Modeling *n*-alkane solubility in supercritical CO₂ via intelligent methods

Reza Songolzadeh^{1,2} · Khalil Shahbazi¹ · Mohammad Madani^{1,3}

Received: 27 June 2020 / Accepted: 6 October 2020 / Published online: 22 October 2020
© The Author(s) 2020

Abstract

Injection of carbon dioxide is a familiar, cost-effective and influential technology of enhancing oil recovery whose application has been limited owing to the low *n*-alkane solubility in supercritical CO₂. Thus, determining the amount of dissolved *n*-alkane in supercritical CO₂ is of importance. Accordingly, in this study, least-squares support vector machine (LSSVM), tuned with two different optimizing algorithms, namely particle swarm optimization (PSO) and cross-validation-assisted Simplex algorithm (CV-Simplex), has been used for this simulation process. Based on the results, the predicted values for dissolved *n*-alkane mole fraction in supercritical CO₂ by PSO–LSSVM model were quite in line with experimental data. Furthermore, the accuracy of these models was compared with Chrastil correlation. Absolute average relative error for PSO–LSSVM, CV-Simplex–LSSVM and Chrastil was calculated to be 3.88%, 13.49% and 18.22% for total dataset, respectively, which leaves PSO–LSSVM as the superior model with the highest accuracy. Finally, the statistical parameters of absolute average relative error, mean square error and determination coefficient equal to 3.88%, 0.0164 and 0.994 for total dataset, respectively, proved that PSO–LSSVM model is an efficient method that can predict *n*-alkane solubility in supercritical CO₂ with high precision within 8.99–45.90 MPa pressure and 308.15–344.15 K temperature range.

Keywords CV-Simplex–LSSVM · PSO–LSSVM · *n*-alkane solubility · Supercritical CO₂

Introduction

Enhanced oil recovery (EOR) is an important field of petroleum industry that is used to increase oil recovery from petroleum reservoirs. CO₂ injection is a cost-effective and favorable method that is widely used as an EOR technology for oil industry. Due to low polarity and high density of CO₂ at supercritical state, it has high amount of solubility in oil and can extract hydrocarbon from crude

oil (Hemmati-Sarapardeh et al. 2013; Zhang et al. 2013; Abedini and Torabi 2014; Cao and Gu 2012). Because heavy fractions of oil have low solubility in supercritical CO₂, EOR applications of CO₂ injection have been limited in oil industry. Another limitation of applying CO₂ extraction phenomenon in EOR is asphaltene precipitation that creates major problems for oil recovery and production. Therefore, finding the solubility of oil fractions in supercritical CO₂ is very important (Daryasafar et al. 2018).

Due to low miscibility of heavy oil in supercritical CO₂, reservoirs with light to medium oil density are more preferable for CO₂ injection (Cao and Gu 2012, 2013; Luo et al. 2012). Cao and Gu investigated the effect of hydrocarbon polarity on the recovered oil density through CO₂ injection process (Cao and Gu 2012). According to the results of their work, the recovered oil is lighter and hydrocarbon solubility becomes lower as hydrocarbon polarity increases. Many researchers have studied phase equilibrium behavior of *n*-alkane and supercritical CO₂ system at high temperature and pressure condition (Ren and Scurto 2007; Gardeler et al. 2002; Yu et al. 2006; Eustaquio-Rincón and Trejo 2001). Choi et al. found dew and bubble points of some normal

✉ Reza Songolzadeh
reza.songolzadeh@put.ac.ir

✉ Mohammad Madani
m_madani70@yahoo.com

¹ Department of Petroleum Engineering, Petroleum University of Technology, Box 8961633967, Ahvaz, Iran

² Reservoir Engineering Studies, Petroleum Engineering Department, Main Office Building of National Iranian South Oil Company (NISOC), Ahvaz, Iran

³ Reservoir Engineering Systems, Petroleum Engineering Department, Main Office Building of National Iranian South Oil Company (NISOC), Ahvaz, Iran

alkanes (hexane, heptane, octane and nonane) contacting supercritical CO₂ for pressure and temperatures up to 153.41 bar and 395.7 K, respectively (Choi and Yeo 1998). By measuring volume expansion of alkane-CO₂ system at different temperatures and pressures in PVT apparatus, Yang et al. demonstrated that system expansion depends on the dispersion status of CO₂ molecules (Yang, et al. 2012). Chandler et al. calculated the capacity factors of solid *n*-alkanes (C₉H₂₀–C₃₆H₇₄) by estimating the solubility of these *n*-alkanes at temperature range of 308.2–348.2 K and pressure range of 100–200 bar (Chandler et al. 1996). Shi et al. studied the effects of system condition and solution properties on *n*-alkane (C₉H₂₀–C₃₆H₇₄) solubility at pressure and temperature range of 9.185–20.862 MPa and 318–343 K (Shi et al. 2015). They showed that CO₂ density and system pressure have direct relation with *n*-alkane solubility; however, temperature and the length of *n*-alkane chain are inversely proportional to it. Wang et al. measured the solubility of polar substitute in supercritical CO₂ by using compounds with aromatic nucleus (Wang et al. 2006). Furuya et al. determined the solubility of heavy *n*-alkane in supercritical CO₂ and at pressures up to 50 MPa (Furuya and Teja 2004). They used these measured values to evaluate Krichevskii parameters and showed that these parameters were directly related to *n*-alkane carbon number. Eustaquio-Rincon et al. performed experimental measurements by using a semi-flow-type setup to find *n*-octadecane solubility in supercritical CO₂ at pressure range of 10–20 MPa and temperature range of 310–353 K (Eustaquio-Rincón and Trejo 2001). The results of their work indicated that *n*-octadecane solubility has direct relation with pressure and it is maximum at the temperature of 313 K.

Expensive and time-consuming experimental methods for determining *n*-alkane solubility in supercritical CO₂ demand a fast and accurate model that can predict the solubility of different *n*-alkane at a wide range of pressure and temperature values. Accordingly, many models have been suggested for predicting *n*-alkane solubility in supercritical CO₂. Peters et al. proposed a model that relates logarithm of solute mole fraction to the logarithm of solvent carbon number at supercritical conditions (Peters et al. 1989). Jha and Madras predicted the solubility of *n*-alkane hydrocarbon in supercritical CO₂ using Peng–Robinson EOS, quadratic mixing rules and an adjustable parameter (Jha and Madras 2004). They proved that there is a linear relation between the adjustable parameter and carbon number of the *n*-alkane largest chain. The limitation of their model is the defined adjustable parameter that needs time-consuming laboratory measurements. Some researchers (Gordillo et al. 1999; Sung and Shim 1999; Keshmiri et al. 2014) have suggested their semiempirical models for simulating *n*-alkane solubility in supercritical CO₂. Chrastil proposed a correlation between supercritical CO₂ density and the solute solubility (Chrastil

1982). The simplicity and accuracy of this model have convinced many researchers to use it in their studies. Moreover, a number of machine learning algorithms have been used in the literature to model solubility of *n*-alkane in supercritical CO₂ (Huang et al. 2019a, b; Sedaghat and Esfandiarian 2019; Abdi-Khanghah et al. 2018). Daryasafar et al. investigated the viability of using artificial techniques for predicting the solubility of *n*-alkane in supercritical CO₂. Based on the results of their study, intelligent tools can be good candidates for predicting *n*-alkane solubility (Daryasafar et al. 2018).

In this work, least-squares support vector machine (LSSVM) technique has been utilized to predict *n*-alkane solubility in supercritical CO₂. Modeling with LSSVM tools has some advantages as high flexibility in solving nonlinear regression, lowering the required time and cost of regression process (Gorjaei et al. 2015; Hou et al. 2009). In this study, two different LSSVM-based models including PSO optimized LSSVM (PSO–LSSVM) and cross-validation Simplex–LSSVM (CV–Simplex–LSSVM) models have been applied to model and predict *n*-alkane solubility in supercritical CO₂. The key purpose of this study is to propose a comprehensive and accurate model that can relate the influential parameters such as molar weight of *n*-alkane, temperature and pressure of the system and CO₂ density to *n*-alkane solubility in supercritical CO₂.

Model development for determining solubility in supercritical CO₂

Least-squares support vector machine (LSSVM)

Support vector machine is one of the artificial intelligence methods that uses statistical calculation through learning process and can be implemented for both classifying and regressing big amount of data (Mesbah et al. 2014; Vapnik 2013; Cortes and Vapnik 1995). SVM is a powerful and reliable tool that solves nonlinear problems by mapping input data from input space to a feature space (Bassir and Madani 2019; Fayyaz et al. 2019). The data transformation into the feature space updates all nonlinear relations of input space to linear relations in the feature space. Another ability of SVM that makes it unique among other artificial intelligent tools is its flexibility in using the deviated data through model training process. Actually, SVM does not discard any deviated data and allows them to make part through training process by a penalty factor (γ), which is a tuning parameter.

In case of large-sized datasets, regression with SVM demands quadratic programming that makes computation difficult and time-consuming (Suykens et al. 2002; Suykens and Vandewalle 1999). Recently, a new modification for SVM, named LSSVM, has been introduced by Suykens

and Vanderwalle (1999) that changes inequality constraints of conventional SVM into equality constraints. As a result, LSSVM solves a linear equation system instead of quadratic programming. Accordingly, LSSVM can solve large-scale data with high convergence speed and accuracy.

Similar to other artificial intelligent tools, a nonlinear regression function f is determined by using map function (ϕ) and evaluating w and b in Eq. 1. LSSVM minimizes J in Eq. 2 subject to linear constrain, Eq. 3, to find w (Suykens and Vandewalle 1999):

$$f(x) = \omega\phi(x) + b \quad (1)$$

$$\min J(w, e) = \frac{1}{2} \|\omega^2\| + \frac{1}{2} \gamma \sum_{i=1}^N e_i^2 \quad (2)$$

$$y_i = \omega\phi(x_i) + b + e_i, \quad i = 1, \dots, N \quad (3)$$

where e_i s are slack variables, ω is weight factor, b is bias, and y_i s are matching data. By applying Lagrange multiplier method for minimizing Eq. 2 with regard to linear constrain of Eq. 3, the regression function is reformulated as below (Gorjaei et al. 2015; Farasat et al. 2013; Zhang and Ge 2019)

$$L_{\text{LSSVM}} = \frac{1}{2} \|\omega^2\| + \gamma \frac{1}{2} \sum_{i=1}^N e_i^2 - \sum_{i=1}^N \alpha_i (\omega\phi(x_i) + b + e_i - y_i) \quad (4)$$

where α_i s are Lagrange multipliers. The optimal conditions for Eq. 4 are

$$\begin{cases} 1. \frac{\partial L_{\text{LSSVM}}}{\partial b} = 0 \rightarrow \sum_{i=1}^N \alpha_i = 0 \\ 2. \frac{\partial L_{\text{LSSVM}}}{\partial \omega} = 0 \rightarrow \omega = \sum_{i=1}^N \alpha_i \phi(x_i) \\ 3. \frac{\partial L_{\text{LSSVM}}}{\partial \alpha_i} = 0 \rightarrow \{(\omega^T \phi(x_i)) + b + e_i\} - y_i = 0 \\ 4. \frac{\partial L_{\text{LSSVM}}}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i \end{cases} \quad i = 1, \dots, N \quad (5)$$

where LSSVM variables including dependent and independent are separable. LSSVM can be formulated as below (Gorjaei et al. 2015):

$$y = \sum_{i=1}^N \alpha_i x_i^T x + b \quad (6)$$

Equation 4 is only valid for linear regression problems. In case of nonlinear regression, $\phi^T(x_i)\phi(x)$ is substituted with $x_i^T x$. Because dot product between input maps ($\phi(x_i)$) makes the explicit computation complex, some kernel functions have been suggested to evaluate the kernel matrix implicitly. The mainly used kernel functions are linear, polynomial and radial basis function (RBF). In this study, regarding lower tuning parameter and better numerical verification (Keerthi

and Lin 2003; Vapnik and Lerner 1963; Na'imi et al. 2014), RBF kernel has been used. RBF kernel is defined as:

$$k(x_i, x) = \phi(x_i)\phi(x) = \exp\left(\frac{-\|x_i - x\|^2}{\sigma^2}\right) \quad i = 1, \dots, N \quad (7)$$

where σ^2 represents RBF kernel width and needs to be tuned by an external optimizer before training LSSVM model (Gorjaei et al. 2015; Zhang et al. 2009).

Parameters optimization

LSSVM at least has two tuning parameters. These parameters need to be optimized before training the model. In this study, two optimization algorithms, namely particle swarm optimization (PSO) and cross-validation Simplex algorithm, were used to optimize the tuning parameters.

Particle swarm optimization

Kennedy and Eberhart suggested an optimization algorithm named particle swarm optimization (PSO) for finding optimal value of a continuous space (Eberhart and Kennedy 1995). PSO applies insect swarm behavior and evolutionary programming theory to find the global best of a function or space. Through optimization process, particles with no volume explore the search space to discover the optimum location. Velocity and position are two specifications of the particles that are updated through each step. In the first step, PSO uses random function to define initial particle position and initial particle velocity. Through each iteration, PSO seeks the minimum value of the space of each neighborhood (local best) and the minimum all over the space (global best). Using the following equations, position and velocity of particles are updated for next step (Eberhart and Kennedy 1995; Shi and Eberhart 1999)

$$v_i^{n+1} = \omega v_i^n + c_1 r_1^n (x_{i\text{Lb}}^n - x_i^n) + c_2 r_2^n (x_{\text{gb}}^n - x_i^n) \quad (8)$$

$$x_i^{n+1} = x_i^n + v_i^{n+1} \quad (9)$$

where v is the velocity of the particles, n denotes iteration number, x is particle position at i iteration, $x_{i\text{Lb}}$ and x_{gb} represent local and global best, ω is moment of inertia, c_1 and c_2 are acceleration factors and usually are about 2, and r_1 and r_2 represent random numbers and change in $[0,1]$ interval.

For the next step, all particles except global best's velocity are updated toward global best of the last step. Through optimization process, particle migration continues until all particles converge to a position with lowest cost value.

Data preparation

In this work, using data mentioned in (Eustaquio-Rincón and Trejo 2001; Shi et al. 2015; Furuya and Teja 2004), a complete data bank was gathered to generate comprehensive models for predicting the solubility of *n*-alkane in supercritical CO₂. To generate a reliable model, it is crucial to use influential parameters as input data. According to Chrastil (1982) (Chrastil 1982), temperature and carbon dioxide density are two parameters affecting the *n*-alkane solubility in solubility in supercritical CO₂. In addition, *n*-alkane carbon number (*n*-alkane molar weight) and pressure are influential in this regard (Yang et al. 2012; Chandler et al. 1996; Shi et al. 2015). Therefore, input data include temperature (K), pressure (MPa), normal alkane molecular weight (g/mol) and carbon dioxide density (kg/m³). A set of 180 input points are used to propose a model for predicting the mole fraction of dissolved *n*-alkane in supercritical CO₂. The summary of the gathered data is tabulated in Table 1.

Optimizing LSSVM model using PSO

Determining an optimum penalty factor γ and tuning kernel parameter are the prerequisites of using LSSVM. One of the recommended kernel functions for nonlinear system is RBF which contains only one tuning parameter σ^2 (Gorjaei et al. 2015). Optimizing γ and σ^2 before starting training process can strongly assist LSSVM in determining the support vectors. For this reason, the data points are divided into three groups, namely training, validating and testing data. The number of these groups must be selected cautiously so that enough training data are available for model training and validating and adequate data are considered for testing the model. After several try and errors, the best division of 90% training data and 10% test data was found for this work. The swarm size of 30, iteration number of 200 and minimum MSE of 0.0001 were defined before starting the optimization.

Through optimization process and at each iteration, LSSVM is trained for all particle positions, γ and σ^2 . MSE of the outputs indicates the accuracy of the trained model for the considered position. Finally, the location of the particle with lowest MSE is the global best of the system. The details of optimization process are demonstrated in Fig. 1.

Optimizing LSSVM model using cross-validation Simplex technique

Simplex is a popular algorithm that has widely been suggested for linear optimization (Klee and Minty 1972). Through Simplex optimization, the minimum of a function is found by searching the function in a limited interval. The interval is updated at each iteration by reflection, contraction and expansion techniques (Nelder and Mead 1965). Interval update continues until function minimum with acceptable error is found. In this study, cross-validation method has been applied as cost function evaluation technique for LSSVM function.

Cross-validation (CV) is one of the essential techniques in sample selection and optimization validation technique (Schaffer 1993; Browne 2000). To evaluate cost value, CV divides training dataset T into K subsets ($T_1, T_2, T_3, \dots, T_K$) and trains LSSVM K times. At each training iteration, one T set is reserved for validating the model and other sets participate in support vector training process. Finally, average MSE of all reserved sets will be considered as the cost value.

Results and discussion

In this study, PSO–LSSVM and CV–Simplex–LSSVM models have been trained to predict the relationship between solution properties and *n*-alkane solubility in CO₂. A total number of 180 experimental data with division of 162 points for model training and 18 points for model testing are used in model preparation. The developed models are compared with Chrastil correlation which is as follows (Chrastil 1982):

$$S = \rho^k \exp\left(\frac{m}{T} + n\right) \quad (10)$$

where T (K) is temperature, ρ (kg/m³) is CO₂ density, k , m , n are characteristic constants of the correlation, and S (kg/m³) is *n*-alkane solubility defined as:

$$S = \frac{x\rho M_1}{(1-x)M_0} \quad (11)$$

Table 1 Range of input and output parameters utilized in this study

Reference	Pressure (MPa)	Temperature (K)	<i>n</i> -alkane molar weight (g/mol)	CO ₂ density (kg/m ³)	Solubility (fraction)
Eustaquio-Ricon and Trejo (2001)	8.987–20	308.15–344.15	226.41–254.5	204.83–864.72	0.003183–0.0648
Furuya and Teja (2004)	14.2–45.9	313	338.65–408.6	767.32–956.6	0.001–0.00106
Shi et al. (2015)	9.185–20.862	313–343	86.18–254.5	214.93–847.44	0.0054–0.0677

Fig. 1 Flowchart of PSO–LSSVM optimization: **a** PSO optimization **b** cost function calculation

where x represents mole fraction of dissolved n -alkenes in supercritical CO_2 , and M_0 and M_1 are, respectively, the molar weight of CO_2 and n -alkane.

As mentioned before, LSSVM with RBF kernel function has two tuning parameters σ^2 and γ that have been determined by two different optimization algorithms (PSO and CV-Simplex) in this work. In PSO optimization, the data were divided into three subgroups, namely training, validating and testing data. Dividing data into these groups is an important part of optimization since if the portion of training data is small, the model will be inaccurate and if it is large, the model will be over-trained. Therefore, data were selected randomly and the portion of each group was determined after multiple runs. Consequently, 59% of data

were chosen for training the model, 31% were kept for model validation, and 10% were reserved for testing the model. For Simplex case model, validity was checked by applying k -fold cross-validation technique. For this purpose, a total number of tenfold were assigned to the model. K -fold CV allows the optimization algorithm to select training and validating data with alike data distribution. To be confident that the validation process is not dependent on the data points or group subdivision, the validation is repeated ten times with different validation groups and average MSE of validation data of these recurrences will be considered as cost value. Therefore, in CV-Simplex optimization, LSSVM model needs to be trained 10 times for each Simplex iteration.

PSO determined the values of tuned σ^2 and γ 4.958 and 7.284×10^5 ; however, CV-Simplex adjusted these parameters to 2.527×10^5 and 2.034×10^7 , respectively. By using these tuned parameters, two different LSSVM models were trained with training data and the accuracy of these models were compared with some statistical parameters including MSE. It is worth mentioning that the determined value for σ^2 by CV-Simplex optimizer was abnormally high, but the results of CV-Simplex–LSSVM model were reasonable. Correlation coefficient (R^2) is another statistical parameter defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i^{\text{ac}} - y_i^{\text{sim}})^2}{\sum_{i=1}^N (y_i^{\text{ac}} - \bar{y}^{\text{ac}})^2} \quad (12)$$

where $\bar{y}^{\text{ac}}$ is average value of actual data, y_i^{sim} is the simulated value for data with index i , and y_i^{ac} is the actual data with index i .

The predicted values of dissolved n -alkane mole fraction in supercritical CO_2 for both PSO–LSSVM and CV-Simplex–LSSVM models are demonstrated in Figs. 2, 3, 4 and 5 for training and testing datasets. The dotted line shown in the figures is the best linear fit for model predicted results, while the solid lines ($Y=X$) represent the prediction with no error. According to these figures, the simulation accuracy of CV-Simplex–LSSVM is not as high as PSO–LSSVM model. Also, the figures show that PSO–LSSVM model can better track actual data for both training and testing data; however, CV-Simplex–LSSVM cannot predict testing as well as training datasets. This difference in accuracy proves the superiority of PSO algorithm.

Fig. 2 PSO–LSSVM predicted mole percent of dissolved n -alkane versus actual data for training dataset

Fig. 3 CV-Simplex–LSSVM predicted mole percent of dissolved n -alkane for training dataset

The determined linear fit for training data of both PSO–LSSVM and CV-Simplex–LSSVM is obtained as, respectively:

$$y = 1.0x : R^2 = 0.997 \quad (13)$$

$$y = 0.98x : R^2 = 0.978 \quad (14)$$

As Eq. 13 shows, the linear fitted line for PSO–LSSVM model completely overlaps $y = x$ line and R^2 of training data for this model is 0.997; however, R^2 of CV-Simplex–LSSVM is 0.978 and the fitted line for correlated outputs is slightly deviated from $y = x$.

The same linear fit lines are calculated and demonstrated in Eqs. 15 and 16 for testing data in PSO–LSSVM and CV-Simplex–LSSVM models, respectively:

$$y = 1.08x - 1.0 : R^2 = 0.94 \quad (15)$$

$$y = 0.9x : R^2 = 0.857 \quad (16)$$

According to PSO–LSSVM calculated R^2 of 0.997 and 0.94 for training and testing datasets, the proposed PSO–LSSVM model has high capability in predicting the solubility of n -alkane in supercritical CO_2 at a wide range of system conditions.

To compare the precision of the proposed model with analytical models of literature, the predicted outputs of the model and Chrastil correlation calculated outputs have been plotted versus actual data in Fig. 6. As this figure and the linear lines show, PSO–LSSVM capability in predicting n -alkane solubility is excellent and its results are reliable.

Fig. 4 PSO–LSSVM predicted mole percent of dissolved *n*-alkane for testing dataset

Fig. 6 Predicted mole percent of dissolved *n*-alkane by Chrastil, CV-Simplex and PSO–LSSVM models versus actual data

Fig. 5 CV-Simplex–LSSVM predicted dissolved mole percent of *n*-alkane for testing dataset

The relative error percent of model outputs calculated by Eq. 17 are shown in Fig. 7. The relative error percent for both train and testing datasets for PSO–LSSVM model change in the range of -20.02% to 20.18% with minimum absolute error of 3.88% ; however, the relative error percent for CV-Simplex–LSSVM model repose in span of -68.0% to 83.01% with minimum absolute error of 13.49% . Chrastil model can also predict the outputs with relative error of -75.92% to 63.68% . According to Fig. 7, the superiority of PSO–LSSVM, compared to other

models, in predicting *n*-alkane solubility in supercritical CO_2 is inevitable.

$$\text{Relative Error Percent} = \frac{y_{ac} - y_{pred}}{y_{ac}} \times 100 \quad (17)$$

To better compare and find the robustness of the developed models, a number of statistical parameters such as AARE, MSE and R^2 of PSO–LSSVM, CV-Simplex–LSSVM and Charstil models are reported in Table 2. The values of these parameters for training and testing datasets prove the superiority of PSO–LSSVM model in comparison with other models, which is due to method that PSO uses to find the minimum by seeking all the space without limiting the search space.

The developed PSO–LSSVM model can be used to accurately predict *n*-alkane solubility in supercritical carbon dioxide without requiring laboratory methods which are inherently expensive and time-consuming. Note that the limitation in utilizing this model is that it is solely applicable within the range of input parameters (See Table 1).

Conclusions

In this work, the capability of implementing intelligent tools for determining *n*-alkane solubility in supercritical CO_2 was studied. For this purpose, least-squares support vector machine (LSSVM) tuned with two different algorithms, namely particle swarm optimization (PSO) and cross-validation-compensated Simplex was trained to predict the molar fraction of *n*-alkane solved in supercritical

Fig. 7 Relative error percent of **a** PSO-LSSVM, **b** CV-Simplex-LSSVM and **c** Chrastil models in predicting *n*-alkane mole percent

Table 2 Comparison between statistical parameters of Chrastil, PSO-LSSVM and CV-Simplex-LSSVM models in predicting *n*-alkane solubility in supercritical CO₂

Model/dataset	AARE (%)	MSE	R ²
<i>Chrastil</i>			
Train	9.89	0.0145	0.962
Test	23.11	0.0901	0.937
Overall	18.22	0.1042	0.958
<i>PSO-LSSVM</i>			
Train	2.45	0.0083	0.997
Test	9.36	0.2657	0.939
Overall	3.88	0.0164	0.994
<i>CV-Simplex-LSSVM</i>			
Train	0.51	0.0489	0.978
Test	31.21	0.4840	0.856
Overall	13.49	0.0720	0.970

CO₂. The considered influential factors for this prediction were temperature, pressure, *n*-alkane molecular weight, and carbon dioxide density. According to the results, PSO-LSSVM model was more powerful and accurate than CV-Simplex tuned LSSVM. By using evolutionary finding technique, PSO could efficiently locate the minimum of a function with local minimums and steep varying curvature. The performance of PSO-LSSVM model was compared with Chrastil model by calculating relative error percent of these models. PSO-LSSVM model with MSE value of 0.016, R² of 0.994 and average relative error percent of 3.88% was proved to be an efficient method for calculating *n*-alkane solubility in supercritical CO₂ at wide

range of pressure and temperature values and can be used with high confidence and reliability.

Funding This study has received no funding at all.

Compliance with ethical standards

Conflict of interest On behalf of all the co-authors, the corresponding authors state that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Abdi-Khanghah M et al (2018) Prediction of solubility of N-alkanes in supercritical CO₂ using RBF-ANN and MLP-ANN. J CO₂ Util 25:108–119
- Abedini A, Torabi F (2014) Oil recovery performance of immiscible and miscible CO₂ huff-and-puff processes. Energy Fuels 28(2):774–784
- Bassir SM, Madani M (2019) A new model for predicting asphaltene precipitation of diluted crude oil by implementing LSSVM-CSA algorithm. J Pet Sci Technol 37(22):2252–2259
- Browne MW (2000) Cross-validation methods. J Math Psychol 44(1):108–132

- Cao M, Gu Y (2012) Physicochemical characterization of produced oils and gases in immiscible and miscible CO₂ flooding processes. *Energy Fuels* 27(1):440–453
- Cao M, Gu Y (2013) Oil recovery mechanisms and asphaltene precipitation phenomenon in immiscible and miscible CO₂ flooding processes. *Fuel* 109:157–166
- Chandler K, Pouillot FL, Eckert CA (1996) Phase equilibria of alkanes in natural gas systems. 3. Alkanes in carbon dioxide. *J Chem Eng Data* 41(1):6–10
- Choi E-J, Yeo S-D (1998) Critical properties for carbon dioxide+n-alkane mixtures using a variable-volume view cell. *J Chem Eng Data* 43(5):714–716
- Chrastil J (1982) Solubility of solids and liquids in supercritical gases. *J Phys Chem* 86(15):3016–3021
- Cortes C, Vapnik V (1995) Support-vector networks. *Mach Learn* 20(3):273–297
- Daryasafar A et al (2018) Connectionist approaches for solubility prediction of n-alkanes in supercritical carbon dioxide. *Neural Comput Appl* 29(1):295–305
- Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In: Proceedings of the sixth international symposium on micro machine and human science, 1995. MHS'95. IEEE
- Eustaquio-Rincón R, Trejo A (2001) Solubility of n-octadecane in supercritical carbon dioxide at 310, 313, 333, and 353 K, in the range 10–20 MPa. *Fluid Phase Equilib* 185(1–2):231–239
- Farasat A et al (2013) Toward an intelligent approach for determination of saturation pressure of crude oil. *Fuel Process Technol* 115:201–214
- Fayyaz K et al (2019) Accurate prediction of electrical conductivity of ionic liquids+ propylene carbonate binary mixtures. *J Mol Liq* 279:400–410
- Furuya T, Teja AS (2004) The solubility of high molecular weight n-alkanes in supercritical carbon dioxide at pressures up to 50 MPa. *J Supercrit Fluids* 29(3):231–236
- Gardeler H, Fischer K, Gmehling J (2002) Experimental determination of vapor–liquid equilibrium data for asymmetric systems. *Ind Eng Chem Res* 41(5):1051–1056
- Gordillo M et al (1999) Solubility of the antibiotic Penicillin G in supercritical carbon dioxide. *J Supercrit Fluids* 15(3):183–190
- Gorjaei RG et al (2015) A novel PSO-LSSVM model for predicting liquid rate of two phase flow through wellhead chokes. *J Nat Gas Sci Eng* 24:228–237
- Hemmati-Sarapardeh A et al (2013) Experimental determination of interfacial tension and miscibility of the CO₂–crude oil system; temperature, pressure, and composition effects. *J Chem Eng Data* 59(1):61–69
- Hou L, Yang Q, An J (2009) An improved LSSVM regression algorithm. In: 2009 international conference on computational intelligence and natural computing. IEEE
- Huang X et al (2019a) CSA-LSSVM model for the estimation of solubility of n-alkane in supercritical CO₂. *J Energy Sour Part A Recovery Util Environ Eff* 41(16):1938–1948
- Huang X et al (2019b) Rigorous modeling of solubility of normal alkanes in supercritical CO₂. *Pet Sci Technol* 37(9):1008–1015
- Jha SK, Madras G (2004) Modeling the solubilities of high molecular weight n-alkanes in supercritical carbon dioxide. *Fluid Phase Equilib* 225:59–62
- Keerthi SS, Lin C-J (2003) Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Comput* 15(7):1667–1689
- Keshmiri K, Vatanara A, Yamini Y (2014) Development and evaluation of a new semi-empirical model for correlation of drug solubility in supercritical CO₂. *Fluid Phase Equilib* 363:18–26
- Klee V, Minty GJ (1972) How good is the simplex algorithm. *J Inequal* 3(3):159–175
- Luo P et al (2012) Propane-enriched CO₂ immiscible flooding for improved heavy oil recovery. *Energy Fuels* 26(4):2124–2135
- Mesbah M et al (2014) Prediction of phase equilibrium of CO₂/cyclic compound binary mixtures using a rigorous modeling approach. *J Supercrit Fluids* 90:110–125
- Na'imi SR, Gholami A, Asoodeh M (2014) Prediction of crude oil asphaltene precipitation using support vector regression. *J Dispers Sci Technol* 35(4):518–523
- Nelder JA, Mead R (1965) A simplex method for function minimization. *Comput J* 7(4):308–313
- Peters C et al (1989) On the relationship between the carbon-number of n-paraffins and their solubility in supercritical solvents. *Fluid Phase Equilib* 52:389–396
- Ren W, Scurto AM (2007) High-pressure phase equilibria with compressed gases. *Rev Sci Instrum* 78(12):125104
- Schaffer C (1993) Selecting a classification method by cross-validation. *Mach Learn* 13(1):135–143
- Sedaghat M, Esfandiarian A (2019) On the prediction of solubility of alkane in carbon dioxide using a novel ANFIS-GA method. *J Energy Sour Part A Recovery Util Environ Eff* 41(23):2861–2867
- Shi Y, Eberhart RC (1999) Empirical study of particle swarm optimization. In: Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406). IEEE
- Shi Q, Jing L, Qiao W (2015) Solubility of n-alkanes in supercritical CO₂ at diverse temperature and pressure. *J CO₂ Util* 9:29–38
- Sung H-D, Shim J-J (1999) Solubility of CI disperse red 60 and CI disperse blue 60 in supercritical carbon dioxide. *J Chem Eng Data* 44(5):985–989
- Suykens JA et al (2002) Basic methods of least squares support vector machines. *Least Sq Support Vector Mach* 71–116
- Suykens JA, Vandewalle J (1999) Least squares support vector machine classifiers. *Neural Process Lett* 9(3):293–300
- Vapnik V (2013) The nature of statistical learning theory. Springer, Berlin
- Vapnik V, Lerner A (1963) Generalized portrait method for pattern recognition. *Autom Remote Control* 24(6):774–780
- Wang B et al (2006) Solubility of chlorobutane, ethyl methacrylate and trifluoroethyl acrylate in supercritical carbon dioxide. *Fluid Phase Equilib* 239(1):63–68
- Yang Z et al (2012) Dispersion property of CO₂ in oil. 1. Volume expansion of CO₂+ alkane at near critical and supercritical condition of CO₂. *J Chem Eng Data* 57(3):882–889
- Yu J, Wang S, Tian Y (2006) Experimental determination and calculation of thermodynamic properties of CO₂+ octane to high temperatures and high pressures. *Fluid Phase Equilib* 246(1–2):6–14
- Zhang X, Ge Z (2019) Local parameter optimization of LSSVM for industrial soft sensing with big data and cloud implementation. *IEEE Trans Ind Inf* 16(5):2917–2928
- Zhang W, Li C, Zhong B (2009) LSSVM parameters optimizing and non-linear system prediction based on cross validation. In: Fifth international conference on natural computation, 2009. ICNC'09. IEEE
- Zhang P, Yang H-J, Xu L (2013) Solubilities and partial molar volumes of new CO₂-philic propane derivatives in supercritical carbon dioxide. *J Chem Thermodyn* 67:234–240

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.