Experimental and Computational Multiphase Flow
Vol. 2, No. 4, 2020, 225–246
https://doi.org/10.1007/s42757-019-0047-5

Performance comparison of bubble point pressure from oil PVT data: Several neurocomputing techniques compared

Hamzeh Ghorbani[1], David A. Wood[2] (✉), Abouzar Choubineh[3], Nima Mohamadian[4], Afshin Tatar[5], Hamed Farhangian[6], Ali Nikooey[3]

1. Young Researchers and Elite Club, Ahvaz Branch, Islamic Azad University, Ahvaz, Iran
2. DWA Energy Limited, Lincoln, United Kingdom
3. Petroleum Department, Petroleum University of Technology, Ahwaz, Iran
4. Young Researchers and Elite Club, Omidiyeh Branch, Islamic Azad University, Omidiyeh, Iran
5. Young Researchers and Elite Club, North Tehran Branch, Islamic Azad University, Tehran, Iran
6. Department of Chemical Engineering, Oil and Gas, Shiraz University, Shiraz, Iran

Abstract

Pressure–Volume–Temperature (PVT) characterization of a crude oil involves establishing its bubble point pressure, which is the pressure at which the first gas bubble forms on a fluid sample while reducing pressure at a stabilized temperature. Although accurate measurement can be made experimentally, such experiments are expensive and time-consuming. Consequently, applying reliable artificial intelligence (AI)/machine learning methods to provide an accurate mathematical prediction of an oil's bubble point pressure from more easily measured characteristics can provide valuable cost and time savings.

This paper develops and compares four neurocomputing models applying algorithms consisting of a Multilayer Perceptron (MLP), a Radial Basis Function trained with a Genetic Algorithm (RBF-GA), a Combined Hybrid Particle Swarm Optimization-Adaptive Neuro-Fuzzy Inference System (CHPSO-ANFIS), and Least Squared Support Vector Machine (LSSVM) tuned with a coupled simulated annealing (CSA) optimizer. Based on a comprehensive analysis, although the four proposed models yield acceptable outputs, the CHPSO-ANFIS model has the best performance with the average absolute relative deviation of 0.846, the standard deviation of 0.0126, the root mean square error of 43.21, and the correlation coefficient of 0.9902. These algorithms are deployed for the accurate estimation of the bubble point pressure from the giant Ahvaz oil field (Iran).

Keywords

crude oil bubble point pressure (BPP)
prediction of BPP
neural network optimization
LSSVM ANFIS MLP RBF learning networks
neurocomputing/machine learning error analysis
tuning network models

Article History

Received: 13 July 2019
Revised: 1 September 2019
Accepted: 1 September 2019

Research Article

1 Introduction

Determination of reservoir fluid properties, e.g., point pressure (BPP), gas oil ratio (GOR), oil formation volume factor (B_o), etc., is one of the key factors for reservoir evaluation, reservoir performance, inflow performance, surface facility design, well test analysis, and material balance calculation (Kloubek, 1972; Elsharkawy et al., 1995; Velarde et al., 1997; Holcomb and Outcalt, 1999; Mishchuk et al., 2000; Valkó and McCain, 2003; Fainerman and Miller, 2004; Sun et al., 2005; Yazaydin and Martin, 2007; Bandyopadhyay and Sharma, 2011; Dixit et al., 2012; Ikiensikimama and Ajienka, 2012; Li and Yang, 2012; Adeleke et al., 2013; Simjoo et al., 2013). The maximum pressure at which the first bubble of gas evolves from the corresponding liquid phase is called BPP (Farasat et al., 2013). BPP is a critical property of reservoir fluids (Standing, 1947; Dindoruk and Christman, 2004; Farasat et al., 2013; Arabloo et al., 2014) which needs to be determined accurately by reservoir engineers.

It cannot be computed directly from compositional data using basic material balance methods, unlike properties such as American Petroleum Institute (API) gravity. BPP can be determined either by conducting laboratory analysis or by applying numerical prediction methods (Velarde et al., 1997; Bandyopadhyay and Sharma, 2011; Farasat et al., 2013). However, although standardized procedures of laboratory analysis provide the most accurate results, such analyses are highly dependent on the quality and validity of reservoir fluid

✉ dw@dwasolutions.com

Nomenclature

API^0	API gravity	FIS	Fuzzy inference system
BPP	Bubble point pressure	LSSVM	Least squared support vector machine
B_o	Oil formation volume factor	ANN	Artificial neural network
AARD%	Average absolute relative deviation	SVM	Support vector machine
err%	Average absolute error	ANFIS	Adaptive neuro-fuzzy inference system
MSE	Mean square error	FCM	Fuzzy C-means
P_b	Bubble point pressure	RBF	Radial basis function networks
PSO	Particle swarm optimization	MLP	Multilayer perceptron networks
R^2	Correlation coefficient	PN	Predictive networks
RMSE	Root mean square error	GA	Genetic algorithm
R_s	Solution gas oil ratio	CSA	Coupled simulated annealing
STD	Standard deviation	CLM	Coupled local minimizer
T	Temperature	SA	Simulated annealing
γ_g	Gas specific gravity	FBPNN	Forward back-propagation neural network
γ_o	Oil specific gravity		

samples, especially for under-saturated reservoirs (Velarde et al., 1997; Bandyopadhyay and Sharma, 2011; Farasat et al., 2013). When experimental measurements are not available (or reliable), empirical correlations or other predictive models are used for estimation of reservoir fluid properties. This is essential for certain types of crude oils, e.g., those with medium specific gravity, asphaltene base oil, paraffinic oils, and/or mixed oils (Proett et al., 2000; Dong et al., 2007; Nnochiri and Lawal, 2010; Bandyopadhyay and Sharma, 2011; Deisman et al., 2013). Thus, many researchers have attempted to find fast and accurate methods for the prediction of BPP and other reservoir fluid properties. The early empirical models, such as those by Standing (1947), Lasater (1958), and Glaso (1980) show the importance of accurate determination of BPP. During the last decade, various graphical and mathematically-derived equations have been published for estimation of BPP. Some of these equations have been developed for specific region or oil types, but many have been presented as potentially being suitable for generic application. Generally, the statistical accuracy of such correlations/equations is unreliable when applied to other datasets.

From the previous studies mentioned, there is a consensus that BPP is a function of solution gas oil ratio (R_s), oil formation volume factor (B_o), gas specific gravity (γ_g), API gravity (API^0) (or oil density γ_o), and temperature (T).

$$P_b = f(\gamma_g, R_s, B_o, \gamma_o \text{or} API, T) \tag{1}$$

Standing (1947) published a proposed correlation for determination of BPP of crude oil systems based on 105 experimentally-derived BPP measurements on 22 hydrocarbon systems in California of U.S. A log–log plot of R_s/γ_g defined the Standing equation with 11.2% error (Eq. (2)). This relationship became the basis of many other methods for estimating BPP.

$$P_b = a_1\left[\left(\frac{R_s}{\gamma_g}\right)^{a_2} \times 10^{a_3 \times T - a_4 \times API} - a_5\right] \tag{2}$$

Lasater (1958) provided 158 analyses for 137 black oil samples from Canada and U.S., to develop a model including non-hydrocarbon components of black oil (H2, N2, H2S) with an average algebraic error of 3.8% and maximum error of 14.7%. That relationship identified that the presence of non-hydrocarbon components increases negative error. Vazquez and Beggs (1980) used 6000 analyses of 600 crude oil samples from different parts of world to provide a BPP relationship. Glaso (1980) also included non-hydrocarbon components (N2, H2, H2S) in North Sea oil samples to develop a BPP prediction relationship.

Al-Marhoun (1988) developed a correlation for estimating BPP based on 160 analyses of 69 hydrocarbon mixtures from Middle East oil fields. McCain (1991) modified the Standing model, based on analysis of 100 samples from around the world, to develop a new BPP prediction model. Dokla and Osman (1992), based on 51 bottom hole samples from reservoirs in the United Arab Emirates reservoirs develop a correlation for BPP. Petrosky and Farshad (1993) used 81 test samples from Gulf of Mexico, Texas, and Louisiana to compare the available BPP relationships. Their comparison showed that Glaso equation to be more accurate than the Standing (1947) or Al-Marhoun (1988) relationships. Kartoatmodjo and Schmidt (1994) used 5392 worldwide samples to present an alternative BPP.

Farshad et al. (1996) presented new constants for the Glaso (1980) equation based on 98 analyses of Colombian

oils. Gharbi and Elsharkawy (1997) used an artificial neural network (ANN) algorithm to model the PVT properties of reservoir fluids. They used 5200 analyses of 500 test samples to develop their ANN model. The correlation coefficient and absolute error for this model are 0.96 and 15.38% respectively. Elsharakawy (1998) developed an ANN algorithm to compute the key fluid properties, i.e., BPP and gas oil ratio (GOR). Dindoruk and Christman (2004) used 100 analyses of Gulf of Mexico to optimize Petrosky and Farshad (1993) model. Al-Marhoun and Osman (2002) applied an ANN algorithm to predict BPP with 283 analyses from one Saudi Arabian field. They reported a correlation coefficient and absolute error of 0.9965 and 5.8915%, respectively, for that model. Boukadi et al. (2007) used 24 experimental data of Northern Omani oil fields to develop a new equation for BPP. Goda et al. (2003) used 160 analyses from Middle East oil fields to develop an ANN technique with 2 hidden layers and 10 neurons in each layer to provide a new BPP model with an average absolute error of 3.0704% and a correlation coefficient of 0.9981.

Malallah et al. (2006) based on 5000 analyses of samples from around the world, proposed graphical alternating conditional expectation (ACE) technique to derive a new BPP equation. El-Sebakhy et al. (2007) determined the relationship between BPP and oil formation volume factor (B_o) by using support vector regression technique (SVR) for 3 different databases. Hemmati and Kharrat (2007) presented a new equation for calculation of BPP based on 287 analyses of 30 oil samples from Iran. Moradi et al. (2010) proposed a new BPP equation based on 1801 analyses, including 1177 datasets from previously published papers and 634 datasets from various unspecified Iranian oil fields. A key feature of their equation is that it spans a wide range of oil gravities (i.e., 6–57 API^0). They reported an absolute deviation error of 16.96% for their equation.

Dutta and Gupta (2010) developed a new model based on 372 analyses involving a feed-forward, back-propagation neural network (FBPNN) technique involving two hidden layers, reporting an absolute error of 7.66%. Ikiensikimama and Ajienka (2012) used 250 analyses for Niger Delta oils and developed new BPP model based on the Standing equation (1947). Asoodeh and Kazemi (2013) proposed a better solution for predicting BPP based on 361 previously published analyses and applying the relationships proposed by Standing (1947), Velarde et al. (1997), and Al-Shammasi (2001). Arabloo et al. (2014) develop a BPP model by using 750 analyses that involved normalizing the input variables. Gomaa (2016) proposed a new BPP prediction equation based on 441 crude oil samples from around the Middle East, reporting the highest correlation coefficient (0.98), the lowest average relative error (–0.56%), the lowest average absolute error (8.12%), and the lowest standard deviation among the other correlations mentioned.

Artificial intelligence (AI)/machine learning is a rapidly developing applied branch of computer science which significantly arguments understanding through machine learning, thereby enhancing human abilities to directly measure, predict, and interpret complex, non-linear data sets (AlAjmi et al., 2015). AI is now routinely and successfully applied to many petroleum engineering systems (Gharbi and Elsharkawy, 1997; Elsharkawy, 1998; Gharbi et al., 1999; Al-Shammasi, 2001; Al-Marhoun and Osman, 2002; Choubineh et al., 2017; Ghorbani et al., 2017, 2018, 2019), providing predictions at various levels of accuracy to metrics that can only be measured precisely on a few samples, for reasons of costs or accessibility.

In this research, four new models: 1) Multilayer Perceptron (MLP), 2) Radial Basis Function trained with Genetic Algorithm (RBF-GA), 3) Hybrid Particle Swarm Optimization-Adaptive Neuro-Fuzzy Inference System (CHPSO-ANFIS), and 4) Least Squared Support Vector Machine (LSSVM) are developed and applied and their performance is compared for predicting BPP as a function of six input variables from a dataset available for the large Ahvaz oil field in Iran. Graphical analysis and statistical error measurement are presented for the BPP predictions derived for each of the four new models, with comparisons made to experimentally-measured values. The results and performance of these four new models are compared with the performance of previously published correlations using statistical error analysis to demonstrate that they are superior.

2 Details of the intelligent models

2.1 Optimization methods

Analytical optimization techniques, applying differential calculus and gradient-descent techniques and many evolutionary optimization algorithms represent well-established and widely used techniques that offer the ability to rapidly establish optimum correlations for complex data sets. It is the hybridization of such techniques with neurocomputing/machine learning algorithms, such as ANN and LSSVM, and fuzzy mathematical algorithms (e.g., ANFIS) that enable these methods to provide artificial intelligent learning systems capable of providing fast, reliable, and accurate predictions from complex systems.

2.2 Coupled Simulated Annealing (CSA)

Simulated annealing (SA), developed by Metropolis et al. (1953) and generalized by Kirkpatrick et al. (1983), approximates the global optimum of a given function and has been applied in a wide range of disciplines, such as syncretistic

and combinatorial optimization. To escape from local optima, SA involves heuristics that promote the broader search for better optima and thereby improving the chance of ultimately locating the global optimal condition. This process is like the physical process of annealing in which all crystal grains suddenly reach the lowest internal energy state when a molten metal is gradually cooled down (Liscic et al., 2010). Instead of using multi-start gradient optimizers, Couple Local Minimizers (CLMs) employ multiple gradient descent optimizers. Couple Simulated Annealing (CSA) (Xavier-de-Souza et al., 2010) that is inspired by an extension of CLMs, is used in this study to improve the accuracy and rapidity of convergence of the LSSVM model (see Section 3.2.4). It does so by seeking to minimize the Root Mean Squared Error (RMSE) between predicted and measured BPP values.

2.2.1 Particle Swarm Optimization (PSO)

Particle swarm optimization (PSO) is a well-established (Kennedy and Eberhart, 1995) and widely-used evolutionary algorithm. PSO is now widely applied to many petroleum engineering challenges (Onwunalu and Durlofsky, 2010; Ahmadi et al., 2013; Atashnezhad et al., 2014). PSO is inspired by the swarm behavior of birds and insects. Because of low memory and CPU requirements, this method achieves rapid computation in short processing periods.

To solve a problem by PSO method, the locations of particles are initially distributed in an arbitrary way throughout the possible solution space. A performance analysis via a fitness test is performed on each particle position in each iteration of the model. The best particle positions for each iteration, plus the global best position found by all iterations performed so far are used to modify the positions of the other particles for the following iteration. Once the fitness score of the best particle meets the stopping criteria or the designated number of iterations is completed, the values associated with the global best particle provide the optimum solution.

2.2.2 Genetic Algorithm (GA)

Genetic algorithms are inspired by the process of natural selection based on Darwin's theory of biological evolution (Darwin, 1859). They represent one of the most widely used evolutionary algorithms in computer science and operational research with many applications in the oil and gas industry, including the optimization of multiple objective (Yasari et al., 2013; Mansouri et al., 2015). There are five main steps involved in classic genetic algorithms: initialization, evaluation, selection, crossover, and mutation. The population is randomly generated and spread as widely as possible across the possible solution space in the initializing iteration. Each member of the population (an individual solution) is evaluated with a fitness test to assess how well it fits with the objective function (in this case BPP prediction). In the selection step the "unfit" individuals (those with low fitness scores) are discarded and only the best individuals of the population are kept for the next iteration. During the crossover step, some aspects of selected individuals are combined and create many new individuals for the evolving population. To introduce some diversity into the population (helping to avoid becoming trapped at local minima) a controllable degree of randomness is introduced into the populations' genetics during the mutation stage. Repeating the last four steps through multiple iterations gradually causes the population to converge on optimum solutions. The algorithm also enables many high-performing solutions to be preserved and compared.

2.3 Predictive Networks (PN)

2.3.1 Multilayer Perceptron Networks (MLP)

The multilayer perceptron (MLP) is one of the most commonly applied neural network (Hush and Horne, 1993; Haykin, 1994) and is used for the approximation of functions relating system output to input variables with complex and poorly correlated relationships. A MLP is a network that consists of one input layer consisting of several input nodes, one or more hidden layers, and one output layer. A MLP is trained using a systematic algorithm applied to a sufficiently representative sample of the dataset being modelled (i.e., the training set). Once trained the MLP is tuned to match with reasonable accuracy outputs with an input dataset by approximating rather than precisely defining functional relationships between the input and output variables.

Selecting the appropriate number of iterations during the training of a MLP can be challenge (Haykin, 1995); both undertraining (lack of sufficient time to finish the learning procedure) and overtraining/overfitting (remembering rather than learning) can result in the poor forecasting outcomes.

2.3.2 Radial Basis Function Networks (RBF)

Radial basis function networks (RBF) (Broomhead and Lowe, 1988) formulate activation functions as RBFs and the output is the sum of radial basis function values relating the input and network parameters.

Although MLP's involve different internal network calculations from Radial Basis Function (RBF), there are similarities between the methods. The main advantage of an RBF is the simplicity of its design. With only three layers, an RBF has a high tolerance of noise associated with the input variables, while retaining a significant learning capacity (Sayahi et al., 2016). The main differences between RBF and MLP networks are:

1) RBF networks involve a simpler design;

2) while the internal architecture of MLP networks can vary (e.g., one or several hidden layers), the structure of RBF networks is fixed;

3) RBF networks tend to focus on local approximations, whereas MLP networks seek global outputs determined by neuron connections; and,

4) clustering in RBF networks is attributed to hyperspheres (points defined in one or more dimensions less than the ambient space), while in MLP networks it is attributed to hypersurfaces (an n–1-dimensional surface embedded in an n-dimensional ambient space).

2.3.3 *Adaptive Neuro-Fuzzy Inference Systems (ANFIS)*

An adaptive neuro fuzzy inference system (ANFIS) is a neural network combined with fuzzy logic that is typically based on a first-order Sugeno fuzzy model (Jang, 1993; Jang et al., 1997). ANFIS essentially hybridizes Artificial Neural Network (ANN) with a fuzzy logic methodology. In recent years the technical has been successfully applied to predicting many complex oil and gas systems (Zoveidavianpoor et al., 2013; Basarir et al., 2014; Yavari et al., 2018). Fuzzy logic deals with the vagueness of human assessments and changes it from qualitative knowledge into more rigorous quantitative analysis. Fuzzy logic is highly adaptable to uncertainty and vagueness, which are common features of many naturally occurring systems and environments. By combining ANN with fuzzy logic, the machine learning input tends to reduce the rate of errors associated with the defined fuzzy membership functions.

A Fuzzy Inference System (FIS) has three main components which are basic IF-THEN rules, reasoning fuzzy inference techniques, and output. In an FIS, by using the fuzzification process, the input values are converted into fuzzy values in a range between 0 and 1. These fuzzy numbers ultimately need to be processed through a defuzzification process to provide meaningful output. The database on which an FIS is based, is a crucial component in determining its decision-making capabilities. FIS development involves defining a universe of relationships and the determination of the number of linguistic terms describing the membership functions to those relationships.

2.3.4 *Least Squares Support Vector Machines (LSSVM)*

Support vector machines (SVM) can be used for classifications and non-linear function estimation and, in some cases, duplicate and/or complement the capabilities of neural networks. Unlike multilayer perceptrons (MLP) and radial basis functions (RBF), the networks developed by SVM involve no constraints regarding local minima and the choice of the number of hidden units (Vapnik, 2013). The advantages of SVM compared to the conventional ANN are:

1) SVM networks involve fewer adjustable parameters;

2) SVM networks do not involve hidden nodes;

3) SVM networks can be generalized with precision and accuracy;

4) overfitting is less likely to occur with SVM models; and,

5) Standard applications of quadratic programming algorithms (i.e., minimizing or maximizing quadratic functions involving several variables subject to linear constraints) typically provide more rapid solutions for SVM networks.

3 Results

3.1 Characterization of Ahvaz field dataset

The accuracy and reliability of the measured experimental data making up the dataset to be optimized clearly influence the achievable accuracy of any model regardless of the algorithms involved. The dataset used in this study include 79 data points from oil samples collected from the Ahvaz oil field, located in south of Iran (Fig. 1). The Ahvaz oil field extends for 67 km and is 6 km wide. It is located within a prolific oil province: north of Ramin oil field, east of Maroun

Fig. 1 Location of Ahvaz oil field in Iran in respect of surrounding oil fields and infrastructure.

oil field, south of Shadegan and Mansouri oil fields, and west of Ab Teymour and Susangerd oil fields. The oil-bearing reservoirs are in the asymmetrical Ahvaz anticline within the Zagros province at undulating elevations below the Aghajari Formation. The base of the oil-bearing part of the anticline lies about 2500 meters below the sea level in the Asmari Formation. The sandstone and limestone reservoirs display an average porosity of 18 percent and oil produced from the Asmari Formation averages about 30 API^0.

Bubble point pressure (BPP defined by the symbol P_b in the subsequent analysis) is a function of temperature, oil formation volume factor and gas specific gravity, solution gas ratio, oil specific gravity and/or API, as defined in Eq. (1). The relationship between these parameters in the Ahvaz field dataset is illustrated graphically in Fig. 2. The bubble point pressure (P_b) shows a moderate positive correlation with temperature (T) and oil formation volume factor (B_o) and a strong positive correlation with the solution gas–oil ratio (R_s). On the other hand, it demonstrates poor correlations with gas specific gravity (γ_g) and oil specific gravity (γ_o) and oil API gravity (the latter two possessing a strong negative correlation between them as oil specific gravity is, of course, involved in the API gravity calculation). Gas specific gravity (γ_g) and oil specific gravity (γ_o) show poor inverse correlations with P_b, while oil API gravity shows a poor positive correlation with P_b. These relationships are summarized by Eq. (3).

$$P_b \propto (T, B_o, R_s, \text{API}) \quad \text{and} \quad P_b \propto \frac{1}{(\gamma_g, \gamma_o)} \tag{3}$$

From this dataset, 80% (63 data) of the samples were selected randomly to be used for training of the fuzzy-PSO algorithm, while the remaining 20% (16 data) of samples became the testing subset to test the validity of the trained BPP prediction models. Statistical analysis of the input variable values for the training and testing data subsets are provided in Tables 1 and 2. For the testing data subset (Table 2) the solution gas–oil ratio with a standard deviation of 131.669 and the oil specific gravity with an average deviation of 0.01013 have the maximum and minimum standard deviations for the input variables, respectively. The bubble point pressure (BPP) mean and standard deviation for testing data subset are 3541.03 and 400.662, respectively.

3.2 Data preparation and subset selection

The values of each metric within each data record are initially normalized using Eq. (4) to produce values between –1 and 1. This removes data biases and speeds up the optimization process associated with each algorithm.

$$X_N = 2 \times \frac{(X - \text{Min}(X))}{(\text{Max}(X) - \text{Min}(X))} - 1 \tag{4}$$

After the normalization, the data is randomly split into two

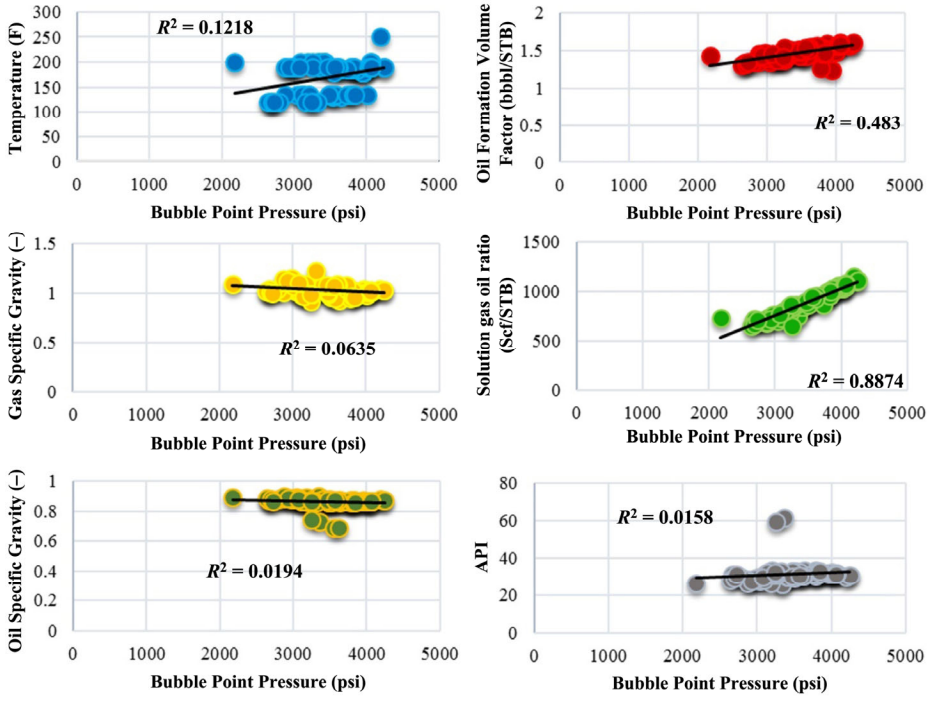

Fig. 2 Relationships of bubble point pressure individually with its input calculation variables: temperature (T), oil formation volume factor (B_o), gas specific gravity, solution gas ratio, oil specific gravity, and oil degrees API gravity for each of 79 data records of the Ahvaz oil field dataset.

Table 1 Statistical summary for the 63 data records of the training subset. Statistical analysis performed using SPSS software

	γ_g	γ_o	API	R_s	T	B_o	P_b
Min	0.92	0.69	25.93	648.25	120.00	1.27	2176.00
Max	1.22	0.90	61.54	1111.30	200.00	1.61	4246.00
Mean	1.03	0.87	31.01	864.30	190.00	1.48	3426.00
Average	1.04	0.86	31.74	871.31	169.82	1.47	3411.10
Variance	0.002984	0.001840	34.75	14690.89	937.00	0.00642	202379.55
Standard deviation	0.0541	0.04250	5.84	120.09	30.33	0.0794	445.7576

Table 2 Statistical summary for the 16 data records of the testing subset. Statistical analysis performed using SPSS software

	γ_g	γ_o	API	R_s	T	B_o	P_b
Min	0.93	0.86	26.00	681	120	1.23	2794
Max	1.15	0.90	33.30	1147	250	1.61	4190
Mean	1.02	0.87	31.42	919	185	1.46	3606
Average	1.02	0.87	30.66	902	170	1.47	3541
Variance	0.00239	0.00010	3.62	18091	1181	0.01	167510
Standard deviation	0.04785	0.01013	1.86	131.66	33.64	0.0934	400.66

data subsets: the ***training subset*** and the ***testing subset***. This process is repeated several times to ensure that data points in each subset have a homogeneous distribution and local aggregation of data points is avoided. 80% of data records (63 data records in total) are used to construct the model as the training subset, while the remaining 20% (16 data records in total) is used to as the testing subset to evaluate and compare the performance of each model.

3.2.1 *Multilayer Perceptron (MLP) model description*

Multilayer Perceptrons (MLP) are artificial neural networks (ANN) connecting multiple layers in a directional network. Cybenko (1989) demonstrated mathematically that an MLP using backpropagation as its supervised learning technique can predict the combined outcomes of a series of non-linear relationships between variables accurately with just one hidden layer. MLP networks with just one hidden layer are therefore utilized here. By varying the MLP network architecture with 4 to 25 neurons in their hidden layer, the optimum performing MLP architecture for application to the Ahvaz dataset was established. Figure 3 compares the RMSE of the various MLP networks evaluated with their hidden layer having different numbers of neurons. As can be seen in Fig. 3, the best performance for the training subset was achieved with an MLP networks consisting of less than 18 neurons.

3.2.2 *Radial Basis Function (RBF) model description*

RBF networks have two main tuning parameters, which are ***spread*** and ***maximum number of neurons*** (MNN). The optimum determination of these parameters improves the

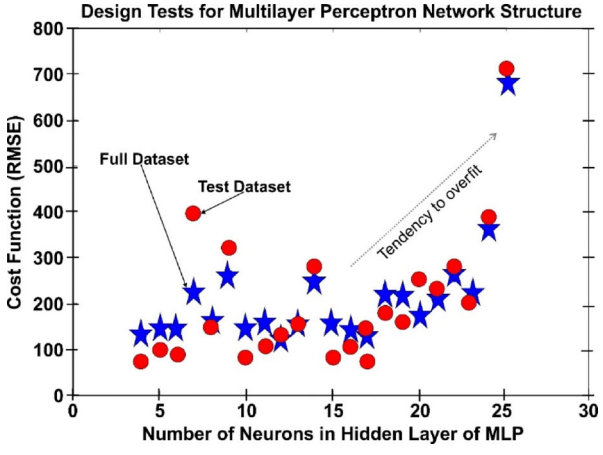

Fig. 3 Performance of different MLP networks with numbers of neurons in the hidden layer varying from 4 to 25. The MLP networks with about less than about 18 neurons in the hidden layer present the best models for the combined performance of the training and testing subsets.

accuracy of a model's performance. The nonlinearity of RBF model makes a trial-and-error approach to finding the optimum model architecture time-consuming. Consequently, a genetic algorithm (GA) is used here to determine the optimum value of RBF spread and MNN tuning parameters. Three control parameters also need to be tuned to optimize performance of the GA in order to solve constrained and unconstrained optimization problems (Vapnik, 2013). GA are widely adopted and hybridized as optimizers and have been applied to optimize many systems in the oil and gas industry (Yasari et al., 2013; Mansouri et al., 2015).

For the model developed, the GA algorithm was initialized with a population size of 50. Calculating the RMSE (for

measured versus predicted BPP) for each individual in the population, enabled the individuals with the highest fitness score to be retained and modified to produce the next generation of the population. This process in conjunction with cross-over and mutation actions was perpetuated over several generations to determine the optimum values for the spread and MNN parameters of the RBF network. Thirty GA generations were evolved and evaluated to determine the optimum value of spread and MNN values of 0.57 and 43, respectively. The RMSE values for the GA generation sequence perpetuated are displayed in Fig. 4, which shows that convergence of the GA during the optimization process occurred after about 12 generations (iterations of the algorithm).

3.2.3 Combined Hybrid Particle Swarm Optimization-Adaptive Neuro-Fuzzy Inference System (CHPSO-ANFIS) model description

The first step is to develop the logic that determines the underlying FIS. There are three common options used to structure FIS: 1) grid partitioning of the data, which is time consuming to build and slow to compute; 2) Sugeno-type FIS structure using the subtractive clustering method; and 3) fuzzy C-means (FCM) clustering. These equate to functions genfis1, genfis2, and genfis3 in the MATLAB toolbox, respectively.

The Sugeno-type FIS structure (genfis2) algorithm was used here to create the FIS. Initially rules and antecedent membership functions are established for the FIS, followed by least-squares estimation to determine each rule's consequent equations. The influence radius is an important parameter in this method, with a range between 0 and 1. Although a value near to zero for the influence radius often results in a better performance, it makes the training process more complex and inefficient. Consequently, a value between 0.7 and 1 was evaluated for the influence radius. In order to be more systematic, a genetic algorithm (GA) was applied instead of a trial and error approach to find optimal values for the influence radius for the evaluated range. The GA determined that an influence radius of 1.0 was optimal for this dataset. This value for influence radius generates FIS with an acceptable accuracy and simplifies the training process. Figure 5 shows the convergence of the GA for a radius influence of 1. The vertical axis indicates the GA fitness (cost) function and the horizontal axis shows the number of generations with the GA method.

Applying an influence radius of 1, four FIS rules were established. The membership functions for those four rules are presented in Fig. 6. As the data is normalized, all membership functions are constrained between variable values of between –1 and 1.

The rules for the six variables shown in Fig. 6 can be expressed in a compressed numerical format based on the membership function indices, for example, as used by the MATLAB fuzzy logic tool box that form might be displayed as follows:

1 1 1 1 1 1, 1 (1): 1
1 2 1 4 3 2, 2 (1): 1
4 3 1 2 2 3, 3 (1): 1
2 1 3 4 4 4, 4 (1): 1

For coding and quick interpretation/model verification purposes, it is often useful to express rules in such compressed forms. Each line describes a single rule. The first six digits of each line refer to the six independent variables in sequence. The first digit after the comma refers to a condition of the output variable. The digit in brackets refers to an optional weight applied to each specific rule (values of 1 in

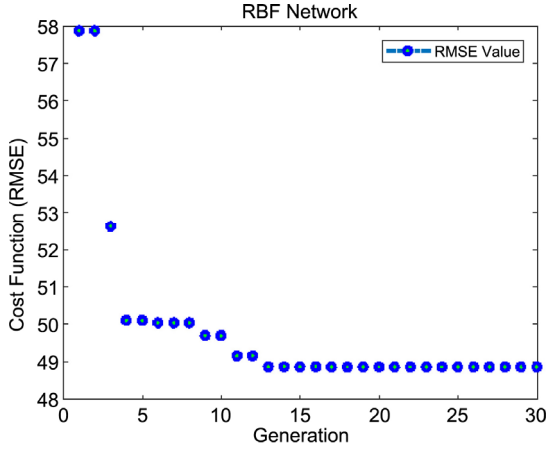

Fig. 4 Convergence of genetic algorithm (GA) to optimum maximum number of neurons (MNN) and spread values used to tune the RBF algorithm.

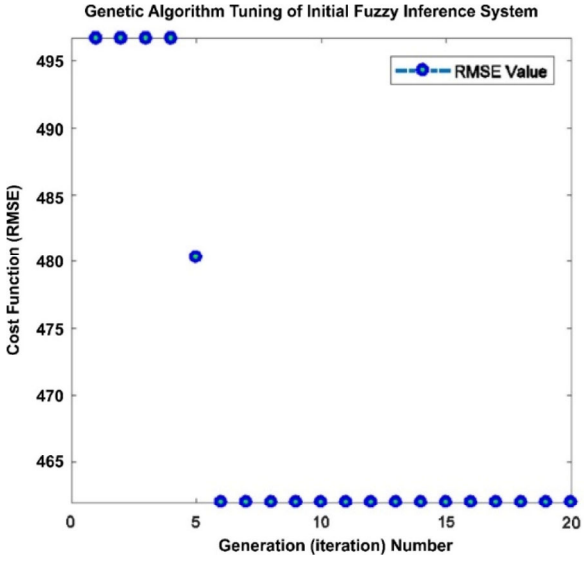

Fig. 5 Convergence of the genetic algorithm (GA) to the optimum value of the cost function for the developing the initial FIS for the ANFIS model.

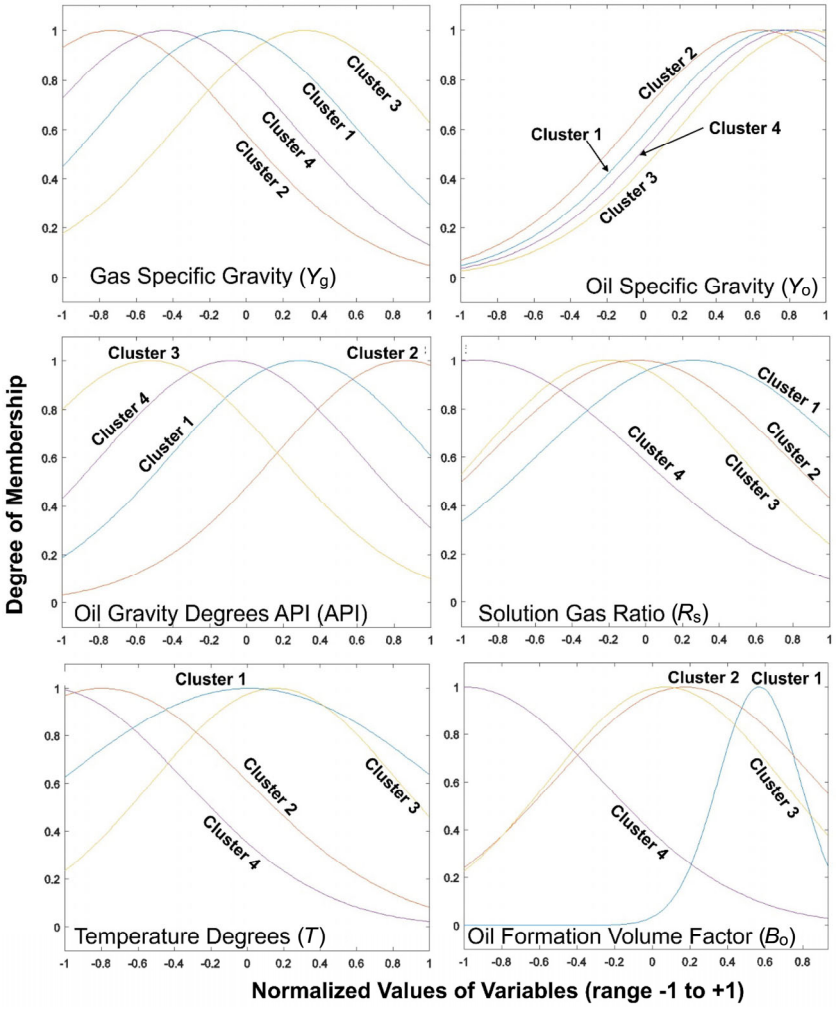

Fig. 6 Membership functions of the initial FIS for data variables used to determine BPP for the Ahvaz field training subset aided by a genetic algorithm.

the brackets of each rule mean that equal weighting is applied to all rules). The number at the right-hand end of the compressed rule format, after the colon, is a code to describe the type of rule being applied; in MATLAB, for example, a "1" refers to an AND rule, and a "2" refers to an OR rule. The values associated with the first six digits in the code refer to the dominant cluster (Figs. 6 and 8) for the specific input variables, which the model described can be one of clusters 1 to 4.

In order to establish the optimum membership functions, the initial FIS is trained by minimizing the RMSE of the objective function. Common ways to achieve this are by back propagation and hybrid methods. An alternative is to apply an optimization method to tune the FIS training process. A particle swarm optimization (PSO) algorithm is applied here to find the optimum tuning parameters for the FIS by minimizing the RMSE of the objective function.

Combining and hybridizing the PSO with ANFIS (CHPSO-ANFIS) involves running the PSO algorithm in two distinct stages. The hybrid stage employs PSO to tune and train the initial FIS (constructed with the aid of the GA) with five runs of the PSO algorithm. In the final stage, the tuned FIS is combined with a subsequent single hybrid PSO run to complete the algorithms combined training. For the hybrid stage of training the PSO application involves just 10 iterations, whereas the final training stage involves the PSO being applied for 250 iterations. Table 3 lists the key control and tuning parameters of the ANFIS training functions combined with the CHPSO method.

The performance of the hybrid training stage of the initial FIS with the CHPSO method is presented in Fig. 7, where the vertical axis shows the RMSE cost function in terms of achieving the objective function (lower RMSE represents better performance) and the horizontal axis shows the number stages over which the hybrid CHPSO-ANFIS is run to further tune the initial FIS. Monitoring the RMSE value for both the training subset and the testing subset over runs involving various numbers of stages to train the network helps to avoid

overfitting. The best performance was achieved by running the CHPSO-ANFIS algorithm with greater than 20 tuning stages (Fig. 7) with no evidence of overfitting.

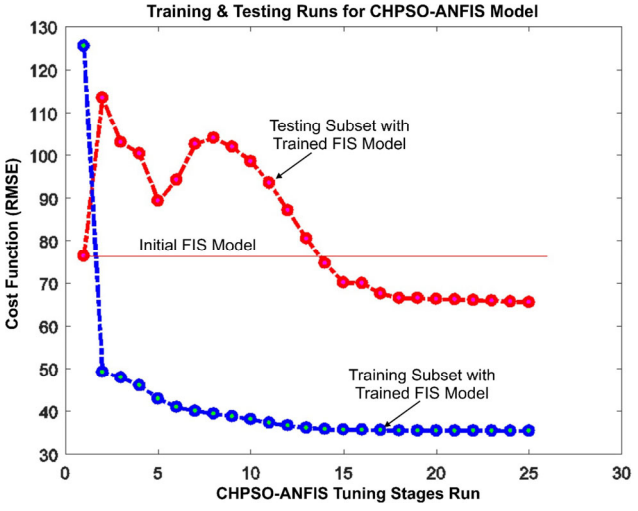

Fig. 7 Performance following training the initial FIS with the PSO during the hybrid stage of CHPSO by varying the number of runs of the PSO (10 iterations for each PSO run at the hybrid stage).

The final membership functions obtained from CHPSO-ANFIS method for each of input data variables are shown in Fig. 8. Clearly, by comparing Figs. 6 and 8, it can be

Table 3 Control parameters for ANFIS training functions applying the CHPSO method

Parameter	Value
(Initial FIS) minimum value of influence radius	0.7
(Initial FIS) maximum value of influence radius	1
(Initial FIS) maximum number of GA iterations	20
(Initial FIS) number of individuals in the initial GA population	30
(Hybrid) number of iterations of PSO in training	10
(Hybrid) training error goal	0
(Hybrid) initial PSO step size	0.01
(Hybrid) PSO step size decline rate	0.9
(Hybrid) PSO step size increase rate	1.1
(Hybrid) number runs selected for the hybrid training	5
(PSO) number of PSO iterations	250
(PSO) number particles in initial PSO population	100
(PSO) number of runs for final PSO optimization training	1
Number of tuning stages evaluated	25

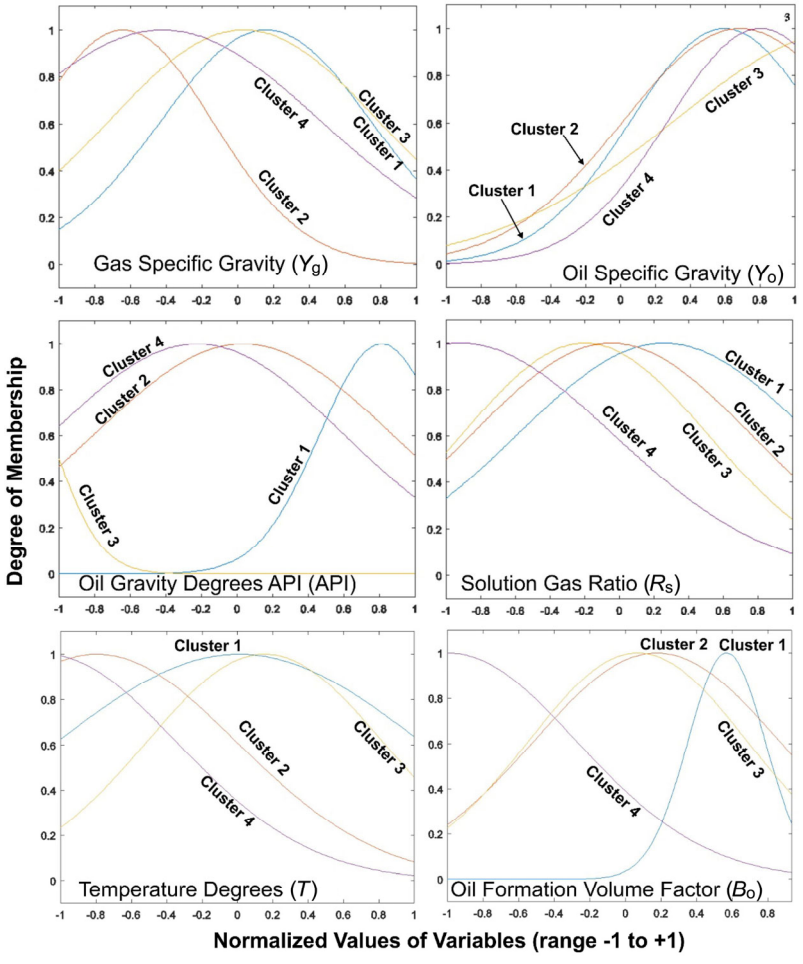

Fig. 8 Membership functions of the CHPSO-trained FIS for data variables used to determine BPP for the Ahvaz field training subset.

observed that following FIS training at the hybrid PSO stage, there is a significant change in the form of some of the membership functions of the FIS. Table 4 reveals that the RMSE value of the trained CHPSO-ANFIS model applied to the whole dataset and each of the training and testing subsets is less than the RMSE value for the initial FIS applied to each of those sets. Table 4 presents the statistical parameters for both the initial FIS and the trained FIS applying the CHPSO method.

3.2.4 *Least Squared Support Vector Machine (LSSVM) model description*

LSSVM models involve two key control parameters, γ and σ^2, that need to be established for specific datasets as they influence the accuracy and globalization ability of the models. γ is a tuning factor related to the error term influencing the LSSVM network structure. σ^2 is the width term associated with the RBF kernel function also used to tune the LSSVM network. Here, a coupled simulated annealing (CSA) algorithm (Xavier-de-Souza et al., 2010) firstly optimizes/tunes the values of the two control parameters (γ and σ^2). The algorithm combination CSA plus LSSVM has been described mathematically and successfully applied to predicting the densities of ionic liquids. The tuning of the CSA-LSSVM algorithm employed here resulted in the determination of optimum values for γ and σ^2 of 32.53 and 29.76, respectively.

Figure 9 provides a flow diagram for the four artificial intelligence/machine learning algorithms developed and

Table 4 Comparison of a range of statistical performance measurements for the initial and CHPSO-trained FIS

Predictor	Dataset	R^2	AARD	STD	RMSE	N
Initial FIS	Train data	0.9173	2.1439	0.0517	125.4860	63
	Test data	0.9768	1.7914	0.0209	76.3412	16
	All data	0.9283	2.0725	0.0472	117.2090	79
CHPSO trained FIS	Train data	0.9934	0.7453	0.0107	35.3778	63
	Test data	0.9775	1.2415	0.0186	65.4885	16
	All data	0.9902	0.8458	0.0126	43.2055	79

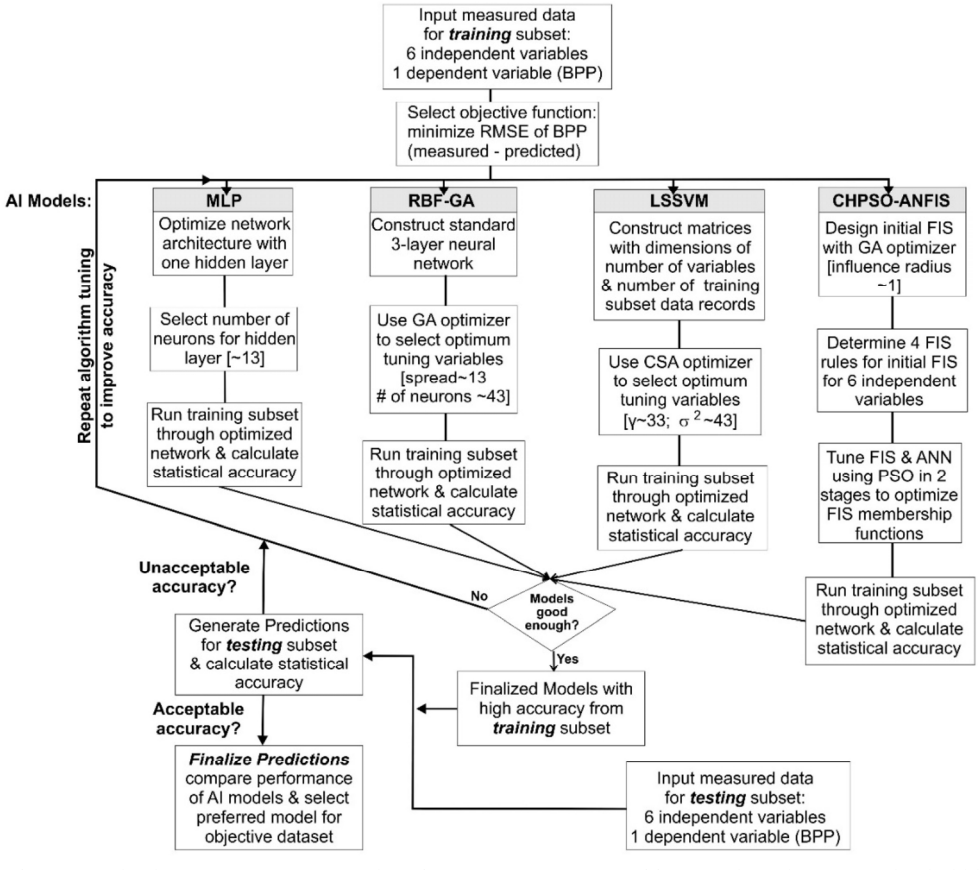

Fig. 9 Sequence of steps involved in training, testing, and performance comparison of four AI algorithms to predict bubble point pressure.

described here, and the sequence of steps involved in their training, evaluation, testing, and comparison of their statistical accuracy when applied to predicting bubble point pressure for the Ahvaz field PVT dataset.

4 Statistical-graphical analysis

4.1 4 Developed AI models applied to the Ahvaz field dataset

A dataset of 79 PVT data from wells drilled in the Ahvaz oil field, located onshore southwest of Iran, is used here to evaluate four developed models for predicting BPP from PVT data. Tables 5 and 6 show detailed measured variable values for each of the 79 individual data records divided into training and testing subsets.

Figure 10 shows cross plots with correlation coefficients displayed of measured versus predicted BPP values for the entire dataset (79 data records) for each of the four AI models developed. Acceptable correlation coefficients with values greater than 0.90 were obtained for all models and the trends for measured versus predicted data approximately follow a 45° line in each case. These results indicate acceptable accuracy is achieved by each of the AI prediction models developed and evaluated.

The ANFIS and RBF models achieved the highest correlation coefficients value of above 0.98. Figure 11 indicates the relative deviations of predicted versus measured BPP values for the four AI models for each record of the training and testing subsets distinguished. The root mean squared error achieved by each of the models for the total dataset was 107.45, 48.84, 43.20, and 115.98 for MLP, RBF, ANFIS, and LSSVM, respectively. Figure 11 is revealing as it identifies those few records that have been poorly predicted by the respective models. In the case of the MLP and LSSVM models it is the data record with the lowest BPP (data record 38 of the training subset) that achieved the worst predictions. A case could be made to exclude data record 38 as an outlier.

Table 5 Ahvaz oil field bubble point pressure data—FPSO training subset (63 records, 80% of full dataset)

Ahvaz oil field PVT training subset								
Data record details		Independent variables						Dependent variable
Data record number	Status	Gas specific gravity (γ_g)	Oil specific gravity (γ_o)	Oil degrees API gravity (API)	Solution gas to oil ratio (R_s, scf/stb)	Temperature (T, °F)	Oil formation volume factor (B_o, bbl/stb)	Bubble point pressure (P_b)
1	Training	1.0400	0.8776	29.7352	1146.70	250	1.5680	4190.0
2	Training	1.0007	0.8670	31.7065	1042.00	180	1.2340	3940.0
3	Training	0.9283	0.8586	33.3032	935.70	130	1.4492	3640.0
4	Training	0.9781	0.8658	31.9327	1002.10	185	1.5678	3870.0
5	Training	1.0287	0.8669	31.7253	1076.00	200	1.5933	4067.0
6	Training	1.0350	0.8705	31.0503	1057.76	180	1.5680	3960.0
7	Training	1.0320	0.8674	31.6312	1023.71	180	1.5680	3950.0
9	Training	1.0260	0.8679	31.5372	979.00	185	1.5398	3794.0
10	Training	0.9740	0.8636	32.3490	1030.20	135	1.4854	4003.0
11	Training	1.0243	0.8707	31.0129	1096.10	190	1.5917	4237.0
13	Training	0.9965	0.8665	31.8006	1024.40	185	1.5948	3874.0
14	Training	1.0535	0.8718	30.8079	944.20	185	1.5312	3630.0
15	Training	0.9436	0.8583	33.3608	992.50	135	1.2712	3786.0
16	Training	0.9569	0.8605	32.9393	912.00	130	1.4298	3528.0
17	Training	1.0141	0.8673	31.6500	977.00	185	1.5285	3788.0
18	Training	1.0535	0.8708	30.9943	937.60	185	1.5495	3632.5
19	Training	0.9802	0.8660	31.8949	901.30	135	1.4300	3539.0
20	Training	1.0375	0.8733	30.5291	968.70	190	1.5250	3783.0
21	Training	1.0076	0.8694	31.2559	958.20	190	1.5264	3826.0
23	Training	1.0770	0.8988	25.9321	753.30	200	1.4415	3340.0
24	Training	1.0015	0.8577	33.4761	839.70	135	1.4137	3202.0
26	Training	1.0534	0.8736	30.4734	864.00	200	1.4897	3426.0

(Continued)

Ahvaz oil field PVT training subset								
Data record details		Independent variables						Dependent variable
Data record number	Status	Gas specific gravity (γ_g)	Oil specific gravity (γ_o)	Oil degrees API gravity (API)	Solution gas to oil ratio (R_S, scf/stb)	Temperature (T, °F)	Oil formation volume factor (B_o, bbl/stb)	Bubble point pressure (P_b)
27	Training	1.0430	0.8737	30.4549	848.70	200	1.4879	3418.0
29	Training	1.0576	0.8728	30.6219	844.20	190	1.4788	3358.0
31	Training	0.9901	0.8608	32.8820	897.39	135	1.4379	3476.5
32	Training	1.0828	0.8691	31.3121	866.30	185	1.5640	3728.4
33	Training	0.9946	0.8709	30.9756	779.90	135	1.3844	3115.0
34	Training	1.0721	0.8782	29.6250	849.30	200	1.4924	3401.0
35	Training	1.0390	0.8741	30.3808	968.90	190	1.5377	3788.0
37	Training	1.0886	0.8954	26.5299	736.60	200	1.4299	3183.0
38	Training	1.0923	0.8947	26.6536	735.00	200	1.4360	2176.0
39	Training	1.0177	0.8908	27.3460	707.10	135	1.3485	3042.0
40	Training	0.9630	0.8610	32.8438	960.40	135	1.4583	3814.0
41	Training	1.0913	0.8984	26.0022	757.60	200	1.4409	3339.0
43	Training	1.0032	0.8682	31.4809	1048.60	190	1.5680	4059.0
44	Training	1.0610	0.8639	32.2921	900.50	190	1.5062	3465.0
46	Training	0.9815	0.8649	32.1027	795.60	135	1.3902	3125.0
47	Training	1.0404	0.8688	31.3683	966.47	135	1.5386	3726.1
48	Training	1.0640	0.8796	29.3686	862.56	190	1.4880	3499.0
49	Training	1.0029	0.8610	32.8438	900.45	135	1.4360	3476.5
50	Training	1.0760	0.8784	29.5883	848.90	200	1.4896	3413.0
71	Training	1.0066	0.8866	28.0985	681.20	135	1.3444	2875.0
73	Training	0.9164	0.7414	29.3551	654.34	200	1.3710	3248.0
106	Training	1.2230	0.8615	32.7500	864.30	120	1.5044	3320.0
107	Training	0.9852	0.8713	30.9000	835.95	190	1.5369	3254.0
135	Training	0.9566	0.8616	32.7293	994.30	135	1.4708	3844.0
136	Training	1.0689	0.6929	30.2800	935.44	180	1.5181	3550.0
137	Training	1.0789	0.6916	30.0500	956.79	190	1.5311	3617.0
159	Training	1.0590	0.8860	28.2065	733.38	200	1.4341	3088.0
161	Training	1.0681	0.8772	29.8087	765.00	190	1.4365	3009.0
171	Training	1.0396	0.8777	29.7168	722.68	120	1.3369	2683.0
172	Training	1.1101	0.8853	28.3328	735.00	190	1.4523	2880.0
186	Training	1.0889	0.8755	30.1219	929.00	190	1.5126	3605.0
187	Training	1.0364	0.8771	29.8271	704.42	120	1.3496	2794.0
188	Training	1.1088	0.8817	28.9854	760.06	190	1.4642	2969.0
201	Training	1.0215	0.8864	28.1345	648.25	120	1.3117	2638.0
202	Training	1.1414	0.8971	26.2305	694.59	190	1.4138	2869.0
238	Training	1.0029	0.8795	29.3869	672.00	120	1.3170	2689.0
239	Training	1.1186	0.8887	27.7213	724.00	190	1.4247	2928.0
350	Training	1.0798	0.8741	30.3808	895.00	190	1.4929	3469.0
377	Training	0.9990	0.8692	31.2934	710.00	120	1.3545	2727.0
400	Training	0.9836	0.8627	32.5199	870.00	120	1.4143	3243.0
401	Training	1.0676	0.8711	30.9383	951.00	190	1.5398	3566.0

Table 6 Ahvaz oil field bubble point pressure data—FPSO testing subset (16 records, 20% of full dataset)

Ahvaz oil field PVT testing subset								
Data record details		Independent variables						Dependent variable
Data record number	Status	Gas specific gravity (γ_g)	Oil specific gravity (γ_o)	Oil degrees API gravity (API)	Solution gas to oil ratio (R_s, scf/stb)	Temperature (T, °F)	Oil formation volume factor (B_o, bbl/stb)	Bubble point pressure (P_b)
8	Testing	1.0130	0.8675	31.6124	1070.00	200	1.6053	4056.0
12	Testing	1.0320	0.8715	30.8637	1111.30	190	1.6069	4246.0
22	Testing	1.0140	0.8697	31.1998	963.10	190	1.5123	3824.0
25	Testing	1.0502	0.8644	32.1974	898.90	190	1.5052	3466.0
28	Testing	0.9944	0.8663	31.8383	795.00	135	1.3865	3112.0
30	Testing	1.0510	0.7330	31.5423	849.00	190	1.4809	3363.0
36	Testing	0.9480	0.8635	32.3680	901.40	135	1.4336	3581.0
42	Testing	1.0024	0.8682	31.4809	1031.90	190	1.5596	4018.0
45	Testing	1.0210	0.8687	31.3871	1066.20	190	1.5814	4064.0
105	Testing	1.0160	0.8683	31.4621	1042.80	190	1.5776	4036.0
160	Testing	1.0080	0.8698	31.1811	699.00	120	1.3281	2673.0
173	Testing	0.9710	0.8732	30.5476	782.64	135	1.3764	3217.0
185	Testing	0.9886	0.8658	31.9327	842.00	120	1.3881	3253.0
349	Testing	0.9677	0.8638	32.3111	802.00	120	1.3764	3150.0
378	Testing	1.1037	0.8770	29.8455	775.00	190	1.4500	3075.0
402	Testing	1.1495	0.8848	28.4231	764.85	190	1.3973	2988.0

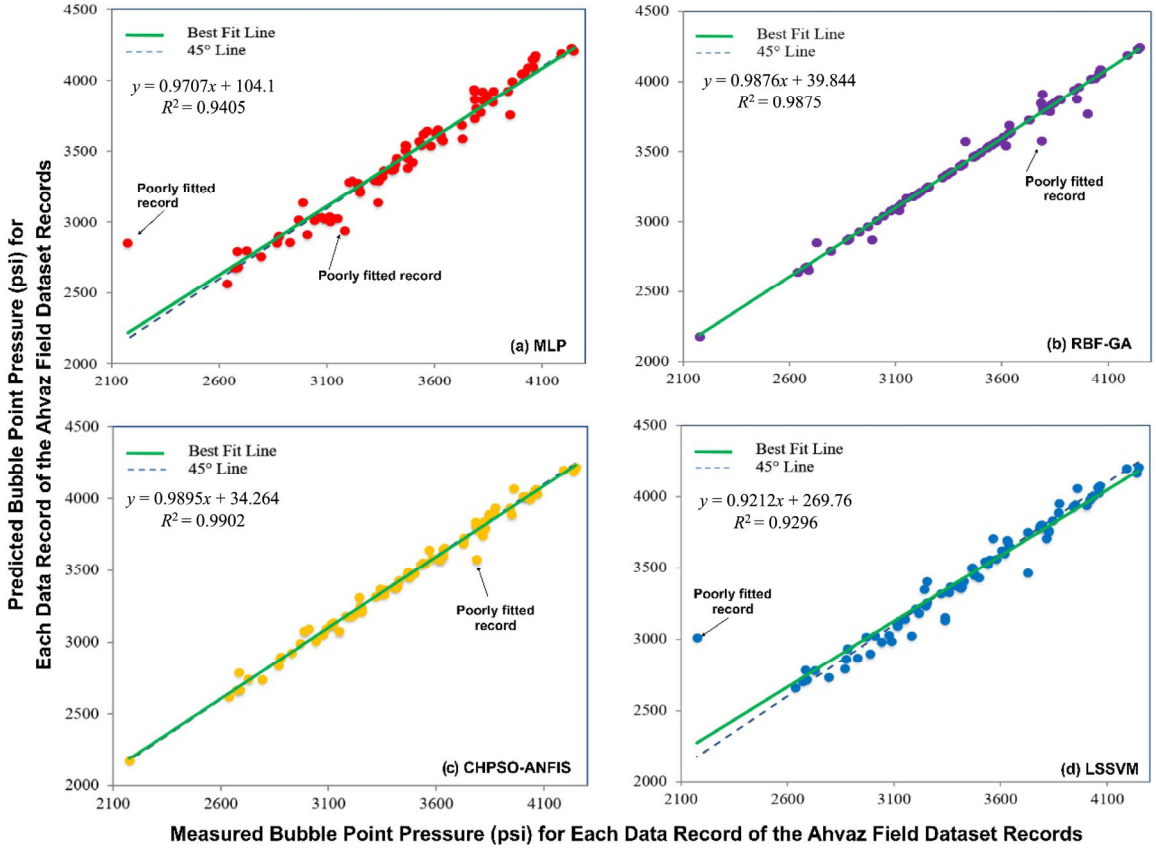

Fig. 10 Predicted BPP values for (a) MLP, (b) RBF, (c) ANFIS, and (d) LSSVM models versus experimentally measured values.

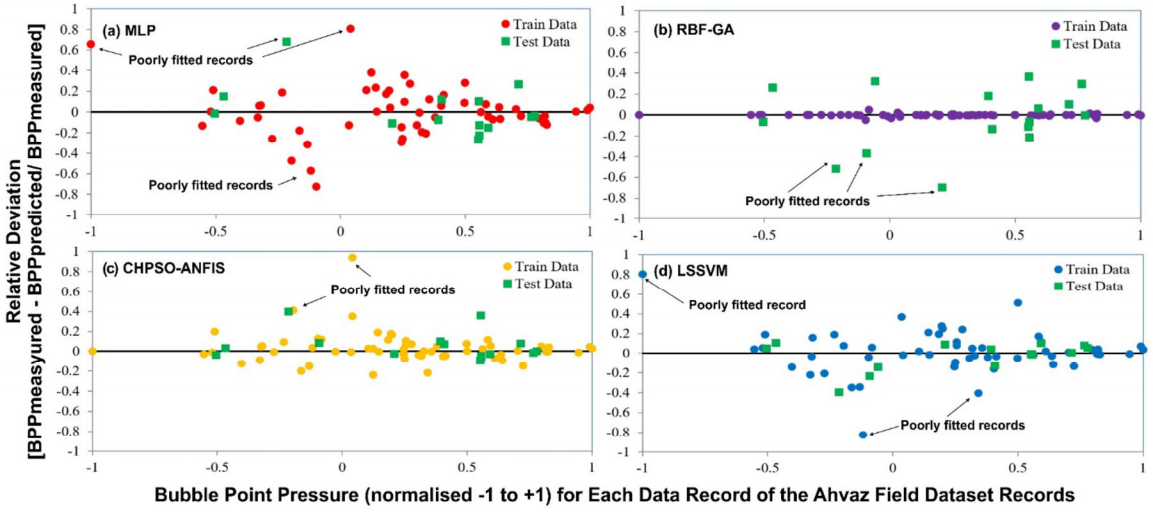

Fig. 11 Relative deviation [($BPP_{measured}$ – $BPP_{predicted}$)/$BPP_{measured}$] of predicted BPP values achieved by (a) MLP, (b) RBF, (c) ANFIS, and (d) LSSVM models with data records of the training and testing subsets distinguished.

However, for the RBF and ANFIS models do manage to successfully predict data record 38. On the other hand, it is data records in the testing subset with mid-range BPP values that achieved the worst predictions for the RBF and ANFIS models (i.e., data record 402 for RBF; data record 71 for ANFIS). Figure 12 illustrates the prediction sequence for individual data records by each of the four models for the training subset, showing clearly that the pattern of prediction by ANFIS and RBF models is close; on the other hand the pattern of prediction by MLP and LSSVM although similar is distinct.

The different manner in which the four AI models fit the data for the Ahvaz field data set suggests that for other field datasets the performance of the MLP, RBF, ANFIS, and LSSVM models might be ranked differently. It is therefore not justified, based on just the results of the Ahvaz field, to conclude that the RBF model will always outperform the other three models in applications to other datasets. A case can be made to use more than one AI model to evaluate the data from other fields.

In addition to the deviation error four other commonly used statistical measures of accuracy were calculated for each model. These are: correlation factor (R^2) (Eq. (5)), Average Absolute Relative Deviation (AARD) (Eq. (6)), Root Mean Squared Error (RMSE) (Eq. (7)), and Standard Deviation (STD) (Eq. (8)). In the Eqs. (5) to (8) λ denotes the BPP.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (\lambda_{i(\text{predicted})} - \lambda_{i(\text{measured})})^2}{\sum_{i=1}^{N} (\lambda_{i(\text{predicted})} - \lambda_{\text{mean(measured)}})^2} \tag{5}$$

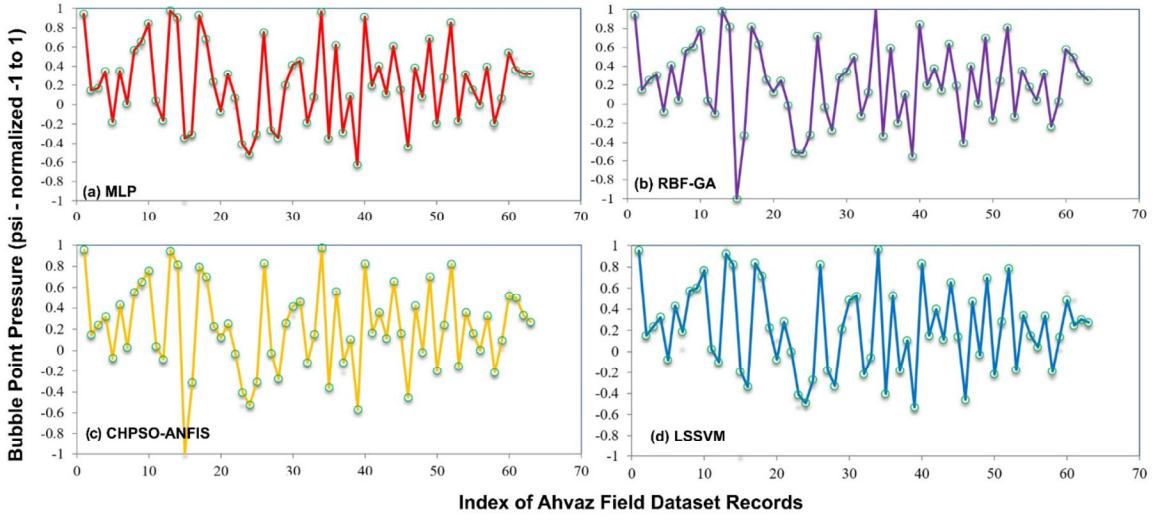

Fig. 12 Simultaneous representations of the predicted BPP values for the training subset applying (a) MLP, (b) RBF, (c) ANFIS, and (d) LSSVM models against index sequence number for each data record of the training subset.

$$\text{AARD}(\%) = \frac{100 \times \sum_{i=1}^{N} \left| \frac{(\lambda_{i(\text{predicted})} - \lambda_{i(\text{measured})})}{\lambda_{i(\text{measured})} \right|}{N} \quad (6)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left[(\lambda_{i(\text{predicted})} - \lambda_{i(\text{measured})}) \right]^2}{N}} \quad (7)$$

$$\text{STD} = \sqrt{\frac{\sum_{i=1}^{n} \left[(\lambda_{i(\text{predicted})} - \lambda_{\text{mean(measured)}}) \right]^2}{N-1}} \quad (8)$$

Table 7 shows the results for the four AI models applying all statistical measures of accuracy calculated. Displaying the highest correlation coefficient and lowest values of AARD%, RMSE, and STD applied to the entire dataset (all 79 records), the RBF model exhibits superior results in comparison with the other models. However, in terms of performance on the testing subset only the LSSVM model (followed by the MLP) outperforms the other models. If data record #40 (lowest BPP) were excluded from the training dataset the overall performance of the LSSVM and MLP models would be improved significantly, rivalling the RBF model. The statistical data (Table 7) suggest that the ANFIS model performs the best overall for this dataset, closely followed by the RBF model. The LSSVM model does though perform well with the testing dataset. However, all AI models perform well and outperforms historically published correlations for Ahvaz field dataset (see Section 4.2).

Figure 13 displays in radar diagrams the results for the R^2 and RMSE statistical measures of accuracy for the four AI models applied to the Ahvaz field PVT dataset with performances for training and testing subsets, and the full dataset distinguished. Table 7 and Fig. 13 identify that the ANFIS and RBF models perform more accurately than the other models in the prediction of BPP for the entire dataset. However, LSSVM and MLP models perform better than the RBF model when applied only to the testing subset.

4.2 Comparison of models developed with published PBB prediction models

Since the first bubble point pressure equation was developed during the 1940s, many researchers have tried to adapt and

Table 7 Statistical parameters values for four artificial intelligence models developed in this study

Intelligent predictor	Dataset	R^2	AARD	STD	RMSE	N
MLP	Train data	0.9378	2.0408	0.0446	109.2357	63
	Test data	0.9506	2.4239	0.0292	100.1158	16
	Total data	0.9504	2.1184	0.0418	107.4511	79
GA-RBF	Train data	0.9999	0.0407	0.0009	3.8321	63
	Test data	0.9381	2.4692	0.0311	108.2706	16
	Total data	0.9875	0.5326	0.0137	48.8456	79
ANFIS	Train Data	0.9934	0.7453	0.0107	35.3778	63
	Test Data	0.9775	1.2415	0.0186	65.4885	16
	Total Data	0.9902	0.8458	0.0126	43.2055	79
LSSVM	Train Data	0.9135	2.0936	0.0533	128.3316	63
	Test Data	0.9923	0.8998	0.0120	39.5873	16
	Total data	0.9295	1.8518	0.0478	115.9780	79

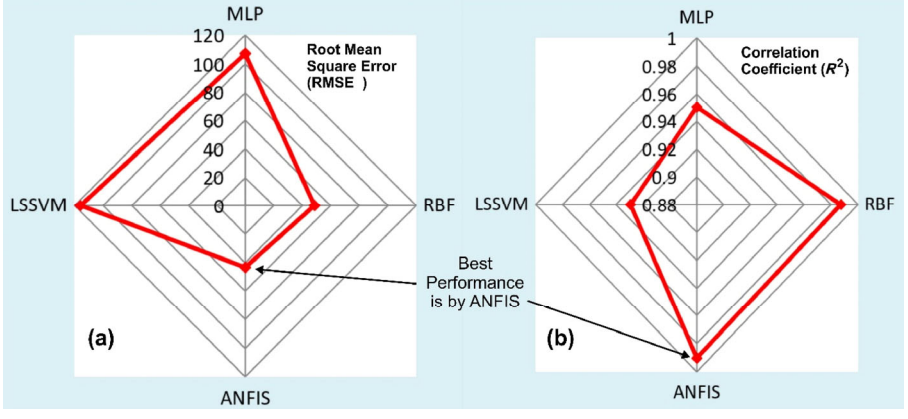

Fig. 13 Statistical accuracy measures for BPP prediction: (a) RMSE, (b) R^2 comparing performance of the four developed AI models applied to the entire Ahvaz field PVT dataset (79 records).

improve upon its performance, as addressed in the introduction. Here we have designed, developed, and evaluated four novel models using artificial intelligence algorithms that achieve low error rates and high accuracy when applied to the Ahvaz field PVT dataset. It is appropriate and instructive to compare the performance, with respect to the Ahvaz field dataset, of the four model developed here in terms of accuracy with those historically proposed BPP prediction models published by other researchers including: Standing (1947), Lasater (1958), Glaso (1980), Vasquez and Beggs (1980) (API⩾30), Al-Marhoun (1988), McCain (1991), Dokla and Osman (1992), Kartoatmojo and Schmidt (1994) (API<30), Farshid et al. (1996), Almehaideb (1997), Velarde et al. (1997), Al-Shammasi (2001), Dindoruk and Christman (2004), Bolondarzadeh et al. (2006), Mehran et al. (2006), Hemmati and Kharrat (2007), Ikiensikimama and Ogbaja (2009).

Appendix 1 lists details for 19 published correlation formulas for calculating the bubble point pressure including the data ranges for which each correlation is applicable. Table 8 compares the statistical accuracy achieved by these correlations applied to the entire Ahvaz field dataset. Figure 14 illustrates the comparative accuracy achieved by the historically published correlations versus the four AI models presented here. Table 8 and Fig. 14 confirm the significantly superior performance of all four AI models compared to historically published BPP prediction correlations when applied to the Ahvaz field dataset. They also identify the ANFIS and RBF neural network model as the most accurate BPP prediction model.

5 Conclusions

Precision in calculations and accurate measurement of fluid PVT properties, such as bubble point pressure, are key concerns for petroleum engineers associated with reserve evaluation, determination of recovery plans and estimation of the quantity and the quality of production fluid associated with oil field reservoir development. To improve the accuracy of predicting bubble point pressure from reservoir PVT data, we have developed models based on artificial intelligence/machine learning algorithms (i.e., MLP, RBF-GA, CHPSO-ANFIS, and LSSVM) and evaluated and compared their performance. The PVT-derived variables, temperature (T), oil formation volume factor (B_o), gas specific gravity (γ_g), solution gas oil ratio (R_s), oil specific gravity (γ_o), and API gravity (API) provide input to each model from multiple records of field datasets divided randomly into training (80%) and testing (20%) subsets. Here, 79 data records from Ahvaz oil field in southern Iran have been used as the PVT dataset to train, test, and compare the four AI models in the prediction

Table 8 Statistical measures of accuracy calculated to compare the performance of the various BPP prediction models applied to the entire Ahvaz field PVT dataset. The superior performance of the AI models developed in this study is highlighted. The four AI models are positioned in the final rows

Correlation	R^2	AARD	STD	RMSE	N
Standing (1947)	0.709431	16.79519	0.075907	604.7738	79
Lasater (1958)	0.449244	10.43311	0.120441	428.0071	79
Glaso (1980)	0.709826	9.539167	0.08411	382.4983	79
Farshad et al. (1996)	0.769647	4.7785	0.076531	224.1987	79
Al-Marhoun (1988)	0.509781	38.00561	0.08116	1350.635	79
Dokla et al. (1992)	0.504278	31.84684	0.071076	1153.006	79
Almehaldeb (1997)	0.507551	25.0963	0.080534	924.2731	79
Dindoruk and Christman (2004)	0.800778	12.29052	0.062495	461.8827	79
Vazquez and Beggs (1980)	0.709826	9.539167	0.08411	382.4983	79
McCain (1991)	0.709431	16.79519	0.075907	604.7738	79
Velarde et al. (1997)	0.787577	29.13315	0.052413	1028.594	79
Al-Shammasi (2001)	0.401844	18.40284	0.11969	742.5581	79
Bolondarzadeh et al. (2006)	0.807687	23.24218	0.055435	818.3258	79
Mehran et al. (2006)	0.512104	26.39188	0.100092	965.609	79
Gomaa (2016)	0.480333	96.84225	0.240478	3411.509	79
MLP	0.940472	2.118443	0.041816	107.4511	79
RBF	0.987530	0.532629	0.013758	48.84562	79
ANFIS	0.990234	0.845837	0.012619	43.20551	79
LSSVM	0.929555	1.851888	0.047890	115.9780	79

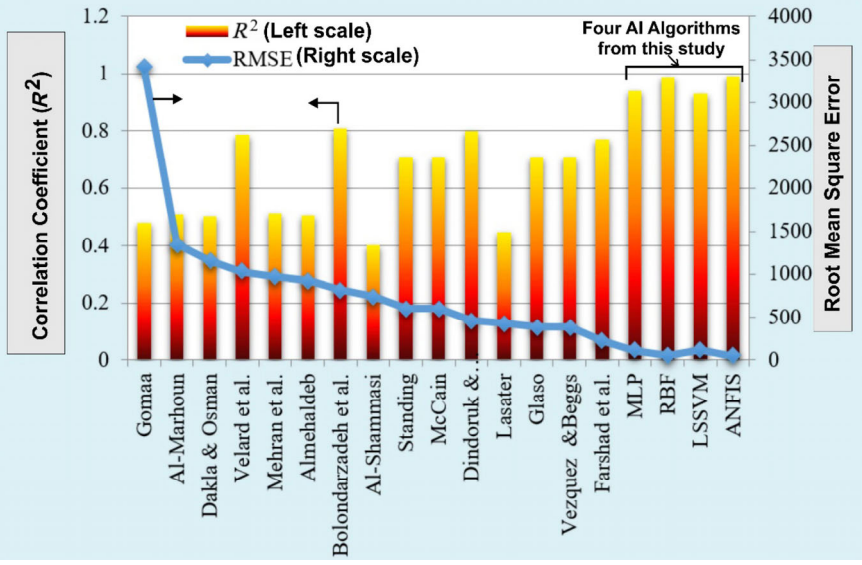

Fig. 14 Comparison of different correlations and the ANFIS and RBF model as the best prediction model in this study as measured by R^2 and RMSE values.

of bubble point pressure. Statistical error and accuracy analyses identify that the CHPSO-ANFIS model performs the best of the four models developed, displaying superior Average Absolute Relative Deviation (0.846), Standard Deviation (0.0126), and Root Mean Square Error (43.21), and Correlation Coefficient (0.9902) when applied to all 79 data records of the Ahvaz field dataset. Moreover, all four of the AI models developed here significantly outperform historically published correlations for predicting bubble point. Notably, the four models developed hybridize established learning network algorithms with high-performing optimization algorithms, both of which have many useful applications when employed on a standalone basis. The results of this study justifying the application of such hybridized AI-based and optimization-based algorithms for predicting bubble point from PVT data. Indeed, they suggest that for practical applications associated with specific fields predictions of bubble point pressure should use AI methods in preference to the traditional approach of adopting one of the less accurate published correlation formulas.

Appendix 1 Published correlations that predict BPP for crude oil. This list is modified and expanded after the earlier compilations of Moradi et al. (2010) and Gomaa (2016)

Publication	Origin	Data No.	Correlation
Standing (1947)	California	105	$P_b = a1^*\left[\left(\frac{R_s}{\gamma_g}\right)^{a2} {}^*10^{a3^*T - a4^*\mathrm{API}} - a5\right]$ $a1 = 18.2, a2 = 0.83, a3 = 0.00091, a4 = 0.0125, a5 = 1.4$
Lasater (1958)	Canada West Mid Continental U.S.A. South America	158	$P_b = \frac{[(p_f)(T + 459.67)]}{\gamma_g}, \gamma_g = \frac{\frac{R_s}{a1}}{\frac{R_s}{a1} + \frac{a2}{M_o}}$ $M_o = a3 - a4^*\mathrm{API} + a5^*\mathrm{API}^2$ $p_f = a6 - a7^*\gamma_g + a5^*\gamma_g^2$ $a1 = 379.3, a2 = 350, a3 = 725.32143, a4 = 16.033, a5 = 0.09524, a6 = 0.38418,$ $a7 = 1.20081, a8 = 9.6486$
Vazquez and Beggs (1980)	World Wide	5008	$P_b = \left[\left(a1^*\frac{R_s}{\gamma_{gs}}\right)^*10^{a3^*\frac{\mathrm{API}}{T+460}}\right]^{\frac{1}{a2}}$ $\gamma_{gs} = \gamma_g{}^*\left(1 + 5.915^*10^{-5}{}^*T_{sep}{}^*\mathrm{API}^*\log\left(\frac{P_{sep}}{114.7}\right)\right)$ If API⩽30, $a1 = 27.64, a2 = 1.0937, a3 = 11.172$ If API>30, $a1 = 56.06, a2 = 1.187, a3 = 10.393$

(Continued)

Publication	Origin	Data No.	Correlation
Glaso (1980)	North Sea	41	$P_b = 10^{a1+a2*\log(G)-a3(\log(G))^2}$ $G = \left(\frac{R_s}{\gamma_g}\right)^{a4} * T^{a5} * \text{API}^{a6}$ $a1 = 1.7669$, $a2 = 1.7447$, $a3 = 0.3021$, $a4 = 0.86$, $a5 = 0.172$, $a6 = -0.989$
Al-Marhoun (1988)	Middle East	160	$P_b = a1 * R_s^{a2} * \gamma_g^{a3} * \gamma_o^{a4} * (T + 460)^{a5}$ $a1 = 18.25.380\text{E-}3$, $a2 = 0.83715$, $a3 = -1.877$, $a4 = 3.143$, $a5 = 1.326$
McCain (1991)	World Wide	100	$P_b = a1 * (C_{pb} - a2)\left[\left(\frac{R_s^{a2}}{\gamma_g}\right) * 10^{a3*T - a4*\text{API}} - a5\right]$ $C_{pb} = \left(\frac{R_s^{a2}}{\gamma_g}\right)^{a3} * 10^{(a4*T - a5*\text{API})}$ $a1 = 18.2$, $a2 = 1.4$, $a3 = 0.83$, $a4 = 0.00091$, $a5 = 0.0125$
Kartoatmojo and Schmidt (1994)	World Wide	5392	$P_b = \left(\frac{R_s}{a1 * \gamma_{g100}^{a2} 10^{a3*\frac{\text{API}}{T+460}}}\right)^{a4}$ If API⩽30, $a1 = 0.05958$, $a2 = 0.7972$, $a3 = 13.1405$, $a4 = 0.998$ If API>30, $a1 = 0.03150$, $a2 = 0.7972587$, $a3 = 11.28$, $a4 = 0.914$
Dokla and Osman (1992)	U.A.E	51	$P_b = a1 * R_s^{a2} * \gamma_g^{a3} * \gamma_o^{a4} * (T + 460)^{a5}$ $a1 = 18.20.83638\text{E}4$, $a2 = 0.724047$, $a3 = -1.01049$, $a4 = 010799$, $a5 = -0.95258$
Farshad et al. (1996)	Colombia	98	$P_b = 10^{\left(a1+a2*\log(G)-a3*[\log(G)]^2\right)}$ $G = \gamma_g^{a4} * R_s^{a5} * 10^{(a6*T - a7*\text{API})}$ $a1 = 0.3058$, $a2 = 1.9013$, $a3 = 0.26$, $a4 = -1.378$, $a5 = 1.053$, $a6 = 0.00069$, $a7 = 0.0208$
Almehaideb (1997)	U.A.E	62	$P_b = a1 + a2 * \frac{R_s * \gamma_o}{(\gamma_g * B_o^{a3})} + a4 * T$ $a1 = -620.592$, $a2 = 6.23087$, $a3 = 1.38559$, $a4 = 2.89868$
Velarde et al. (1997)	World Wide	2097	$P_b = a1 * [R_s^{a2} * \gamma_{gs}^{a3} * 10^x - a4]^{a5}$ $x = (a6 * T^{a7}) - (a8 * \text{API}^{a9})$ $a1 = 18.2$, $a2 = 0.83$, $a3 = 0.00091$, $a4 = 0.0125$, $a5 = 1.4$
Al-Shammasi (2001)	World Wide	1661	$P_b = \gamma_o^{a1} * (\exp(a2 * \gamma_g * \gamma_o)) * (R_s * (T + 460) * \gamma_g)^{a3}$ $a1 = 5.527215$, $a2 = -1.841408$, $a3 = 0.783716$
Dindoruk and Christman (2004)	Gulf of Mexico	99	$P_b = a8 * \left(\left(\frac{R_s^{a9}}{\gamma_g^{a10}}\right) * 10^A + a11\right)$ $P_b = \frac{a1 * T^{a2} + a3 * \text{API}^{a4}}{\left(a5 + 2\frac{R_s^{a6}}{\gamma_g^{a7}}\right)^2}$ $a1 = 1.4282\text{E-}10$, $a2 = 2.84459$, $a3 = -6.7489\text{E-}04$, $a4 = 1.22522$, $a5 = 0.03338$, $a6 = -0.27294$, $a7 = -0.084226$, $a8 = 1.869979$
Mehran et al. (2006)	Iran	387	$P_b = a1 * R_s^{a2} * \gamma_g^{a3} * \gamma_o^{a4} * T^{a5}$ $a1 = 3.146$, $a2 = 0.8035$, $a3 = -1.3114$, $a4 = 3.3925$, $a5 = 0.3466$
Bolondarzadeh et al. (2006)	Iran	166	$P_b = a1 * \left(\left(\frac{a2 * R_s^{a3}}{a4 * \gamma_g^{a5}}\right) * \left(\frac{a6 * T^{a7}}{a8 * \text{API}^{a9}}\right) - a10\right)$ $a1 = 27.16$, $a2 = 3.4394$, $a3 = 0.57102$, $a4 = 0.56807$, $a5 = 922092$, $a6 = 3.7387$, $a7 = 0.2304$, $a8 = 6.27605$, $a9 = 0.42469$, $a10 = 30.28$

(Continued)

Publication	Origin	Data No.	Correlation
Hemmati and Kharrat (2007)	Iran	287	$P_b = 10.4566^* \left[\left(\frac{R_s}{\gamma_g} \right)^x * 10^{0.0008T - 0.0098\gamma_o} - 8.6817 \right]$ $x = a1 + a2^*B_o + a3^*\gamma_g + a4^*B_{ob}{}^2 + a5^*\gamma_g{}^2 + \frac{a6}{\gamma_g * B_o}$ $a1 = 18.2,\ a2 = 0.83,\ a3 = 0.00091,\ a4 = 0.0125,\ a5 = 1.4$
Ikiensikimama and Ogbaja (2009)	Niger Delta	250	$Q = \frac{\frac{R_s}{a1}}{\left(\frac{R_s}{a1}\right) + \left(a2^* \frac{\gamma_o}{S}\right)}, S = (a3 - a4^*\text{API})^{a5}$ $P_b^* = a6 + a7^*Q + a8^*Q^{a9}$ $P_b = \frac{(P_b^*)(T + a10)}{\gamma_g}$ $a1 = 336.0064009,\ a2 = 6.7063984,\ a3 = 0.677706662,\ a4 = 47.57094,\ a5 = 1.5309356,$ $a6 = 0.243181,\ a7 = -2.316548,\ a8 = 10.60567,\ a9 = 1.518030,\ a10 = 635.4152$
Moradi et al. (2010)	World Wide	1801	$P_b = a5^*P_b^* + a6^*P_b^{*2} + a7^*P_b^{*3}$ $P_b^* = a1^*\text{LOG}(\text{API}) * \left(\frac{141.5}{\gamma_o} - 131.5 \right)^{a2} * \left(\left(\text{EXP}\left(a3^*\gamma_g * \gamma_g\right) \right) * \left(R_s * (T + 460) * \gamma_g \right)^{a4} \right)$ $a1 = 1.10382622244782,\ a2 = 6.20868199092533,\ a3 = -1.84068688374902,$ $a4 = 0.688750576134232,\ a5 = -65.853149,\ a6 = 0.0004066890,\ a7 = -0.00000015472455$
Gomaa (2016)	Middle East	441	$P_b = a1\gamma_g{}^{a2} * \gamma_o{}^{a3} * R_s{}^{a4} * T^{a5} \exp\left(a1\gamma_g T\right)$ $a1 = 0.000221989,\ a2 = 0.07921662,\ a3 = 2.882096242,\ a4 = 0.787046015,\ a5 = 1.968344769,$ $a6 = -0.002128156$

Acknowledgements

The authors wish to express special thanks to the National Iranian Oil Company (NIOC) and to Dr. Jamshid Moghadasi, Mr. Saeed Kooti, Mr. Pejman Ghazaeipour Abaghoei from NIOC for their advice.

References

Adeleke, N., Ityokumbul, M. T., Adewumi, M. 2013. Blockage detection and characterization in natural gas pipelines by transient pressure-wave reflection analysis. *SPE J*, 18: 355–365.

Ahmadi, M. A., Zendehboudi, S., Lohi, A., Elkamel, A., Chatzis., I. 2013. Application of hybrid genetic algorithm with particle swarm optimization and neural network for reservoir permeability prediction. *J Geophys Prospect*, 61: 582–598.

AlAjmi, M. D., Alarifi, S. A., Mahsoon, A. H. 2015. Improving multiphase choke performance prediction and well production test validation using artificial intelligence: A new milestone. In: Proceedings of the SPE Digital Energy Conference and Exhibition: SPE-173394-MS.

Al-Marhoun, M. A. 1988. PVT correlations for middle east crude oils. *J Petrol Technol*, 40: 650–666.

Al-Marhoun, M. A., Osman, E. A. 2002. Using artificial neural networks to develop new PVT correlations for saudi crude oils. In: Proceedings of the Abu Dhabi International Petroleum Exhibition and Conference: SPE-78592-MS.

Almehaideb, R. A. 1997. Improved PVT correlations for UAE crude oils. In: Proceedings of the Middle East Oil Show and Conference: SPE-37691-MS.

Al-Shammasi, A. A. 2001. A review of bubblepoint pressure and oil formation volume factor correlations. *SPE Reserv Eval Eng*, 4: 146–160.

Arabloo, M., Amooie, M. A., Hemmati-Sarapardeh, A., Ghazanfari, M. H., Mohammadi, A. H. 2014. Application of constrained multi-variable search methods for prediction of PVT properties of crude oil systems. *Fluid Phase Equilibr*, 363: 121–130.

Asoodeh, M., Kazemi, K. 2013. Estimation of bubble point pressure: Using a genetic integration of empirical formulas. *Energ Source Part A*, 35: 1102–1109.

Atashnezhad, A., Wood, D. A., Fereidounpour, A., Khosravanian, R. 2014. Designing and optimizing deviated wellbore trajectories using novel particle swarm algorithms. *J Nat Gas Sci Eng*, 21: 1184–1204.

Bandyopadhyay, P., Sharma, A. 2011. Development of a new semi analytical model for prediction of bubble point pressure of crude oils. *J Petrol Sci Eng*, 78: 719–731.

Basarir, H., Tutluoglu, L., Karpuz, C. 2014. Penetration rate prediction for diamond bit drilling by adaptive neuro-fuzzy inference system and multiple regressions. *Eng Geol*, 173: 1–9.

Bolondarzadeh, A., Hashemi, S., Solgani, B. 2006. The new PVT generated correlations of Iranian oil properties. In: Proceedings of the 4th Iranian Petroleum Engineering Student Conference.

Boukadi, F. H., Bermani, A. S., Hashmi, A. 2002. PVT empirical models for saturated Omani crude oils. *Petroleum Science and Technology*,

20: 89–100.

Broomhead, D. S., Lowe, D. 1988. Radial basis functions, multi-variable functional interpolation and adaptive networks (No. RSRE-MEMO-4148). Royal Signals and Radar Establishment Malvern (United Kingdom).

Choubineh, A., Ghorbani, H., Wood, D. A., Robab Moosavi, S., Khalafi, E., Sadatshojaei, E. 2017. Improved predictions of wellhead choke liquid critical-flow rates: Modelling based on hybrid neural network training learning based optimization. *Fuel*, 207: 547–560.

Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function. *Math Control Signals Syst*, 2: 303–314.

Darwin, C. 1859. *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*. London: John Murray.

Deisman, N., Khajeh, M., Chalaturnyk, R. J. 2013. Using geological strength index (GSI) to model uncertainty in rock mass properties of coal for CBM/ECBM reservoir geomechanics. *Int J Coal Geol*, 112: 76–86.

Dindoruk, B., Christman, P. G. 2004. PVT properties and viscosity correlations for gulf of Mexico oils. *SPE Reserv Eval Eng*, 7: 427–437.

Dixit, N., Zeng, D. L., Kalonia, D. S. 2012. Application of maximum bubble pressure surface tensiometer to study protein–surfactant interactions. *Int J Pharmaceut*, 439: 317–323.

Dokla, M., Osman, M. 1992. Correlation of PVT properties for UAE crudes (includes associated papers 26135 and 26316). *SPE Formation Eval*, 7: 41–46.

Dong, J., Feldmann, G., Huang, J., Wu, S., Zhang, N., Comerford, S. A., Gayyed, M. F., Anders, R. A., Maitra, A., Pan, D. 2007. Elucidation of a universal size-control mechanism in drosophila and mammals. *Cell*, 130: 1120–1133.

Dutta, S., Gupta, J. P. 2010. PVT correlations for Indian crude using artificial neural networks. *J Petro Sci Eng*, 72: 93–109.

El-Sebakhy, E., Sheltami, T., Al-Bokhitan, S., Shaaban, Y., Raharja, P., Khaeruzzaman, Y. 2007. Support vector machines framework for predicting the PVT properties of crude-oil systems. In: Proceedings of the SPE Middle East Oil and Gas Show and Conference: SPE-105698-MS.

Elsharkawy, A. M. 1998. Modeling the properties of crude oil and gas systems using RBF network. In: Proceedings of the SPE Asia Pacific Oil and Gas Conference and Exhibition: SPE-49961-MS.

Elsharkawy, A. M., Elgibaly, A. A., Alikhan, A. A. 1995. Assessment of the PVT correlations for predicting the properties of Kuwaiti crude oils. *J Petrol Sci Eng*, 13: 219–232.

Fainerman, V. B., Miller, R. 2004. Maximum bubble pressure tensiometry—An analysis of experimental constraints. *Adv Colloid Interfac*, 108–109: 287–301.

Farasat, A., Shokrollahi, A., Arabloo, M., Gharagheizi, F., Mohammadi, A. H. 2013. Toward an intelligent approach for determination of saturation pressure of crude oil. *Fuel Process Technol*, 115: 201–214.

Farshad, F., LeBlanc, J. L. Garber, J. D., Osorio, J. G. 1996. Empirical PVT correlations for colombian crude oils. In: Proceedings of the SPE Latin America/Caribbean Petroleum Engineering Conference: SPE-36105-MS.

Gharbi, R. B., Elsharkawy, A. M. 1997. Neural network model for estimating the PVT properties of middle east crude oils. In: Proceedings of the Middle East Oil Show and Conference: SPE-37695-MS.

Gharbi, R. B., Elsharkawy, A. M., Karkoub, M. 1999. Universal neural-network-based model for estimating the PVT properties of crude oil systems. *Energ Fuel*, 13: 454–458.

Ghorbani, H., Moghadasi, J., Wood, D. A. 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.

Ghorbani, H., Wood, D. A., Choubineh, A., Tatar, A., Abarghoyi, P. G., Madani, M., Mohamadian, N. 2018. Prediction of oil flow rate through an orifice flow meter: Artificial intelligence alternatives compared. *Petroleum*, https://doi.org/10.1016/j.petlm.2018.09.003.

Ghorbani, H., Wood, D. A., Moghadasi, J., Choubineh, A., Abdizadeh, P., Mohamadian, N. 2019. Predicting liquid flow-rate performance through wellhead chokes with genetic and solver optimizers: An oil field case study. *J Petrol Explor Prod Technol*, 9: 1355–1373.

Glaso, O. 1980. Generalized pressure–volume–temperature correlations. *J Petrol Technol*, 32: 785–795.

Goda, M. H., Shokir, E. M., Fattah, K. A., Sayyouh, M. H. 2003. Prediction of the PVT data using neural network computing theory. In: Proceedings of the Nigeria Annual International Conference and Exhibition: SPE-85650-MS.

Gomaa, S. 2016. New bubble point pressure correlation for middle east crude oils. *International Advanced Research Journal in Science, Engineering and Technology*, 3: 1–9.

Haykin, S. 1994. *Neural Networks: a Comprehensive Foundation*. Prentice Hall PTR Upper Saddle River.

Hemmati, M. N., Kharrat, R. 2007. A correlation approach for prediction of crude oil PVT properties. In: Proceedings of the SPE Middle East Oil and Gas Show and Conference: SPE-104543-MS.

Holcomb, C., Outcalt, S. 1999. Near-saturation (P, ρ, T) and vapor-pressure measurements of NH_3, and liquid-phase isothermal (P, ρ, T) and bubble-point-pressure measurements of NH_3+H_2O mixtures. *Fluid Phase Equilibr*, 164: 97–106.

Hush, D. R., Horne, B. G. 1993. Progress in supervised neural networks. *IEEE Sig Proc Mag*, 10: 8–39.

Ikiensikimama, S. S., Ajienka, J. A. 2012. Impact of PVT correlations development on hydrocarbon accounting: The case of the Niger Delta. *J Petrol Sci Eng*, 81: 80–85.

Ikiensikimama, S. S., Ogboja, O. 2009. New bubblepoint pressure empirical PVT correlation. In: Proceedings of the Nigeria Annual International Conference and Exhibition: SPE-128893-MS.

Jang, J. S. R., Sun, C. T., Mizutani, E. 1997. *Neuro-Fuzzy and Soft Computing*. Prentice Hall: 335–368.

Jang, J.-S. R. 1993. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE T Syst Man Cyb*, 23: 665–685.

Kartoatmodjo, T., Schmidt, Z. 1994. Large data bank improves crude physical property correlations. *Oil and Gas Journal*, 92(27).

Kennedy, J., Eberhart, R. 1995. Particle swarm optimization. In: Proceedings of ICNN'95 - International Conference on Neural Networks, 4: 1942–1948.

Kirkpatrick, S., Gelatt Jr., C. D., Vecchi, M. P. 1983. Optimization by simulated annealing. *Science*, 220: 671–680.

Kloubek, J. 1972. Measurement of the dynamic surface tension by the maximum bubble pressure method. III. Factors influencing the measurements at high frequency of the bubble formation and an extension of the evaluation to zero age of the surface. *J Colloid Interf Sci*, 41: 7–16.

Lasater, J. A. 1958. Bubble point pressure correlation. *J Petrol Technol*, 10: 65–67.

Li, H., Yang, D. T. 2012. Phase behaviour of C_3H_8-n-C_4H_{10}-heavy oil systems at high pressures and elevated temperatures. In: Proceedings of the SPE Heavy Oil Conference Canada: SPE-157744-MS.

Liscic, B., Tensi, H. M., Canale, L. C. E., Totten, G. 2010. *Quenching Theory and Technology*, 2nd edn. CRC Press.

Malallah, A. M., Gharbi, R., Algharaib, M. 2006. Accurate estimation of the world crude oil PVT properties using graphical alternating conditional expectation. *Energy Fuels*, 20: 688–698.

Mansouri, V., Khosravanian, R., Wood, D. A., Aadnoy, B. S. 2015. 3-D well path design using a multi objective genetic algorithm. *J Nat Gas Sci Eng*, 27: 219–235.

McCain Jr., W. D. 1991. Reservoir-fluid property correlations - State of the art. *SPE Reservoir Eng*, 6: SPE-18571-PA.

Mehran, F., Movagharnejad, K., Didanloo, A. 2006. New correlation for estimation of formation vilume factor and bubblepoint pressure for Iranian oil fields. In: Proceedings of the 1st Iranian Pet. Eng. Conference.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E. 1953. Equation of state calculations by fast computing machines. *J Chem Phys*, 21: 1087–1092.

Mishchuk, N. A., Fainerman, V. B., Kovalchuk, V. I., Miller, R., Dukhin, S. S. 2000. Studies of concentrated surfactant solutions using the maximum bubble pressure method. *Colloid Surface A*, 175: 207–216.

Moradi, B., Malekzadeh, E., Amani, M., Boukadi, F. H., Kharrat, R. 2010. Bubble point pressure empirical correlation. In: Proceedings of the Trinidad and Tobago Energy Resources Conference: SPE-132756-MS.

Nnochiri, M. O., Lawal, K. A. 2010. How variable fluid PVT model affects the performance of an integrated production system. In: Proceedings of the SPE EUROPEC/EAGE Annual Conference and Exhibition: SPE-130881-MS.

Onwunalu, J. E. O., Durlofsky, L. J. 2010. Application of a particle swarm optimization algorithm for determining optimum well location and type. *Computat Geosci*, 14: 183–198.

Petrosky Jr., G. E., Farshad, F. F. 1993. Pressure–volume–temperature correlations for gulf of Mexico crude oils. In: Proceedings of the SPE Annual Technical Conference and Exhibition: SPE-26644-MS.

Proett, M. A., Chin, W. C., Mandal, B. 2000. Advanced dual probe formation tester with transient, harmonic, and pulsed time-delay testing methods determines permeability, skin, and anisotropy. In: Proceedings of the International Oil and Gas Conference and Exhibition in China: SPE-64650-MS.

Sayahi, T., Tatar, A., Bahrami, M. 2016. A RBF model for predicting the pool boiling behavior of nanofluids over a horizontal rod heater. *Int J Therm Sci*, 99: 180–194.

Simjoo, M., Dong, Y., Andrianov, A., Talanana, M., Zitha, P. L. J. 2013. Novel insight into foam mobility control. *SPE J*, 18: 416–427.

Standing, M. B. 1947. A pressure–volume–temperature correlation for mixtures of California oils and gases. *Drilling and Production Practice*, API-47-275: 275–287.

Sun, H., Fang, W., Guo, Y., Lin, R. 2005. Investigation of bubble-point vapor pressures for mixtures of an endothermic hydrocarbon fuel with ethanol. *Fuel*, 84: 825–831.

Tadeusiewicz, R. 1995. Neural networks: A comprehensive foundation. *Control Eng Pract*, 3: 746–747.

Valkó, P. P., McCain Jr., W. D. 2003. Reservoir oil bubblepoint pressures revisited; solution gas–oil ratios and surface gas specific gravities. *J Petrol Sci Eng*, 37: 153–169.

Vapnik, V. 2013. *The Nature of Statistical Learning Theory*. Springer Science & Business Media.

Vasquez, M., Beggs, H. D. 1980. Correlations for fluid physical property prediction. *J Petrol Technol*, 32: 968–970.

Velarde, J., Blosingame, T. A., McCain Jr., W. D. 1997. Correlation of black oil properties at pressures below bubble point pressure - A new approach. In: Proceedings of the Annual Technical Meeting: PETSOC-97-93.

Xavier-de-Souza, S., Suykens, J. A. K., Vandewalle, J., Bolle, D. 2010. Coupled simulated annealing. *IEEE T Syst Man Cy B*, 40: 320–335.

Yasari, E., Pishvaie, M. R., Khorasheh, F., Salahshoor, K., Kharrat, R. 2013. Application of multi-criterion robust optimization in water-flooding of oil reservoir. *J Petrol Sci Eng*, 109: 1–11.

Yavari, H., Sabah, M., Khosravanian, R. Wood, D. A. 2018. Application of adaptive neuro-fuzzy inference system and mathematical ROP models for prediction of drilling rate. *Iranian Journal of Oil & Gas Science and Technology*, 7: 73–100.

Yazaydin, A. Ö., Martin, M. G. 2007. Bubble point pressure estimates from Gibbs ensemble simulations. *Fluid Phase Equilibr*, 260: 195–198.

Zoveidavianpoor, M., Samsuri, A., Shadizadeh, S. R. 2013. Adaptive neuro fuzzy inference system for compressional wave velocity prediction in a carbonate reservoir. *J Appl Geophys*, 89: 96–107.