International Journal of Computational Intelligence and Applications
Vol. 10, No. 3 (2011) 269–293

DOI: 10.1142/S1469026811003100

Imperial College Press
www.icpress.co.uk

SUPPORT VECTOR REGRESSION AND FUNCTIONAL NETWORKS FOR VISCOSITY AND GAS/OIL RATIO CURVES ESTIMATION

AMAR KHOUKHI[*,‡], MUNIRUDEEN OLOSO[*,§], MOSTAFA ELSHAFEI[*,¶], ABDULAZEEZ ABDULRAHEEM[†,‖] and ABDULAZIZ AL-MAJED[†,**]

[*]*Systems Engineering Department*
King Fahd University of Petroleum and Minerals, Dhahran, KSA

[†]*Petroleum Engineering Department*
King Fahd University of Petroleum and Minerals, Dhahran, KSA
[‡]*amar@kfupm.edu.sa*
[§]*ajadioloso@yahoo.co.uk*
[¶]*elshafei@kfupm.edu.sa*
[‖]*aazeez@kfupm.edu.sa*
[**]*aamajed@kfupm.edu.sa*

Received 10 October 2010
Revised 22 December 2010

In oil and gas industry, prior prediction of certain properties is needed ahead of exploration and facility design. Viscosity and gas/oil ratio (GOR) are among those properties described through curves with their values varying over a specific range of reservoir pressures. However, the usual single point prediction approach could result into curves that are inconsistent, exhibiting scattered behavior as compared to the real curves. Support Vector Regressors and Functional Networks are explored in this paper to solve this problem. Inputs into the developed models include hydrocarbon and non-hydrocarbon crude oil compositions and other strongly correlating reservoir parameters. Graphical plots and statistical error measures, including root mean square error and average absolute percent relative error, have been used to evaluate the performance of the models. A comparative study is performed between the two techniques and with the conventional feed forward artificial neural networks. Most importantly, the predicted curves are consistent with the shapes of the physical curves of the mentioned oil properties, preserving the need of such curves for interpolation and ensuring conformity of the predicted curves with the conventional properties.

Keywords: Reservoir characterization; viscosity; gas/oil ratio (GOR); Artificial Neural Networks; Support Vector Regressors; Functional Networks.

1. Introduction

1.1. *Motivation*

Reservoir fluid properties are very important in petroleum engineering computations such as material balance calculations, well test analysis, reserve estimation, inflow performance calculations, fluid flow in porous media, evaluation of new formation for potential development, numerical reservoir simulations, design

of production equipment, and planning future enhanced oil recovery projects. At every stage of the petroleum exploration and production business, *a priori* knowledge of how the fluids will behave under a wide range of pressure and temperature conditions, particularly in terms of their volumetric and thermo physical properties, is required. The relationships of Pressure-Volume-Temperature (PVT) properties for oil and gas are traditionally estimated using empirical studies. Ideally, those properties could be measured in laboratories. The problem with those measurements is the availability of those laboratory tests and the right samples collected from the well-bore or well surface. The need for the prediction of PVT properties was then accomplished through the use of equations of states (EOS), which are derived from the basic mass, energy, and chemical balance equations. Because the EOS are derived for pure substances, correction factor(s) are always added when used on practical data. To overcome these problems, empirically derived correlations between those properties and well data had been developed based on available data for different regions in the world. During the last several years, neural networks have been used to obtain better prediction models than the empirical ones and they have shown a significant prediction improvement.

Unfortunately, the developed neural networks correlations are often limited and global correlations are usually less accurate compared to local correlations. Nevertheless, the achievements of neural networks opened the door to computational intelligence techniques to play a major role in oil and gas industry. To improve prediction accuracy, computational intelligence techniques, such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Adductive Networks and Genetic Algorithms (GA) among others, have been applied.

Some of these properties described as curves are estimated through single or multi-data point prediction. However, the usual single or multi-data point predictions could compromise the original shape of the curves. Henceforth, prediction techniques for entire curves are needed to elaborate for some of these properties.

Two of such PVT properties that need to be presented as curves are oil viscosity and gas oil ratio (GOR). These two properties vary with pressure and their prediction is defined over a specified range of pressures. Fluid viscosity is a measure of its internal resistance to flow. This property is highly needed for many calculations and applications in the petroleum industry such as oil recovery estimation, multi-phase flow calculation, gas-lifting and pipeline design. The GOR is basically determined through separator calculations. In this paper, two main advanced neural networks techniques are implemented for oil viscosity and gas oil ratio (GOR) curves prediction.

2. Related Works

2.1. *Viscosity correlations*

A good number of empirical correlations have been developed in the literature to estimate crude oil viscosity at, below and above bubble point pressure (P_b). P_b is

the pressure at which the light hydrocarbon components in the oil starts to change to the gas phase, and hence appear as bubbles in the oil sample. Correlations based on soft computing techniques have recently been developed to predict this important PVT property. Graphical correlations[1] were developed for predicting viscosity at different reservoir pressures using a data set from U.S.A. oil samples. The authors correlated the under-saturated oil viscosity with the viscosity at $P_b(\mu_{ob})$, and pressure above bubble point P_b, while using oil gravity and temperature of range 100–220° F to develop the dead oil (at the standard temperature and pressure) viscosity (μ_{od}) correlation. In the same vein, a graphical correlation[2] was presented for predicting oil viscosity at P_b. The correlation was developed as a function of GOR using viscosity data of 457 crude oil samples from Canada and U.S.A.

Correlations for μ_{ob} and μ_{od} were developed, with a data set of 2073 oil viscosity measurements used to develop the μ_{ob} correlation while 460 dead oil observations was used to develop μ_{od} correlation.[3] Another correlation was developed in Ref. 4, for μ_{od} in the temperature range of 50–300° F. The variable μ_{od} was correlated as a function of American Petroleum Institute (API) oil gravity scale and temperature. A large set of PVT measurements was used[5] to develop an under-saturated oil viscosity correlation as a function of μ_{ob}, P_b and reservoir pressure. Also, a μ_{od} correlation[6] based on a modification of Beggs and Robinson's correlation was presented. Correlations for μ_{ob} was used to predict oil viscosity below and above bubble point were developed in Ref. 7. Total data points of 150, 1503 and 1691 were used to develop the three correlations. The viscosity data used in the study were from Saudi crude oil. For μ_{ob} correlation, gas relative density, solution GOR, relative temperature and oil relative density, independent variables were used. For viscosity above and below P_b correlations, the correlating variables used were μ_{ob}, reservoir pressure and P_b. Correlations were also developed[8] for μ_{od}, μ_{ob} and under-saturated oil viscosity using light crude oil data of Libya. μ_{od} was correlated as a function of stock tank oil gravity and temperature. The μ_{ob} was correlated with API oil gravity, μ_{od} and P_b, while the under-saturated oil viscosity was correlated with pressure, P_b, μ_{ob}, μ_{od} and API oil gravity.

In Ref. 9, new empirical correlations for μ_{od}, μ_{ob} and under-saturated oil viscosity were suggested. The μ_{od} was correlated as a function of API oil gravity and reservoir temperature, μ_{ob} as a function of dead oil viscosity and solution GOR, and under-saturated oil viscosity as a function of μ_{ob}, P_b and reservoir pressure. A total of 126 laboratory PVT analyses from Texas and Louisiana in U.S. were used to develop oil viscosity correlations. Other correlations for μ_{ob} and under-saturated oil viscosity based on UAE crude oil were presented in Ref. 10. The correlated μ_{ob} with solution GOR, reservoir temperature, gas specific gravity and API oil gravity using 57 data points. An under-saturated oil viscosity correlation was developed as a function of P_b, μ_{ob}, reservoir pressure and solution GOR using 328 data points.

Other correlation functions were introduced in Ref. 11 for μ_{od}, μ_{ob} and under-saturated oil viscosity for Gulf of Mexico crude oils based on 100 PVT laboratory

reports. The authors correlated μ_{od} with temperature, pressure and solution GOR at P_b, and API oil gravity. The μ_{ob} was correlated with μ_{od} and solution GOR, while the under saturated oil viscosity was correlated with μ_{ob}, P_b, reservoir pressure and solution GOR.

In the same vein, a number of researchers have developed viscosity correlations using soft computing techniques. A Radial Basis Function Network model[12] was used for predicting oil viscosity using reservoir pressure and temperature, API oil gravity and gas gravity as the network inputs. A universal neural-network-based model for estimating PVT properties of crude oil systems was introduced in Ref. 13. Another ANN model[14] was developed using all the data points of a reservoir to obtain the viscosity curve. A PVT data of 650 reservoir fluids from around the world was used to develop the viscosity correlation model and also for some other PVT properties. An ANN correlation model[15] was introduced to predict brine viscosity using temperature and salinity as the network inputs. A total of 1040 data points were used to build the model. Another neural network was constructed[16] to predict viscosity below P_b for Pakistani crude oil. The correlating parameters were: pressure, reservoir temperature P_b, oil formation volume factor, solution GOR, gas specific gravity and API gravity. More recently, an oil viscosity correlation was presented in Ref. 17 for Iranian crude oil using genetic algorithms. The input parameters were the pressure, temperature, and reservoir fluid GOR and oil density. An excellent study had been reported recently,[18] implementing a support vector regression technique on PVT Correlations for Indian Crude oil.

In all abovementioned prediction techniques, including empirical correlations and soft computing models, data points are being predicted and the shapes of the resulting estimated curves may not be consistent with the experimental ones.

2.2. *Gas/oil ratio correlations*

Ordinarily, GOR correlations are derived from P_b correlation. However, a precedent was set in Ref. 5 where a regression analysis was used to obtain an empirical correlation for R_s. They used 5008 data points to perform the regression analysis. The solution GOR was correlated as a function of pressure, gas relative density, oil API gravity and temperature. Other researchers have developed general correlations for GOR including Refs. 11, 19 and 20. They have all developed separate empirical correlations for GOR.

Several P_b correlations had been published. Good reviews of these correlations can be found in Refs. 21–24. Generally, P_b correlation is developed as a function of solution GOR, gas specific gravity, oil API gravity and reservoir temperature. Though solution GOR correlation can be obtained from any of the existing P_b correlation, some possible complexities in solving such resulting solution GOR correlation have been observed in Ref. 11, when there is the need to have separate correlation for solution GOR.

3. PVT Data Acquisition and Pre-Processing

At the implementation phase, it is important to make sure that the input data values fall in a natural domain. Such a quality control step is a must to have very accurate and reliable results at the end.

The following are the most common domains for the input/output variables, gas–oil ratio, bubble point pressure, API oil gravity, relative gas density, reservoir temperature and oil formation volume factor that are used in both input and output layers of modeling schemes for PVT analysis (a nomenclature is provided in Appendix 1).

- Gas oil ratio which varies from 151 to 1332, scf/stb.
- Bubble point pressure, starting from 210, and ending with 2985, psia.
- Reservoir temperature with its range from 100°F to 250°F.
- API gravity which changes between 21.4 and 47.6.
- Gas relative density, changing from 0.744 to 1.367.
- Bubble point/gas-saturated oil viscosity 0.88 to 6.49 (cP).
- Dead oil viscosity varies from 0.305 to 1.91 (cP).

The implementation studies of the presented work were achieved based on three databases say, data sets A, B and C. Data set A consists of the hydrocarbon and non-hydrocarbon components, and some other properties of the crude oil. Data set B consists of the viscosity-pressure measurements to generate viscosity curves for the corresponding wells in data set A, while data set C consists of GOR pressure measurements to generate gas/oil ratio curves for corresponding wells in A. These data were taken from Middle East crude oil reservoirs.

Initially, there were 106 data points in the set A. In preprocessing the data, we applied two different outlier-detection methods on the data set A before utilizing it. These are Cook's distance method and Chauvenet's criterion.[25] The former method was implemented using STATISTICA software while details on the latter method can be found in Ref. 25. Only data points that were detected to be outliers by the two methods have been declared as outliers and removed.

Eventually, seven data points were declared as outliers. After the removal of the outliers from data set A, it was reduced to 99 data points. As recommended and is usually done, the predictors were normalized within the interval [0 1] using formula (1) below. This makes the input data dimensionless and ensures that the predictors are independent of the measurement units.

$$x_i^{\text{new}} = \frac{(x_i^{\text{old}} - \min(x_i))}{(\max(x_i) - \min(x_i))}, \quad (i = 1, 2, \ldots, n). \tag{1}$$

The data set A was then divided into training and testing sets. The training set consists of 70% (approx. 70 data points) while the testing set consists of 30% (approx. 29 points). The idea of curve prediction using artificial neural network

technique was first introduced in Ref. 26 using ANN, in Ref. 27 using RBF, and Refs. 28 and 29 using a hybrid ANN with differential evolution.

4. Problem Statement

Typical viscosity and GOR curves are shown in Figs. 1 and 2. Each curve represents the variation of viscosity or GOR for the corresponding oil well. The significance of such a curve is compromised, if single point or multi-point based prediction is utilized. Equations (2) and (3) can be used to represent any crude oil viscosity curve and Eq. (4) is used to represent GOR curves.

$$\mu = \mu_{od} + (\mu_{ob} - \mu_{od})\left(\frac{P - P_d}{P_b - P_d}\right)^{\beta} \quad \text{for } P < P_b, \tag{2}$$

$$\mu = \mu_{ob} + \alpha(P - P_b) \quad \text{for } P \geq P_b, \tag{3}$$

$$R_s = R_{sb}\left(\frac{P - P_b}{P_b - P_d}\right)^{\tau}, \tag{4}$$

where α and β are viscosity curve coefficients and τ is the fitting GOR curve coefficient. The statistical distribution of the fitting coefficients is shown in Table 1.

From Eqs. (2) and (3), three parameters μ_{ob}, α and β are needed to generate the viscosity curve. The first parameter μ_{ob} is determined from the laboratory PVT analyses while α and β are to be generated from the curve fitting. Two

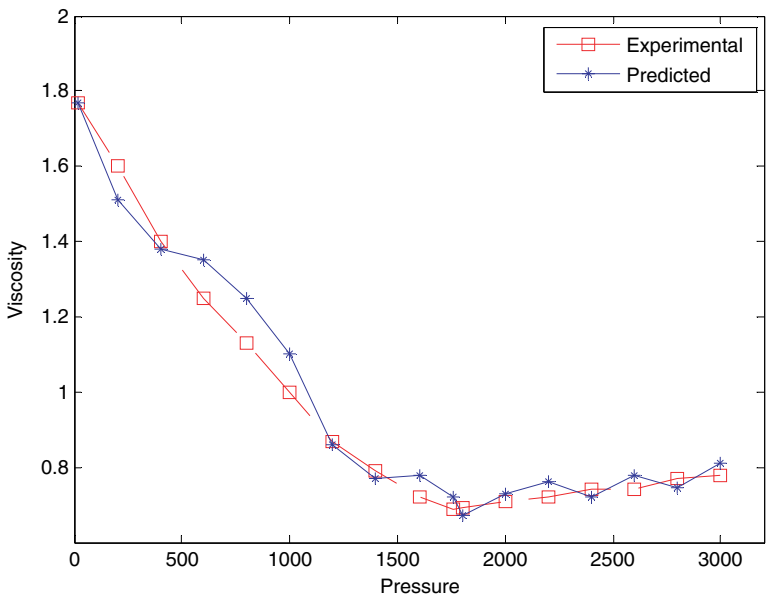

Fig. 1. Typical result from single or multi-data point prediction for viscosity curve.

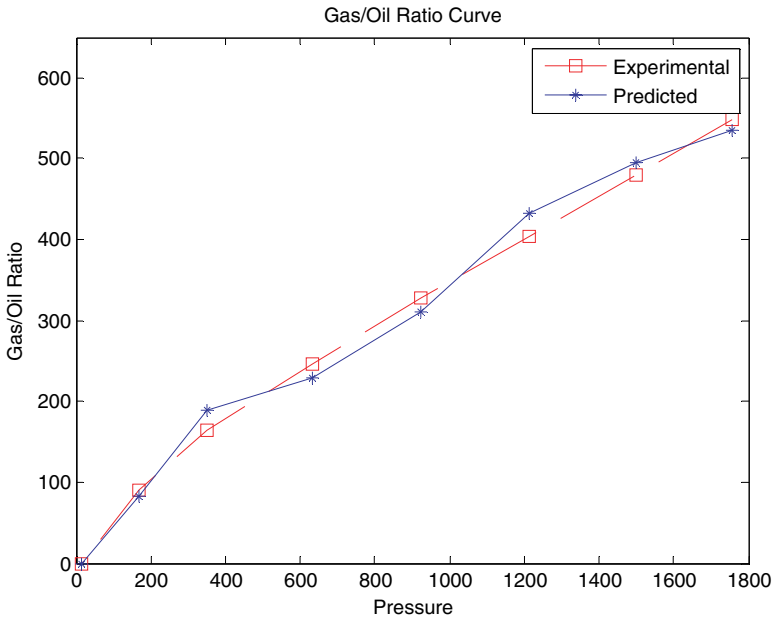

Fig. 2. Typical result from single or multi-data point prediction for GOR curve.

Table 1. Statistical distribution of the fitting coefficients.

Parameter	Max. Value	Min. Value
α	1.91E-04	1.48E-05
β	0.894	0.1251
τ	0.941657	0.42174

parameters, R_{sb} and τ are needed to generate GOR curves, R_{sb} is determined from PVT laboratory analyses while τ is obtained from curve fitting.

5. Approach

5.1. *Support vector regression*

Support vector machine modeling schemes and methods are among the most successful and effective algorithms in both machine learning and data mining communities. It has been widely used as a robust tool for classification and regression. An overview can be found in Refs. 30 and 31. Support Vector Regression (SVR) is a regression version of Support Vector Machines (SVMs) (Fig. 3). Unlike classification problems where the outputs are either 1 and 0 or 1 and -1, the outputs in the regression problems are real numbers. This makes it a bit difficult to model this type of information which has infinite possibilities. With the introduction of Vapnik's

ε-insensitive loss function, SVM has been extended to solve nonlinear regression estimation problems, leading to techniques known as SVR. These have been shown to exhibit excellent performance.[32] SVR has been found to be very robust to predict complex nonlinear relationship problems in many applications, including problems such as optical character recognition, text categorization, and face detection in images.[33] In the case of regression, a margin of tolerance $\in$ is set in approximation to the SVM which would have already being inferred from the problem. As shown in Figs. 3 and 4, SVMs map input vectors to a higher dimensional space, where a maximal separating hyperplane is constructed.[34–36]

The kernel function is responsible for transforming the data set into hyperplane. The variables of the kernel must be computed accurately since they determine the structure of high-dimensional feature space which governs the complexity of the final solution.

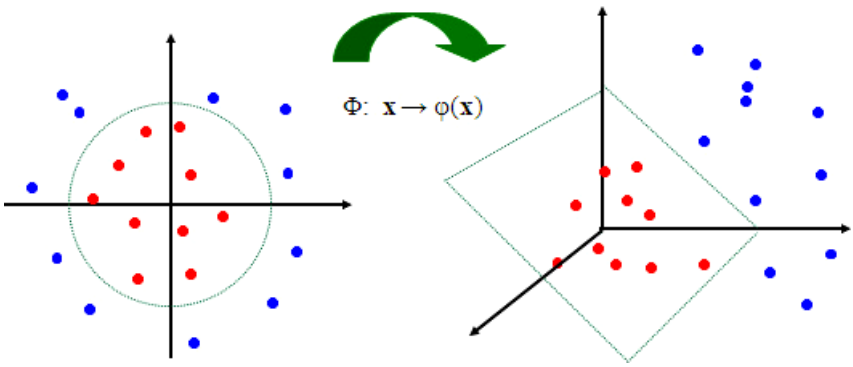

Fig. 3. The original input space mapped to a higher feature space with a separable training set.

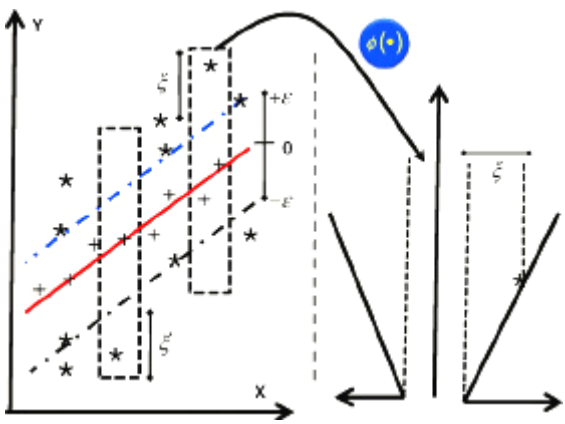

Fig. 4. Soft margin loss setting for a linear dimensional SVR (Schölkopf and Smola, 2002).

After the selection of a kernel, the other highly influential parameters in any SVR model based on the observation are "C" and "kernel option". For polynomial kernel, kerneloption denotes the degree of the kernel polynomial while it denotes kernel bandwidth for "Gaussian". "C" is the trade-off between achieving minimal training error and complexity of the model. The kernel functions and SVR options used in this study appear in Appendix 2.

5.2. *Functional networks*

Functional networks were introduced as a powerful alternative to neural networks.[37,38] Unlike neural networks, functional networks have the advantage of using domain knowledge in addition to data knowledge. The network initial topology can be derived based on the modeling of the properties of the real world. Once this topology is available, functional equations allow one to obtain a much simpler equivalent topology. Although functional networks can also deal with data only, the class of problems where functional networks are most convenient is the class where the two sources of knowledge about domain and data are available. In functional networks, neural functions are to be learned instead of weights (Figs. 5 and 6). To learn these neural functions, a set of linearly independent functions are to be used. These are called basis functions. Possible basis functions are: polynomial, exponential, Fourier and logarithm functions or their combinations. The selection of the basis function along with the possible learning method is essential in developing the FN model. To learn (parametric) functional networks, one can choose different sets of linearly independent functions for the approximation of the neuron functions.

At the same time, there is a need to select the best model according to some criterion of optimality. The Minimum Description Length Principle (MDLP) is one of the model selection principles we can use as discussed in Ref. 37. This consists of finding the minimum information required to store the given training set using a functional network model. Therefore, it was demonstrated that the best functional network model for a given problem corresponds to that with the minimum description length value.[38,39] The code length $\mathrm{L}(x)$ of x is defined as the amount of memory needed to store the information x.

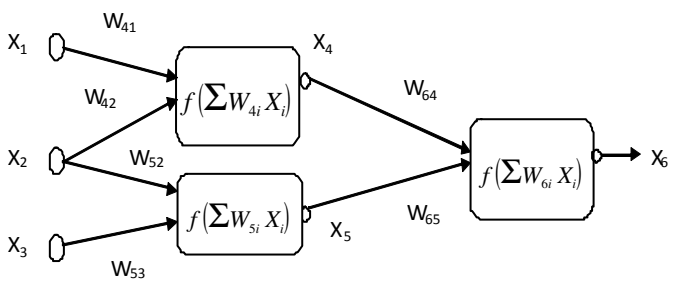

Fig. 5. A standard neural network.

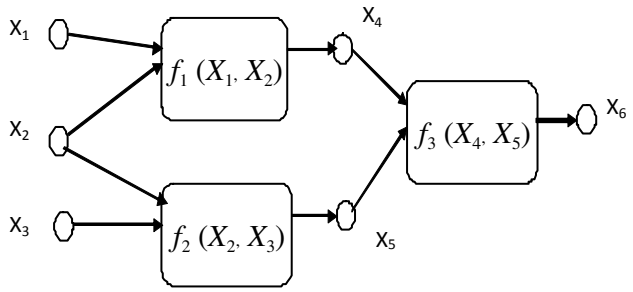

Fig. 6. A standard functional network.

The output is given by:

$$y = f_1(x_1) + f_2(x_2) + f_3(x_3) + \cdots + f_{12}(x_{12}). \tag{5}$$

For the case at hand, the different families of functions used for each parameter appear in Appendix 3. The MDLP was used to optimize the network and select the best model. It can be noted that in some cases, some functions are zero, this means that the corresponding input to that node does not really affect the predicting output at that instance.

5.3. *Statistical quality measures*

The performance and accuracy of SVR and FN as well as a feed-forward neural network (FFNN) are compared. In doing so, two commonly used statistical techniques have been adopted along with the graphical plots of the predicted curves (only sample plots are shown here for comparison). These are the root mean square error (RMSE) of the training and testing wells Eq. (6) and average absolute percent relative error (AAPRE) of the training and testing wells Eq. (8).[40,41]

The formulas for the two statistical measures RMSE and AAPRE are given as follows:

(1) Root mean square error

$$\text{RMSE} = \sqrt{\frac{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}{n}}. \tag{6}$$

(2) Average absolute percent relative error

$$E_i = \left(\frac{x_i - y_i}{y_i}\right) \times 100; \quad (i = 1, 2, 3, \ldots, n), \tag{7}$$

$$\text{AAPRE} = \frac{1}{n}\sum_{i}^{n} |E_i|. \tag{8}$$

A good model should have low *RMSE* and *AAPRE* values.

In Eqs. (6)–(8), x's are the predicted values, y's are the actual/experimental values and n is the total number of data points in all training wells (70 data points) or testing wells (29 data points).

6. Results and Discussion

6.1. *Viscosity curve prediction*

In the performed study, a comparison is done with the artificial neural networks to assess the performance of SVR, as compared to FN and ANN. A feedforward neural network (FFNN) model is developed for five aforementioned prediction variables. A number of trials were made viz: selecting the number of hidden layers, number of neurons in each hidden layer and the training algorithm. For μ_{ob} and R_{sb}, we eventually used two hidden layers with thirteen and six nodes respectively. Hence, we have 12-13-6-1 FFNN structure, (12 input neurons, 13 nodes in the first hidden layer, 6 nodes in the second hidden layer and 1 output neuron), for each case parameter. For the three fitting variables, we used 12-12-5-1 FFNN architecture. In all cases, tangent sigmoid transfer function and Levenberg-Marquardt training optimization were eventually used, and the best network out of 1000 runs in each case was taken.

Also, a sample of prediction plots of training and testing wells from the two frameworks are shown in Figs. 7 through 12. The statistical performance measures

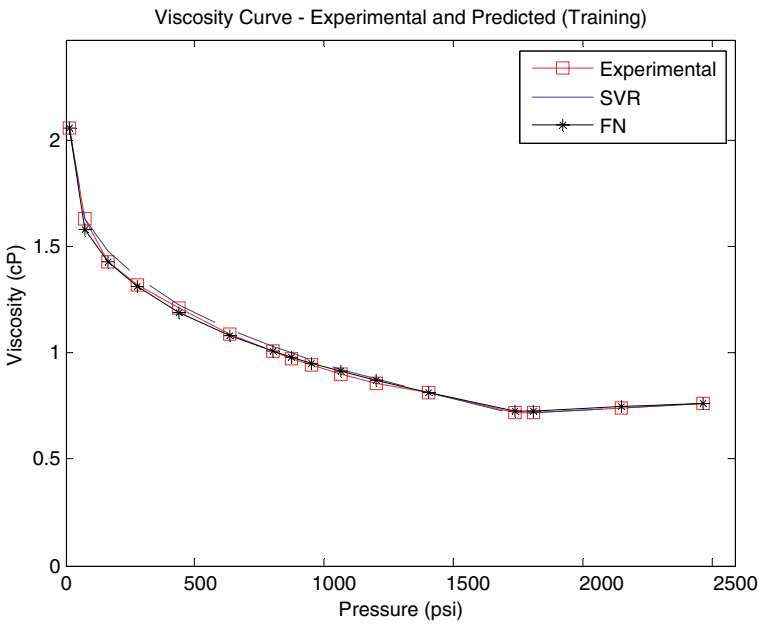

Fig. 7. Viscosity vs. pressure plot for sample well TR1.

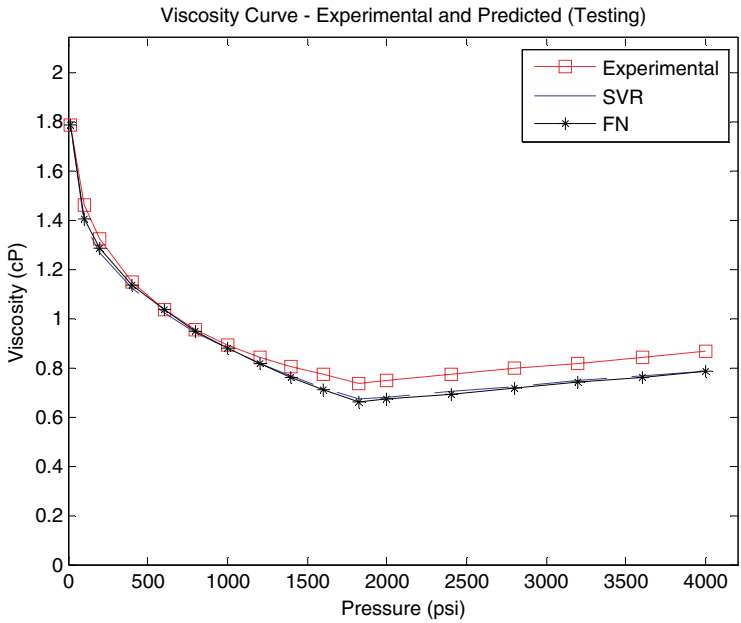

Fig. 8. Viscosity vs. pressure plot for sample well TS1.

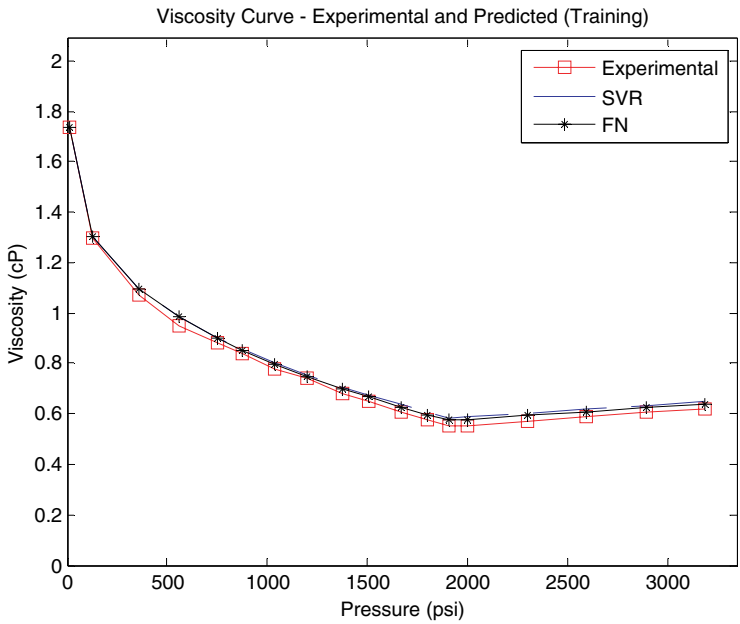

Fig. 9. Viscosity vs. pressure plot for sample well TR2.

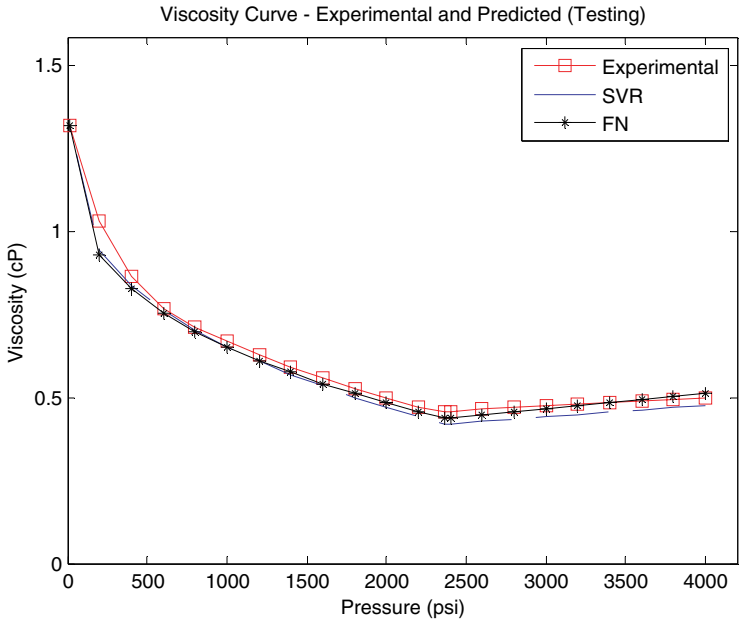

Fig. 10. Viscosity vs. pressure plot for sample well TS2.

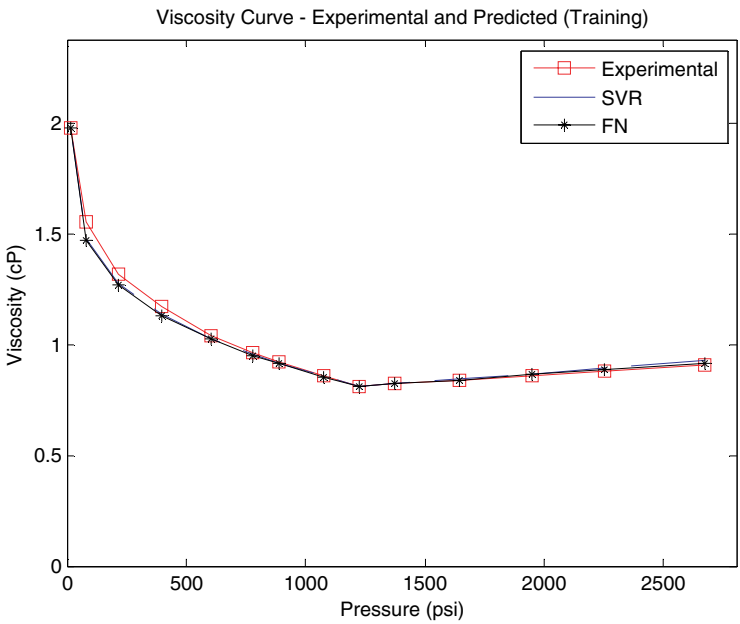

Fig. 11. Viscosity vs. pressure plot for sample well TR3.

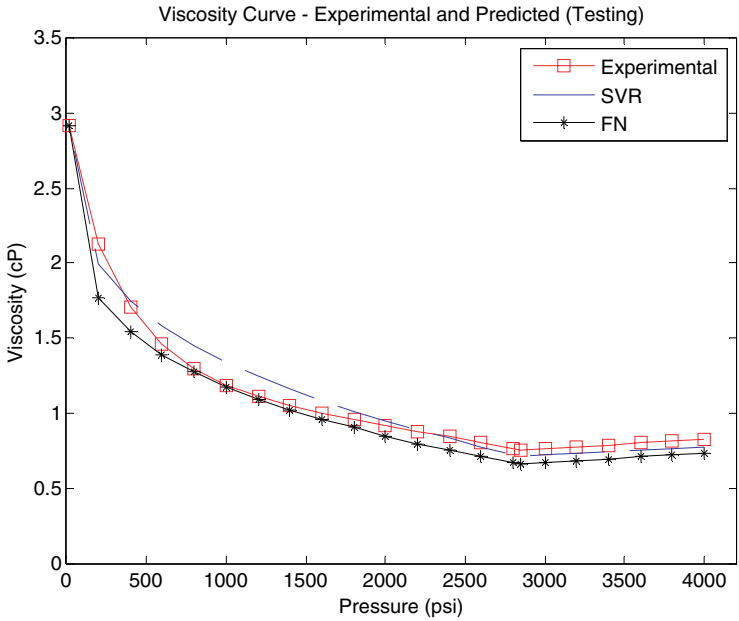

Fig. 12. Viscosity vs. pressure plot for sample well TS3.

Table 2. Statistical performance measures of SVR, FN and FFNN models for viscosity curves prediction.

Model	SVR		FN		FFNN	
	RMSE	AAPRE%	RMSE	AAPRE%	RMSE	AAPRE%
Training	0.07495	6.3953	0.067648	5.400661	0.08664	8.33945
Testing	0.07659	8.5969	0.079412	8.551437	0.08712	10.24569

for the two frameworks are shown in Table 2. For this pair of techniques, the performance of both frameworks, SVR and FN, are very competitive. While FN performance is better than that of SVR in the training phase with lower RMSE and AAPRE, which are 0.06765 and 5.4% respectively, against those of SVR which are 0.07495 and 6.3953% respectively, SVR performance is very competitive with that of FN for the testing wells. For the testing phase, FN has lower AAPRE of 8.5514%, against that of SVR which is 8.5969%, while SVR has lower RMSE, 0.0765, against that of FN which is 0.07941. Also, from Table 2, the RMSE and APPRE for the FFNN predictions are the highest for both training and testing. In essence, the results of SVR and FN are very competitive for viscosity curve prediction, while both clearly outperform FFNN. The predicted curves from the two SC techniques show good matching with the experimental curves for both training and testing wells with little deviation in some testing wells. Table 3 shows a sample of predicted viscosity curve parameters by SVR FN and FFNN models.

Table 3. Sample predicted viscosity curve parameters by SVR, FN and FFNN models.

	Actual	SVR	FN	FFNN
Training: α	7.19E-05	5.79E-05	5.7914E-05	5.89E-05
β	0.6688	0.626296	0.625281	0.50256
μ_{ob}	0.69	0.686544	0.71764	0.6628
Testing: α	3.87E-05	4.88E-05	4.32E-05	5.38E-05
β	0.3388	0.3386	0.3190	0.4889
μ_{ob}	0.58	0.5736	0.5549	0.5228

Table 4. Time complexity of all models for viscosity curve prediction.

CPU Time (seconds)		
Model	Training	Testing
SVR	2.6208	0.00103
FN	2.318	0.0936
FFNN	11.466	0.0468

Comparison based on time to complete development of each model is shown in Table 4. The training time may not be necessary or could be traded off, since after development of the model, only the testing phase will be utilized. And from this viewpoint, it is clear that SVR is the best of the three techniques.

6.2. *Gas/oil ratio curve prediction*

The SVR and FN frameworks were implemented to predict the required variables, τ and R_{sb} for gas/oil ratio curve prediction. Similar to the previous cases, only sample training and testing plots of the predicted gas/oil ratio curves are shown in Figs. 13 through 18. The predicted curves from these two techniques show good matching with the experimental curves for training and testing wells. Table 5 shows the statistical measures for evaluating the performance of SVR, FN and FFNN techniques in predicting gas/oil ratio curves. In this case, unlike the viscosity curve prediction where performances of both SVR and FN are very competitive, SVR has better average performance than FN in both training and testing phases, based on the statistical measures used for evaluation. For the training, SVR has RMSE 19.0043 and AAPRE of 7.5279%, while FN has RMSE of 21.6942 and AAPRE of 8.4167%. For testing, SVR has RMSE of 30.0170 and AAPRE of 9.0757%, while FN has RMSE of 32.8196 and AAPRE of 10.2012%. Table 6 shows that FFNN predictions are the worst with the highest RMSE and AAPRE.

Based on the preceding analysis, though performance of FN is also good, SVR framework gives better performance in predicting gas/oil ratio than FN. This is also evident from the sample predicted curves. Table 6 shows a sample of predicted GOR curve parameters by SVR, FN and FFNN models.

Table 7 shows the computational time for both training and testing of SVR, FN and FFNN models for gas/oil ratio curve prediction. It is noteworthy that SVR is

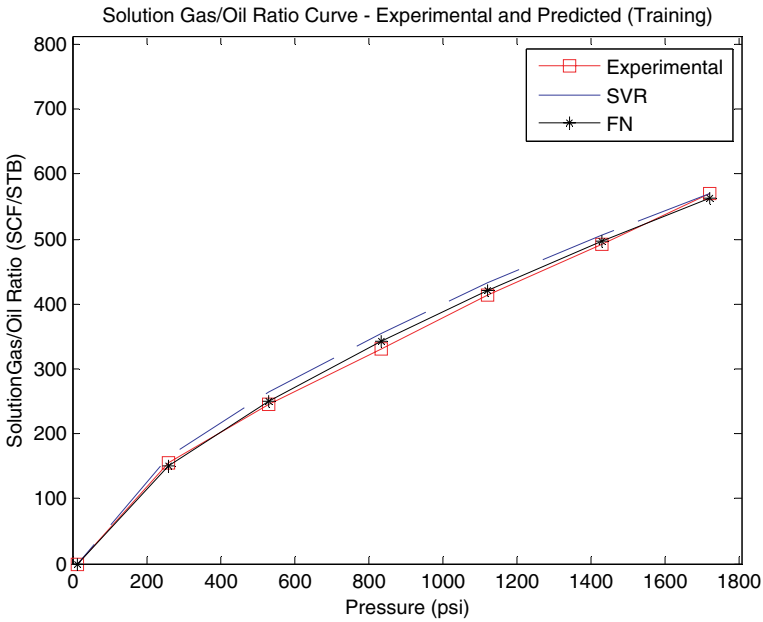

Fig. 13. Gas/oil ratio vs. pressure plot for sample well TR1.

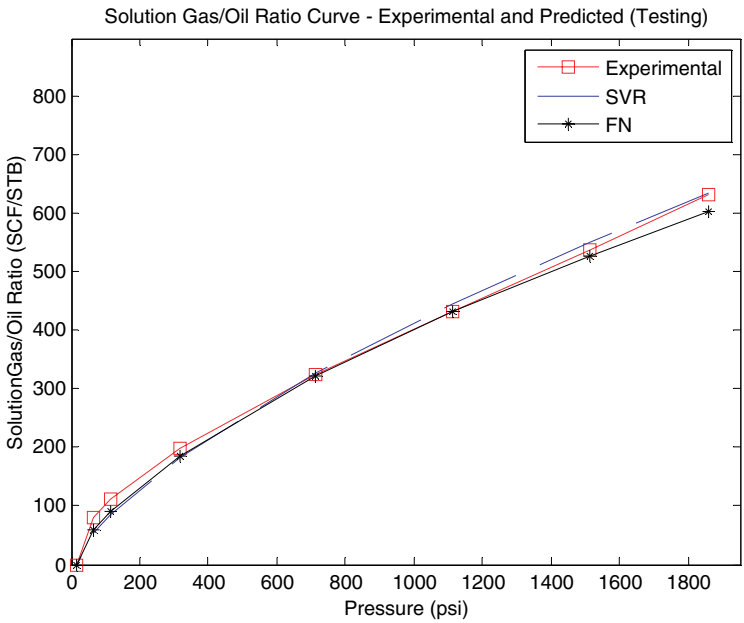

Fig. 14. Gas/oil ratio vs. pressure plot for sample well TS1.

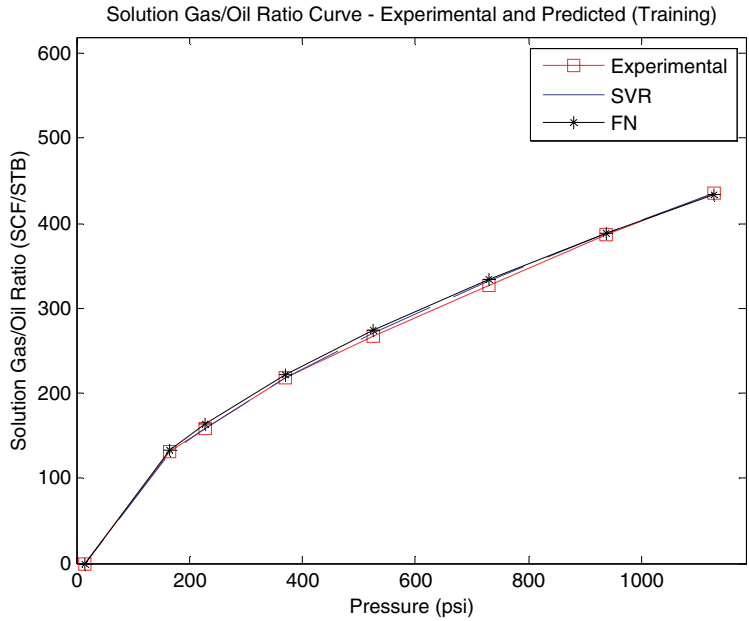

Fig. 15. Gas/oil ratio vs. pressure plot for sample well TR2.

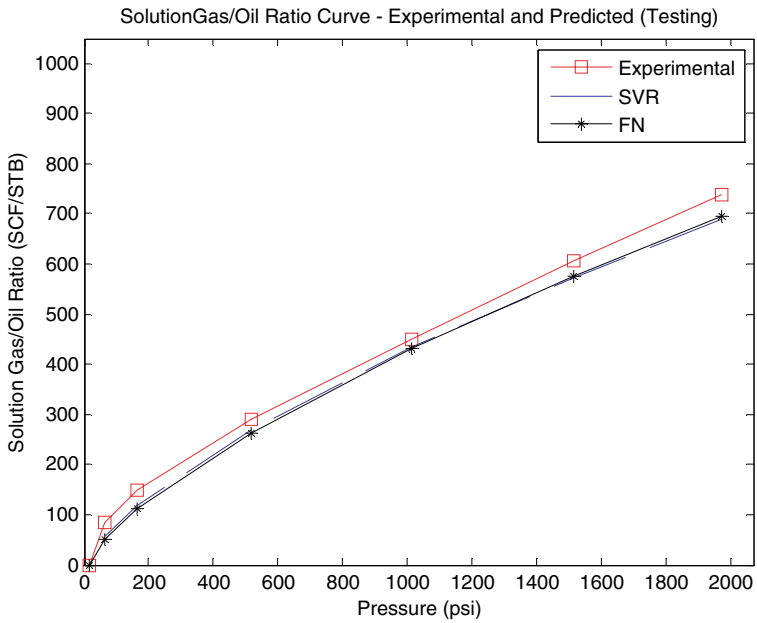

Fig. 16. Gas/oil ratio vs. pressure plot for sample well TS2.

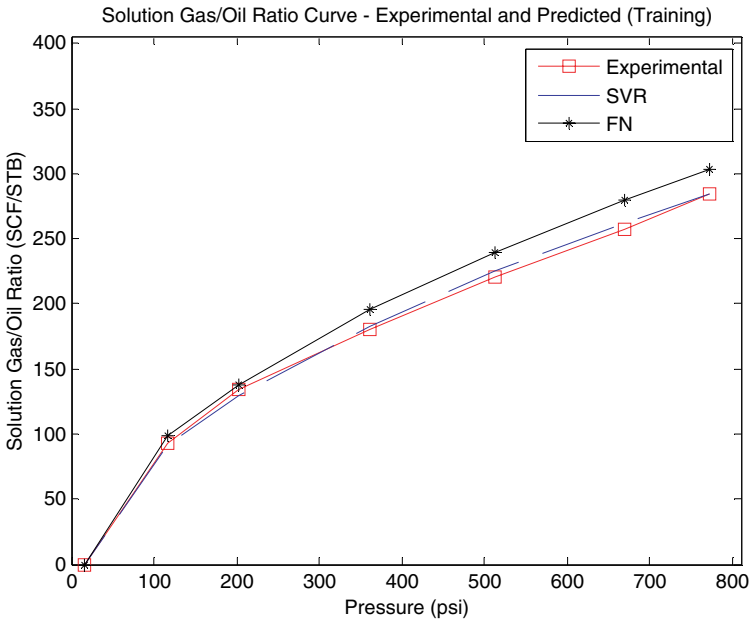

Fig. 17. Gas/oil ratio vs. pressure plot for sample well TR3.

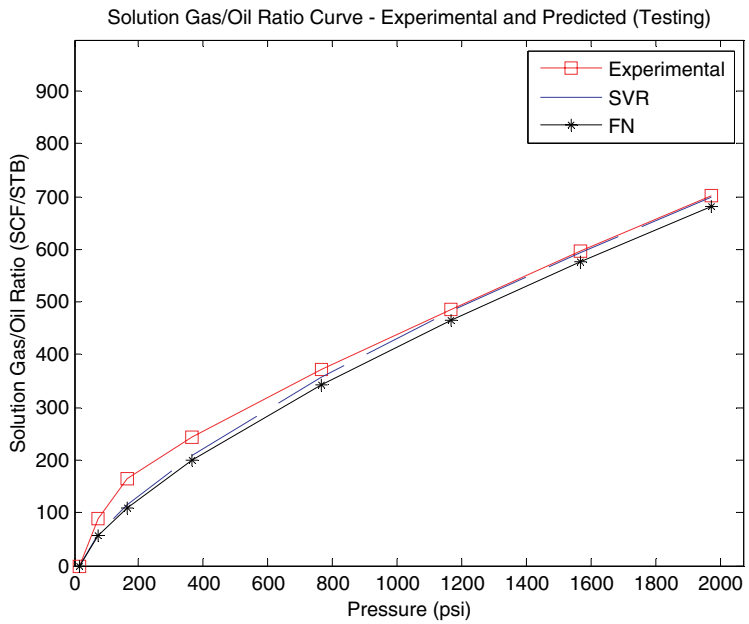

Fig. 18. Gas/oil ratio vs. pressure plot for sample well TS3.

Table 5. Statistical performance measures of SVR, FN and ANN models for gas/oil curve prediction.

Model	SVR		FN		FFNN	
	RMSE	AAPRE%	RMSE	AAPRE%	RMSE	AAPRE%
Training	19.0043	7.5279	21.6942	8.4167	23.5217	8.804342
Testing	30.0170	9.0757	32.8196	10.2012	39.0161	12.701

Table 6. Sample predicted gas/oil curve parameters by SVR, FN and ANN models.

	Actual	SVR	FN	FFNN
Training: τ	0.7762	0.6655	0.6813	0.6558
R_{sb}	558	584.0497	593.5556	599.3255
Testing: τ	0.61900	0.6397	0.5983	0.6468
R_{sb}	689	689.0279	669.7892	692.3316

Table 7. Time complexity of all models for gas/oil ratio curve prediction.

CPU Time (seconds)		
Model	Training	Testing
SVR	3.0264	0.0001
FN	2.4804	0.0624
FFNN	7.566	0.078

Table 8. Gas/oil ratio correlations.

Model	Training	Testing
SVR	0.997233	0.987923
FN	0.994975	0.964899

Table 9. Viscosity correlations.

Model	Training	Testing
SVR	0.986636	0.921167
FN	0.991126	0.920655

more demanding than FN at the training and testing with a significant difference, whereas the FFNN is the most demanding at training. Furthermore, Tables 8 and 9 show the correlation coefficient as to relate the statistical significance of the results which also illustrate that SVR performs better than FN for predicting gas/oil Curve and that SVR and FN are competitive regarding viscosity curve prediction.

7. Conclusion

In this paper, we have presented two advanced computational intelligence techniques to predict crude oil Pressure-Volume-Temperature (PVT) properties that need to be represented as curves over a specified range of reservoir pressures.

Instead of the usual single or multi-data points prediction, which could distort the consistency of the curve's shape, an efficient approach for predicting such PVT properties curves has been introduced and implemented. In all predictions, we have implemented different independent neural network techniques, viz: Support Vector Regression, Functional Networks and Feedforward Neural Network. The viscosity and gas/oil ratio curves prediction problems were formulated and implemented using these approaches. Simulation of these results had been reported and comparisons between the three techniques were discussed on both viscosity and solution GOR curve predictions. Interestingly, the shapes of the predicted viscosity and solution GOR curves are consistent with the physical law and the experimental curves. This makes the use of such predicted curve practicable for use, contributing thereby to enhanced exploration and production processes through cost reduction and improving human operator conditions.

Acknowledgment

This work was supported by King Fahd University of Petroleum and Minerals (Grant No. SB100014), and partially by King Abdulaziz City for Science and Technology under Grant MSTP-KACST-08-OIL82-4, Saudi Arabia.

Appendix 1: Nomenclature and Abbreviations

P:	Pressure (psi)
P_b:	Bubble point pressure (psi)
P_{od}:	Pressure at dead oil viscosity (psi)
R_s:	Solution gas/oil ratio, SCF/STB (m^3/m^3)
R_{sb}:	Bubble point solution gas/oil ratio, SCF/STB (m^3/m^3)
T:	Temperature (°F)
V:	Volume (m^3)
μ_a:	Viscosity above bubble point (cP)
μ_b:	Viscosity below bubble point (cP)
μ_o:	Oil viscosity (cP)
μ_{ob}:	Bubble point/gas-saturated oil viscosity (cP)
μ_{od}:	Dead oil viscosity (cP)
Res_Temp:	Reservoir temperature (°F)
Mol_N_2:	Mole fraction of N_2 (mol%)
Mol_CO_2:	Mole fraction of CO_2 (mol%)
Mol_H_2S:	The mole fraction H_2S (mol%)
PVT:	Pressure-Volume-Temperature
EOS:	Equations of States
GOR:	Gas/Oil Ratio,
RMSE:	Root Mean Square Error
AAPRE:	Average Absolute Percent Relative Error

Appendix 2: Kernel of SVR

For the five predicting variables, using Matlab, the selected option for optimal relevant variables are stated as follows.

(1) $\alpha : C = 10000$; lambda = 1e-7; epsilon = 0.09; kernel option = 0.9; kernel = 'poly'; verbose = 1.
(2) $\beta : C = 60$; lambda = 1e-7; epsilon = 0.08; kernel option = 0.8; kernel = 'poly'; verbose = 1.
(3) $\mu_{ob} : C = 40000$; lambda = 1e-7; epsilon = 0.001; kernel option = 0.994; kernel = 'Gaussian'; verbose = 1.
(4) $\tau : C = 100000$; lambda = 1e-7; epsilon = 0.001; kernel option = 2.8; kernel = 'poly'.
(5) $R_{sb} : C = 500000$; lambda = 1e-7; epsilon = 0.001; kernel option = 0.12; kernel = 'Gaussian'; verbose = 1.

Appendix 3: Families of Functions Used for Each Parameter for FN Model

For α, polynomial family of degree 3 was used and $f_1 \cdots f_{12}$ are

$$
\begin{aligned}
f_1(x_1) &= -0.78629 - 0.00026x_1 + 0.00012x_1^2 - 4.5 \times 10^{-5}x_1^3;\\
f_2(x_2) &= -0.0002x_2; \quad f_3(x_3) = -0.00019x_3 + 7.6 \times 10^{-7}x_3^2;\\
f_4(x_4) &= -0.00021x_4 + 2.29 \times 10^{-7}x_4^2;\\
f_5(x_5) &= -1.7 \times 10^{-5}x_5^2 + 4.93 \times 10^{-7}x_5^3;\\
f_6(x_6) &= -4.7 \times 10^{-6}x_6^2 + 3.72 \times 10^{-8}x_6^3;\\
f_7(x_7) &= 2.58 \times 10^{-4}x_7 - 5.3 \times 10^{-6}x_7^2;\\
f_8(x_8) &= 0.03466x_8 - 0.0005x_8^2 + 2.4 \times 10^{-6}x_8^2;\\
f_9(x_9) &= -1.7 \times 10^{-7}x_9 + 4.38 \times 10^{-11}x_9^2;\\
f_{10}(x_{10}) &= 5.44 \times 10^{-5}x_{10} - 8.7 \times 10^{-7}x_{10}^2;\\
f_{11}(x_{11}) &= -4.2 \times 10^{-5}x_{11} + 2.25 \times 10^{-7}x_{11}^2 - 4 \times 10^{-10}x_{11}^3;\\
f_{12}(x_{12}) &= -0.00011x_{12} + 3.3 \times 10^{-5}x_{12}^2 - 2.3 \times 10^{-6}x_{12}^3.
\end{aligned}
$$

For β, polynomial family of degree 3 gave the best result and $f_1 \cdots f_{12}$ are

$$
\begin{aligned}
f_1(x_1) &= -119.266 + 1.37266x_1^2 - 0.4728x_1^3; \quad f_2(x_2) = 0.80142x_2;\\
f_3(x_3) &= 0.752408x_3 + 0.02248x_3 - 0.00108x_3^2;\\
f_4(x_4) &= = 0.80746x_4 + 0.000544x_4^2;\\
f_5(x_5) &= 1.83579x_5 - 0.09053x_5^2 + 0.002752x_5^3;\\
f_6(x_6) &= 1.72097x_6 - 0.02352x_6^2 + 0.000207x_6^3;\\
f_7(x_7) &= 0.01923x_7^2 - 0.00052x_7^3; \quad f_8(x_8) = 0;\\
f_9(x_9) &= 0.000295x_9 - 1.6 \times 10^{-6}x_9^2 + 2.67 \times 10^{-10}x_9^3;
\end{aligned}
$$

$$f_{10}(x_{10}) = 3.2953x_{10} - 0.09306x_{10}^2 + 0.000853x_{10}^3;$$
$$f_{11}(x_{11}) = -0.33375x_{11} + 0.00175x_{11}^2 - 3 \times 10^{-6}x_{11}^3;$$
$$f_{12}(x_{12}) = -0.5895x_{12} + 0.070413x_{12}^2.$$

For μ_{ob}, polynomial family of degree 3 gave the best result and $f_1 \cdots f_{12}$ are

$$f_1(x_1) = -25.5869x_1;$$
$$f_2(x_2) = 0.162028x_2 - 0.105347x_2^2 + 0.004261x_2^3;$$
$$f_3(x_3) = 0.032553x_3 - 0.00362x_3^2; \quad f_4(x_4) = 0;$$
$$f_5(x_5) = 0; \quad f_6(x_6) = 0;$$
$$f_7(x_7) = 1.9566x_7 - 0.07975x_7^2 + 0.001032x_7^3;$$
$$f_8(x_8) = 0; \quad f_9(x_9) = -3.9 \times 10^{-7}x_9^2 + 8.62x_9^3;$$
$$f_{10}(x_{10}) = 1.70833x_{10} - 0.04799x_{10}^2 + 0.000443x_{10}^3;$$
$$f_{11}(x_{11}) = -0.13176x_{11} + 0.000663x_{11}^2 - 1.1 \times 10^{-6}x_{11}^3;$$
$$f_{12}(x_{12}) = 0.152996x_{12} + 0.01277x_{12}^2.$$

For τ, logarithm family gave the best result and $f_1 \cdots f_{12}$ are

$$f_1(x_1) = -14589.5 - 9.80634\log(x+2) + 15.5367\log(x+3);$$
$$f_2(x_2) = 107.4192\log(x_2+3) - 331.197\log(x_2+4) + 239.503\log(x_2+5);$$
$$f_3(x_3) = 77.258\log(x_3+3) - 251.037\log(x_3+4) + 188.3491\log(x_3+5);$$
$$f_4(x_4) = -903.557\log(x_4+3) + 949.7861\log(x_4+4);$$
$$f_5(x_5) = 1.94 \times 10^5\log(x_5+2) - 7.1 \times 10^5\log(x_5+3) + 8.7 \times 10^5\log(x_5+4) - 349768\log(x_5+5);$$
$$f_6(x_6) = 3105700\log(x_6+2) - 10000000\log(x_6+3) + 10858916\log(x_6+4) - 3895946\log(x_6+5);$$
$$f_7(x_7) = -2.3 \times 10^7\log(x_7+2) + 7.59 \times 10^7\log(x_7+3) - 8.5 \times 10^7\log(x_7+4) + 3.13 \times 10^7\log(x_7+5);$$
$$f_8(x_8) = 1.2 \times 10^4\log(x_8+2) - 12235.2\log(x_8+4);$$
$$f_9(x_9) = 1.05 \times 10^9\log(x_9+2) - 3.2 \times 10^9\log(x_9+3) + 3.16 \times 10^9\log(x_9+4) - 1.1 \times 10^9\log(x_9+5);$$
$$f_{10}(x_{10}) = -1.2502\log(x_{10}+4); \quad f_{11}(x) = 0;$$
$$f_{12}(x_{12}) = -12.0186\log(x_{12}+2) + 14.2783\log(x_{12}+3).$$

For R_{sb}, logarithm family gave the best result and $f_1 \cdots f_{12}$ are

$$f_1(x_1) = 251353 + 3409.847\log(x_1+2) - 6184.59\log(x_1+3);$$
$$f_2(x_2) = -17732.5\log(x_2+2) + 66166.75\log(x_2+3) - 54857.3\log(x_2+4);$$
$$f_3(x_3) = 10315.9\log(x_3+3) - 14498.5\log(x_3+4);$$
$$f_4(x_4) = -1.3 \times 10^7\log(x_4+2) + 2.68 \times 10^7\log(x_4+3) - 1.4 \times 10^3\log(x_4+4);$$

$$f_5(x_5) = 2.63 \times 10^6 \log(x_5 + 2) - 5.9 \times 10^6 \log(x_5 + 3) + 3.3144 \times 10^6 \log(x_5 + 4);$$
$$f_6(x_6) = -2.2 \times 10^7 \log(x_6 + 2) + 4.7 \times 10^7 \log(x_6 + 3) - 2.5 \times 10^7 \log(x_6 + 4);$$
$$f_7(x_7) = 1.65 \times 10^6 \log(x_7 + 3) - 1.71 \times 10^6 \log(x_7 + 4);$$
$$f_8(x_8) = 2056.883 \log(x_8 + 4);$$
$$f_9(x_9) = 1.47 \times 10^9 \log(x_9 + 2) - 2.9 \times 10^9 \log(x_9 + 3) + 1.48 \times 10^9 \log(x_9 + 4);$$
$$f_{10}(x_{10}) = -5 \times 10^5 \log(x_{10} + 3) + 5.119 \times 10^5 \log(x_{10} + 4); \quad f_{11}(x_{11}) = 0;$$
$$f_{12}(x_{12}) = 84618.57 \log(x_{12} + 2) - 246973 \log(x_{12} + 3) + 169459 \log(x_{12} + 4).$$

References

1. C. Beal, The viscosity of air, water, natural gas, crude oil and its associated gases at oil field temperature and pressures, *Trans. AIME* **165** (1946) 94–115.
2. J. Chew and C. A. Jr. Connally, A viscosity correlation for gas–saturated crude oils, *Trans. AIME* **216** (1959) 23–25.
3. H. D. Beggs and J. R. Robinson, Estimating the viscosity of crude oil system, *J. Pet. Tech.* **9** (1975) 1140–1149.
4. O. Glaso, Generalized pressure-volume temperature correlations, *J. Pet. Tech.* **32**(5) (1980) 785–795.
5. M. Vazuquez and H. D. Beggs, Correlation for fluid physical property prediction, *J. Pet. Tech.* **32**(6) (1980) 968–970.
6. E. O. Egbogah and T. Ng. Jack, An improved temperature-viscosity correlation for crude oil systems, *J. Pet. Sc. & Eng.* **5** (1990) 197–200.
7. S. A. Khan, M. A. Al-Marhoun, S. O. Duffuaa and S. A. Abu-Khamsin, Development of viscosity correlations for crude oils, *Fifth SPE Middle East Oil Show*, Bahrain, March (1987), pp. 7–10.
8. R. Labedi, Improved correlations for predicting the viscosity of light crudes, *J. Pet. Sc. & Eng.* **8** (1992) 221–234.
9. G. E. Jr. Petrosky and F. F. Farshad, Viscosity, correlations for Gulf of mexico crude oils, *Paper SPE 29468 Presented at Production Operations Symposium* (Oklahoma City, OK, U.S.A.), 2–4 April, 1995.
10. R. A. Almehaideb, Improved PVT correlations for UAE crude oils, *SPE Middle East Oil & Gas Show and Conference* (Bahrain, March 1997), pp. 15–18.
11. B. Dindoruk and P. G. Christman, PVT properties and viscosity correlations for Gulf of Mexico oils, *SPE Res. Eng.* **7**(6) (2004) 427–437.
12. A. M. El-Sharkawy, Modeling the properties of crude oil and gas systems using RBF network, *SPE Asia Pacific Oil & Gas Conference* (Perth, Australia, October 1998), pp. 12–14.
13. Ridha B. Gharbi, Adel M. Elsharkawy and Mansour Karkoub, Universal Neural-Network-Based Model for estimating the PVT properties of crude oil systems, *Energy Fuels* **13**(2) (1999) 454–458, DOI: 10.1021/ef980143v.
14. N. Varotsis, V. Gaganis, J. Nighswander and P. Guieze, A novel non-iterative method for the prediction of the PVT behavior of reservoir fluids, *SPE Annual Technical Conference and Exhibition* (Houston, Texas, October 1999), pp. 3–6.
15. E. Osman and M. Al-Marhoun, Artificial neural networks models for predicting PVT properties of oil field brines, *14th SPE Middle East Oil & Gas Show and Conf.* (Bahrain, 12–15 March, 2005).

16. M. A. Ayoub, D. M. Raja and M. A. Al-Marhoun, Evaluation of below bubble point viscosity correlations and construction of a new neural network model, *SPE Asia Pacific Oil & Gas Conference and Exhibition* (Indonesia, 30 October–1 November, 2007).
17. Y. Hajizadeh, Viscosity prediction of crude oils with genetic algorithms, *SPE Latin American and Caribbean Petroleum Engineering Conference* (Argentina, April 2007), pp. 15–18.
18. S. Dutta and J. P. Gupta, PVT correlations of Indian crude using support vector regression, *Energy Fuels* **23**(11) (2009) 5483–5490, DOI: 10.1021/ef900518f.
19. F. Farshad, J. L. LeBlanc and J. D. Garber, Empirical PVT correlations for Colombian crude oils, *SPE Fourth Latin American and Caribbean Petroleum Engineering Conf.* (Tobago, April 1996), pp. 23–26.
20. G. E. Petrosky and F. F. Farshad, Pressure volume temperature correlations for gulf of Mexico crude oils, *SPE Res. Eval. & Eng.* **1**(5) (1998) 416–420.
21. A. A. Al-Shammasi, A review of bubble point pressure and oil formation volume factor correlations, *SPE Res. Eval. & Eng.* **4**(2) (2001) 146–160.
22. M. A. Al-Marhoun, Evaluation of empirically derived PVT properties for Middle East crude oils, *J. Pet. Sc. & Eng.* **42** (2004) 209–221.
23. S. S. Ikiensikimama, J. Madu and L. Dipeolu, Black oil empirical PVT correlations screening for the Niger delta crude, *30th Annual SPE International Technical Conference and Exhibition* (Nigeria, 31 July–2 August, 2006).
24. M. N. Hemmati and R. Kharrat, Evaluation of empirically derived PVT properties for Middle East crude oils, *Scientia Iranica* **14**(4) (2007) 358–368.
25. J. P. Holman, *Experimental Methods for Engineers* (Seventh Edition, McGraw-Hill, 2001).
26. M. A. Al-Marhoun, A. A. Abdul Raheem, S. Nizamuddin and S. Shujath Ali, Prediction of crude oil viscosity curve using empirical derived correlations and artificial intelligence techniques, submitted, *J. Pet. Sc. & Eng.*, 2010.
27. M. A. Oloso, A. Khoukhi, A. Abdulraheem and M. Elshafei, Prediction of crude oil viscosity and gas/oil ratio curves using advances to neural networks. 2009 *SPE/EAGE Reservoir Characterization and Simulation Conference* (Abu Dhabi, UAE, October 2009), pp. 19–21.
28. A. Khoukhi, M. Oloso, A. Abdulraheem and M. Elshafei, Evolutionary neural networks for estimating viscosity and gas/oil ratio curves, *Proc. of the 21st IASTED Int'l Conf. Modelling & Simulation MS2010* (Banff, Alberta, Canada, July 15–17, 2010).
29. A. Khoukhi, M. Oloso, A. Abdulraheem and M. Elshafei, Viscosity and gas/oil ratio curves estimation using advances to neural networks, 2011, 7th International Workshop on Systems, Signal Processing and their Applications (WOSSPA) (Tipaza, Algeria, May 9–11, 2011).
30. C. Cortes and V. Vapnik, Support-vector networks, *Mach. Learn.* **20** (1995) 273–297.
31. P. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, Cambridge, MA, 2002).
32. A. E. El-Sebakhy, A fast and efficient algorithm for multi-class support vector machines classifier. ICICS2004: 28–30 November, *IEEE Computer Society* (2004), pp. 397–412.
33. W. T. A. Littman, *Introduction to Support Vector Machines, Machine Learning Course 536. In Tom Mitchell*, Text book: Machine learning, McGraw Hill, 1997. Department of Computer Science, Rutgers University (The State University of New Jersey, USA, 2003).

34. C. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines* (Cambridge Univ. Press, Cambridge, U.K., 2000).
35. M. S. Mohamad, S. Deris, R. M. Illias, A hybrid of genetic algorithm and support vector machine for features selection and classification of gene expression microarray, *Intl. Journal of Computational Intelligence and Applications IJCIA* **5**(1) (2005) 91–107.
36. L. Zhang, W. Zhou and L. Jiao, Support vector machines based on the orthogonal projection kernel of father wavelet, *Intl. Journal of Computational Intelligence and Applications IJCIA* **5**(3) (2005) 283–303.
37. E. Castillo, Functional networks, *Neural Processing Letters* **7** (1998) 151–159.
38. E. Castillo, A. Cobo, J. M. Guteirrez and R. E. Pruneda, Functional networks: A new network-based methodology, *Computer-Aided Civil and Infrastructure Engineering* **15**(2) (2000) 90–106.
39. Available: <http://www.cs.rutgers.edu/∼mlittman/courses/ml03/>.
40. A. Khoukhi and S. Al-Bukhitan, Crude oil PVT properties prediction using hybrid Genetic-Neuro-Fuzzy techniques, *International Journal of Oil Gas and Coal Technology* **4**(1) (2011) 47–63.
41. A. Khoukhi and S. Al-Bukhitan, Data-Driven Genetic-Neuro-Fuzzy systems to crude oil PVT properties prediction, *NAFIPS10* (Toronto, Canada, July 12–14, 2010), pp. 1–7.