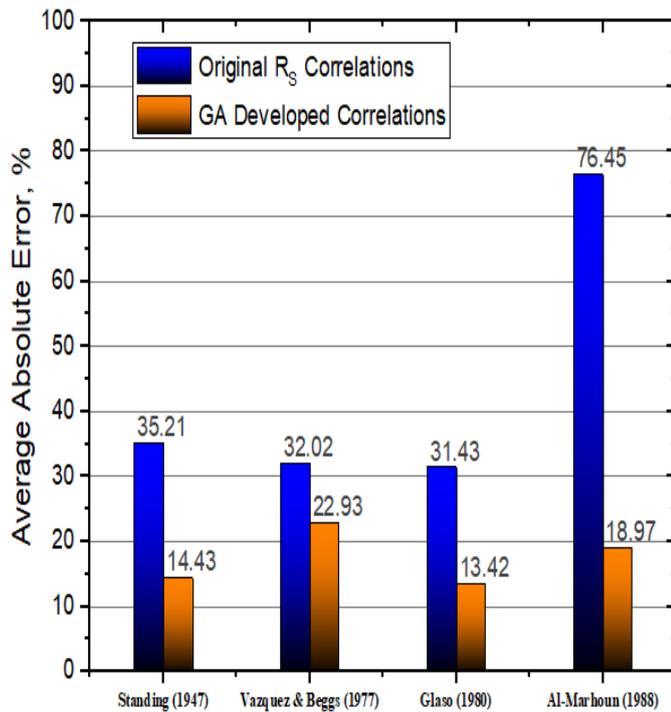


Figure 25 Result of applying the genetic algorithm on published correlation for solution gas-oil ratio.

Oil Formation Volume Factor

The genetic algorithm model is used to calculate B_o which validated by using the other 30% of the data. The average absolute error for oil formation volume factor correlations has been decreased. For Standing: the average absolute error decreased from 2.35% to 1.65%, for Vazquez & Beggs: the average absolute error decreased from 14.7% to 3.85%, for Glaso: the average absolute error decreased from 7.35% to 2.1%, for Al-Marhoun: the average absolute error decreased from 4.31% to 1.19%, and for Kartoatmodjo & Schmidt: the average absolute error decreased from 22.34% to 1.96%. Table 7 illustrates the least average absolute error for these correlations. Figure (26) shows the comparison between the least average absolute error of these correlations and after using the genetic algorithm on them.

Table 7 Validation of genetic algorithm oil formation volume factor correlation.

Correlations	Average absolute error	
	Before using GA	After using GA
Standing	2.35 %	1.65 %
Vazquez & Beggs	14.7 %	3.85 %
Glaso	7.35 %	2.1 %
Al-Marhoun	4.31 %	1.19 %
Kartoatmodjo & Schmidt	22.34 %	1.96 %

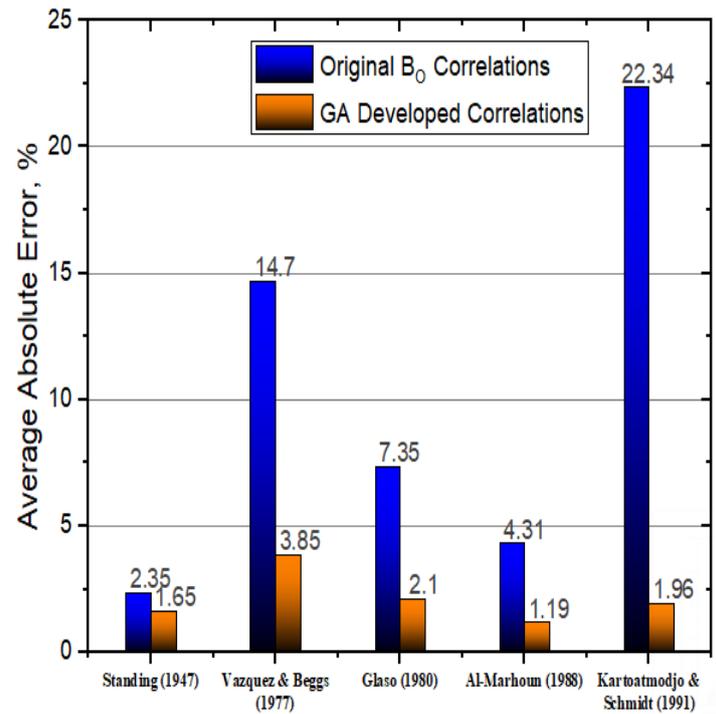


Figure 26 Result of applying the genetic algorithm on published correlation for oil formation volume factor.

Conclusion

PVT analysis is very crucial and valuable for the engineers of the oil and gas reservoir to observe the phase behavior of oil and gas, and producing the calculations of material balance. Conventionally, PVT experiments are conducted through high temperature – high pressure (PVT) cells, but in the absence of experimental facilities, Researchers have recourse to empirically derived correlations. In this paper, novel and further accurate correlations to predict crude oil properties by using the genetic algorithm (GA) is applied. The data which came from the middle east is divided into two sections, the first one who has 70% of data to build initiation, and the second one, which has 30% of the data to make validation for models. Based on the results obtained, the following can be concluded:

- New correlations for estimating the oil formation volume factor, bubble point pressure, and solution gas-oil ratio for crude oil have been developed using the genetic algorithm technique.
- The average absolute error of genetic algorithm correlations was lower for this study than for estimations based on other published empirical correlations such as Standing, Lasater, Vazquez & Beggs, Glaso, Al-Marhoun, and Kartoatmodjo & Schmidt.
- This development will assist the reservoir engineer in predicting all crude oil characteristics with high accuracy.

Conflicts of interest

There are no conflicts to declare.

References

- [1] Dake, L., *Fundamentals of Reservoir Engineering* Amsterdam Elsevier Scientific Publishing Company, 1978.
- [2] Alramahi, B.A., K.A. Alshibli, and A.M. Attia, *Influence of Grain Size and Consolidation Pressure on Porosity of Rocks*, in *Site Characterization and Modeling*. 2005. p. 1-13.
- [3] Alshibli, K.A., B.A. Alramahi, and A.M. Attia, *Assessment of spatial distribution of porosity in synthetic quartz cores using microfocus computed tomography (μ CT)*. Particulate science and technology, 2006. **24**(4): p. 369-380.
- [4] McCain Jr, W., *The Properties of Petroleum Fluids, second edition*. Tulsa, Oklahoma: PennWell Publishing Company, 1990.
- [5] Danesh, A., *PVT and phase behaviour of petroleum reservoir fluids*. 1998: Elsevier.
- [6] Standing, M. *A pressure-volume-temperature correlation for mixtures of California oils and gases*. in *Drilling and Production Practice*. 1947. American Petroleum Institute.
- [7] Lasater, J., *Bubble point pressure correlation*. Journal of Petroleum Technology, 1958. **10**(05): p. 65-67.
- [8] Vazquez, M. and H.D. Beggs. *Correlations for fluid physical property prediction*. in *SPE Annual Fall Technical Conference and Exhibition*. 1977. Society of Petroleum Engineers.
- [9] Glaso, O., *Generalized pressure-volume-temperature correlations*. Journal of Petroleum Technology, 1980. **32**(05): p. 785-795.
- [10] Al-Marhoun, M.A., *PVT correlations for Middle East crude oils*. Journal of Petroleum Technology, 1988. **40**(05): p. 650-666.
- [11] Kartoatmodjo, T.R. and Z. Schmidt, *New correlations for crude oil physical properties*. 1991.
- [12] Gharbi, R. and A.M. Elsharkawy. *Neural network model for estimating the PVT properties of Middle East crude oils*. in *Middle East Oil Show and Conference*. 1997. Society of Petroleum Engineers.
- [13] Dindoruk, B. and P.G. Christman. *PVT properties and viscosity correlations for Gulf of Mexico oils*. in *SPE annual technical conference and exhibition*. 2001. Society of Petroleum Engineers.
- [14] Hemmati, M.N. and R. Kharrat. *A correlation approach for prediction of crude oil PVT properties*. in *SPE Middle East Oil and Gas Show and Conference*. 2007. Society of Petroleum Engineers.
- [15] Mazandarani, M.T. and S.M. Asghari. *Correlations for predicting solution gas-oil ratio, bubblepoint pressure and oil formation volume factor at bubblepoint of Iran crude oils*. in *European Congress of Chemical Engineering, Copenhagen*. 2007.
- [16] Ikiensikimama, S.S. and O. Ogboja. *New bubblepoint pressure empirical PVT correlation*. in *Nigeria Annual International Conference and Exhibition*. 2009. Society of Petroleum Engineers.
- [17] Sadiq, D.J. and O.F. Hassn, *New Correlation for Oil Formation Volume Factor at and Below Bubble Point Pressure*. Journal of Engineering, 2009. **15**(4): p. 4347-4355.
- [18] Hassan, O.F., *Correlation for solution gas-oil ratio of Iraqi oils at pressures below the bubble point pressure*. Iraqi Journal of Chemical and Petroleum Engineering, 2011. **12**(2): p. 1-8.
- [19] El-Hoshoudy, A. and S. Desouky, *PVT Properties of Black Crude Oil*, in *Processing of Heavy Crude Oils-Challenges and Opportunities*. 2019, IntechOpen.
- [20] Holland, J., *Adaptation in natural and artificial systems*. Univ of Michigan press. Ann Arbor, 1975. **228**.
- [21] Pelikan, M., K. Sastry, and E. Cantú-Paz, *Scalable optimization via probabilistic modeling*. 2006: Springer.
- [22] Coley, D.A., *An introduction to genetic algorithms for scientists and engineers*. 1999: World Scientific Publishing Company.
- [23] Goldberg, D.E. *Genetic algorithms and rule learning in dynamic system control*. in *Proceedings of the First International Conference on Genetic Algorithms and Their Applications*. 1985.
- [24] Davis, L. and S. Coombs, *Optimizing network link sizes with genetic algorithms*. Modelling and simulation methodology, Knowledge Systems' Paradigms, 1989: p. 317-331.
- [25] Holland, J., *Hidden order: how adaptation builds complexity reading*. Mass: Addison Wesley, 1995.
- [26] Gen, M., R. Cheng, and D. Wang. *Genetic algorithms for solving shortest path problems*. in *Proceedings of 1997 IEEE International Conference on Evolutionary Computation (ICEC'97)*. 1997. IEEE.
- [27] Chun, J.-S., H.-K. Jung, and S.-Y. Hahn, *A study on comparison of optimization performances between immune algorithm and other heuristic algorithms*. IEEE transactions on magnetics, 1998. **34**(5): p. 2972-2975.
- [28] Jensen, M.T., *Generating robust and flexible job shop schedules using genetic algorithms*. IEEE Transactions on evolutionary computation, 2003. **7**(3): p. 275-288.
- [29] Takahashi, R. *Solving the traveling salesman problem through genetic algorithms with changing crossover operators*. in *Fourth International Conference on Machine Learning and Applications (ICMLA'05)*. 2005. IEEE.
- [30] Nguyen, H.D., et al., *Implementation of an effective hybrid GA for large-scale traveling salesman problems*. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2007. **37**(1): p. 92-99.
- [31] De Ghetto, G. and M. Villa. *Reliability analysis on PVT correlations*. in *European Petroleum Conference*. 1994. Society of Petroleum Engineers.
- [32] Nelson, R.M., M. Kierczak, and Ö. Carlborg, *Higher order interactions: detection of epistasis using machine learning and evolutionary computation*, in *Genome-Wide Association Studies and Genomic Prediction*. 2013, Springer. p. 499-518.

Appendix Collected Data

T, (°F)	P _b , (Psia)	B _o , (bbl/STB)	R _s , (SCF/STB)	Y _g , (Dimensionless quantity)	Y _{API} , °API	Y _o , (Dimensionless quantity)
248	1680	1.468	557	1.1955	37.2	0.83877
248	1415	1.432	486	1.2468	37.2	0.83877
248	1215	1.404	433	1.2955	37.2	0.83877
248	1015	1.378	381	1.3539	37.2	0.83877
248	815	1.352	328	1.4272	37.2	0.83877
248	615	1.322	273	1.5264	37.2	0.83877
248	415	1.292	215	1.6611	37.2	0.83877
248	4197	2.365	2371	0.8253	37.2	0.82749
248	1725	1.522	663	1.3205	39.5	0.83235
248	1515	1.493	603	1.3692	38.5	0.83235
248	1315	1.465	547	1.4241	38.5	0.83235
248	1115	1.438	490	1.4923	38.5	0.83235
248	915	1.409	432	1.5775	38.5	0.83235
248	715	1.38	376	1.6801	38.5	0.83235
248	515	1.35	316	1.818	38.5	0.83235
248	315	1.314	251	2.0083	38.5	0.83235
248	183	1.278	192	2.2297	38.5	0.83235
229	1316	1.375	435	1.403	40.5	0.82267
229	1065	1.35	379	1.4905	40.5	0.82267
229	865	1.329	335	1.5762	40.5	0.82267
229	665	1.306	288	1.6918	40.5	0.82267
229	465	1.282	239	1.8545	40.5	0.82267
229	265	1.25	182	2.0949	40.5	0.82267
229	163	1.227	145	2.3	40.5	0.82267
222	2949	1.94	1321	1.2613	29	0.88162
222	2615	1.44	1210	1.3003	29	0.88162
222	2215	1.753	1074	1.3595	29	0.88162
222	1815	1.681	937	1.4356	29	0.88162
222	1415	1.61	802	1.5338	29	0.88162
222	1015	1.541	670	1.664	29	0.88162
222	615	1.467	506	1.8954	29	0.88162
222	298	1.386	340	2.252	29	0.88162
232	1525	1.46	550	1.3428	39	0.82991
232	1315	1.431	496	1.3898	39	0.82991
232	1115	1.403	446	1.4407	39	0.82991
232	915	1.376	395	1.5022	39	0.82991
232	715	1.348	342	1.5808	39	0.82991
232	515	1.32	288	1.6839	39	0.82991
232	315	1.286	228	1.8442	39	0.82991
232	185	1.253	180	2.037	39	0.82991
188	1717	1.3994	556	1.2595	42.6	0.81275
188	1515	1.373	509	1.3058	42.6	0.81275
188	1315	1.354	462	1.3614	42.6	0.81275

188	1115	1.335	419	1.4231	42.6	0.81275
188	915	1.318	378	1.4938	42.6	0.81275
188	715	1.298	330	1.5954	42.6	0.81275
188	515	1.275	280	1.7311	42.6	0.81275
188	315	1.247	225	1.9298	42.6	0.81275
188	171	0.215	165	2.245	42.6	0.81275
296	283	2.619	1977	1.4071	39.9	0.82555
296	2615	2.475	1757	1.4613	39.9	0.82555
296	2315	2.331	1536	1.5337	39.9	0.82555
296	2015	2.203	1340	1.6191	39.9	0.82555
296	1715	2.092	1169	1.7167	39.9	0.82555
296	1415	1.995	1018	1.8277	39.9	0.82555
296	1115	1.91	884	1.9523	39.9	0.82555
296	815	1.832	760	2.0951	39.9	0.82555
296	515	1.747	628	2.2811	39.9	0.82555
296	249	1.633	470	2.5585	39.9	0.82555
296	152	1.599	379	2.7812	39.9	0.82555
296	104	1.504	317	2.98	39.9	0.82555
281	4975	2.713	2496	1.1545	31.9	0.86597
281	1445	1.981	1458	1.1888	31.9	0.86597
281	3315	1.777	1074	1.441	31.9	0.86597
281	2615	1.658	827	1.6839	31.9	0.86597
281	1915	1.552	615	1.922	31.9	0.86597
281	1215	1.449	407	2.5098	31.9	0.86597
281	615	1.351	248	3.4445	31.9	0.86597
237	1226	1.418	470	1.5337	39.4	0.82797
237	105	1.401	433	1.5922	39.4	0.82797
237	915	1.385	398	1.6561	39.4	0.82797
237	765	1.369	362	1.7323	39.4	0.82797
237	615	1.35	325	1.8241	39.4	0.82797
237	465	1.33	285	1.9424	39.4	0.82797
237	315	1.305	241	2.0908	39.4	0.82797
237	1295	1.349	357	1.2435	39.5	0.82749
237	1165	1.335	330	1.2758	39.5	0.82749
237	1015	1.318	299	1.3184	39.5	0.82749
237	865	1.303	268	1.3687	39.5	0.82749
237	715	1.287	236	1.4307	39.5	0.82749
237	565	1.268	202	1.5137	39.5	0.82749
237	415	1.248	166	1.6281	39.5	0.82749
237	265	1.225	126	1.7897	39.5	0.82749
237	162	1.2	92	1.97	39.5	0.82749
254	1475	1.804	885	1.6334	42	0.81462
254	1315	1.771	821	1.6891	42	0.81462
254	1115	1.73	744	1.7673	42	0.81462
254	915	1.685	666	1.8614	42	0.81462
254	715	1.639	588	1.9736	42	0.81462
254	515	1.588	505	2.1152	42	0.81462
254	315	1.523	411	2.2987	42	0.81462

254	195	1.461	333	2.465	42	0.81462
254	135	1.411	276	2.5868	42	0.81462
254	95	1.351	213	2.708	42	0.81462
246	1737	1.524	635	1.3362	38	0.83089
246	1515	1.491	567	1.3907	38	0.83089
246	1315	1.463	515	1.422	38	0.83089
246	1115	1.436	468	1.4985	38	0.83089
246	915	1.41	414	1.5786	38	0.83089
246	715	1.383	360	1.6812	38	0.83089
246	515	1.353	302	1.8202	38	0.83089
246	315	1.319	24	2.01	38	0.83089
246	172	1.28	181	2.2408	38	0.83089
248	1482	1.511	582	1.4367	38.1	0.83432
248	1265	1.476	519	1.5069	38.1	0.83432
248	1065	1.449	466	1.5795	38.1	0.83432
248	865	1.421	413	1.6682	38.1	0.83432
248	665	1.392	360	1.7782	38.1	0.83432
248	465	1.358	302	1.9308	38.1	0.83432
248	265	1.312	230	2.1583	38.1	0.83432
248	55	1.276	180	2.342	38.1	0.83432
252	1460	1.821	936	1.6433	43.8	0.80719
252	1265	1.777	850	1.7173	43.8	0.80719
252	1065	1.733	768	1.8015	43.8	0.80719
252	865	1.685	683	1.905	43.8	0.80719
252	665	1.637	601	2.0267	43.8	0.80719
252	465	1.584	517	2.1753	43.8	0.80719
252	265	1.514	416	2.3873	43.8	0.80719
252	170	1.459	347	2.5477	43.8	0.80719
252	115	1.404	278	2.688	43.8	0.80719
244	1569	1.456	452	1.3248	37.5	0.83728
244	1315	1.423	474	1.3929	37.5	0.83728
244	1115	1.398	423	1.4575	37.5	0.83728
244	915	1.371	371	1.5385	37.5	0.83728
244	715	1.344	318	1.6421	37.5	0.83728
244	515	1.313	261	1.7871	37.5	0.83728
244	315	1.277	199	1.9937	37.5	0.83728
244	187	1.241	147	2.209	37.5	0.83728