



Predicting liquid flow-rate performance through wellhead chokes with genetic and solver optimizers: an oil field case study

Hamzeh Ghorbani¹ · David A. Wood² · Jamshid Moghadasi³ · Abouzar Choubineh⁴ · Peyman Abdizadeh⁵ · Nima Mohamadian⁶

Received: 12 March 2018 / Accepted: 8 August 2018 / Published online: 18 August 2018
© The Author(s) 2018

Abstract

None of the various published models used to predict oil production rates through wellhead chokes from fluid composition and pressures can be considered as a universal model for all regions. Here, a model is provided to predict liquid production-flow rates for the Reshadat oil field offshore southwest Iran, applying a customized genetic optimization algorithm (GA) and standard Excel Solver non-linear and evolutionary optimization algorithms. The dataset of 182 records of wellhead choke measurements spans liquid flow rates from < 100 to 30,000 stock tank barrels/day. Each data record includes measurements of five variables: liquid production rate (QL), wellhead pressure, choke size, basic sediment and water, and gas–liquid ratio. 70% of the dataset (127 data records) was used for training purposes to establish the prediction relationships, and 30% of the dataset (55 data records) was utilized for independently testing the accuracy of the derived relationships as predictive tools. The methodology applying either the customized GA or standard Solver optimization algorithms, demonstrates significant improvements in QL-prediction accuracy with the lowest APD (−7.72 to −2.89), AAPD (7.33–8.51), SD (288.77–563.85), MSE (91,871–316,429), and RMSE (303.1–562.52); and the highest R^2 (greater than 0.997) compared to six previously published liquid flow-rate prediction models. As a general result, the novel methodology is easily applied to other field/reservoir datasets, to achieve rapid practical flow prediction applications, and is consequently of worldwide significance.

Keywords Flow-rate prediction · Evolutionary optimization algorithms · Non-linear optimization · Choke size · Liquid production rate · Wellhead flow-rate variables

Abbreviations

<i>A</i>	Proportionality constant	BS&W	Basic sediment and water
AAPD	Absolute average percent deviation	<i>D</i>	Basic sediment and water term exponent
ANN	Artificial neural network	<i>F</i>	Fitness function
APD	Average percent deviation	GA	Genetic algorithm
<i>B</i>	Bean or choke size exponent	GLR	Producing gas–liquid ratio at standard conditions (SCF/STB)
bbl	Barrel	GRG	Generalized reduced gradient

✉ David A. Wood
dw@dwasolutions.com

Hamzeh Ghorbani
hamzehghorbani68@yahoo.com

Jamshid Moghadasi
j.moghadasi@put.ac.ir

Abouzar Choubineh
abouzar68choubineh@gmail.com

Peyman Abdizadeh
p.abdizadeh1362@gmail.com

Nima Mohamadian
nima.0691@gmail.com

¹ Young Researchers and Elite Club, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran

² DWA Energy Limited, Lincoln, UK

³ Petroleum Engineering Department, Petroleum Industry University, Ahvaz, Iran

⁴ Petroleum Department, Petroleum University of Technology, Ahvaz, Iran

⁵ Iranian Association of Chemical Engineers, North Tehran Branch, Tehran, Iran

⁶ Young Researchers Club, Petroleum Department, Azad University, Omidiyeh Branch, Omidiyeh, Iran

m	Total number of iterations specified for the GA
MSE	Mean squared error
n	Number of data records in the training data subset
P_{wh}	Wellhead pressure, psig
PD_i	Percent deviation of data record i
Q_{avg}	Average value of liquid flow rate
Q_i	i th value of liquid flow rate
Q_L	Gross liquid flow rate, STB/day
$Q_{L(\text{predicted})}$	Predicted production liquid flow rate, STB/day
$Q_{L(\text{measured})}$	Measured production liquid flow rate, STB/day
R^2	Correlation coefficient
RMSE	Root-mean-square error, (STB/D) ²
S	Choke or bean size, 1/64 in.
SD	Standard deviation
σ	Standard deviation of the Gaussian distribution applied for the mutation operator of the genetic algorithm
SCF	Standard cubic foot
STB	Stock tank barrel
T_{avg}	Average fluid temperature of all data records
T_i	i th fluid temperature input value

Introduction

Chokes are key equipment installed at the wellhead of almost all producing oil, gas, and gas condensate wells. Wellhead chokes control and stabilize production flow rates of single or multiple phases, which is essential to prevent reservoir damage by creating a back pressure on the reservoir. Wellhead chokes also maintain the integrity and safety of surface production equipment (downstream of the wellhead), prevent water or gas coning within the reservoir, limit sand production from the reservoir in the flow stream, and provide flexible adjustments that can be used to adjust production rates and ultimate resource-recovery rates for a wide range of reservoir conditions (Guo 2007; Nasriani and Kalantariasl 2011; Mirzaei-Paiaman and Nourani 2012; Mirzaei-Paiaman and Salavati 2013).

Flow through a wellhead choke can be critical or sub-critical (Zarenezhad and Aminian 2011). From the perspective of controlling solid and fines production in a well production flow stream, the critical flow rate represents a production threshold rate above which the production of solids contained within the produced flow stream is uniform. Maintaining sub-critical flow conditions can be important in reservoirs prone to sand or fines production. When the flow rate exceeds this threshold (i.e., becomes critical), the production of sand and fines increases significantly.

Critical flow typically occurs when the pressure upstream of the wellhead is at least 70% higher than the pressure downstream of the wellhead, or when the ratio of downstream pressure to upstream pressure is 0.588 or less. If the ratio of downstream pressure to upstream pressure is greater than 0.588, sub-critical flow conditions prevail (Beiranvand et al. 2012). When critical flow conditions prevail, the flow rate is primarily a function of upstream pressure, gas oil ratio (GOR), and the choke aperture diameter. In critical flow conditions pressure changes in the flow lines downstream of the wellhead do not affect the flow rate therein. Since the liquid production rate through wellhead chokes is affected by changes in their aperture diameter (i.e., choke size), modeling and simulating the rate of flow through chokes of various aperture diameters for specific oil/gas field conditions enables production engineers to better understand and control production flow conditions.

Studies on wellhead chokes and their impacts on production flow rates began with Tangren et al. (1949) and their analysis of flow rate through restrictions. That analysis focused only on critical flow conditions and revealed that when gas bubbles are added to incompressible fluids, they prevent the upstream pressure from being transferred downstream of the restriction. Gilbert (1954) conducted pioneering work on production well-test data and analyzed 260 test datasets for choke sizes from 6/64 to 18/64 in. to derive an experimental relationship for critical flow (Eq. 1)

$$P_{wh} = \frac{CR^m Q}{S^n} \quad (1)$$

where C , m , and n are experimental coefficients, which can be calculated when there is sufficient data (Al-Ajmi et al. 2015). P_{wh} is wellhead pressure (psig). R is gas to liquid ratio (SCF/STB). Q is flow rate (STB/day). S is choke size (1/64 in.).

Gilbert's equation became the basis for many subsequent studies and adaptations beginning with Baxendell (1958), and with Ros (1960) addressing sub-critical as well as critical flow regimes. Achong (1961) claimed an improved relationship based upon a dataset of 104 well tests through chokes with 1/2 to 4 in. Poettmann and Beck (1963) adapted Ros's model for field application, proposing a relationship for critical flow conditions that depends on upstream and downstream pressures.

Two relationships for critical and sub-critical flows were derived by Fortunati (1972), applying the Guzov and Medvedive log for sub-critical flow. Ashford (1974) proposed a model addressing two-phase critical flow through wellhead chokes further developing the model of Ros (1960). Abdul-Majeed (1988), based on data from 155 well tests from eastern Baghdad oilfields developed an experimental model,

based on the Gilbert equation, demonstrating an absolute error of about 6%.

Using 97 well-test datasets and with choke diameters ranging from 24/64 to 128/64 in., Al-Attar (2008) developed an equation that predicts choke performance for sub-critical conditions. Beiranvand et al. (2012) proposed two formulas for estimating the liquid critical flow rate, based on 748 experimental data records including measured data for the input parameters of wellhead pressure, gas–liquid ratio, surface wellhead choke size, and the percentage of water, sediment, and emulsion. Mirzaei-Paiaman and Salavati (2013) investigated the effect of gas-specific gravity and oil-specific gravity on liquid flow rate. According to a statistical analysis of their results, they concluded that neither gas-specific gravity nor oil-specific gravity had a significant effect on flow rate. 704 test datasets from 31 wells in field A were used by Al-Ajmi et al. (2015) to develop a model for flow rate through chokes applying an artificial neural network (ANN) algorithm. Their model resulted in a calculated absolute error for that dataset of 13.92%. Choubineh et al. (2017a, b) improved predictions of wellhead choke liquid critical flow rates, applying a model based on a neural network hybridized with a training-learning-based optimization algorithm, to 113 data points from 12 oil wells in south Iran, achieving an average relative error of 2.09%. Ghorbani et al. (2017) optimized gas flow predictions for 92 datasets from the Pazanan gas condensate field (Iran) applying a firefly algorithm to successfully minimize the mean square error between measured and predicted gas flow rates from a wellhead test dataset.

Our objective here is to provide a new model for estimating oil flow through wellhead chokes improving upon the Gilbert's equation and derivatives of it. Based on this formula, the liquid flow rate is as a function of three parameters such as wellhead pressure, choke size, and gas–liquid ratio. Since almost all models available in the literature used these three variables to predict the flow rate, we decided to include a new parameter named basic sediment and water (BS&W) to develop a model with new parameters and better performance. Also, it is necessary to say that the temperature did not have any significant impact on the flow rate based on relevancy factor formula; not including it in the new model (Discussion section). Another novelty is that it uses a mean square error (MSE) function test as the single-objective function for the GA and Solver algorithms, and then applies the optimized formula to a testing data subset, that is evaluated in terms of a set of statistical accuracy and correlation measures. A dataset of 182 wellhead choke measurements from wells drilled in the Reshadat oil field, located offshore southwest of Iran, is used to evaluate models to predict liquid flow rates through wellhead chokes (see Appendices 1, 2 and 3 for details of the reservoir and individual data record values).

Flow-rate prediction model incorporating optimization with a genetic algorithm

Our objective is to build optimization models that can estimate the liquid flow rate through wellhead chokes as accurately as possible. To achieve this objective, suitable mathematical tools are required. In recent decades, numerous optimization algorithms have been published and refined to find the best solution for optimization problems. We have selected a genetic optimization algorithm (GA) and standard Excel Solver non-linear and evolutionary optimization algorithms because of their simplicity, transparency, and general availability. The procedures for setting up and applying these algorithms are explained here.

A standard genetic algorithm (GA) is applied in this study as an optimization tool to assist in rapidly establishing the most accurate flow-rate prediction models in a novel easy to apply, transparent methodology that can be easily adapted to datasets (populations) of various sizes from other oil fields.

Genetic algorithms (GA) are robust stochastic evolutionary algorithms widely applied for solving optimization challenges in many scientific and industrial fields. The GA concept was initially proposed and exploited in the 1960s and 1970s (Holland 1975), and further developed with improved computation power in the 1980s and 1990s (Goldberg 1989; Jefferys 1993; Mitchell 1996). GA methodologies have continued to be refined in the past decade (e.g., Gen and Cheng 2008; Sivanandam and Deepa 2008) and are now frequently applied to provide multi-objective optimization solutions for various oil and gas engineering (Mansouri et al. 2015) and portfolio (Wood 2016) challenges.

The GA, therefore, offers a well-established, evolutionary optimization method suitable for application to many non-linear optimization problems. Genetic algorithms are easy to apply in a consistent and transparent manner (Fig. 1), involving relatively few control parameters.

The methodology adopted here for predicting liquid flow rate involves a sequence of 12 steps, described in detail in a flow diagram (Fig. 2). It incorporates a customized genetic optimization algorithm, which optimizes a fitness function defined by Eqs. 2 and 3

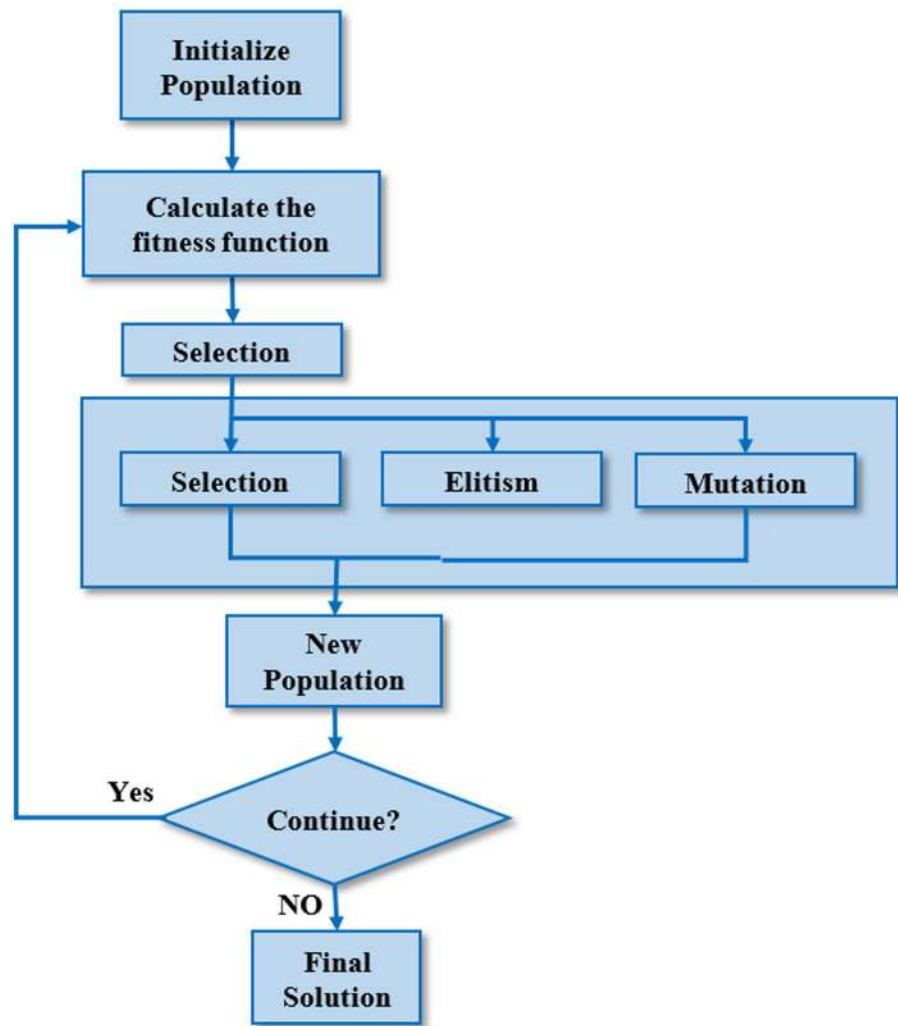
$$\text{Fitness}(f) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n) \quad (2)$$

where

$$\hat{y}_i = \frac{P_{\text{wh}}(D_{64})^C(1 - \text{BS\&W}\%)^D}{A(\text{GLR})^B} \quad (3)$$

y_i = measured liquid flow rate $Q_{L(\text{measured})}$, \hat{y}_i = predicted liquid flow rate $Q_{L(\text{predicted})}$, n number of data records sampled.

Fig. 1 Flow diagram for the sequence involved in a generic genetic algorithm for optimization analysis



A proportionality constant. B gas liquid ratio exponent. C = choke size exponent. D BS&W exponent.

The coefficients A , B , C , and D in Eq. (3) are the independent variables evaluated by the GA with Eq. (2) as its objective function.

A key control parameter for the GA (applied in step 8, Fig. 2) is a shrink factor applied (Eq. 4) as part of Gaussian mutation. This factor causes the standard deviation a Gaussian distribution supplying random numbers to progressively shrink as GA generations (iterations) advance

$$\sigma_{t+1} = \sigma_t \left(1 - \frac{t}{m}\right) \quad (4)$$

where σ is the standard deviation of the gaussian distribution applied for the mutation operator in that generation. t is a specific generation. $t + 1$ is the next generation. m is the total number of generations (iterations) specified for the GA. The impact of the shrink factor is to reduce the scale of change induced by mutation as generations progress toward convergence.

The control and behavioral parameters for the genetic algorithm are given in Table 1.

Excel's Solver provides standard evolutionary and non-linear optimization

The Solver optimizer is a standard function in Microsoft's Excel software and readily applied with options to select linear (simplex), non-linear (generalized reduced gradient, GRG) or evolutionary algorithms to apply. We compare the customized genetic algorithm developed for this study with the Solver options applying GRG and evolutionary algorithms to the same field dataset.

The Solver optimizer is easily set up for the flow prediction model by: (1) specifying a range of cells in a spreadsheet specifying the input data variable range; (2) specifying a range of cells in the spreadsheet specifying the constraints to place on the formula coefficient ranges to be tested; (3) specifying a single cell in the spreadsheet as the statistical

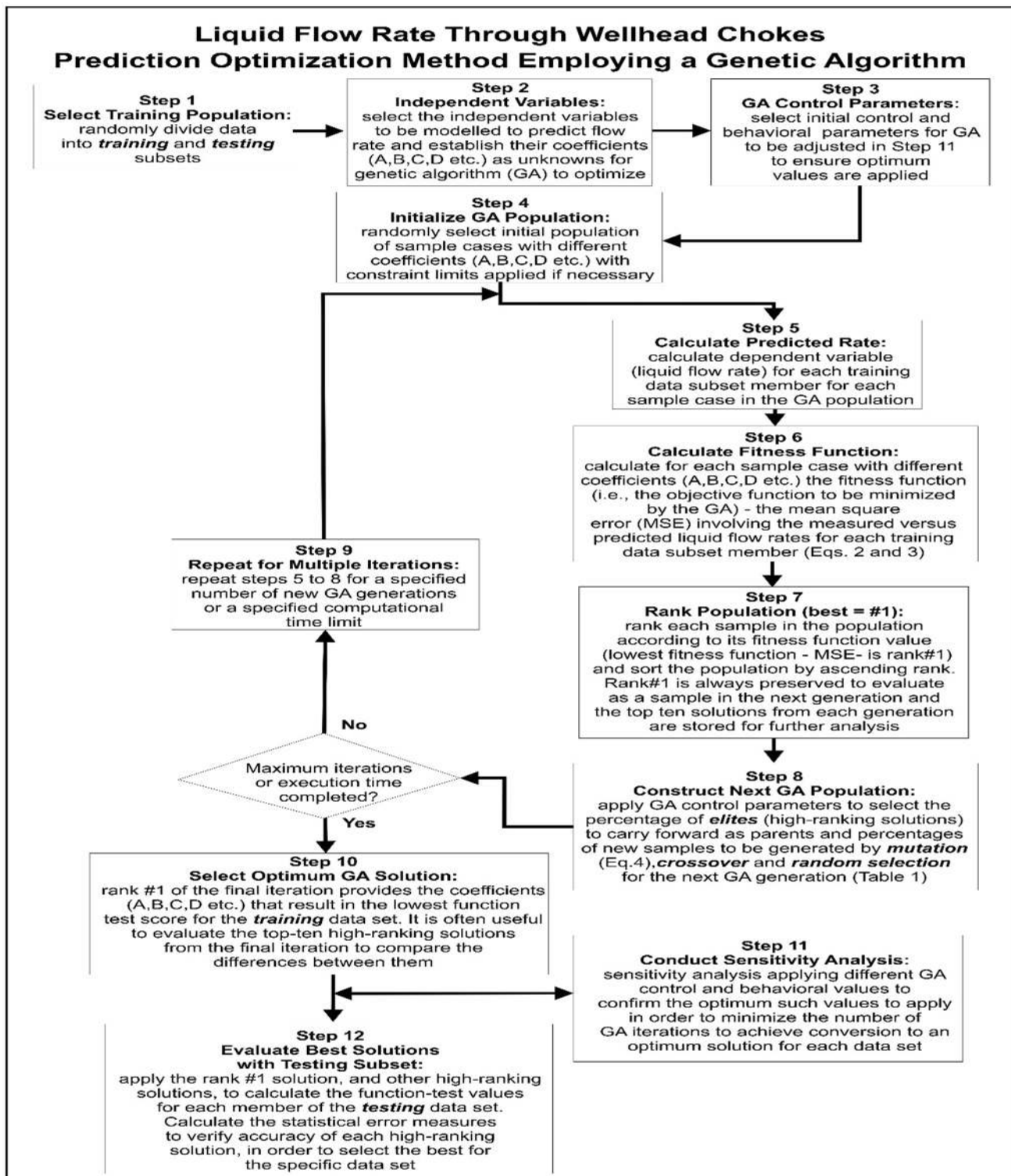


Fig. 2 Flow diagram for the customized methodology to optimize wellhead choke oil flow-rate prediction incorporating a genetic algorithm

precision variable to be minimized by Solver (i.e., the target cell and objective function), and (4) selecting which solver algorithm to apply (e.g. GRG or evolutionary in the context of the current model).

Restrictions on optimization applying Excel's Solver are:

1. Only one objective function variable can be specified for each run of Excel's Solver making it a single-dimen-

Table 1 Genetic algorithm control and behavioral parameters applied

Parameter	Value
Initial population	200
Crossover percent	90%
The number of elites	10
Migrated population	10%
Selection method	Tournament selection
Crossover method	Selection based on binary vector
Mutation method	Gaussian mutation

sional optimizer. This is the way that Microsoft have configured the Solver algorithm in Excel so it cannot be used directly for multi-objective optimization. This is not an issue for the flow-rate prediction model considered here because it does have just one objective function (i.e., liquid flow rate).

2. These specified input variable range must be linked via formula(s) in the spreadsheet to the single-specified objective function.
3. It is important before running the optimizer to test, by some running deterministic cases, that the mathematical relationship between the specified input variables and the objective function is working appropriately. This involves defining cell formula(s) to define the flow-rate predictions for each data record (involving the unknown coefficient values applied to the formula) and the single statistical precision measure to be minimized as the objective function. It also involves testing the constraint ranges applied to the coefficients in the formula. If this is not done, there is a risk that the optimizer will not function as expected.
4. Excel's Solver optimizer only provides a single (optimum) result, if it finds one. In some optimization problems, it is useful to retain and evaluate several high-performing solutions rather than just one. A customized evolutionary algorithm, such as the genetic algorithm applied in this study, can provide multiple high-ranking solutions, if required.

Table 2 Distribution of variable values for the Reshadat oil field dataset

Wellhead choke test variable	Number of points		Range		Mean	
	Training data	Test data	Training data	Test data	Training data	Test data
QL, STB/D	127	55	253–34,450	205–29,400	9930.28	9273
Pwh, psig	127	55	133–842	294–881	526.01	510.16
GLR, Scf/STB	127	55	36–761	42–885	249.24	257.27
BS&W%	127	55	0.05–66	0.02–50	18.997	20.55
D64, inch	127	55	25.6–64	38.4–51.2	59.06	44.80

Analysis of Reshadat oil field data

In this study, 182 datasets were collected from 7 wells penetrating 3 distinct reservoir zones (#1, #2, and #3) that are located adjacent to one another in the Reshadat oil field. The Reshadat oil field is located 110 km south–west of Lavan Island. It was discovered in 1965 and first brought onstream in 1968. It has undergone a significant renovation and extended development since 2008 with five new platforms installed and many new wells drilled in recent years to increase oil production rates. Fluid properties of the three reservoir zones are shown in “Appendix 2” (training subset) and “Appendix 3” (testing subset).

All the collected data are divided randomly into two groups: 70% (127 datasets) were used for training; and, 30% for (55 datasets) for testing. The data include wellhead pressure, gas–liquid ratio, percentage of BS&W [base sediment and water, which incorporates produced water (free water), water in produced sediment/solids, and some water derived from emulsions with oil], choke size, and oil production rate. The ranges and mean values associated with each of these variables in the Reshadat field dataset are listed in Table 2; a full data listing is provided in “Appendix 2”.

Developing a new model for predicting liquid flow rates through wellhead chokes

Oil flow rate through a wellhead choke is a function of: (a) choke size; (b) wellhead pressure; (c) percentage of water produced (expressed as BS&W); and (d) gas–liquid ratio (Beiranvand et al. 2012), which are shown in the following equations:

$$Q_{L(\text{predicted})} = f(P_{\text{wh}}, D_{64}, \text{BS\&W}, \text{GLR}) \quad (5)$$

$$X = \frac{P_{\text{wh}} \times (D_{64}) \times (1 - \text{BS\&W}\%) }{(\text{GLR})} \quad (6)$$

where the units of these variables are: liquid production rate (Q_L) is measured in stock tank barrels (STB/D). Wellhead

pressure (P_{wh}) is measured in pounds per square inch gauge (Psig). Choke size (D₆₄) is measured in 1/64 in. Base sediment and water (BS&W) is measured as a percentage of liquid production. Gas–liquid ratio (GLR) is measured in Scf/STB.

To analyze the X function and find the best relationship between parameters, BS&W, D₆₄, GLR, and P_{wh} to predict $Q_{L(\text{predicted})}$ accurately, $Q_{L(\text{measured})}$ is evaluated versus the calculated value of X (Eq. 6) for each dataset record (Fig. 3). The Q_L versus X cross plot for the Reshadat field dataset (Fig. 3) reveals a correlation coefficient of 0.8901 between these parameters, indicating a good positive correlation with the equation for the best-fit straight line through the data shown in the equation below:

$$X = 0.0184 Q_{L(\text{measured})} - 8.027 \quad (7)$$

The accuracy of the prediction of liquid flow rate [$Q_{L(\text{predicted})}$] can be improved by adding coefficients A , B , C , and D to Eq. (6) to form the expression forming part of the GA fitness test. This relationship (repeated as Eq. 8) enables various values of A , B , C , and D to be applied to it to test the accuracy of the predicted versus measured values of Q_L for each record in the dataset

$$Q_{L(\text{predicted})} = \frac{P_{\text{wh}}(D_{64})^C(1 - \text{BS\&W}\%)^D}{A(\text{GLR})^B} \quad (8)$$

The GA applied uses Eq. (8) as part of its objective function to minimize the fitness function to rapidly find optimum values for coefficients A , B , C , and D that minimize the error between $Q_{L(\text{predicted})}$ and measured $Q_{L(\text{measured})}$ for

the entire dataset. The GA objective function is essentially a mean-square-error (MSE) function and can be expressed in expanded form as the following equations:

$$\text{Fitness}(f) = \text{MSE} = \frac{1}{n} \sum_{i=1}^n ((Q_L)_{\text{Measured}_i} - (Q_L)_{\text{Predicted}_i})^2 \quad (9)$$

$$\text{Fitness}(f) = \frac{1}{n} \sum_{i=1}^n \left((Q_L)_{\text{Measured}_i} - \left(\frac{P_{\text{wh}}(D_{64})^C(1 - \text{BS\&W}\%)^D}{A(\text{GLR})^B} \right)_i \right)^2 \quad (10)$$

where $Q_{L(\text{predicted})}$ = predicted liquid production rate. $Q_{L(\text{measured})}$ = measured liquid production rate. n = number of data records sampled.

To establish appropriate control and behavioral parameters for the GA that speed its convergence toward acceptable and repeatable minima for the fitness function f (Eq. 10), it is necessary to run a series of trials applying different values to these parameters. Each problem and dataset have their own particular characteristics, which mean that GA control parameter values and methods that work well for one dataset may be sub-optimal for another. For the Reshadat field dataset, trials led to the selection of the GA control and behavioral parameters and the GA selection, crossover, and mutation methods as listed in Table 1. Applying those values to the GA, it was then used to determine the optimum values for the A , B , C ,

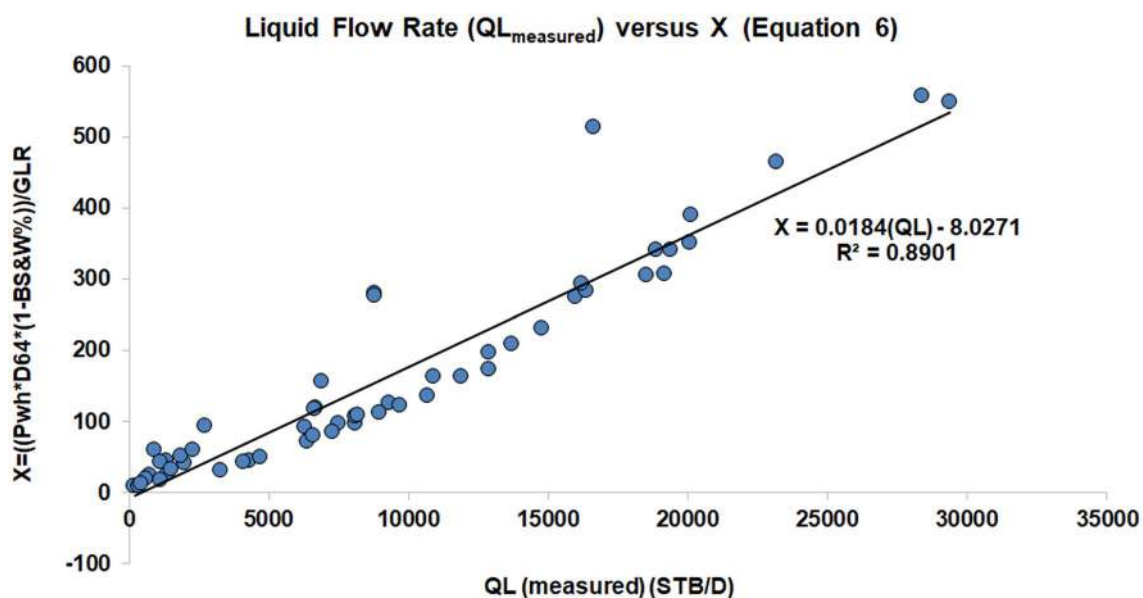


Fig. 3 Liquid production rate $Q_{L(\text{measured})}$ versus parameter X calculated with Eq. (6) displaying the best-fit straight line (Eq. 7) and correlation coefficient

and *D* coefficients to use in Eq. (8), by minimizing fitness function *f* for the 70% of the dataset records allocated for training. Exactly, the same approach was used when applying Excel’s Solver optimizers (GRG and evolutionary) to the Reshadat field dataset.

Statistical measures used to measure the accuracy of optimum solutions

Once the GA, by minimizing the fitness function, has selected optimum values for coefficients *A*, *B*, *C*, and *D* to apply in Eq. (8), it is necessary to establish the accuracy with which Eq. (8) can match *Q_L*(predicted) values with *Q_L*(measured) values for each record the test section of the dataset (i.e., the 30% of the entire Reshadat field data records selected randomly to form the test subset).

The following six statistical error measures for accuracy, precision, and correlation (expressed as Eqs. 11 to 17) were calculated for the optimum solution values for coefficients *A*, *B*, *C*, and *D* (Eq. 8) applied to the test subset with results shown in Table 3.

Percent deviation for dataset record *i* (PD_{*i*}):

$$PD_i = \frac{Q_{L(Measured)} - Q_{L(Predicted)}}{Q_{L(Measured)}} \times 100 \tag{11}$$

Average percent deviation (APD):

$$APD = \frac{\sum_{i=1}^n PD_i}{n} \tag{12}$$

Absolute average percent deviation (AAPD):

$$AAPD = \frac{\sum_{i=1}^n |PD_i|}{n} \tag{13}$$

Standard deviation (SD):

$$SD = \sqrt{\frac{\sum_{i=1}^n (D_i - D_{imean})^2}{n - 1}} \tag{14}$$

where *D_i* is *Q_L*(*measured*) – *Q_L*(*predicted*) for each dataset record *i*. *D_{imean}* is the mean of the *D_i* values

$$D_{imean} = \frac{1}{n} \sum_{i=1}^n ((Q_L)_{Measured_i} - (Q_L)_{Predicted_i})$$

Mean square error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n ((Q_L)_{Measured_i} - (Q_L)_{Predicted_i})^2 \tag{15}$$

Also used for the GA fitness function *f* (Eqs. 7 and 8).

Table 3 Optimum values for coefficients *A*, *B*, *C*, and *D* applied to Eqs. 6 and 8 and other published liquid flow prediction (Eq. 19) for wellhead choke relationships

Wellhead choke flow relationships	Coefficients				APD%	AAPD%	SD	MSE	RMSE	R squared (R ²)
	A	B	C	D						
(Coefficients = 1) (Eq. 3)	1	1	1	1	98.01	98.01	7318.64	1.36E+08	11643.97	0.8901
This work—genetic algorithm (optimized coefficients) (Eq. 5)	1.3522	0.74042	1.7056	-0.164	-2.89	7.33	563.85	316,429	562.52	0.9970
Solver (GRG)	16.4957	0.7136	2.29	0.015	-6.27	8.43	345.40	117,313	342.50	0.9987
Solver (evolutionary)	17.412	0.709	2.3	0.00001	-7.72	8.51	288.77	91,871	303.10	0.9989
Gilbert (1954)	10.00	0.546	1.84	0	24.60	32.79	3960.55	30,002,360	5477.44	0.9821
Baxendell (1958)	9.552	0.546	1.93	0	-11.72	26.57	2190.60	5,768,234	2401.71	0.9823
Ros (1960)	17.40	0.500	2.00	0	-3.57	26.46	2817.55	10,718,503	3273.91	0.9696
Achong (1961)	3.82	0.650	1.88	0	-30.76	30.93	579.72	1,066,113	1032.52	0.9946
Mirzaei-Paiaman and Salavati (2013)	11.41	0.533	1.92	0	2.99	27.13	2955.87	13,327,660	3650.70	0.9793
Beiranvand et al. (2012)	26.17	0.5154	2.151	0.5297	0.37	19.29	2711.39	9,357,695	3059.03	0.9359

Root-mean-square error (RMSE):

$$\text{RMSE} = \sqrt{\text{MSE}} \quad (16)$$

Coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^N (Q_{L(\text{Predicted})} - Q_{L(\text{Measured})})^2}{\sum_{i=1}^N \left(Q_{L(\text{Predicted})} - \frac{\sum_{i=1}^n Q_{L(\text{Measured})}}{n} \right)^2} \quad (17)$$

The statistical measures (Eqs. 11–17) are also reported in Table 3 for our GA evaluations, using the same method described for our proposed Eq. (8), of the liquid flow-rate prediction equations proposed and historically published by Gilbert (1954), Baxendell (1958), Ros (1960), and Achong (1961) (shown here as Eq. 19 with different values for coefficients A , B , and C) applied to the Reshadat field dataset.

Re-arranging Gilbert's formula (Eq. 1) to predict liquid flow rate, and renaming the coefficients as A , B , and C to match those coefficients expressed in Eq. (8) provides the equation below:

$$Q = \frac{P_{\text{wh}} S^B}{AR^C} \quad (18)$$

Using the same symbols for the variables as Eq. (5), Gilbert's formula for predicting flow (Eq. 18) can be expressed as Eq. (19). This omits the BS&W term (not considered by Gilbert) from Eq. (8) [which is the same as applying a value of zero to exponent D in (Eq. 8)]. The correlations derived by Baxendell (1958), Ros (1960), Achong (1961), Beiranvand et al. (2012), and Mirzaei-Paiaman and Salavati (2013) involve different values for the coefficients A , B , and C applied to the following equation (Table 3):

$$Q_{L(\text{predicted})} = \frac{P_{\text{wh}} (D_{64})^C}{A(\text{GLR})^B} \quad (19)$$

where Gilbert (1954) derived values of coefficients $A = 10$; $B = 0.546$; $C = 1.84$; Baxendell (1958) correlation involves values of coefficients $A = 9.56$; $B = 0.546$; $C = 1.93$; Ros (1960) correlation involves values of coefficients $A = 17.4$; $B = 0.5$; $C = 2.00$; Achong (1961) correlation involves values of coefficients $A = 3.82$; $B = 0.65$; $C = 1.88$; Mirzaei-Paiaman and Salavati (2013) correlation involves values of coefficients $A = 11.41$; $B = 0.553$; $C = 1.92$; BS&W coefficient $D = 0$ in all five of the above cases.

Beiranvand et al. (2012) correlation involves values of coefficients $A = 26.17$; $B = 0.5154$; $C = 2.151$; $D = 0.5297$.

Table 3 reveals the significant superiority of our proposed Eq. (8), in terms of accuracy, compared to the other liquid flow-rate prediction formulas evaluated. APD, AAPD, SD, MSE, and RMSE values are all much lower for optimization of the Reshadat oil field dataset than for the

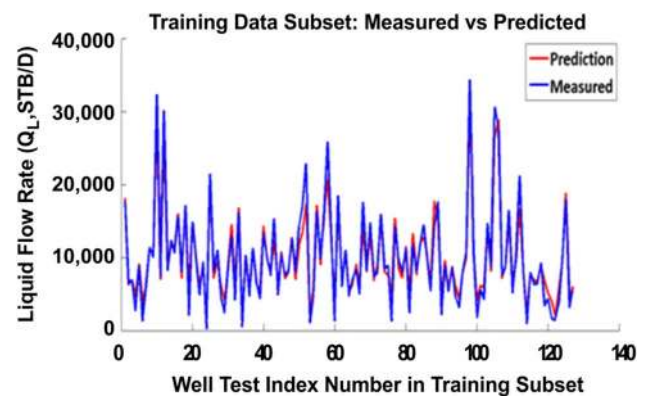


Fig. 4 Liquid flow-rate (Q_L) prediction reliability and accuracy for the Reshadat oil field training data subset applying the GA optimized Eq. (8)

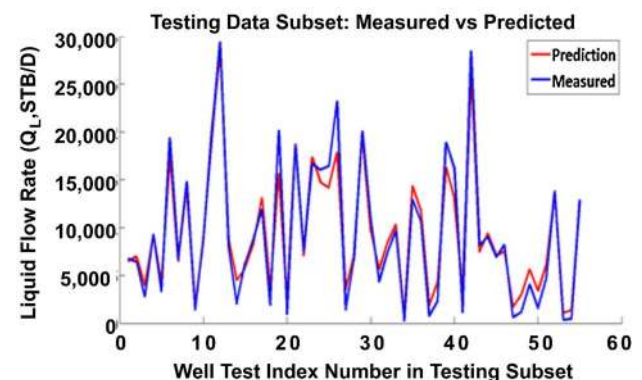


Fig. 5 Liquid flow-rate (Q_L) prediction reliability and accuracy for the Reshadat oil field testing data subset applying the GA optimized Eq. (8)

other published formulas applying Eq. (19). All formulas evaluated show high correlation coefficients ($>> 0.9$ or 90%), which is hardly surprising as Eq. (6) (with A , B , C , and D coefficients all equal to 1) achieves a correlation coefficient of 0.89 (89%). However, the high correlation coefficients reveal little about accuracy of the predictions [i.e., $Q_{L(\text{predicted})}$ versus $Q_{L(\text{measured})}$]. As can be seen from Fig. 3 and Table 3, the values of X calculated by Eq. (6) include orders of magnitude of error in relation to $Q_{L(\text{measured})}$. Although the Gilbert (1954), Baxendell (1958), Ros (1960), Achong (1961), Beiranvand et al. (2012), and Mirzaei-Paiaman and Salavati (2013) formulas also significantly outperform Eq. (6) in terms of accuracy, their levels of accuracy (i.e., APD, AAPD, SD, MSE, and RMSE values) they are much inferior in accuracy to those achieved by our proposed Eq. (8) and the Solver (GRG and evolutionary) solutions (Table 3).

Figures 4 and 5 show a comparison of $Q_{L(\text{predicted})}$ versus $Q_{L(\text{measured})}$ for each data point training subset (127 randomly

selected data records) and testing subset (55 randomly selected data records), respectively, of the Reshadat oil field dataset (total of 182 wellhead test data records).

Figures 4 and 5 show excellent agreement between $Q_{L(\text{predicted})}$ versus $Q_{L(\text{measured})}$ for both training and testing datasets using Eq. (8), which is confirmed by Fig. 6.

Figure 7 demonstrates the ability of the GA optimization method to transform the highly inaccurate functional

relationship expressed as Eq. (6) into the highly accurate prediction formula expressed as Eq. (8).

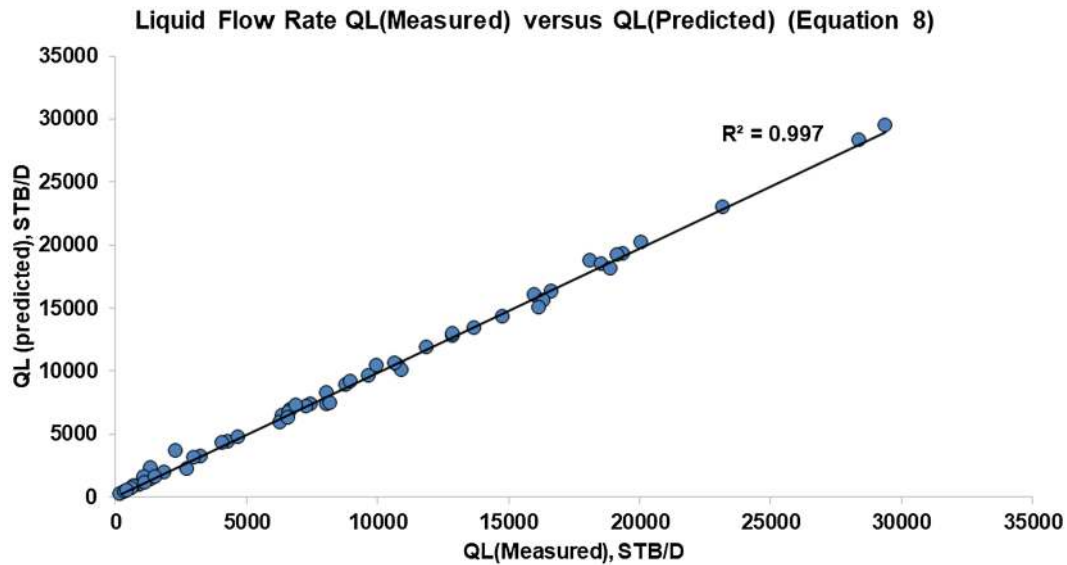


Fig. 6 Liquid flow rate $Q_{L(\text{predicted})}$ versus $Q_{L(\text{measured})}$ reveals a strong correlation and high reliability and accuracy for the Reshadat oil field applying the GA optimized Eq. (8) to the test data subset (55 data records)

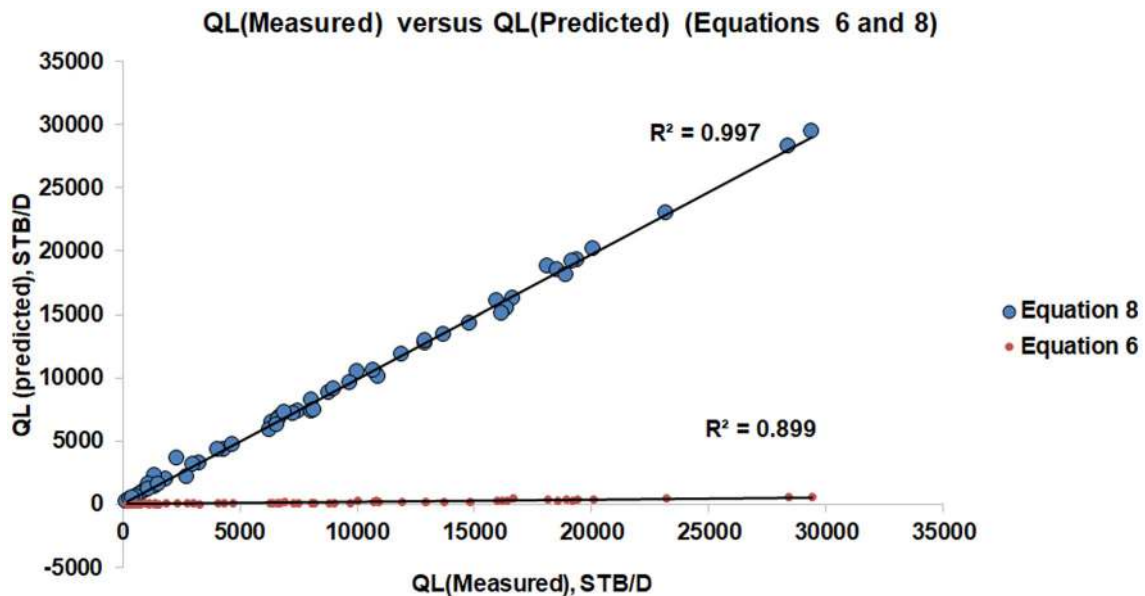


Fig. 7 Liquid flow rate $Q_{L(\text{predicted})}$ versus $Q_{L(\text{measured})}$ reveals a strong correlation for the Reshadat oil field applying Eq. (6) and the GA optimized Eq. (6) to the test data subset (55 data records). However, Eq. (6) provides highly unreliable and inaccurate predictions in comparison to Eq. (8)

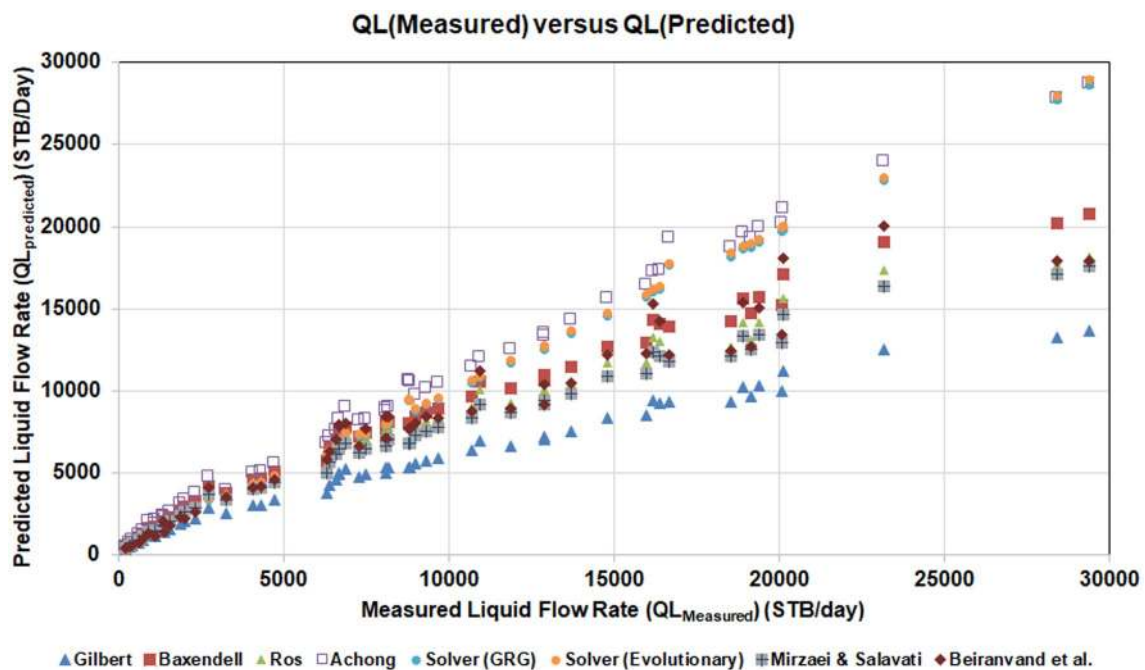


Fig. 8 Liquid flow rate $Q_{L(\text{predicted})}$ versus $Q_{L(\text{measured})}$ achieved by applying the two Excel Solver optimization algorithms (non-linear and evolutionary) and the published liquid flow-rate prediction formulas of Gilbert (1954), Baxendell (1958), Ros (1960), Achong

(1961), Beiranvand et al. (2012), and Mirzaei-Paiaman and Salavati (2013) to the Reshadat oil field applying Eq. (19) with appropriate coefficients to the test data subset (55 data records)

Performance comparisons with published correlations

Figure 8 compares the relative performance for the Reshadat oil field dataset of applying Eq. (19) with the *A*, *B*, *C*, and *D* (where appropriate) coefficient values of Gilbert (1954), Baxendell (1958), Ros (1960), Achong (1961), Beiranvand et al. (2012), Mirzaei-Paiaman and Salavati (2013) and Excel Solver (GRG and evolutionary). All show good correlations (coefficient of determination ≥ 0.93) between measured and predicted liquid flow rates, but not very impressive accuracy (Table 3). However, the performance of all of published correlations are substantially inferior in terms of accuracy to Eq. (8) (GA solution) and the two Solver solutions, as measured by a range of statistical metrics (Table 3).

Figures 9 and 10 show the percent deviation (PD_{*i*}) for each data record in the test data subset applying the GA optimization algorithm to Eqs. (6) and (8) (Fig. 9) and with various coefficient values from published studies and Solver algorithms applied to Eq. (19) (Fig. 10) for predicting liquid flow rate. The PD_{*i*} measure is useful in revealing where in the production rate range significant errors occur. In the case of Eq. (6), high errors occur across the entire production rate range of the dataset. In the case of Eq. (8), significant errors (beyond the plus or minus 20% range) only occur in the prediction of flow rates in four data records with flow rates

below 3000 STB/D (see Figs. 5, 9). This is not surprising as higher percentage errors are much more likely to be calculated using Eq. (11) for low rate tests. What is encouraging for Eq. (8) is that so few data records suffer from such high PD_{*i*} errors, even among the data records with low measured flow rates.

In contrast to Fig. 9, the PD_{*i*} result comparisons for the published Eq. (19) applying *A*, *B*, *C*, and *D* (where appropriate) coefficient values of Gilbert (1954), Baxendell (1958), Ros (1960), Achong (1961), Beiranvand et al. (2012), and Mirzaei-Paiaman and Salavati (2013) plotted in Fig. 10. These reveal, for these published correlations, predictions of multiple records with flow rates less than 3000 STB/D display PD_{*i*} values of less than minus 50%. This suggests that these formulas are systematically over-estimating liquid flow-rate predictions for the lower flow-rate data records. On the other hand, Gilbert, Baxendell, and Ros formulas are all under-estimating liquid flow-rate predictions for the higher flow-rate data records (> 5000 STB/D). The Gilbert formula performs the worst for this dataset by under-estimating liquid flow-rate predictions for flow rates > 3000 STB/D to a progressively higher degree from about 25% at 5000 STB/D to > 50% at 30,000 STB/D. Figure 10 reveals that for flow rates above 3000 STB/D, the Achong formula performs the best of these published formulas with low (negative) PD_{*i*} errors but performs the worst at flow rates < 3000 STB/D,

Fig. 9 Percent deviation (PD_i) calculated by Eq. (11) for the Reshadat oil field applying Eq. (6) and the GA optimized Eq. (8) to the test data subset (55 data records). Equation (6) displays very high PDi values (reflecting high prediction errors), whereas Eq. (8) displays very low PDi values (reflecting low prediction errors)

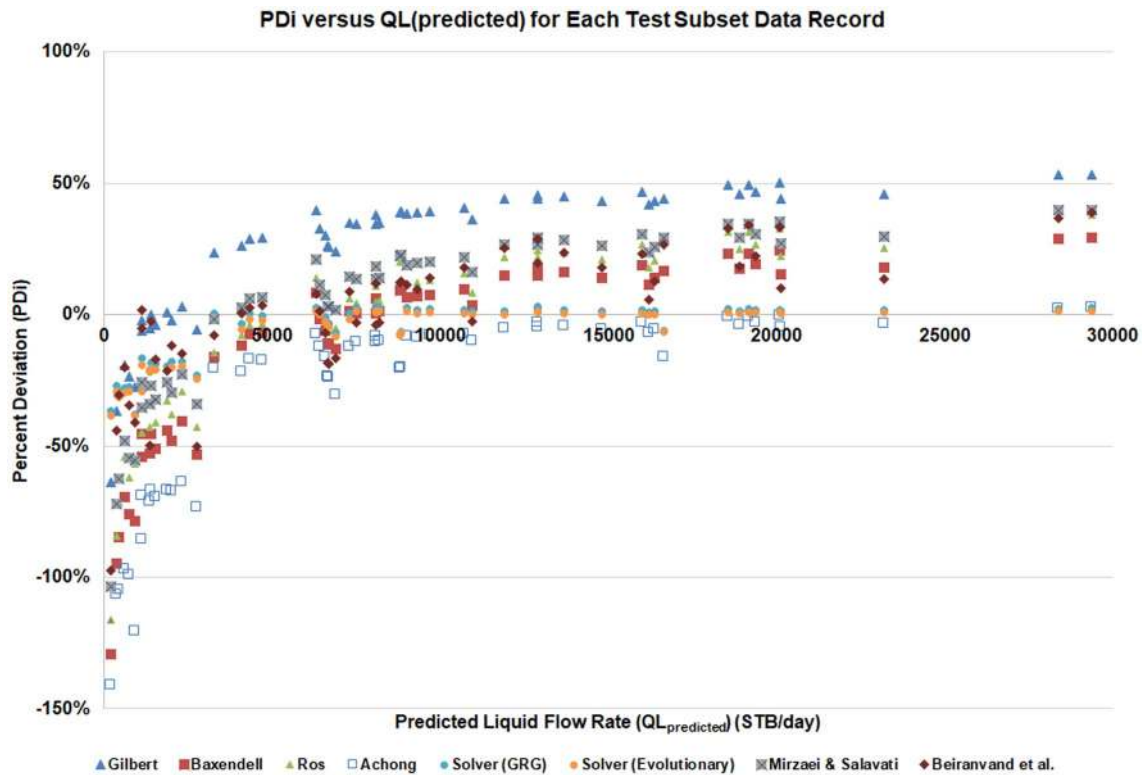
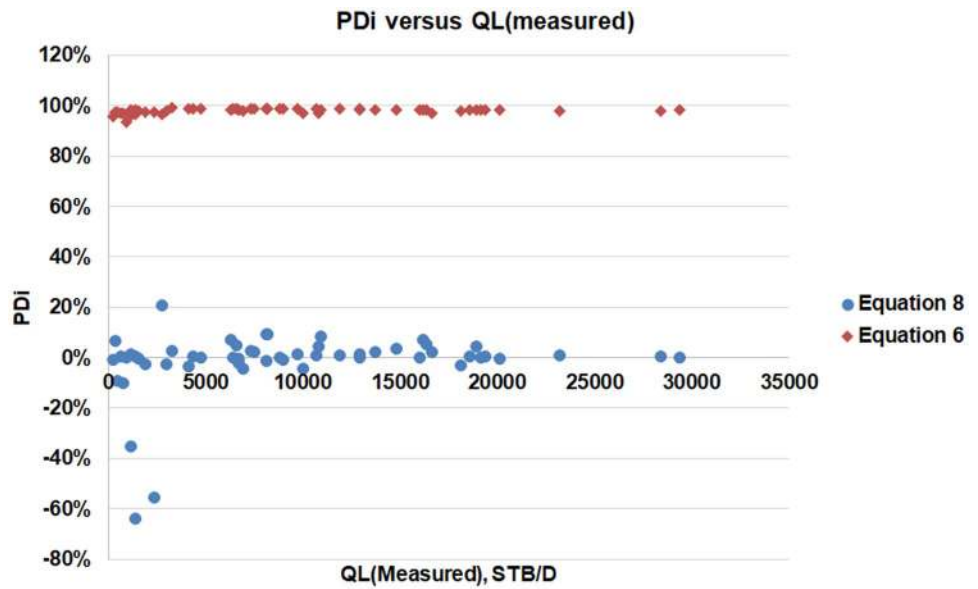


Fig. 10 Percent deviation (PD_i) calculated by Eq. (11) for the Reshadat oil field applying various coefficients to Eq. (19) for the published liquid flow-rate prediction formulas of Gilbert (1954), Baxendell (1958), Ros (1960), Achong (1961), Beiranvand et al. (2012),

and Mirzaei-Paiaman and Salavati (2013) to the test data subset (55 data records). The Achong (1961) formula displays lower PDi values than the other published formulas (reflecting lower prediction errors), but these are higher than for our Eq. (8) (Fig. 9)

particularly when compared to the Excel Solver (GRG and evolutionary) solutions.

Figures 11, 12, and 13 confirm the superior performance in terms of statistical accuracy of Eq. (8) optimized with evolutionary and/or non-linear algorithms compared

Fig. 11 A comparison between root-mean-square error (RMSE) and absolute average percent deviation (AAPD%) for the Reshadat oil field applying GA and Solver optimization to Eqs. (6), (8), and (19) for the published liquid flow-rate prediction formulas proposed by this study and by Gilbert (1954), Baxendell (1958), Ros (1960), Achong (1961), Beiranvand et al. (2012), and Mirzaei-Paiaman and Salavati (2013) to the test data subset (55 data records)

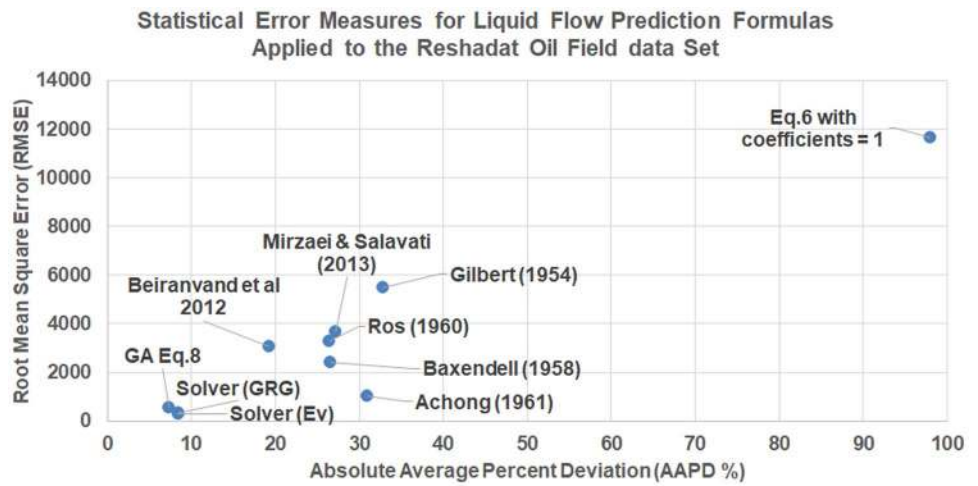
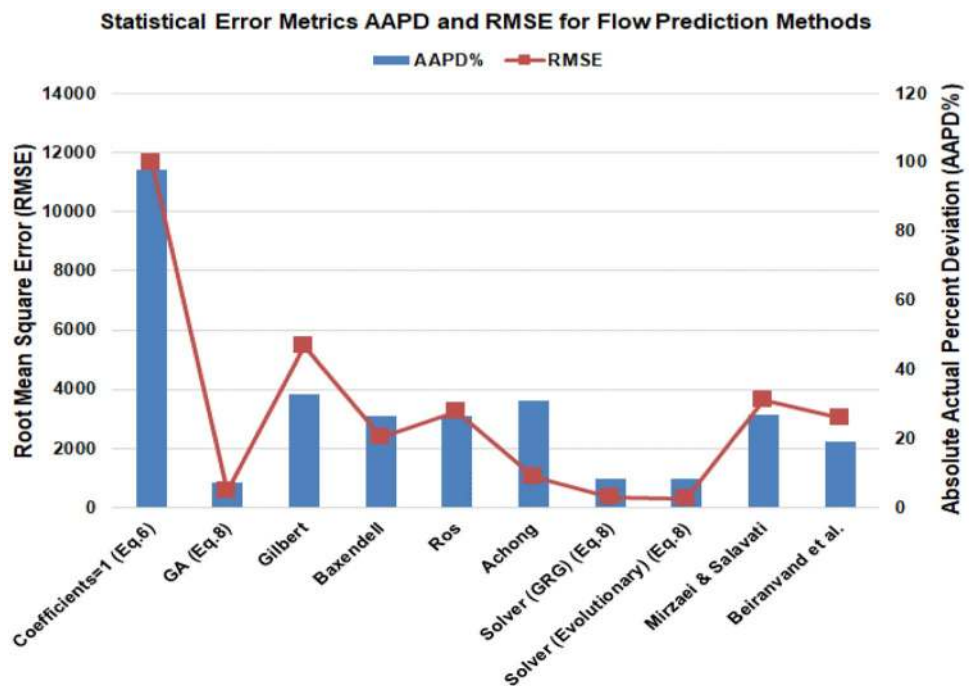


Fig. 12 A comparison between root-mean-square error (RMSE, Eq. 16) and absolute average percent deviation (AAPD%, Eq. 13) for the Reshadat oil field for the various flow-rate prediction methods identified in Fig. 11



to the other formula published for predicting liquid flow rate from wellhead test data records for the Reshadat oil field dataset.

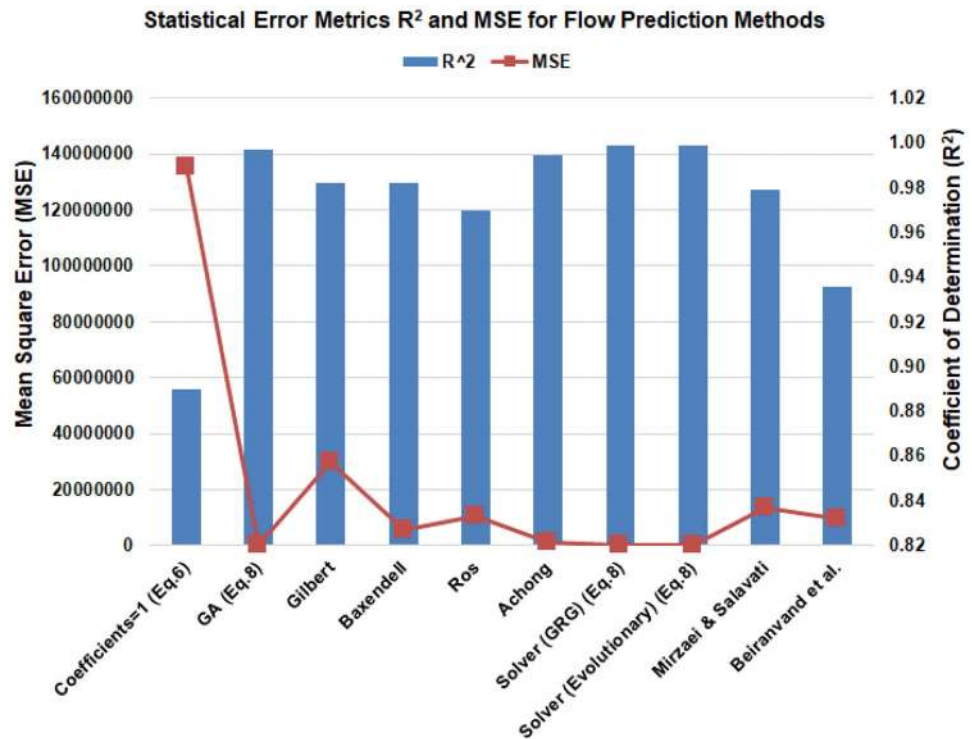
Discussion

The analysis presented for flow-rate prediction for the Reshadat field dataset indicate that both evolutionary and non-linear optimization algorithms can provide highly accurate results by applying the proposed methodology. Although there is almost negligible difference between the level of accuracy achieved by the customized genetic algorithm (GA) and Excel Solver’s GRG and evolutionary

optimizers methods is the same approximately, the Solver optimizers display slightly higher levels of accuracy. The preciseness of the GA is shown to be fit for purpose acceptable but is slightly outperformed by the Solver solutions.

We consider the main reason for this slight difference between GA and the Solver solutions is due to the several behavioral and control setting parameters associated with the GA (Table 1). The developed GA requires that values have to be selected for the behavioral and control setting, including: initial population, crossover percent, the number of elites, migrated population, selection method, and crossover method. The values for these metrics are usually determined by trial and error or tuned to a specific dataset. This can be time consuming and mean that when applied

Fig. 13 A comparison between mean square error (MSE, Eq. 15) and coefficient of determination (R^2 , Eq. 17) for the Reshadat oil field for the various flow-rate prediction methods identified in Fig. 11



to a slightly different set of data, slightly higher errors are incurred in the prediction (e.g., tuning the control metrics for the training data subset and then applying them to the testing data subset). One advantage of the Solver optimizers is that they can be setup and applied rapidly avoiding the need to tune any behavioral or control metrics.

The methodology has the potential to add additional relevant variables, if required. We have not considered additional variables, such as fluid temperature, because other studies have suggested that additional variables to those considered have very small impacts on liquid flow rates from oil reservoir production or well tests. For instance, Choubineh et al. (2017a, b) concluded that the effects of temperature on liquid flow rate is not remarkable. Furthermore, sensitivity analysis performed on the Reshadat field dataset suggests that it is best not to include fluid temperature in the liquid flow-rate performance model. Relevancy factor (r) as defined by Choubineh et al. (2017a, b) was calculated using Eq. (20) to investigate the dependence of the liquid flow rate on fluid temperature. The r value calculated is 0.11 that shows it does not have any significant effect on liquid flow rate

$$r = \frac{\sum_{i=1}^N (T_i - T_{ave})(Q_i - Q_{ave})}{\sqrt{\sum_{i=1}^N (T_i - T_{ave})^2 \sum_{i=1}^N (Q_i - Q_{ave})^2}} \quad (20)$$

where T_i is the i th fluid temperature input value; T_{ave} is the average fluid temperature of all data records; Q_i is the i th value of liquid flow rate; Q_{ave} is the average value of liquid

flow rate; and, N is the number of all data records in the dataset.

The QA and Solver algorithms presented are configured as single-objective optimization calculations. It is possible to expand these algorithms as multi-objective algorithms by using a function test based on assigning a score to each of the statistical measures of accuracy identified in “Statistical measures used to measure the accuracy of optimum solutions”. The multi-objective optimization algorithms are then configured to maximize the combined function test score. This approach is deemed unnecessary for the Reshadat field dataset, because the GA and Solver algorithms achieve acceptable levels of accuracy configured to minimize MSE. However, in datasets where a single-objective statistical measure does not achieve sufficient levels of accuracy between predicted and measured liquid flow rates, applying a multi-objective algorithm configured as described is likely to improve that accuracy.

Conclusions

A new liquid flow-rate prediction model based on 182 collected data records of well tests through production chokes from seven wells drilled in the Reshadat oil field, offshore Southwest Iran, is developed and evaluated. The novel liquid flow-rate-prediction formula employed (Eq. 8) involves four independent variables and when optimized with either a customized genetic algorithm or Excel’s Solver optimizers

demonstrates high accuracy in predicting flow rate for the Reshadat oil field data. For the genetic algorithm it achieves the following values for statistical accuracy measures: average percent deviation = -2.89% ; absolute average percent deviation = 7.33% ; standard deviation = 563.85 ; mean squared error = $316,429$; root-mean-square error = 562.52 ; and coefficient of determination = 0.9970 .

The optimization methodology applied divided the dataset into training and testing subsets and minimizes as the objective function the mean square error between measured and predicted flow rates. This approach is easy and flexible to setup train and test and is readily adaptable for application to other well-test datasets.

The two Solver algorithms achieved lower MSE values than the GA algorithm, with the Solver evolutionary algorithm achieving the lowest MSE value of $91,871$. This confirms that both non-linear and evolutionary optimization algorithms are almost equally effective when applied with the proposed methodology to the Reshadat field dataset. Applying Eq. (8) with optimized values for its coefficients

derived from non-linear and evolutionary algorithms, fine-tuned for specific field data, should enable production engineers to significantly improve their flow-rate predictions for this field compared to other methods.

Acknowledgements The authors thank the National Iranian Oil Company (NIOC) and its subsidiary company, National Iranian South Oil Company (NISOC), for their support during this study and for providing the Reshadat oil field dataset.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix 1

See Table 4.

Table 4 Fluid properties of oil reservoirs in the Reshadat field

Fluid property	Reservoir zone			Reservoir conditions		
	#1	#2	#3	#1	#2	#3
Bubble point pressure, psig	2713	2052	2273	Under-saturated oil at reservoir condition		
Reservoir initial pressure, psig	3500	3746	2573			
Oil formation volume factor, –	1.39	1.72	1.41	3500 psig and 190°F	3746 psig and 182 °F	2573 psig and 188 °F
Oil viscosity, cp	4.3	3.9	3.85			
Crude oil density, °API	20–23.7	21.01–22	19–24			
Gas oil ratio, Scf/STB	338	268	304			
Reservoir temperature, °F	190	182	188			
Oil compressibility, 1/psi	10.2×10^{-6}	13.04×10^{-6}	9×10^{-6}			

Appendix 2

See Table 5.

Table 5 Reshadat oil field wellhead test data—genetic algorithm training subset (127 records—70% of database)

Reshadat oil field wellhead choke training data subset										Reshadat oil field wellhead choke training data subset										Reshadat oil field wellhead choke training data subset									
Variable type	Independent			Dependent	Variable type	Independent			Dependent	Variable type	Independent			Dependent	Variable type	Independent			Dependent										
	Choke size	Pwh	BS&W			GLR	Q_L	(STB/D)			(SCF/STB)	Choke size	Pwh			BS&W	GLR	Q_L		(STB/D)	(SCF/STB)	Choke size	Pwh	BS&W	GLR	Q_L	(STB/D)	(SCF/STB)	
Number	(1/64) in.	(Psi)	(%)	(SCF/STB)	Number	(1/64) in.	(Psi)	(%)	(SCF/STB)	Number	(1/64) in.	(Psi)	(%)	(SCF/STB)	Number	(1/64) in.	(Psi)	(%)	(SCF/STB)	Number	(1/64) in.	(Psi)	(%)	(SCF/STB)					
1	64	450	28	73	17,700	44	61.44	456	30	386	4917	87	64	457	24	393	5417												
2	64	538	12.4	400	6400	45	64	508	21	177	10,500	88	64	454	40	102	14,200												
3	64	540	15	356	6870	46	64	571	9	343	7489	89	64	649	6.5	121	17,700												
4	51.2	424	38.5	406	2715	47	64	525	14	269	8200	90	38.4	580	24	300	2175												
5	51.2	600	7	125	8910	48	64	523	22	144	12,720	91	64	499	24	222	8905												
6	38.4	392	53	351	1289	49	64	616	0.1	335	8000	92	51.2	783	0.05	369	5355												
7	64	510	16	380	6180	50	64	708	0.4	188	14,200	93	64	493	20	230	8630												
8	51.2	621	4	92	11,390	51	64	675	0.4	136	17,065	94	61.44	435	32	405	4554												
9	64	501	24	189	10,049	52	64	720	0.1	99	22,870	95	61.44	433	30	661	3160												
10	64	530	20	39	32,365	53	61.44	133	28	646	1004	96	64	549	15	299	7992												
11	64	568	10	348	7400	54	61.44	421	34	390	4632	97	64	535.1	12.4	189	10,720												
12	64	556	20	46.6	30,240	55	64	439	30	80	16,300	98	64	527	26	36	34,450												
13	64	543	16	280	8300	56	64	512	11	201	9800	99	64	464	18	149	11,000												
14	64	479	18	131	12,400	57	64	423	30	82	15,300	100	30.72	330	40	84	1675												
15	64	511	24	178	10,670	58	58.88	784	0.4	69	25,878	101	64	466	22	388	5578												
16	64	439	27	84	15,700	59	51.2	655	0.3	76	13,740	102	61.44	417	34	415	4250												
17	64	600	4	333	8045	60	30.72	342	40	133	1248	103	64	603	13.5	142	14,760												
18	51.2	519	15	41	17,100	61	64	549	16	91	18,500	104	64	573	10.5	279	8700												
19	30.72	342	38	67	2039	62	64	467	24	348	6050	105	64	559	18	46	30,650												
20	64	467	24	98	14,870	63	64	549	15	188	11,000	106	64	462	34	41	27,600												
21	64	480	24	192	9600	64	51.2	833.8	0.1	477	4700	107	64	580	10.3	352	7500												
22	61.44	446	28	372	4920	65	64	371	9	205	7000	108	64	582	16	284	8700												
23	64	537	18	242	9458	66	64	494	24	238	8860	109	64	616	13.5	126	16,450												
24	25.6	395	12	761	253	67	61.44	456	30	381	5002	110	61.44	455	30	364	5118												
25	64	490	26	63	21,500	68	64	677	0.8	131	17,600	111	64	518	12.5	194	10,100												
26	64	593	5.5	331	7934	69	64	561	14	296	8200	112	64	670	3.3	99	21,270												

Table 5 (continued)

Reshadat oil field wellhead choke training data subset				Reshadat oil field wellhead choke training data subset				Reshadat oil field wellhead choke training data subset			
Variable type	Independent	Dependent	Variable type	Independent	Dependent	Variable type	Independent	Dependent	Variable type	Independent	Dependent
Data record	Choke size	GLR	Data record	Choke size	GLR	Data record	Choke size	GLR	Data record	Choke size	GLR
Number	(1/64) in. (Psi)	(%)	Number	(1/64) in. (Psi)	(%)	Number	(1/64) in. (Psi)	(%)	Number	(1/64) in. (Psi)	(%)
		(SCF/STB)			(SCF/STB)			(SCF/STB)			(SCF/STB)
27	64	722.1	0.3	275	11,000	70	64	628	10.5	150	14,800
28	61.44	421	34	387	4520	71	64	560	9	347	7345
29	38.4	650	18.6	299	2390	72	64	482	26	244	8000
30	64	580	15	297	8500	73	64	549	18	111	16,000
31	64	419	32	102	13,000	74	51.2	378	3.8	69	8500
32	61.44	411	32	414	4210	75	64	664	0.1	336	8900
33	64	488	27.5	93	16,200	76	43.52	455	66	670	1310
34	30.72	324	44	463	490	77	64	444	30	97	14,300
35	64	558	9	208	10,300	78	64	551	11	249	9700
36	61.44	436	30	408	4668	79	64	602	27	325	8002
37	64	528	18	172	11,300	80	64	517	24	173	11,300
38	64	511	15	368	6350	81	38.4	602	28	269	2399
39	46.08	832	2.4	362	4392	82	64	420	32	118	11,800
40	64	426	30	102	13,400	83	51.2	372	3.8	74	8020
41	64	541	17	214	9900	84	64	460	22	133	11800
42	51.2	775	0.2	225	7530	85	56.32	748	0.2	122	14,500
43	64	673	0.4	157	15,370	86	64	506	23	197	9834

Statistical summary for the 127 data records of the testing subset

	Choke size	Pwh	BS&W%	GLR	QL
Min	25.6	324	0.2	39	253
Max	64	832	53	761	32,365
Mean	58.23	520.14	19.83	244.76	9949.23

Appendix 3

See Table 6.

Table 6 Reshadat oil field wellhead test data—accuracy testing subset (55 records—30% of database)

Reshadat oil field wellhead choke testing data subset						Reshadat oil field wellhead choke testing data subset					
Variable type	Independent				Dependent	Variable type	Independent				Dependent
Data record	Choke size	Pwh	BS&W	GLR	Q_L	Data record	Choke size	Pwh	BS&W	GLR	QL
Number	(1/64) in.	(Psi)	(%)	(SCF/STB)	(STB/D)	Number	(1/64)in.	(Psi)	(%)	(SCF/STB)	(STB/D)
1	56.32	727	0.05	347	6700	29	64	444	26	60	20,100
2	64	505	21	358	6400	30	64	689	4	261	10,940
3	38.4	881	0.05	362	2750	31	61.44	420	30	410	4330
4	64	502	18	210	9320	32	64	475	26	267	7300
5	61.44	443	28	652	3286	33	64	513	24	206	9700
6	64	575	16	91	19,399	34	25.6	350	12	885	205
7	56.32	722	0.02	348	6667	35	64	429	32	108	12,900
8	64	559	18	127	14,800	36	64	5D3	27	173	10,700
9	38.4	462	0.8	400	1370	37	30.72	439	32	400	735
10	51.2	378	4.8	66	8800	38	38.4	583	28	272	2320
11	64	452	30	66	19,170	39	64	602	11.5	100	18,900
12	64	500	28	42	29,400	40	64	693	0.2	151	16,200
13	64	473	24	239	8100	41	38.4	420	50	480	1120
14	38.4	619	39	363	1999	42	64	492	24	43	28,410
15	64	325	12	200	6300	43	64	593.1	8	327	8100
16	51.2	384	4.8	68	8830	44	64	511	22	229	9000
17	64	430	30	119	11,900	45	51.2	763	0.1	250	6900
18	38.4	502	24	292	1875	46	64	588.7	8	319	8200
19	64	724	0.1	119	20,150	47	30.72	309	40	310	618
20	25.6	529	18	189	925	48	30.72	294	42	123	1136
21	64	442	28	67	18,555	49	61.44	413	32	410	4100
22	64	554	10	332	7500	50	38.4	507	36	390	1545
23	51.2	513	18	42	16,666	51	61.44	441	30	385	4724
24	64	475	18	91	16,000	52	64	461	24	108	13,700
25	64	626	9	129	16,380	53	30.72	332	38	742	359
26	64	740	0.9	101	23,200	54	30.72	312	42	489	446
27	38.4	392	46	316	1390	55	64	479	20	125	12,900
28	64	569	15	391	6600						

Statistical summary for the 55 data records of the testing subset

	Choke size	Pwh	BS&W%	GLR	QL
Min	25.6	294	0.02	42	205
Max	64	881	50	885	29400
Mean	54.46	510.16	20.55	257.27	9273.09

References

- Abdul-Majeed GH (1988) Correlations developed to predict two phase flow through wellhead chokes. In: Proc. annual technical meeting, Calgary, PETSOC-88-39-26, 12–16 June. <https://doi.org/10.2118/88-39-26>
- Achong IB (1961) Revised bean and performance formula for Lake Maracaibo wells. Shell internal report
- Al-Ajmi MD, Alarifi SA, Mahsoon AH (2015) Improving multiphase choke performance prediction and well production test validation using artificial intelligence: a new milestone. SPE 173394. In: SPE digital energy conference and exhibition, The Woodlands. <https://doi.org/10.2118/173394MS>

- Al-Attar H (2008) Performance of wellhead chokes during sub-critical flow of gas condensates. *J Pet Sci Eng* 60(3):205–212. <https://doi.org/10.1016/j.petrol.2007.08.001>
- Ashford FE (1974) An evaluation of critical multiphase flow performance through wellhead chokes. *J Pet Technol* 26(8):843–850. <https://doi.org/10.2118/4541-PA>
- Baxendell PB (1958) Producing wells on casing flow: an analysis of flowing pressure gradients. *Trans AIME* 213:202–207
- Beiranvand MS, Mohammadmoradi P, Aminshahidy B, Fazelabdolabadi B, Aghahoseini S (2012) New multiphase choke correlations for a high flow rate Iranian oil field. *Mech Sci* 3(1):43–47. <https://doi.org/10.5194/ms-3-43-2012>
- Choubineh A, Ghorbani H, Wood DA, Moosavi SR, Khalafi E, Sadatshojaei E (2017a) Improved predictions of wellhead choke liquid critical-flow rates: modeling based on hybrid neural network training learning based optimization. *Fuel* 207:547–560
- Choubineh A, Khalafi E, Kharrat R, Bahreini A, Hosseini AH (2017b) Forecasting gas density using artificial intelligence. *Pet Sci Technol* 35(9):903–909
- Fortunati F (1972) Two-phase flow through well head chokes. Paper no. SPE 3742. SPE European Meeting, Amsterdam. <https://doi.org/10.2118/3742-MS>
- Gen M, Cheng R (2008) *Network models and optimization*. Springer, Berlin
- Ghorbani H, Moghadasi J, Wood DA (2017) Prediction of gas flow rates from gas condensate reservoirs through wellhead chokes using a firefly optimization algorithm. *J Nat Gas Sci Eng* 45:256–271. <https://doi.org/10.1016/j.jngse.2017.04.034>
- Gilbert WE (1954) Flowing and gas-lift well performance. *Drilling and production practices*. *Drill Prod Pract* 20:126–157 (API)
- Goldberg DE (1989) *Genetic algorithms in search, optimization & machine learning*. Addison Wesley, Boston
- Guo B (2007) *Petroleum production engineering, a computer-assisted approach*. Elsevier, Oxford
- Holland JH (1975) *Adaptation in natural and artificial systems*. University of Michigan Press, Ann Arbor (2nd edn: MIT Press, Cambridge, 1992)
- Jefferys ER (1993) Design applications of genetic algorithms, SPE 26367, Inc. <https://doi.org/10.2118/26367-MS>
- Mansouri V, Khosravian R, Wood DA, Aadnoy BS (2015) 3-D well path design using a multi objective genetic algorithm. *J Nat Gas Sci Eng* 27:219–235
- Mirzaei-Paiaman A, Nourani M (2012) Positive effect of earthquake waves on well productivity: case study: Iranian carbonate gas condensate reservoir. *Sci Iran* 19:1601–1160. <https://doi.org/10.1016/j.scient.2012.05.009>
- Mirzaei-Paiaman A, Salavati S (2013) A new empirical correlation for sonic simultaneous flow of oil and gas through wellhead chokes for Persian oil fields. *Energy Sources Part A Recov Util Environ Eff* 35(9):817–825
- Mitchell M (1996) *An introduction to genetic algorithms*. MIT Press, Cambridge
- Nasriani HR, Kalantariasl A (2011) Two-phase flow choke performance in high rate gas condensate wells. In: Paper presented at the SPE Asia Pacific oil and gas conference and exhibition. SPE 145576. <https://doi.org/10.2118/145576-MS>
- Poettmann FH, Beck RL (1963) New charts developed to predict gas-liquid flow through chokes. *World Oil* 184(3):95–100
- Ros NCJ (1960) An analysis of critical simultaneous gas/liquid flow through a restriction and its application to flow metering. *Appl Sci Res Sect A* 9(1):374–389. <https://doi.org/10.1007/BF00382215>
- Sivanandam SN, Deepa SN (2008) *Introduction to genetic algorithms*. Springer, Berlin, p 442
- Tangren RF, Dodge CH, Seifert HS (1949) Compressibility effects in two-phase flow. *J Appl Phys* 20(7):637–645. <https://doi.org/10.1063/1.1698449>
- Wood DA (2016) Asset portfolio multi-objective optimization tools provide insight to value, risk and strategy for gas and oil decision makers. *J Nat Gas Sci Eng* 33:196–216
- Zarenezhad B, Aminian A (2011) An artificial neural network model for design of wellhead chokes in gas condensate production fields. *Pet Sci Technol* 29(6):579–587. <https://doi.org/10.1080/10916460903551065>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.