Oil Formation Volume Factor Determination Through a Fused Intelligence

Amin GHOLAMI

Reservoir Engineering Division, Iranian Offshore Oil Company, Tehran, Iran
e-mails: amingholamiput@yahoo.com, AGHolami@iooc.co.ir

A b s t r a c t

Volume change of oil between reservoir condition and standard surface condition is called oil formation volume factor (FVF), which is very time, cost and labor intensive to determine. This study proposes an accurate, rapid and cost-effective approach for determining FVF from reservoir temperature, dissolved gas oil ratio, and specific gravity of both oil and dissolved gas. Firstly, structural risk minimization (SRM) principle of support vector regression (SVR) was employed to construct a robust model for estimating FVF from the aforementioned inputs. Subsequently, an alternating conditional expectation (ACE) was used for approximating optimal transformations of input/output data to a higher correlated data and consequently developing a sophisticated model between transformed data. Eventually, a committee machine with SVR and ACE was constructed through the use of hybrid genetic algorithm-pattern search (GA-PS). Committee machine integrates ACE and SVR models in an optimal linear combination such that makes benefit of both methods. A group of 342 data points was used for model development and a group of 219 data points was used for blind testing the constructed model. Results indicated that the committee machine performed better than individual models.

Key words: PVT, oil formation volume factor (FVF), alternating conditional expectation (ACE), support vector regression (SVR).

1. INTRODUCTION

Oil formation volume factor (FVF) is defined as the ratio of the volume of oil (plus the gas in solution) at the prevailing reservoir temperature and pressure to the volume of oil at standard conditions (Ahmed 2000). FVF value has significance in calculating various parameters such as the depletion rate, oil in place, predicting the future of the reservoir, optimizing the rate of production, designing of production operation and facilities (Bagheripour *et al.* 2013). Rigorous depiction of FVF through differential vaporization test on bottom-hole or recombined surface samples is very time, cost and labor intensive (Zargar *et al.* 2014). Furthermore, sampling for experiments is limited to early producing life of reservoir which imposes a restraint to laboratory measurements (Dake 1988). Early attempts of researchers for presenting a practical, cheap and accurate way of determining FVF from available PVT data led to several empirical correlations (Katz 1942, Knopp and Ramsey 1960, Vazquez and Beggs 1970, Glaso 1980, Al-Marhoun 1988, Farshad *et al.* 1996, Petrosky and Farshad 1993, Omar and Todd 1993, Al-mehaideb 1997, Al-Shammasi 1999, Dindoruk and Christman 2001, El-Banbi *et al.* 2006, Hemmati and Kharrat 2007, Elmabrouk *et al.* 2010). A striving competition between intelligent systems and empirical correlations *versus* exactness and generalization has been done to show superiority of intelligent systems (Asoodeh and Kazemi 2013, Kazemi *et al.* 2013, Asoodeh and Bagheripour 2012a, 2013a; Bagheripour and Asoodeh 2013, 2014; Gholami *et al.* 2014a, b; Afshar *et al.* 2014). Hitherto, some scientists utilized intelligent systems for formulating oil FVF to available PVT data (Gharbi and Elsharkawy 1996, Elsharkawy 1998, Elsharkawy and Gharbi 2000, Al-Marhoun and Osman 2002, Dutta and Gupta 2010). The quest for higher accuracy forced researchers not to satisfy themselves with individual intelligent systems but to develop integrated models such as committee machines for enhancing precision of final prediction (Asoodeh and Bagheripour 2012b, Asoodeh 2013, Asoodeh *et al.* 2014a, b; Gholami *et al.* 2014c, d; Bagheripour *et al.* 2014). In this study, two sophisticated models, including support vector regression (SVR) and alternating conditional expectation (ACE) were employed to construct a strong formulation between PVT data and oil FVF. Several researches showed high performance of ACE and superiority of SVR to traditional networks (Shokir 2007, Al-Anazi and Gates 2010, Rafiee-Taghanaki *et al.* 2013, Na'imi *et al.* 2014, Asoodeh and Bagheripour 2013b, Asoodeh *et al.* 2014b, Gholami *et al.* 2014c, Fattahi *et al.* 2014, Bagheripour *et al.* 2015). At next stage, results of SVR and ACE models were combined by means of genetic algorithm-pattern search technique in an optimal linear combination of committee machine. This strategy was successfully applied to open source oil samples. Results indicated that

the committee machine significantly enhanced accuracy of final prediction compared with individual ACE and SVR models.

2. THEORY: COMMITTEE MACHINE WITH SVR AND ACE

Committee machine is a parallel framework, as shown in Fig. 1, which gathers outputs of different models and combines them in an optimal linear structure by means of hybrid genetic algorithm-pattern search (GA-PS) tool. The GA-PS assigns a weight factor showing involvement of each model in overall estimation of target. In this study, PVT data, including reservoir temperature, dissolved gas oil ratio, and specific gravity of both oil and dissolved gas are introduced to support vector regression (SVR) and alternating conditional expectation (ACE) models for estimating oil formation volume factor (FVF). Outputs of SVR and ACE models are then input in committee machine. The GA-PS subsequently extracts involvement weights of each model such that mean square error (MSE) of prediction reaches its global minimum. This process consequently enhances the accuracy of final prediction. A brief introduction of SVR, ACE, and GA-PS is brought in the following paragraphs.
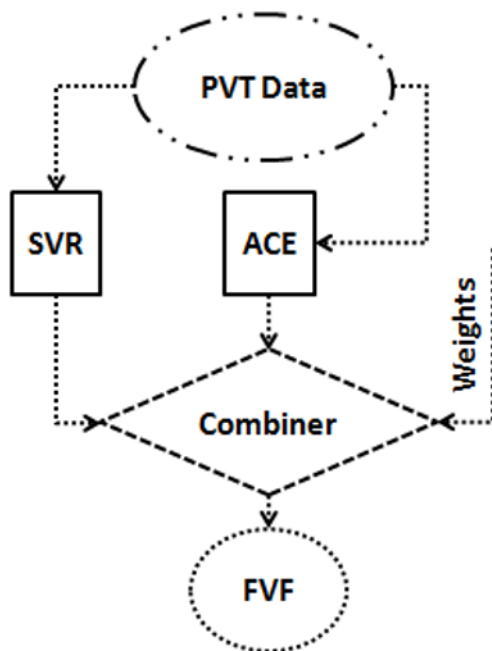

Fig. 1. Schematic diagram of committee machine used in this study.

2.1 Alternating conditional expectation

Alternating conditional expectation (ACE) is a nonparametric regression algorithm invented by Breiman and Friedman (1985). It is widely used for situations where underlying dependency between input/output data space is inexact or functional form between them is unidentified. This method transforms input/output data space to a higher correlated data space and develops the quantitative formulation between them in that space. It approximates optimal transformation of input/output data through minimizing error variance between transformed output data and sum of transformed input data in an alternative minimization process. When optimal transformations are achieved, a simple curve fitting can demonstrate optimal quantitative formulation between input/output data. In current part, a concise description about ACE formulation is given. More details about ACE are brought in a work by Breiman and Friedman (1985). General form of linear regression for formulation between p independent variables, $X_1, X_2, \ldots, X_p$, and a response variable Y is

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \varepsilon \ . \tag{1}$$

Here, $(\beta_i, \ i = 0 - p)$ are the regression coefficients which must be determined accurately, and ε is an error term. In ACE, transformations of Y and $X_1, X_2, \ldots, X_p$ are substituted in regression equation for making formulation between independent variables and response variable. Indeed, independent variables and response variable are firstly transformed into higher correlated data space and then, in this space, the underlying dependency between those is computed.

Based on aforementioned regression form, the non-parametric ACE algorithm is defined as the following equation:

$$\theta(Y) = \alpha + \sum_{i=1}^{p} \phi(X_i) + \varepsilon \ , \tag{2}$$

where $\theta(Y), \phi_1(X_1), \ldots, \phi_p(X_p)$ are the arbitrary measurable mean-zero functions of $Y, X_1, X_2, \ldots, X_p$, respectively. Hence, the main objective in ACE is to find the optimal transformation $\phi_i^*(X_i), \ i = 1, \ldots, p$ and $\theta^*(Y)$ which concluded the maximum correlation between transformed dependent variable and sum of transformed predicted variables. Breiman and Friedman (1985) suggested that the aforementioned objective is achieved through minimizing the value of the error variance (ε^2). The value of the error variance (ε^2) of a linear regression of the transformed dependent variable on the

sum of transformed independent variables (under the constraint, $E[\theta^2(Y)] = 1$) is given by the following equation:

$$\varepsilon^2(\theta, \phi_i, \ldots, \phi_p) = E\left(\left[\theta(Y) - \sum_{i=1}^{p}\phi(X_i)\right]\right)^2 \bigg/ E\theta^2(Y) . \tag{3}$$

By implementing minimization of the value of ε^2 with respect to $\theta(Y)$ and $\phi_k(X_k)(i = 1, 2, \ldots, k)$ through a series of single-function minimizations, the following equation for response variable and predictor variable is computed, respectively:

$$\theta(Y) = E\left[\sum_{i=1}^{p}\phi(X_i)\,|\,Y\right]\bigg/\left\|E\left[\sum_{i=1}^{p}\phi(X_i)\,|\,Y\right]\right\|, \tag{4}$$

$$\phi_{j,1}(X_j) = E\left[\theta(Y) - \sum_{i=1}^{p}\phi_i(X_i)\,|\,X_k\right]. \tag{5}$$

Performing of iteration process of minimizing ε^2 leads to determining the real-valued measurable zero-mean functions $\phi_i(X_i)$, $i = 1, \ldots, p$ and $\theta(Y)$, which is equivalent to optimal transformation $\phi_i^*(X_i)$, $i = 1, \ldots, p$ and $\theta^*(Y)$. In the transformed space, the response and predictor variables are related as following.

$$\theta^*(Y) = \sum_{i=1}^{p}\phi_i^*(X_i) + e^* , \tag{6}$$

where e^* is the error not captured by the use of the ACE transformations and is assumed to have a normal distribution with zero mean.

2.2 Support vector regression

Support vector regression, invented by Vapnik (1995), is a supervised model inspired by statistical learning theory. SVR utilizes structural risk minimization (SRM) in conjunction with empirical risk minimization (ERM) that fortifies it with the highest generalization owing to constructing a structural model that would be as smooth as possible. SVR nonlinearly maps input/output data to a higher-dimensional feature space by means of kernel functions such that a linear relationship between input/output data exists in feature space. Linear hyperplane in feature space produces a nonlinear regression hypersurface in original data space. Therefore, nonlinear relationship between input/output data is extracted. Here brief description of SVR

method is given. For more study about SVR, refer to Al-Anazi and Gates (2010). The primary aim of SVR regression is to discover linear relation between n-dimensional input vectors $x \in R^n$ and output variables $y \in R$ as follows:

$$f(x) = w^T x + b \, , \tag{7}$$

where w and b are the slope and offset of the regression line, respectively. To find the relation between n-dimensional input vectors, the values of regression parameters (w and b) must be determined. For attaining the aforementioned objective, minimizing of the following equation is essential

$$R = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{l}\left|y_i - f(x_i)\right|_\varepsilon \, . \tag{8}$$

The loss function used in this strategy is ε-insensitive. This loss function, introduced by Vapnik (1995), is expressed by the following equation:

$$\left|y_i - f(x_i)\right|_\varepsilon = \begin{cases} 0 & \text{if} \quad \left|y_i - f(x_i)\right| \le \varepsilon \\ \left|y_i - f(x_i)\right| - \varepsilon & \text{otherwise} \end{cases} \tag{9}$$

This problem can be reformulated in a dual space by

Maximize

$$L_p\left(\alpha_i, \alpha_i^*\right) = -\frac{1}{2}\sum_{i,j=1}^{l}\left(\alpha_i - \alpha_i^*\right)\left(\alpha_j - \alpha_j^*\right)x_i^T x_j - \varepsilon\sum_{i=1}^{l}\left(\alpha_i + \alpha_i^*\right) + \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right)y_i$$

subject to

$$\begin{cases} \sum_{i=1}^{l}\left(\alpha_i - \alpha_i^*\right) = 0 \\ 0 \le \alpha_i \le C, \quad i = 1,\dots,l \\ 0 \le \alpha_i^* \le C, \quad i = 1,\dots,l \end{cases} \tag{10}$$

After calculation of Lagrange multipliers, α_i and α_i^*, training data points from which those meeting the conditions $\alpha_i - \alpha_i^* \neq 0$ will be employed to construct the decision function. The total number of the points with prior criteria will be considered as the number of support vectors. Hence, the best linear hyper surface regression is given by:

$$f(x) = w_o{}^T x + b = \sum_{i=1}^{l} \left(\alpha_i - \alpha_i{}^* \right) x_i{}^T x + b \tag{11}$$

in which the desired weight vector of the regression hyper plane is given by:

$$w_o = \sum_{i=1}^{l} \left(\alpha_i - \alpha_i{}^* \right) x_i \ . \tag{12}$$

In the case of the nonlinear regression, learning problem is again formulated in the same way as in the linear case. The only difference between linear and nonlinear regression is the implanting of kernel function in regression function. Hence the nonlinear hyperplane regression function becomes:

$$f(x) = \sum_{i=1}^{l} \left(\alpha_i - \alpha_i{}^* \right) K\left(x_i, x \right) + b \ . \tag{13}$$

In above equation, $K(x_i, x)$ is kernel function which is defined as follows:

$$k\left(x_i, x_j \right) = \Phi^T \left(x_i \right) \Phi \left(x_j \right) \qquad i, j = 1, \dots, l \ , \tag{14}$$

where, $\Phi(x_i)$ and $\Phi(x_j)$ are projection of the x_i and x_j in feature space, respectively.

2.3 Hybrid genetic algorithm-pattern search technique

Genetic algorithm (GA) is an optimization approach which starts with a random population of chromosome-like solutions and evolves to better solutions by applying genetic operations. Genetic algorithm discovers the global minimum of fitness function (function which its global minimum is desired). Therefore, a function meant to be solved should be rearranged such that the global minimum of the rearranged function and the desired point of original function are the same. Evaluation of each chromosome (solution) produces the corresponding fitness score which in turn is used for selection procedure and forming the succeeding population after applying genetic operations. This process continues until the desired chromosome is achieved. For better performance of genetic algorithm, a pattern search technique is integrated with GA. This means that, after each generation, all chromosomes are enhanced by means of pattern search technique. In the pattern search technique, the algorithm searches a set of points, called a mesh, around the current chromosome. The mesh is formed by adding the current chromosome to a scalar multiple of a set of vectors called a pattern. After assessment of all points according to fitness function, the best solution in the mesh is re-

placed by current chromosome. Before the run of GA-PS, the number of regulation parameters must be adjusted. These parameters include population type, population size, initial range, scaling function, selection function, elite preservation, crossover fraction, mutation function, crossover function, hybrid function, generations, stall generations, fitness tolerance, and time limit. Population type specifies the data type of the input to the fitness function. Population size determines the number of individuals which are in each generation. Initial range limits the range of the points in the initial population through setting the lower and upper bounds. Scaling function changes the raw fitness determined by virtue of the fitness function to values in a range of that is fit for the selection function. Selection function specifies how the genetic algorithm chooses parents for the next generation. Elite preservation specifies the number of individuals that are guaranteed to survive to the next generation. Crossover fraction specifies the fraction of the next generation. Mutation function specifies how the genetic algorithm makes small random changes in the individuals in the population to create mutation children. Crossover function specifies how the genetic algorithm combines two individuals, or parents, to form a crossover child for the next generation. Hybrid function is another minimization function that runs after the genetic algorithm terminates. Generations specify the maximum number of iterations for the genetic algorithm to perform. This algorithm stops if the weighted average change in the fitness function value over stall generations is less than function tolerance. The algorithm runs until the cumulative change in the fitness function value over stall generations is less than or equal to function tolerance. The algorithm stops if there is no improvement in the best fitness value for an interval of time in seconds specified by time limit. More details about GA-PS tool are available in MATLAB user's guide (MATLAB User's Guide 2011), Mohaghegh (2000), Asoodeh and Bagheripour (2012b, 2013c), Asoodeh *et al.* (2014c).

3.   INPUT/OUTPUT DATA SPACE

Generalization of intelligence based model is mainly a function of range of dataset employed for its construction. Moreover, owing to infeasibility to incorporate *a priori* knowledge into this group of models, its performance depends on the reliability of data employed for its construction. Hence, gathering of data is an important step in model development using intelligence based model. Dataset which employed in current study for building predictive model for estimation of formation volume factor of crude oil from production data is borrowed from papers available in literature (Al-Marhoun 1988, Bello *et al.* 2008, Dokla and Osman 1990, Mahmood and Al-Marhoun 1996, Moghadam *et al.* 2011, Obomanu and Okpobiri 1987, Omar and Todd 1993). The dataset consists of production data (reservoir temperature, solu-

tion gas oil ratio, reservoir oil gravity (API), and dissolved gas relative density) and the corresponding value of FVF. Out of 561 data points, 342 data points are used for training of model and 219 data points are employed for evaluating the reliability of constructed model. Statistical description of employed dataset is given in Table 1. As seen in the table, crude oils studied herein have a wide range of reservoir and production conditions.

Table 1

Statistical description of employed dataset

Parameter	Minimum	Maximum	Average
Solution gas oil ratio (SCF/STB)	169.53	1608.26	701.43
Dissolved gas relative density [%]	0.91	1.71	1.19
Reservoir oil gravity [^{o}API]	19.30	43.58	27.92
Reservoir temperature [^{o}F]	90.00	260.00	181.37
Formation volume factor [bbl/STB]	1.15	2.01	1.42

4. RESULTS AND DISCUSSION

4.1 ACE model

At the first stage of this study, an alternating conditional expectation algorithm is employed to construct a model meant to estimate oil formation volume factor from available PVT data. ACE transforms input/output data such that error variance between transformed output (FVF) and sum of transformed input data (PVT data) is minimized. After a nonparametric transformation of each input/output dataset is evaluated, a functional form is approximated to each transformation by use of a simple curve fitting tool. Optimal transformation of PVT data (inputs) and FVF (output) is depicted in Fig. 2. In the next step, sum of transformed input data is evaluated. Once again, a simple curve fitting between output and sum of transformations produces functional form for estimating formation volume factor from PVT data. To evaluate performance of constructed ACE model, unseen test data were input to it and FVF was estimated. Figure 3 shows crossplot between actual FVF and ACE predicted FVF along with residual of predictions. This figure indicated that ACE was successful in estimation of FVF.

4.2 SVR model

At the next stage of this study, an epsilon support vector regression algorithm was used for formulating available PVT data to FVF. Firstly, all available data were scaled in range of [–1 1] and subsequently all data were transformed to feature space using radial basis kernel function (RBF) owing
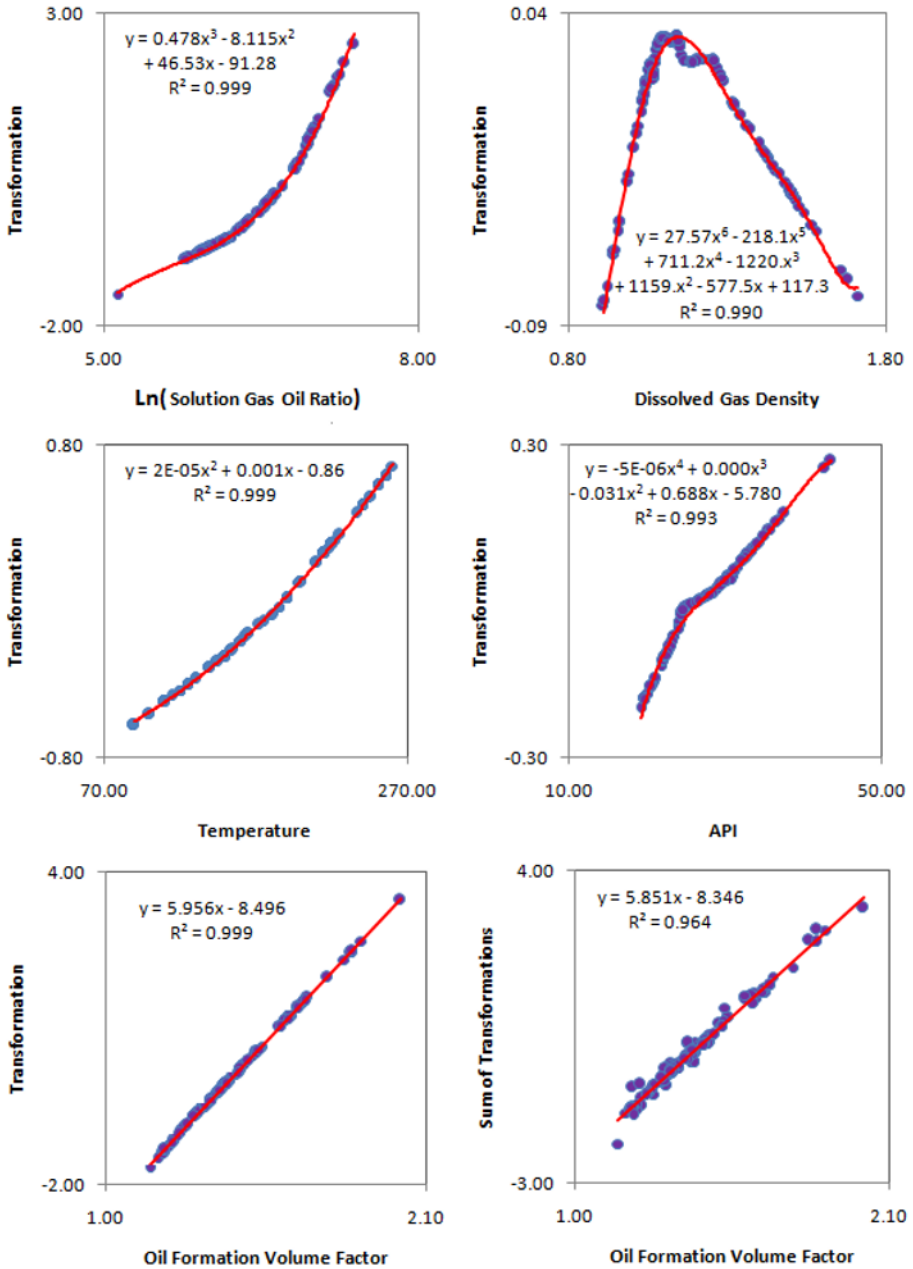
Fig. 2. Optimal transformation of input/output data and crossplot of sum of transformations vs. oil FVF.
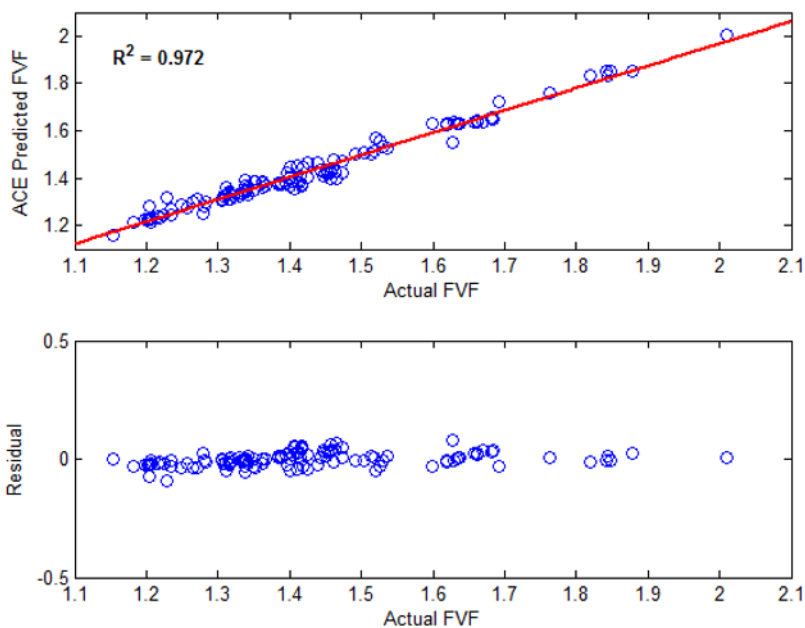
Fig. 3. Graph evaluating the performance of ACE model *versus* correlation coefficient and residuals of prediction.
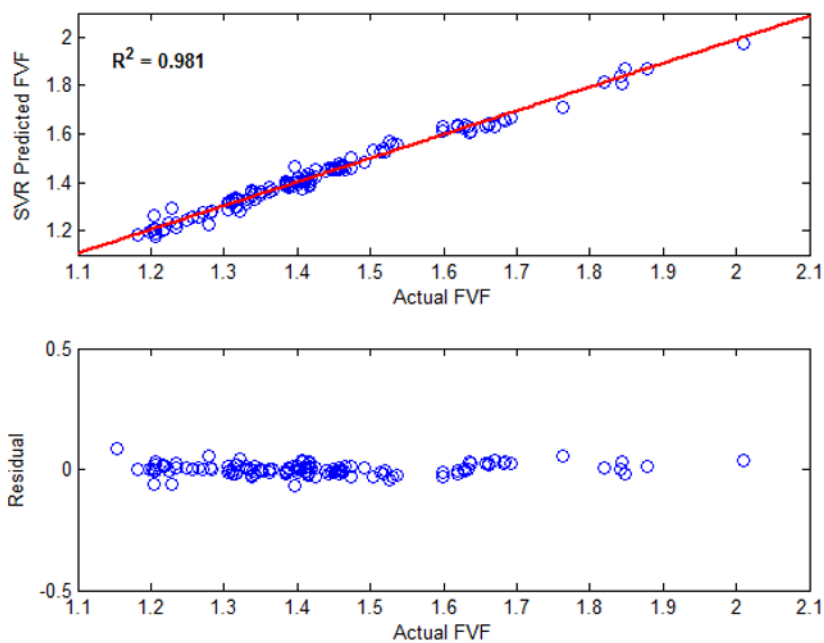

Fig. 4. Graph evaluating the performance of SVR model *versus* correlation coefficient and residuals of prediction.

Table 2

The parameter values corresponding to the optimum ε-SVR model

	C	Gamma	Epsilon
Search ranges	0.01-1000000	0.000001-20	0.0001-100
Optimum value for ε-SVR model	2456.650274	0.0567989	0.0023314

to fewer parameters to be tuned and low computational cost (Keerthi and Lin 2003). Parameters involved in SVR and kernel function (*i.e.*, epsilon, Gamma, and C) were determined through a thorough surveying using combination of grid search and pattern search techniques, as You *et al.* (2014) suggested. Optimum values of these parameters are shown in Table 2. In the present model, 342 Lagrange multiplier pairs were employed where 311 support vectors among them were used for model construction. After the SVR model was built, unseen test data were used for blind testing performance of SVR model. Figure 4 shows assessment of SVR model using concepts of correlation coefficient and residuals. This figure proves there is a satisfying match between predicted and actual FVF.

4.3 Committee machine with ACE and SVR

In the latter stage of present study, a committee machine with ACE and SVR models was constructed to combine their outputs in an optimal linear structure. Committee machine reaps the benefits of both ACE and SVR models through assigning a weight of contribution to each model such that accuracy of final prediction is enhanced. Therefore, estimated FVF from committee machine will be simply of the following form:

$$FVF_{CM} = w_1 \times SVR + w_2 \times ACE \ . \tag{15}$$

To extract the optimal weight of contribution of each model (w_1 and w_2), MSE function of committee machine was introduced to hybrid genetic algorithm-pattern search technique. The GA-PS tool starts with a population of randomly generated pairs of probable solutions (pairs of w_1 and w_2) in a chromosome-like structure. By applying different genetic operators over each population, a new generation of enhanced chromosomes (solutions) is obtained. This process continues until the desired pair of (w_1 w_2) is achieved. This process is shown in Fig. 5. Regulations of genetic algorithm before running are shown in Table 3. Performance of committee machine with ACE and SVR is assessed using concepts of correlation coefficient, relative error and residual analysis of prediction (Fig. 6). High value of correlation coeffi-
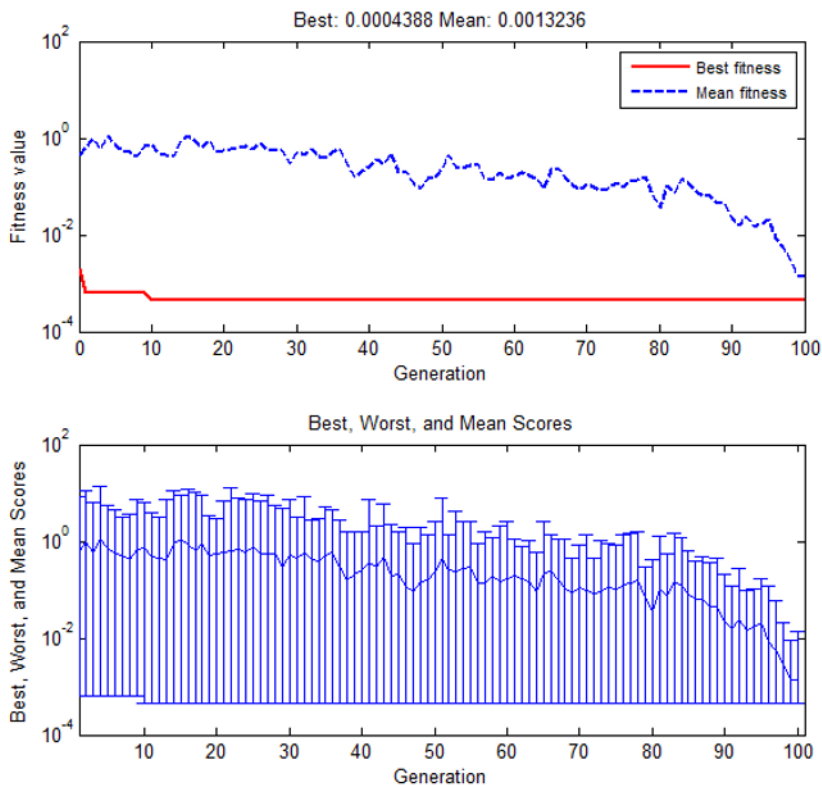
Fig. 5. Graph showing mean, best and worst fitness scores of FVF fitness function during 100 generations.

Table 3

Regulations done before the run of genetic algorithm

Parameter/setting	Type/value	Parameter/setting	Type/value
Population type	Double vector	Mutation function	Gaussian
Population size	20 chromosomes	Crossover function	Scattered
Initial range	[-1 1]	Hybrid function	Pattern search
Scaling function	Proportional	Generations	100
Selection function	Roulette	Stall generations	100
Elite preservation	2	Fitness tolerance	1.0 E -6
Crossover fraction	0.85	Time limit	Infinity

cient, low value of relative errors, low value of residuals, and concentration of residuals for most samples in close proximity of zero are evidences of supreme performance of committee machine.
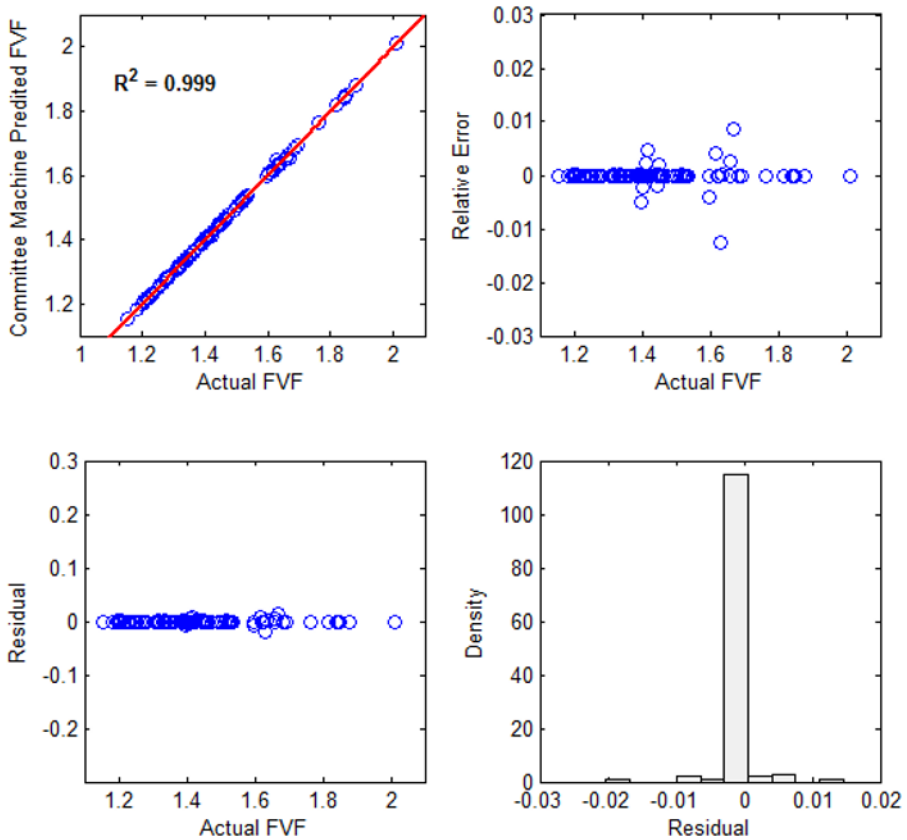
Fig. 6. Graph evaluating the performance of committee machine model using concepts of correlation coefficient, relative error, and residual analysis of prediction.

4.4 Comparison of models

Zargar *et al.* (2014) proposed a fuzzy logic model for estimating FVF from available PVT data. Performing an error distribution analysis, they concluded that the fuzzy model is an accurate model for FVF estimation. Bagheripour *et al.* (2013) made a comparison between traditional and stochastically optimized neural network for estimation of FVF. They showed that the use of genetic algorithm instead of back-propagation algorithm enhances the precision of overall FVF prediction. In this section, results of the mentioned works are compared with models presented in current study. Table 4 provides an opportunity to compare MSE and *R*-square factors of different models. Results show that committee machine surpasses other methods and provides more reliable results relative to Zargar *et al.* (2014), Bagheripour *et al.* (2013), and individual SVR and ACE models.

Table 4

Comparing different models for estimation of FVF *versus* *R*-square and MSE

Method	Results	*R*-square	MSE
Present study	Committee machine	0.999	0.0004389
	SVR	0.981	0.00051017
	ACE	0.972	0.00081161
Zargar *et al.*	Fuzzy Logic	0.992	0.0004767
Bagheripour *et al.*	Optimized neural network	0.982	0.0004822
	Neural network	0.931	0.0022669

5. CONCLUSIONS

This study proposed an accurate, cheap and rapid way for estimating oil formation volume factor from available PVT data. In situations where sampling is not applicable owing to the fact that the producing life of reservoir is too long or in situations where sampling are not desired on account of costs and time-consumption, the proposed strategy is an appealing alternative. ACE and SVR are featured to exactly extract underlying dependency between oil formation volume factor and available PVT data. Genetic algorithm is a sophisticated approach for taking part as combiner of committee machine. It is capable of finding optimal linear combination of SVR and ACE models for enhancing accuracy of final prediction. In situations where multiple options are available for solving a problem, committee machine is a great idea for enhancing accuracy of final prediction by little additional computation. Comparing current study with previous ones revealed superiority of committee machine using concepts of *R*-square and MSE.

References

Afshar, M., A. Gholami, and M. Asoodeh (2014), Genetic optimization of neural network and fuzzy logic for oil bubble point pressure modeling, *Korean. J. Chem. Eng*. **31**, 3, 496-502, DOI: 10.1007/s11814-013-0248-8.

Ahmed, T. (2000), *Reservoir Engineering Handbook*, 2th ed., Gulf Professional Publishing, Burlington, 863 pp.

Al-Anazi, A.F., and I.D. Gates (2010), Support vector regression for porosity prediction in a heterogeneous reservoir: A comparative study, *Comput. Geosci.* **36**, 12, 1494-1503, DOI: 10.1016/j.cageo.2010.03.022.

Al-Marhoun, M.A. (1988), PVT correlations for Middle East crude oils, *J. Petrol. Technol.* **40**, 5, 650-666, DOI: 10.2118/13718-PA.

Al-Marhoun, M.A., and E.A. Osman (2002), Using artificial neural networks to develop new PVT correlations for Saudi crude oils, SPE Paper 78592, DOI: 10.2118/78592-MS.

Almehaideb, R.A. (1997), Improved PVT correlations for UAE crude oils, SPE Paper 26644, DOI: 10.2118/37691-MS.

Al-Shammasi, A.A. (1999), Bubble Point pressure and oil formation volume factor correlations, SPE Paper 53185.

Asoodeh, M. (2013), Prediction of Poisson's ratio from conventional well log data: A committee machine with intelligent systems approach, *Energy Sources A* **35**, 10, 962-975, DOI: 10.1080/15567036.2011.557693.

Asoodeh, M., and P. Bagheripour (2012a), Estimation of bubble point pressure from PVT data using a power-law committee with intelligent systems, *J. Pet. Sci. Eng.* **90-91**, 1-11, DOI: 10.1016/j.petrol.2012.04.021.

Asoodeh, M., and P. Bagheripour (2012b), Prediction of compressional, shear, and stoneley wave velocities from conventional well log data using a committee machine with intelligent systems, *Rock. Mech. Rock. Eng.* **45**, 1, 45-63, DOI: 10.1007/s00603-011-0181-2.

Asoodeh, M., and P. Bagheripour (2013a), Neuro-fuzzy reaping of shear wave velocity correlations derived by hybrid genetic algorithm-pattern search technique, *Open Geosci.* **5**, 2, 272-284, DOI: 10.2478/s13533-012-0129-4.

Asoodeh, M., and P. Bagheripour (2013b), Fuzzy classifier based support vector regression framework for Poisson's ratio determination, *J. Appl. Geophys.* **96**, 7-10, DOI: 10.1016/j.jappgeo.2013.06.006.

Asoodeh, M., and P. Bagheripour (2013c), Core porosity estimation through different training approaches for neural network: Back-propagation learning vs. genetic algorithm, *Int. J. Comput. Appl.* **63**, 5, 11-15, DOI: 10.5120/10461-5172.

Asoodeh, M., and K. Kazemi (2013), Estimation of bubble point pressure using a genetic integration of empirical formulas, *Energy Sources A* **35**, 12, 1102-1109, DOI: 10.1080/15567036.2011.574195.

Asoodeh, M., A. Gholami, and P. Bagheripour (2014a), Asphaltene precipitation of titration data modeling through committee machine with stochastically optimized fuzzy logic and optimized neural network, *Fluid. Phase Equilibr.* **364**, 67-74, DOI: 10.1016/j.fluid.2013.12.016.

Asoodeh, M., A. Gholami, and P. Bagheripour (2014b), Oil-CO_2 MMP determination in competition of Neural Network, support vector regression and committee machine, *J. Disper. Sci. Technol.* **35**, 4, 564-571, DOI: 10.1080/01932691.2013.803255.

Asoodeh, M., A. Gholami, and P. Bagheripour (2014c), Renovating scaling equation through hybrid genetic algorithm-pattern search tool for asphaltene precipitation modeling, *J. Disper. Sci. Technol.* **35**, 4, 607-611, DOI: 10.1080/01932691.2013.825209.

Bagheripour, P., and M. Asoodeh (2013), Fuzzy ruling between core porosity and conventional well logs: subtractive clustering vs. genetic algorithm-pattern search, *J. Appl. Geophys.* **99**, 35-41, DOI: 10.1016/j.jappgeo.2013.09.014.

Bagheripour, P., and M. Asoodeh (2014), Genetic implanted fuzzy model for water saturation determination, *J. Appl. Geophys.* **103**, 232-236, DOI: 10.1016/j.jappgeo.2014.02.002.

Bagheripour, P., M. Asoodeh, and A. Asoodeh (2013), Oil formation volume factor modeling: Traditional NN vs. stochastically optimized neural network, *Open Geosci.* **5**, 4, 508-513, DOI: 10.2478/s13533-012-0154-3.

Bagheripour, P., A. Gholami, and M. Asoodeh (2014), Support vector regression between PVT data and bubble point pressure, *J. Petrol. Explor. Prod. Technol.* **5**, 3, 227-231, DOI: 10.1007/s13202-014-0128-8.

Bagheripour, P., A. Gholami, M. Asoodeh, and M. Vaezzadeh-Asadi (2015), Support vector regression based determination of shear wave velocity, *J. Petrol. Sci. Eng.* **125**, 95-99, DOI: 10.1016/j.petrol.2014.11.025.

Bello, O., K. Reinicke, and P. Patil (2008), Comparison of the performance of empirical models used for the prediction of the PVT properties of crude oils of the Niger delta, *Petrol. Sci. Technol.* **26**, 5, 593-609, DOI: 10.1080/10916460701204685.

Breiman, L., and J.H. Friedman (1985), Estimating optimal transformations for multiple regression and correlation, *J. Am. Stat. Assoc.* **80**, 391, 580-598, DOI: 10.1080/01621459.1985.10478157.

Dake, L.P. (1988), *Fundamental of Reservoir Engineering*, 17th ed., Elsevier Science.

Dindoruk, B., and P.G. Christman (2001), PVT properties and viscosity correlations for Gulf of Mexico oils, SPE Paper 26644, DOI: 10.2118/71633-MS.

Dokla, M.E., and M.E Osman (1990), Correlation of PVT properties for U.A.E. crudes.

Dutta, S., and J.P. Gupta (2010), PVT correlations for Indian crude using artificial neural networks, *J. Petrol. Sci. Eng.* **72**, 1-2, 93-109, DOI: 10.1016/j.petrol.2010.03.007.

El-Banbi, A.H., K.A. Fattah, and M.H. Sayyouh (2006), New modified black-oil correlations for gas condensate and volatile oil fluids, SPE Paper 26644, DOI: 10.2118/102240-MS.

Elmabrouk, S., A. Zekri, and E. Shirif (2010), Prediction of bubble point pressure and bubble point oil formation volume factor in the absence of PVT analysis, SPE Paper 26644.

Elsharkawy, A.M. (1998), Modeling the properties of crude oil and gas system using RBF network, SPE Paper 26644, DOI: 10.2118/49961-MS.

Elsharkawy, A.M., and R.B.C. Gharbi (2000), Comparing classical and neural regression techniques in modeling crude oil viscosity, *Adv. Eng. Softw.* **32**, 3, 215-224, DOI: 10.1016/S0965-9978(00)00083-1.

Farshad, F.F., J.L. Leblance, J.D. Garber, and J.G. Osorio (1996), A new correlation for bubble point pressure according to the separator conditions, SPE Paper 26644.

Fattahi, H., A. Gholami, S. Amiribakhtiar, and S. Moradi (2014), Estimation of asphaltene precipitation from titration data: a hybrid support vector regression with harmony search, *Neural Comput. Appl.* **26**, 4, 789-798, DOI: 10.1007/s00521-014-1766-y.

Gharbi, R.B., and A.M. Elsharkawy (1996), Neural network model for estimating the PVT properties of Middle East crude oils, SPE Paper 56850.

Gholami, A., M. Asoodeh, and P. Bagheripour (2014a), Fuzzy assessment of asphaltene stability in crude oils, *J. Dispersion. Sci. Technol.* **35**, 4, 556-563, DOI: 10.1080/01932691.2013.800457.

Gholami, A., M. Asoodeh, and P. Bagheripour (2014b), Smart determination of difference index for asphaltene stability evaluation, *J. Dispersion. Sci. Technol.* **35**, 4, 572-576, DOI: 10.1080/01932691.2013.805654.

Gholami, A., M. Asoodeh, and P. Bagheripour (2014c), How committee machine with SVR and ACE estimates bubble point pressure of crudes, *Fluid Phase Equilibr.* **382**, 139-149, DOI: 10.1016/j.fluid.2014.08.033.

Gholami, A., S. Moradi, M. Asoodeh, P. Bagheripour, and M. Vaezzadeh-Asadi (2014d), Asphaltene precipitation modeling through ACE reaping of scaling equations, *Sci. Chin. Chem.* **57**, 12, 1774-1780, DOI: 10.1007/s11426-014-5253-1.

Glaso, O. (1980), Generalized pressure-volume-temperature correlations, *J. Petrol. Technol.* **34**, 85-95, DOI: 10.2118/8016-PA.

Hemmati, M.N., and R. Kharrat (2007), A correlation approach for prediction of crude oil PVT properties, SPE Paper 26644, DOI: 10.2118/104543-MS.

Katz, D.L. (1942), Prediction of the shrinkage of crude oils, SPE Paper 26644.

Kazemi, K., S. Moradi, and M. Asoodeh (2013), A neural network based model for prediction of saturation pressure from molecular components of crude oil, *Energy Source A* **35**, 11, 1039-1045, DOI: 10.1080/15567036.2011.584127.

Keerthi, S.S., and C.J. Lin (2003), Asymptotic behavior of support vector machines with Gaussian kernel, *Neural Comput.* **15**, 7, 1667-1689, DOI: 10.1162/089976603321891855.

Knopp, C.R., and L.A. Ramsey (1960), Correlation of oil formation volume factor and solution gas-oil ratio, *J. Petrol. Technol.* **12**, 8, 27-29, DOI: 10.2118/1433-G.

Mahmood, M.A., and M.A. Al-Marhoun (1996), Evaluation of empirically derived PVT properties for Pakistani crude oils, *J. Petrol. Sci. Eng.* **16**, 4, 275-290, DOI: 10.1016/S0920-4105(96)00042-3.

MATLAB User's Guide (2011), Fuzzy logic, neural network & GA and direct search toolboxes, MATLAB CD-rom, Mathworks, Inc.

Moghadam, J.N., K. Salahshoor, and R. Kharrat (2011), Introducing a new method for predicting PVT properties of Iranian crude oils by applying artificial neural networks, *Petrol Sci. Technol.* **29**, 10, 1066-1079, DOI: 10.1080/10916460903551040.

Mohaghegh, S. (2000), Virtual-intelligence applications in petroleum engineering: Part 2 – evolutionary computing, *J. Petrol. Technol.* **52**, 10, 40-46, DOI: 10.2118/61925-JPT.

Na'imi, S.R., A. Gholami, and M. Asoodeh (2014), Prediction of crude oil asphaltene precipitation using support vector regression. *J. Disper. Sci. Technol.* **35**, 4, 518-525, DOI: 10.1080/01932691.2013.798585.

Obomanu, D.A., and G.A. Okpobiri (1987), Correlating the PVT properties of Nigerian crudes, *J. Energy Resour. Technol.* **109**, 4, 214-217, DOI: 10.1115/1.3231349.

Omar, M.I., and A.C. Todd (1993), Development of new modified black oil correlations for Malaysian crudes, SPE-25338-MS, Asia Pacific Oil and Gas Conference, 8-10 February, Singapore, DOI: 10.2118/25338-MS.

Petrosky, J., and F. Farshad (1993), Pressure volume temperature correlation for the Gulf of Mexico, SPE, Annual Technical Conference and Exhibition, 3-6 October, Houston, USA, SPE-26644-MS, DOI: 10.2118/26644-MS.

Rafiee-Taghanaki, S., M. Arabloo, A. Chamkalani, M. Amani, M.H. Zargari, and M.R. Adelzadeh (2013), Implementation of SVM framework to estimate PVT properties of reservoir oil, *Fluid Phase Equilibr.* **346**, 25-32, DOI: 10.1016/j.fluid.2013.02.012.

Shokir, E.M. (2007), CO_2-oil minimum miscibility pressure model for impure and pure CO_2 streams, *J. Petrol. Sci. Eng.* **58**, 1-2, 173-185, DOI: 10.1016/j.petrol.2006.12.001.

Vapnik, V. (1995), *The Nature of Statistical Learning Theory*, Springer, New York.

Vazquez, M., and H.D. Beggs (1970), Correlation for fluid physical property prediction, *J. Petrol. Technol.* **32**, 06, 103-107, DOI: 10.2118/6719-PA.

You, Z., Z. Yin, K. Han, D. Huang, and X. Zhou (2014), A semi-supervised learning approach to predict synthetic genetic interactions by combining functional and topological properties of functional gene network, *BMC Bioinformatics* **11**, 343-355, DOI: 10.1186/1471-2105-11-343.

Zargar, G., P. Bagheripour, and M. Asoodeh (2014), Fuzzy modeling of volume reduction of oil due to dissolved gas run off and pressure release, *J. Petrol. Explor. Prod. Technol.* **4**, 4, 439-442, DOI: 10.1007/s13202-014-0099-9.