

Committee machine reaping of three well-known models: established between saturation pressure and gas chromatography data

Parisa Bagheripour · Mojtaba Asoodeh

Received: 26 April 2014 / Accepted: 13 June 2014 / Published online: 5 July 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract Saturation pressure is critical parameter of reservoir fluids which significantly affects petroleum engineering calculations. Accurate measurement of saturation pressure from laboratory experiment is very time, cost, and labor intensive. Therefore, it is favorable in most cases to achieve this parameter from empirical correlations. Three well-known models for estimation of saturation pressure from gas chromatography data include Elsharkawy model (EM), Soave–Redlich–Kwong (SRK), and Peng–Robinson (PR) equations of states (EOSs). This model proposes a novel approach, called committee machine, to reap beneficial advantages aforementioned three models through the combination of them. Committee machine produces a sophisticated model which performs in cooperation of EM, SRK and PR EOSs. Results indicated that CM model enhanced the accuracy of final prediction and performed more satisfyingly compared with individual model acting alone.

Keywords Gas chromatography · Saturation pressure · Peng–Robinson equations of state · Redlich–Kwong equations of state · Committee machine

Introduction

Saturation pressure is the crucial characteristic of reservoir fluids which plays a significant role in different petroleum

engineering calculations such as material balance, reserve estimation, well testing, reservoir simulation, and production planning [1]. Since accurate measurement of saturation pressure is highly expensive and time-consuming, it is more favorable to estimate saturation pressure from empirical models. Several researchers have proposed empirical correlations to estimate saturation pressure from PVT data [3–6]. Asoodeh and Kazemi [7] integrated three of these correlations in a power-law structure and developed an enhanced model. Quantitative formulation between gas chromatography data and saturation pressure is a potent tool for estimation of saturation pressure in a cheap, fast, and precise way. Soave–Redlich–Kwong (SRK) and Peng–Robinson (PR) equations of states can be employed for such estimations (EOSs). Elsharkawy [8] developed an empirical model to estimate saturation pressure from gas chromatography data. He claimed his model outperformed SRK and PR EOSs. In this study, a novel strategy is proposed to develop a quick, convenient to use, and accurate method for estimation of saturation pressure from gas chromatography data. It utilizes committee machine concept to integrate Elsharkawy, SRK EOS and PR EOS models and produces a sophisticated model reaping all benefits of these models. A hybrid genetic algorithm-pattern search (GA-PS) tool is embedded in the structure of committee machine to extract optimal linear combination of Elsharkawy, SRK EOS and PR EOS models. Results indicated that committee machine performed more satisfyingly compared with individual model performing alone.

Committee machine

Committee machine is an integrated method that produces final output through optimal linear combination of

P. Bagheripour (✉)
Department of Petroleum Engineering, Gachsaran Branch,
Islamic Azad University, Gachsaran, Iran
e-mail: p.bagheripour@gmail.com

M. Asoodeh
Islamic Azad University, Birjand Branch, Birjand, Iran
e-mail: asoodeh.mojtaba@gmail.com

other methods' outputs. Inspired by divide and conquer principle, committee machine breaks down the problem into computationally simpler parts and reaps the benefits of each part through integrating solutions to these parts. The idea behind the committee machine is to fuse the knowledge acquired by different methods to arrive at an overall output that is more accurate than that of any of the individual method acting alone [9]. Supposing K different methods producing prediction vector of O_i such that O_1 are predictions corresponding to the first method, O_2 predictions corresponding to the second method and so on. Knowing the target vector of T , error of prediction is simply evaluated from following equation.

$$e_i = O_i - T, \quad i = 1, \dots, K \quad (1)$$

The expectation of the squared error for the i^{th} method (O_i) is:

$$E_i = \xi[(O_i - T)^2] = \xi[e_i^2], \quad i = 1, \dots, K \quad (2)$$

where $\xi[\cdot]$ denotes the expectation. Therefore, average mean square error (MSE) of prediction for K methods is computed as below:

$$E_{\text{avg.}} = \frac{1}{K} \sum_{i=1}^K E_i = \frac{1}{K} \sum_{i=1}^K \xi[e_i^2] \quad (3)$$

Extending above equation releases:

$$\begin{aligned} E_{\text{avg.}} &= \frac{\xi}{K^2} (e_1^2 + e_2^2 + \dots + e_K^2) \times (1 + 1 + \dots + 1) \\ &= \frac{\xi}{K} (e_1^2 + e_2^2 + \dots + e_K^2) \end{aligned} \quad (4)$$

Assuming average method as the simplest committee machine, following equation results in outputs of committee machine:

$$O_{\text{CM}} = \frac{1}{K} \sum_{i=1}^K O_i \quad (5)$$

Hence, MSE of committee machine is defined as:

$$\begin{aligned} E_{\text{CM}} &= \xi \left[(O_{\text{CM}} - T)^2 \right] = \xi \left[\left(\frac{1}{K} \sum_{i=1}^K O_i - T \right)^2 \right] \\ &= \xi \left[\left(\frac{1}{K} \sum_{i=1}^K e_i \right)^2 \right] \end{aligned} \quad (6)$$

Rewriting above equation as follow:

$$E_{\text{CM}} = \frac{\xi}{K^2} (e_1 \times 1 + e_2 \times 1 + \dots + e_N \times 1)^2 \quad (7)$$

Considering following Cauchy's inequality:

$$(a_1 b_1 + a_2 b_2 + \dots + a_K b_K)^2 \leq (a_1^2 + a_2^2 + \dots + a_K^2)(b_1^2 + b_2^2 + \dots + b_K^2) \quad (8)$$

Comparing Eqs. 4 and 7 with Cauchy's inequality, it is simply deduced that error of committee machine is less than or equal to the average error of all methods acting alone.

$$E_{\text{CM}} \leq E_{\text{avg.}} \quad (9)$$

In this study, Elsharkawy, SRK and PR EOSs models are employed to form inputs of committee machine. To develop the committee machine, a hybrid GA-PS tool is embedded in its structure. Hybrid GA-PS extracts optimal coefficients corresponding to each model to constitute an optimal linear combination of them. To extract optimal coefficients assigned to each model, following fitness function is introduced to GA-PS tool.

$$\text{MSE}_{\text{CM}} = \frac{1}{K} \sum_{j=1}^K (w_1 \text{El}_j + w_2 \text{SRK}_j + w_3 \text{PR}_j - T_j)^2 \quad (10)$$

where, K is number of training data; w_j ($j = 1, 2$, and 3) are assigned coefficient to each model and T is corresponding target value. GA-PS finds appropriate "w" for each model such that MSE of committee machine reaches its minimum value. Constructed committee machine will have better performance compared with individual models (Eqs 1 through 9). For more study about hybrid GA-PS technique, refer to Asoodeh and Bagheripour [10]. Genetic algorithm could be used for following purposes:

- (i) For curve fitting and finding optimal match function over an input/output data space
- (ii) For constructing an optimal linear combination of different models which have an overlapping to enhance accuracy of final prediction (current study)
- (iii) For aggregating different separate models which are constructed on separate subspaces of a specific problem
- (iv) As an alternative for training of intelligent systems (e.g., neural network, fuzzy logic, support vector machine and, etc.)

Results and discussion

Modeling and testing data for this study are originated from crude oil samples of Middle East, North Sea, and North America which are collected by Elsharkawy [8]. These data contain information about mole fraction of hydrocarbon and non-hydrocarbon component ($\text{C}_1\text{-C}_6$, C_{7+} , H_2S , CO_2 ,

Fig. 1 Run of hybrid genetic algorithm-pattern search tool for MSE fitness function of committee machine. This plot shows the best and mean fitness values of fitness functions after 1,000 generations for saturation pressure prediction

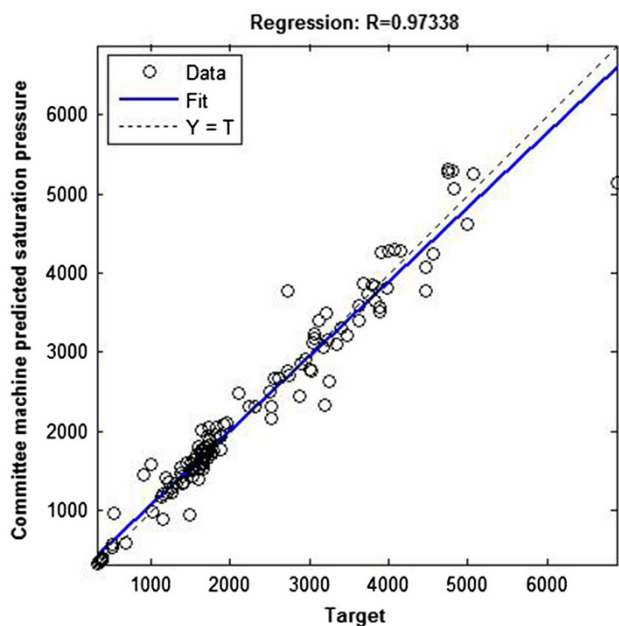
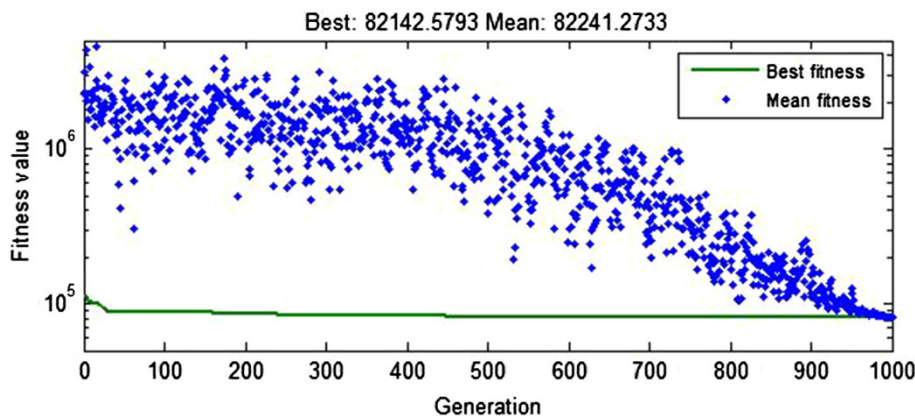


Fig. 2 Crossplot showing correlation coefficient between measured and committee machine predicted saturation pressure. High value of correlation coefficient, i.e., 0.97338 proves the robustness of committee machine modeling

and N_2), specific gravity of C_{7+} , molecular weight of C_{7+} , reservoir temperature, and saturation pressure. A statistical analysis on these datasets exists in the work by Kazemi et al. [11]. Estimated values of saturation pressure from Elsharkawy, SRK EOS, and PR EOS exist in the work by Elsharkawy [8]. In this stage of study, Eq. 10 was introduced to hybrid GA-PS technique as fitness function for committee machine construction meant to predict saturation pressure through combination of empirical models. GA-PS algorithm was regulated to 1,000 generations, crossover fraction of 72 % with scattered crossover function, and initial range of [0 1]. Run of GA-PS (Fig. 1) leads to finding global minimum of MSE function of committee machine. Figure 1 indicates that global minimum of MSE

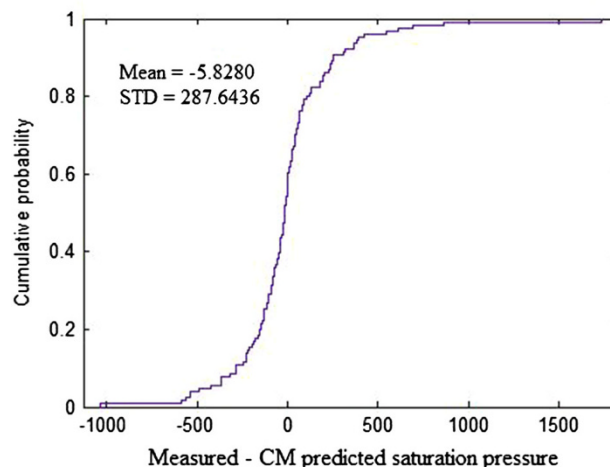


Fig. 3 Cumulative probability of error distribution for committee machine model meant to estimate saturation pressure. Small values of mean and standard deviation (*STD*) reveal high performance of committee machine modeling. This distribution indicates that 68 % of predicted values have errors in range of -5.8280 ± 287.6436

function of committee machine is 82,142.6 Psi^2 . Committee machine offers the following equation for estimating saturation pressure through optimal linear combination of Elsharkawy, SRK and PR equation of states models.

$$\text{Saturation pressure} = 0.2140\text{EI} + 0.0855\text{SRK} + 0.7716\text{PR} \tag{11}$$

After construction of committee machine, test data were input to it to evaluate performance of constructed model. Figure 2 shows crossplot between measured and committee machine predicted saturation pressure. High value of correlation coefficient in this figure puts a confirmation on success of committee machine in estimating of saturation pressure. Figure 3 shows cumulative probability of error distribution for committee machine predictions. This figure explicitly indicates that error for most of data points is located in close proximity of zero. 68 % of data points have errors in range of -5.8280 ± 287.6436 which is an

Table 1 Comparing committee machine (CM) with its elements, including Elsharkawy (El) empirical model, Soave–Redlich–Kwong (SRK) equation of state (EOS) and Peng–Robinson (PR) EOS using correlation coefficient (R) and root mean square error (RMSE) concepts

Methods	CM	El	SRK EOS	PR EOS
R	0.9734	0.9150	0.9120	0.9430
RMSE	286.6052	371.5972	381.7997	359.4215

acceptable error range comparing with range of saturation pressure data, i.e., [313–6880]. Eventually, committee machine was compared with its elements, including Elsharkawy empirical model and SRK and PR equation of states models. Table 1 indicates results of such a comparison. Results indicate superiority of committee machine to its elements. In almost all petroleum-related problems, where performing a real task/job might last for days, months, and even years, algorithm running time is of less importance. Running the genetic algorithm may take a couple of minutes which is negligible. In the other hand, committee machine by little additional computation enhances accuracy of final prediction, significantly.

Conclusions

Owing to the significance of saturation pressure, different empirical correlations are established between gas chromatography data and saturation pressure. Elsharkawy (El) empirical model, SRK EOS, and PR EOS are among the most successful models estimating saturation pressure. This study proved that committee machine concept, inspired by divide and conquer principle can effectively enhance precision of final prediction through combining aforementioned empirical models. Hybrid GA-PS tool, embedded in committee machine was employed to extract optimal coefficients corresponding to each model indicating its involvement in overall prediction of saturation pressure. Hybrid GA-PS was capable of optimizing MSE of

prediction for committee machine such that performance function of committee machine is located in its global minimum. The propounded methodology enables to have all beneficial advantages of the three priceless empirical correlations, including El, SRK EOS, and PR EOS in one model. Implementation of the proposed methodology can effectively save time and money.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Asoodeh M, Bagheripour P (2012) Estimation of bubble point pressure from PVT data using a power-law committee with intelligent systems. *J Petrol Sci Eng* 90–91:1–11
2. Standing MB (1947) A Pressure–volume–temperature correlation for mixture of California oils and gases. American Petroleum Institute
3. Lasater JA (1958) Bubble point pressure correlation. *Trans AIME* 213:379–381
4. Glaso O (1980) Generalized pressure–volume–temperature correlations. *J Petrol Technol* 32(5):785–795
5. Al-Marhoun MA (1988) PVT correlations for Middle East crude oils. *J Petrol Technol* 40(5):650–666
6. McCain WD Jr (1991) Reservoir fluid property correlation-state of art. *SPE Reserv Eng* 6(2):266–272
7. Asoodeh M, Kazemi K (2013) Estimation of bubble point pressure using a genetic integration of empirical formulas. *Energy Sour Part A* 35:1102–1109
8. Elsharkawy AM (2003) An empirical model for estimating the saturation pressure of crude oils. *J Petrol Sci Eng* 38:57–77
9. Haykin S (1991) *Neural networks: a comprehensive foundation*. Prentice-Hall, Englewood Cliffs, p 842
10. Asoodeh M, Bagheripour P (2012) 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
11. Kazemi K, Moradi S, Asoodeh M (2013) A neural network based model for prediction of saturation pressure from molecular components of crudeoils. *Energy Sour Part A* 35:1039–1045