Integration of LSSVM technique with PSO to determine asphaltene deposition

Ali Chamkalani a,b,*, Sohrab Zendehboudi c,**, Alireza Bahadori d, Riaz Kharrat a, Reza Chamkalani e, Lesley James f, Ioannis Chatzis g

a Department of Petroleum Engineering, Petroleum University of Technology, Ahwaz, Iran
b Oil and Gas Engineering Department, Pars Oil and Gas Company (POGC), Asalouyeh, Iran
c Department of Chemical Engineering, Massachusetts Institute of Technology (MIT), Cambridge, MA, USA
d School of Environment, Science and Engineering, Southern Cross University, Lismore, NSW, Australia
e Energy Engineering Department, Faculty of Shahid Abbaspour, Shahid Beheshti University, Tehran, Iran
f Faculty of Engineering and Applied Science, Memorial University of Newfoundland, St. John’s, NL, Canada
g Department of Chemical Engineering, University of Waterloo, Waterloo, ON, Canada

Article info
Article history:
Received 8 April 2013
Accepted 1 October 2014
Available online 13 October 2014

Keywords:
scaling equation
least square support vector machine
asphaltene deposition particle
swarm optimization
predictive tool

Abstract
Asphaltene deposition is a recognized phenomenon in petroleum industry with undesirable outcomes so that it may lead to wellbore plugging and formation damage, resulting in a large amount of remedial costs to decrease its negative impacts on oil production. Therefore, it has attracted lots of research interests in the literature. In this study, an attempt is made to introduce the least square support vector machine (LSSVM) for prediction of asphaltene deposition. This technique with high capabilities which captures the complex nature of asphaltene could be inferred as a scaling model. As there is no a standard procedure to determine the main parameters of the LSSVM model, the particle swarm optimization (PSO) technique is employed to synchronously optimize the LSSVM parameters. The modeling results clearly demonstrate that the optimized LSSVM is able to handle the nonlinearities well and attain satisfactory results. The comparison of available predictive equations for asphaltene deposition confirms that the LSSVM technique linked with PSO exhibits higher robustness and greater precision with an R^2 of 0.989 for the testing phase.

1. Introduction
It is of importance to predict amount of asphaltene precipitation at various operational conditions throughout oil production and transportation stages. While dealing with this serious problem, two essential questions that arise in industrial and research sectors are “How much” and “When” asphaltene will precipitate. Over time, a number of scientists have tried to find acceptable answers for these questions; however no explicit and accurate responses have been provided until now. The lack of adequate knowledge comes from fuzzy and complex nature of asphaltene and a variety of factors involved in this occurrence.

2. Asphaltene thermodynamic models
The models which have been introduced to investigate the phase behavior of asphaltene are mainly classified into two groups, depending on how asphaltene is stabilized in crude oil. The first group supposes that asphaltene is stable in crude oil. However, if the solubility falls under a specific threshold, the asphaltene will initiate to precipitate. Equation of state (EOS)-based models and Flory-Huggins-regular-solution-based models are two important representatives of this category (Hirschberg et al., 1984; Cimino et al., 1995; Nghiem and Coombe, 1998; Akbarzadeh et al., 2002).

In the second category, the resin plays an influential role to accompany asphaltene as a colloid in stabilization. Most common models belonging to this group consist of the solid-asphaltene colloidal model developed by Leontaris and Mansoori (1988), the reversible micellization model developed by Victorov and Firoozabadi (1996) and the model introduced by Wu et al. (2000) which is called McMillan–Mayer-SAFT based theory. Due to the thermodynamics approach of the scaling equation (e.g., Rassamdana et al., 1996), it is taken into account as a model in the second category.
2.1. Scaling equation

Some studies (e.g., Rassamdana et al., 1996; Andersen and Speight, 1999; Yang et al., 1999) found that the current conventional thermodynamic models cannot appropriately predict the asphaltene precipitation without the assistance of data fitting. Park and Mansoori (1988) concluded that asphaltene formation appears to be similar to the aggregation/gelation phenomena, somehow. Hence, it might be described through the scaling/fractal theory. On the basis of this finding, Rassamdana et al. (1996), and Rassamdana and Sahimi (1996) developed a scaling approach to illustrate the behavior of asphaltene precipitation. To obtain a scaling equation, they defined two new parameters as functions of three main variables including, solvent to oil dilution ratio \( R \) in \((\text{mL/g})\), weight percentage of asphaltene precipitated \( W \), and solvent molecular weight \( M \) as given below:

\[
X = \frac{R}{M} \quad (1)
\]

\[
Y = \frac{W}{R^2} \quad (2)
\]

in which \( z \) and \( z' \) refer to two adjustable parameters that should be carefully adjusted to determine the best scaling fit to the real data. Regardless of oil and precipitant used in the tests, \( z' \) and \( z_0 \) were obtained to be \(-2\) and \(2\), respectively, according to the study conducted by Rassamdana et al. (1996). In general, the following scaling equation is appropriate to describe asphaltene deposition behavior:

\[
Y = A_1 + A_2X + A_3X^2 + A_4X^3, \quad X \geq X_c \quad (3)
\]

in which \( A_1, A_2, A_3 \) and \( A_4 \) denote the scaling coefficients, and \( X_c \) stands for the value of \( X \) at the onset point of asphaltene precipitation.

Hu et al. (2000) carried out a systematic investigation on the application of scaling equation which has been suggested by Rassamdana et al. (1996) for asphaltene precipitation/deposition. They assessed the predictive potential of the scaling model by comparing to the asphaltene precipitation data reported in the open sources. They found that the scaling equation is a powerful and attractive tool to investigate/model asphaltene precipitation and deposition through a simple way. The universality of exponents \( z \) and \( z' \) was examined in their study, implying that \( z' \) introduces a universal constant \( (z' = -2) \), whereas the exponent \( z \) is strongly dependent on the oil composition but independent of the type of solvent \((n\text{-alkane})\) utilized. Using the experimental data, the optimum value of \( z \) was found within the range of \(0.1–0.5\).

Despite the simplicity and sufficient accuracy of the Rassamdana et al. scaling equation, the model is restricted to the invariable temperature. Therefore, as the effect of temperature is ignored in the scaling equation, it is not applicable to correlate the asphaltene precipitation to temperature as an important variable when dealing with oil mixture at different temperatures. To remove this shortcoming, Rassamdana and Sahimi (1996) tailored their scaling model by considering the effect of the temperature parameter in the scaling equation. Applying the modification, two new variables \( x \) and \( y \) were introduced as follows:

\[
x = \frac{X}{T^{\frac{1}{c_1}}} \quad (4)
\]

\[
y = \frac{Y}{X^{c_2}} \quad (5)
\]

In Eqs. (4) and (5), \( x \) and \( y \) introduce the same variables as defined in Eqs. (1) and (2). \( C_1 \) and \( C_2 \) are also the adjustable parameters. It was claimed that an acceptable fit to the laboratory data is achieved when \( C_1 \) and \( C_2 \) are set on 0.25 and 1.6, respectively.
Based on the study conducted by Rassamdana and Sahimi (1996), a new scaling equation in the form of a 3rd-order polynomial equation was then obtained as follows:

\[ y = b_1 + b_2 x + b_3 x^2 + b_4 x^3, \quad x \geq x_c \]  

(6)

In addition, a generalized corresponding states principle (CSP) was generated by Hu et al. (2004) to forecast the asphaltene precipitation outputs. The proposed CSP theory leads to a scaling model for asphaltene precipitation where the impact of n-alkane solvents is investigated. They defined X and Y whose expressions involve the effects of various parameters and system characteristics.

Hu and Guo (2001) experimentally studied the effects of molecular weight of n-alkane solvents (or precipitants), temperature, and solvent to oil ratio on the amount of asphaltene precipitated while using Chinese crude oil systems. The amount of asphaltene precipitation at the temperature interval of 293–338 K was quantified where seven n-alkanes (as precipitants) were employed. It was noticed that their experimental data points are not appropriately correlated to Eq. (6) by setting \( C_1 = 0.25 \) and \( C_2 = 1.6 \) that were proposed by Rassamdana and Sahimi (1996). According to their results, the best values for \( C_1 \) and \( C_2 \) are 0.5 and 1.6, respectively.

Ashoori et al. (2003) obtained a novel scaling model through making modification on relationship X (e.g., Eq. (4)) as the following:

\[ X = \frac{R}{T^m M^p} \]  

(7)

\[ Y = \frac{W}{R^q} \]  

(8)

Indeed, X is changed through including the temperature parameter; however the relationship for Y remains unchanged. In Eq. (7), the exponent \( n \) is a constant, varying from 0.10 to 0.25 for different cases. This constant (e.g., \( n \)) is equal to 0.15 for their particular case under study. Two other constants, \( z \) and \( z' \), keep the same values as those introduced in the first scaling model (e.g., \( z = 0.25 \) and \( z' = -2 \)).

Khaksar Menshad et al. (2008) developed a scaling equation by introducing pressure as a spare factor to find the weight percentage of asphaltene precipitation.

Bagheri et al. (2009) investigated various characteristics of different crude oils such as gas/oil ratio (SCF/STB), ratio of solvent to oil (cm³/g), mole fraction of C₁₁, resin content (wt%), extent of asphaltene precipitated (wt%), and asphaltene content (wt%) through an experimental work. They then introduced a new scaling model for estimation of asphaltene precipitation at a variety of dilution ratios.

Soltani et al. (2009) obtained a model for live oils under non-isothermal conditions to cover a wide pressure range between upper onset and lower onset pressures. A number of parameters such as temperature, pressure, critical temperature, bubble point pressure, onset pressure, dilution ratio, weight percent of asphaltene precipitation (at critical pressure) were considered to obtain a logarithmic equation to calculate the amount of asphaltene precipitation in weight percentage.

A scaling equation for asphaltene precipitation during gas injection operation was developed by Moradi et al. (2012). Important parameters (e.g., reservoir pressure and temperature, bubble point pressure of oil sample, mole percent of precipitant, and gas molecular weight) are incorporated in the equation.

3. Particle swarm optimization

On the basis of the stochastic optimization technique, an important population, named as, the particle swarm optimization (PSO) technology was launched by Eberhart and Kennedy (1995) which conceptually simulates the social behavior of groups, such as birds and fishes moving in to achieve their goal in a N-dimensional space (Chamkalanl et al., 2013a). Every character flies in the search space toward its optimum with a velocity which is controlled by its information as well as that provided by other companions within its neighborhood. At the time segment of \( t \), the \( i \)th individual is described as \( X_i(t) = (x_{i1}(t), x_{i2}(t), ..., x_{im}(t)) \). A set of locations of \( m \) particles is recognized as \( X = \{X_1, ..., X_m\} \). The best individual is recorded and expressed as \( P_i(t) = (P_{i1}, P_{i2}, ..., P_{im}) \). The index of the best individual among entire the particles throughout the searching history (global model) and best particle among all the particles in its own experience are represented by the symbols g and l, respectively. The velocity for particle \( i \) at time element \( t \) is expressed as \( v_i(t) = v_i(t + 1) = x_i(t) + r_1 (P_{Best} - x_i(t)) + r_2 (g_{Best} - x_i(t)) \) where \( r_1 \) and \( r_2 \) represent the positives constants, \( r_1 \) and \( r_2 \) decribe two random numbers between zero and one, and \( w \) is the symbol for inertial weight.

The linear relationship of instantaneous inertial weight is presented as follows (Shi and Eberhart, 1999):

\[ w_t = w_{max} - \frac{w_{max} - w_{min}}{T_{max}} \]  

(11)

The only alteration for the neighborhood \( l_{Best} \) system is to replace \( P_i \) with \( P_o \) for \( p_o \) in the velocity equation. In the global model, this relationship is utilized to revise a particle’s new velocity based on the information from its earlier velocity, the particle’s previous best location \( (p_{Best}) \), and the global best location \( (g_{Best}) \). The calculation methodology for the local model is almost the same and the only difference is that the best experience of neighbor particles is employed rather than the global best experience.

4. Least squares support vector machines

The support vector machine (SVM) is generally known as a strong mathematical approach to create accurate and comprehensive correlation between the variables (or parameters) of a certain mathematical problem (Chamkalanl et al., 2013b, 2013c, 2012; Chamkalanl, 2011).

A modified version of SVM named least-squares-SVM (LS-SVM) was introduced by Suykens and Vandewalle (1999). Like SVM, LS-SVM has a variety of applications in both regression and classification cases. LS-SVM normally lowers the run time and exhibits more adaptivity. Moreover, LSSVM attracted great attention because it employs an equality constraint-based formulation rather than quadratic programming techniques.

In LS-SVMs, the regression is expressed as given below:

\[ \min_{w, b} \frac{1}{2} w^T w + C \sum_{i=1}^{l} e_i \]  

s.t. \( y_i (w^T \phi(x) + b) \geq 1 - e_i \) \( i = 1, ..., l \)  

(12)

The Lagrangian equation is defined as follows:

\[ L(w, b, e, a) = \frac{1}{2} w^T w + C \sum_{i=1}^{l} e_i^2 - \sum_{i=1}^{l} a_i \left[ y_i (w^T \phi(x_i) + b) - 1 + e_i \right] \geq 0 \]  

(13)

Provided Lagrangian multipliers \( a_i \in \mathbb{R} \), if the Lagrangian equation is differentiated with respect to \( w, b, a_i \) and \( e_i \), the following...
relationships are obtained after applying the conditions:
\[
\begin{align*}
\frac{dw}{dt} &= 0 \rightarrow w = \sum_{i=1}^{l} \alpha_i y_i \phi(x_i) \\
\frac{d\alpha_i}{dt} &= 0 \rightarrow \sum_{i=1}^{l} \alpha_i y_i = 0 \\
\frac{d\alpha_i}{dt} &= 0 \rightarrow \alpha_i = C e_i, \quad i = 1, ..., l \\
\frac{dy_i}{dt} &= y_i(w^T \phi(x_i) + b) - 1 + e_i, \quad i = 1, ..., l
\end{align*}
\]
(14)

Omitting \(e\) and \(w\) leads to the Karush–Kuhn–Tucker (KKT) conditions (Fletcher, 2000)
\[
\begin{bmatrix}
0 & y^T \\
Y & \Omega + C^{-1}I
\end{bmatrix} = \begin{bmatrix}
0 \\
I
\end{bmatrix}
\]
(15)

Table 1
Characteristics and composition of degassed Caoqiao crude oil and separator gas.

<table>
<thead>
<tr>
<th>Component</th>
<th>Degassed oil</th>
<th>Separator gas</th>
</tr>
</thead>
<tbody>
<tr>
<td>CO\textsubscript{2}</td>
<td>0</td>
<td>2.96</td>
</tr>
<tr>
<td>N\textsubscript{2}</td>
<td>0</td>
<td>1.18</td>
</tr>
<tr>
<td>C\textsubscript{1}</td>
<td>0</td>
<td>89.37</td>
</tr>
<tr>
<td>C\textsubscript{2}</td>
<td>0</td>
<td>3.34</td>
</tr>
<tr>
<td>i-C\textsubscript{4}</td>
<td>0</td>
<td>2.10</td>
</tr>
<tr>
<td>n-C\textsubscript{4}</td>
<td>0</td>
<td>0.32</td>
</tr>
<tr>
<td>n-C\textsubscript{5}</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td>n-C\textsubscript{6}</td>
<td>0.58</td>
<td>0.15</td>
</tr>
<tr>
<td>n-C\textsubscript{7}</td>
<td>1.2</td>
<td>0.12</td>
</tr>
<tr>
<td>C\textsubscript{7+}</td>
<td>98.06</td>
<td></td>
</tr>
<tr>
<td>C\textsubscript{11+}</td>
<td>87.16</td>
<td></td>
</tr>
<tr>
<td>C\textsubscript{1+}, molecular weight (g/mol)</td>
<td>503.6</td>
<td></td>
</tr>
<tr>
<td>C\textsubscript{1+}, density (at 293 K)</td>
<td>0.9526</td>
<td></td>
</tr>
<tr>
<td>Reservoir Temperature</td>
<td>343</td>
<td></td>
</tr>
<tr>
<td>Bubble point pressure at 343 K (MPa)</td>
<td>9.8</td>
<td></td>
</tr>
<tr>
<td>Gas Oil Ratio, GOR, (m\textsuperscript{3}/m\textsuperscript{3})</td>
<td>30.2</td>
<td></td>
</tr>
<tr>
<td>Saturates (wt%)</td>
<td>38.0</td>
<td></td>
</tr>
<tr>
<td>Aromatics (wt%)</td>
<td>47.6</td>
<td></td>
</tr>
<tr>
<td>n-C\textsubscript{5} asphaltenes (wt%)</td>
<td>7.26</td>
<td></td>
</tr>
<tr>
<td>Resins (wt%)</td>
<td>18.8</td>
<td></td>
</tr>
</tbody>
</table>

In Eq. (15), \(C\) implies a positive constant, \(b\) stands for the bias, and \(y = (y_1, ..., y_l)^T\)
\[
\Omega = \begin{bmatrix}
y_i y_j K(x_i, x_j)
\end{bmatrix}, \quad 1 \leq i, j \leq l
\]
(16)

Applying the Mercer condition (Mercer, 1909; Courant and Hilbert, 1970), the following relationship is found as the final result:
\[
y(x) = \text{sign} \left( \sum_{i=1}^{l} \alpha_i y_i K(x_i, x_j) + b \right)
\]
(17)
In (17), the dot product \(K(x_i, x_j)\) refers to the kernel function. Kernel functions result in determination of the dot product within a high-dimension feature space where the low-dimension space input data are used without the transfer function \(\phi\). The conditions stated by (Vapnik, 1995) should be also met by the kernel function. The most usual kernel functions are polynomial, Gaussian-like or some particular sigmoids (Chamkalani et al., 2013b). In the current study, RBF is defined as \(K(x_i, x_j) = \exp(-||x_i - x_j||^2/2\sigma^2)\) where \(\sigma\) introduces a positive real number, taken into account as the kernel function.

5. Results and discussion

In this study, an optimized least square support vector machine (LSSVM) was proposed to provide a model for prediction of asphaltene precipitation. This kind of predictive models is different from scaling equations in terms of theory and application range. Thus, there are no concerns on complex nature of asphaltenes particles and their agglomeration while employing the LSSVM model (Chamkalani et al., 2013c). Furthermore, LSSVM was built on statistical and risk minimization, and not on the real molecular model. It means that the molecular interpretation is not possible through the LSSVM technique (Chamkalani et al., 2013c).

![Fig. 1. The training and testing data for different precipitants employed for model development.](image-url)
In addition, the previous models as benchmarks were selected to examine the LSSVM accuracy and generalization performance. To build the model, about 70% and 30% of all data were assigned to the training and testing phases, respectively.

This algorithmic scheme is adopted here to construct a predictive tool to estimate the precipitated asphaltene, considering the influence of n-alkane precipitants. Based on the available data, it was decided to choose known models such as Rassamdana and Sahimi (1996), Hu and Guo (2001), and Ashoori et al. (2003) for comparison purpose.

5.1. Data acquisition

Experimental data including temperature, molecular weight, dilution ratio, and asphaltene precipitated percentage were taken from the literature (Hu et al., 2004). The properties of oil for model development are listed in Table 1. Solvent to oil (or dilution) ratio, molecular weight, and temperature were introduced as the inputs where the output parameter is the percentage of asphaltene precipitated. The n-alkanes precipitants such as n-C5, n-C6, n-C7, n-C8, n-C9, n-C10, and n-C12 were used in the experiments.

Fig. 2. Developed heuristic training-based LSSVM using PSO.
The temperature range was 293–332 K, as well. All data used for both training and testing stages are given in Fig. 1.

5.2. LS-SVM and its parameters optimization with PSO

In general, least-squares support vector methods conduct non-linear regression through assigning the input parameters into a high dimensional feature space and then a conventional linear regressive case in this space is solved. Applying the LS-SVMs technique to model the nonlinear systems offers a number of benefits. For instance, intrinsic regularization in the LS-SVMs model like other kernel-based prediction methods gives efficient capabilities to conduct complexity control. To achieve optimum conditions, both the prediction problem and optimal representation of the model move in the target direction. As the solution of linear systems is normally asked, the prediction approach employing the LS-SVMs model seems to be cost-effective, numerically. Regardless of determination of a small number of tuning parameters, the case in the support vector techniques is convex and has an exact solution. According to Min and Lee (2005), the optimal parameter search acts as an important element in the SVM method while developing a prognostic tool with excellent predictive accuracy and stability. These parameters include; namely,

1. Regularization parameter $C$, which decides the tradeoff cost between the lowest level of the model complexity and the minimum magnitude of error for the training phase.
2. Parameter sigma ($\sigma^2$) involved in the kernel function that represents mapping of the non-linear relationship of the input system into the higher dimensional space (Duan et al., 2003).

Selecting the proper values for the LS-SVMs model is a difficult task and generally carried out through the trial and error procedure, while using smart technique can solve this problem in an effective and quick way. Search methods such as CSA (Chamkalani et al., 2013b, 2013c, 2012; Chamkalani, 2011), GA (Huang and Wang, 2006; Lorena and de Carvalho, 2008), SA (Sartakhti et al., 2012), ICA (Zendehboudi et al., 2013), PSO (Shafiei et al., 2013; Wei et al., 2011; Gang and Zhuping, 2011; Guo et al., 2008), ACO (Niu et al., 2010; Zhang et al., 2010; Huang, 2009), and ABC (Sulaiman et al., 2012) have been implemented to handle the problem of parameters selection. In this study, the magnitudes of the SVM parameter are determined by using the PSO technique. It has been proved that the PSO is a powerful tool to optimize the objective functions. Like other population and evolutionary-based algorithms, the optimization process is conducted without obtaining the derivatives of the target functions, in spite of gradient-based search models. Due to this strength, the derivatives of the function are not required in case the target equations are non-smooth. The proposed heuristic training-based LSSVM using PSO is shown in Fig. 2. We generally describe this optimization process as the following steps (Chamkalani et al., 2013a):

1. Assign the initial positions and velocities to the particles. Then, the maximum number of iterations, and the magnitudes of the

Table 2

<table>
<thead>
<tr>
<th>PSO parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Population size</td>
<td>40</td>
</tr>
<tr>
<td>Initial Inertia</td>
<td>0.4</td>
</tr>
<tr>
<td>Weight ($W_{\text{max}}$)</td>
<td>0.9</td>
</tr>
<tr>
<td>Final Inertia</td>
<td>0.9</td>
</tr>
<tr>
<td>Weight ($W_{\text{max}}$)</td>
<td>0.9</td>
</tr>
<tr>
<td>Learning factor ($q_1$)</td>
<td>1.4</td>
</tr>
<tr>
<td>Learning factor ($q_2$)</td>
<td>2</td>
</tr>
</tbody>
</table>

Fig. 3. Real and estimated asphaltene deposition weight percent for four different models over training data [ASP stands for the asphaltene precipitation].
inertia weight and learning parameters should be introduced. Accordingly, the particle is composed of $C$ and $\sigma^2$.

(2) Compute the fitness of each particle.

(3) Revise the location and speed (or velocity) of particles through employing Eqs. (9) and (10) to determine the updated $g_{\text{best}}$ and $p_{\text{best}}$.

(4) Check the stop condition. If the requirement (e.g., number of iterations) is met, the process is terminated and the loop returns to Step 3.

In the PSO optimization of LSSVM parameters, the LSSVM searching parameters are set as $C = [10^{-8}, 10^6]$ and $\sigma^2 = [10^{-8}, 10^0]$, respectively. A number of coefficients exist in the PSO model and their extents should be adjusted so that a better convergence rate is achieved. The values of coefficients for the particle swarm are listed in Table 2.

Based on several practical cases reported in the literature, the radial basis functions are appropriate when the general smoothness assumptions are considered, particularly if adequate knowledge of the data is not accessible. The kernel function considered for the LSSVM technique is the Gaussian radial basis function in this paper. Following the modeling procedure, the optimal values of LSSVM parameters obtained by PSO are $C = 193867.25201176$ and $\sigma^2 = 5736.47558314$.

Therefore, the PSO-LSSVM technique is employed for estimation of asphaltene deposition in this study. Although, this hybrid intelligence model is not new, utilization of the technique in this engineering direction is prime and nobody has previously introduced the PSO-LSSVM in petroleum industry, especially asphaltene subject.

5.3. Accuracy and precision benchmarks

Scaling methodology is on the basis of the assumption that asphaltenes appear to be solid particles suspended in crude oil, in the form of colloids, that are stabilized or peptized by the resin molecules adsorbed.

Dissolution of resin molecule of micelles causes asphaltene particles to aggregate. The titration experiments substantiate this by showing that addition of paraffinic compounds change the solubility of asphaltene in oil. This change is the result of solvent effect on asphaltene and resin interaction. As the dilution ratio, $R$, raises, a sharp increase in the amount of precipitated asphaltene is initially observed and then it moves toward stability. If a solvent exhibits strong tendency towards resin but does not like asphaltene, just asphaltene precipitation occurs. The main reason for this behavior is that the asphaltene molecules demonstrate lower interaction, compared to that between resin and asphaltene. The greater attraction between

Fig. 4. Residual values of the models versus number of samples for training data.
the solvent and resin implies the presence of stronger solvent (or diluent). In this case, the capability of resin molecules to cause asphaltene stabilization becomes weak, leading to asphaltene precipitation. Due to the solvent suitability of heavy normal alkanes for resins and inconsistency with asphaltene, the dilution ratio at the onset point of precipitation lowers as an increase in the carbon number of heavy normal-alkane solvents is experienced. According to Fig. 1 and the following results, we can claim that our model has intelligently captured the scaling nature of solution so that it gives the closest predictions to the experimental data.

The predicted and observed asphaltene percentages for all four models considered in the current study are depicted over training phase in Fig. 3. The figure shows how the model predicts the outputs and how close the predicted and observed values are to each other. As clear from Fig. 3, the PSO-LSSVM results satisfactorily capture the observed trend of asphaltene precipitated during production operation.

To project the comparison in a sensible manner, the residual values of all four models are presented in Fig. 4. Furthermore, the standard deviation (SD) as an index of variability that reflects a fixed distance from the mean of the distribution was determined.

![Fig. 5. Comparison of four predictive models throughout testing phases.](image)

![Fig. 6. The residual values of forecasting models for prediction of asphaltene deposition percent with their corresponding standard deviations.](image)
for all predictive models. The LSSVM technique shows the lower standard deviations. On the other hand, its residual trend does not exhibit an oscillating manner (e.g., high fluctuation in the residual values) compared to the other models.

In order to examine the forecasting ability of models, the testing data were fed to the models to yield the weight percentage of asphaltene precipitated. The comparison between the models in terms of predictive precision is shown in Fig. 5. As seen from the figure, it can be concluded from the data dispersion that the PSO-LSSVM model exhibits better performance with respect to other models as the data mostly lie around the 45° line while using the PSO-LSSVM, implying a very good closeness between the real data and predictive outcomes. Residual values for testing stage were calculated as summarized in Fig. 6. Among all models on scaling equations published in the literature, the strategy presented in this study to predict asphaltene precipitation offers an excellent ability to forecast asphaltene deposition that leads to more accurate value for this important parameter throughout oil production. This advantage might be resulted from the ability of the integrated PSO-LSSVM model to capture both linear and nonlinear patterns existing in the asphaltene deposition. In general, the PSO-based model leads to very low generalization errors on all data sets employed for prediction purpose. PSO also assists the LSSVM technique to find the proper parameters as a strong correlation is found between the input variables and output parameter.

The good performance of the hybrid approach (e.g., LSSVM-PSO) is also confirmed in Table 3 where the test data are utilized to predict asphaltene deposition through available models and then a comparison is made between the observed and predicted target values. In general, Table 3 includes various dilution ratios, molecular weights, and temperatures to generalize the results. As it is clear from the table, LSSVM-PSO surpasses the other methods in terms of accuracy.

Four indexes namely, $R^2$, MSE, ARE, and AARE are used to evaluate the forecasting ability of four models for both training and testing phases as listed in Table 4. Again, it confirms that the PSO-LSSVM model predicts the target parameter well so that its accuracy in term of coefficient of determination, mean square error, and average relative error is considerably greater than other models reported in this study. Appendix A provides the corresponding relationships for the above statistical parameters.

### Table 3

Asphaltene deposition percentage for employed models over testing data.

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Target</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dilution ratio (mL/g)</td>
<td>Molecular weight (g/mol)</td>
<td>Temperature (K)</td>
</tr>
<tr>
<td>9.75352</td>
<td>128</td>
<td>293</td>
</tr>
<tr>
<td>20.6338</td>
<td>142</td>
<td>293</td>
</tr>
<tr>
<td>24.4014</td>
<td>170</td>
<td>293</td>
</tr>
<tr>
<td>24.3756</td>
<td>72</td>
<td>308</td>
</tr>
<tr>
<td>14.3278</td>
<td>72</td>
<td>308</td>
</tr>
<tr>
<td>24.3724</td>
<td>86</td>
<td>308</td>
</tr>
<tr>
<td>2.80985</td>
<td>86</td>
<td>308</td>
</tr>
<tr>
<td>14.3568</td>
<td>100</td>
<td>308</td>
</tr>
<tr>
<td>14.418</td>
<td>170</td>
<td>308</td>
</tr>
<tr>
<td>5.05906</td>
<td>170</td>
<td>308</td>
</tr>
<tr>
<td>24.2972</td>
<td>72</td>
<td>323</td>
</tr>
<tr>
<td>3.52535</td>
<td>72</td>
<td>323</td>
</tr>
<tr>
<td>20.6336</td>
<td>86</td>
<td>323</td>
</tr>
<tr>
<td>20.5645</td>
<td>128</td>
<td>323</td>
</tr>
<tr>
<td>9.71198</td>
<td>128</td>
<td>323</td>
</tr>
<tr>
<td>14.3088</td>
<td>142</td>
<td>323</td>
</tr>
<tr>
<td>14.3278</td>
<td>114</td>
<td>323</td>
</tr>
<tr>
<td>9.56989</td>
<td>128</td>
<td>323</td>
</tr>
<tr>
<td>9.64158</td>
<td>170</td>
<td>323</td>
</tr>
<tr>
<td>5.51971</td>
<td>170</td>
<td>323</td>
</tr>
</tbody>
</table>

### Table 4

The criteria used to investigate performance and accuracy.

<table>
<thead>
<tr>
<th></th>
<th>PSO-LSSVM</th>
<th>Rassamdana and Sahimi</th>
<th>Hu and Guo</th>
<th>Ashoori et al.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>0.991</td>
<td>0.989</td>
<td>0.846</td>
<td>0.875</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0244</td>
<td>0.0268</td>
<td>0.4429</td>
<td>0.3254</td>
</tr>
<tr>
<td>ARE</td>
<td>0.0324</td>
<td>0.0727</td>
<td>0.1437</td>
<td>0.1598</td>
</tr>
<tr>
<td>AARE</td>
<td>0.0788</td>
<td>0.0892</td>
<td>0.3464</td>
<td>0.2381</td>
</tr>
</tbody>
</table>

In sum, the hybrid of PSO and LSSVM model demonstrates an excellent ability to forecast asphaltene deposition that leads to more accurate value for this important parameter throughout oil production. This advantage might be resulted from the ability of the integrated PSO-LSSVM model to capture both linear and nonlinear patterns existing in the asphaltene deposition. In general, the PSO-based model leads to very low generalization errors on all data sets employed for prediction purpose. PSO also assists the LSSVM technique to find the proper parameters as a strong correlation is found between the input variables and output parameter.

### 6. Conclusions

This study introduces a LSSVM-based particle swarm optimization model to forecast the percentage of asphaltene deposition. The main parameters such as temperature, molecular weight, and dilution ratio contribute to the predictive model in the form of input variables. The developed model uses the PSO to select the appropriate values for the hyper parameters of the LSSVM technique in order to improve the forecasting accuracy of the asphaltene deposition model. For comparison purpose, other three models including Rassamdana and Sahimi (1996), Hu and Guo (2001) and Ashoori et al. (2003) were discussed in the current...
study. The key conclusions drawn from the study are namely: firstly, the forecasting intelligence asphaltene deposition models have better performance than the regression models reported in the literature because of high non-linear fitting capacity of intelligence forecasting tools. Secondly, the PSO can obtain more appropriate magnitudes for the parameters of the LSSVM model, which can eventually improve the accuracy of the predictive model in determination of the asphaltene deposition percentage. Thirdly, compared with the other deterministic models, the MSE ARE, and AARE values attained for the LSSVM model are much smaller. The $R^2$ values also show the considerable difference between the reported forecasting models in terms of predictive performance. This means that the PSO-LSSVM model is superior over other models to predict asphaltene deposition. It is expected that the proposed PSO-LSSVM can be integrated with the black oil simulators to increase the accuracy of asphaltene deposition calculations.

Appendix A

The following statistical parameters are presented to examine the performance of the proposed models:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_{\text{observed}} - y_{\text{calculated}})^2}{\sum_{i=1}^{n} (y_{\text{observed}} - \bar{y})^2}$$

Mean square error (MSE) = \frac{\sum_{i=1}^{n} (y_{\text{calculated}} - y_{\text{observed}})^2}{n}

Mean square error (MSE) = \frac{\sum_{i=1}^{n} \left(\frac{y_{\text{calculated}} - y_{\text{observed}}}{y_{\text{observed}}}\right)^2}{n}

Average relative error (ARE) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_{\text{calculated}} - y_{\text{observed}}}{y_{\text{observed}}}\right)

Average absolute relative error (AARE) = \frac{1}{n} \sum_{i=1}^{n} \left|\frac{y_{\text{calculated}} - y_{\text{observed}}}{y_{\text{observed}}}\right|

Standard deviation (SD) = \left(\frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2\right)^{1/2}

where $y_{\text{observed}}$ and $y_{\text{calculated}}$ represent the real and calculated magnitudes of $y$, respectively. $\bar{y}$ also stands for the mean value of observed data.

References


