Fuel Processing Technology 115 (2013) 201–214

Contents lists available at SciVerse ScienceDirect

Fuel Processing Technology

journal homepage: www.elsevier.com/locate/fuproc

Toward an intelligent approach for determination of saturation pressure of crude oil

Amir Farasat [a,b], Amin Shokrollahi [a], Milad Arabloo [a], Farhad Gharagheizi [c,d], Amir H. Mohammadi [c,e,*]

[a] Department of Chemical and Petroleum Engineering, Sharif University of Technology, Tehran, Iran
[b] Reservoir Engineering Division, Iranian Offshore Oil Company, Tehran, Iran
[c] Thermodynamics Research Unit, School of Chemical Engineering, University of KwaZulu-Natal, Howard College Campus, King George V Avenue, Durban 4041, South Africa
[d] Department of Chemical Engineering, Buinzahra Branch, Islamic Azad University, Buinzahra, Iran
[e] Institut de Recherche en Génie Chimique et Pétrolier (IRGCP), Paris Cedex, France

ARTICLE INFO

Article history:
Received 12 April 2013
Received in revised form 18 June 2013
Accepted 19 June 2013
Available online xxxx

Keywords:
Bubble point pressure
Crude oil
Support vector machine
Empirical correlations
Isothermal swelling test
Computer program

ABSTRACT

Bubble point pressure is a crucial PVT parameter of reservoir fluids, which has a significant effect on oil field development strategies, reservoir evaluation and production calculations. This communication presents a new mathematical model to calculate the saturation pressures of crude oils as a function of temperature, hydrocarbon and non-hydrocarbon reservoir fluid compositions, and characteristics of the heptane-plus fraction. The model was developed and tested using a total set of 130 experimentally measured compositions and saturation pressures of crude oil samples from different geographical locations covering wide ranges of crude oil properties and reservoir temperatures. In-depth comparative studies have been carried out between this new model and five well known predictive models for estimation of saturation pressure of crude oils. The results show that the developed model is more accurate and reliable with the average absolute relative deviation of 4.7% and correlation coefficient of 0.992. In addition, it is shown that the proposed model correctly captures the physical trend of changing the saturation pressure as a function of the input variables. Finally, the applicability domains of the proposed model and quality of the existing experimental data were examined by outlier diagnostics.
© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Knowledge of reservoir fluid properties is important in petroleum engineering computations, including material balance calculations, well test analysis, reserve estimates, inflow performance calculations, surface facility design, and numerical reservoir simulation [1–4]. These properties can be obtained either by conducting a laboratory study on reservoir fluids or estimation from models or correlations. Among all PVT properties, the bubble point pressure (P_b) is one of the most essential factors in reservoir and production computations [5–7].

The bubble point pressure of a hydrocarbon system is defined as the maximum pressure at which the first bubble of gas begins to evolve from the oil. This important property can be measured experimentally for a crude oil system by conducting a constant-composition expansion (CCE) experiment. As bubble point pressure is used, either directly or indirectly, in almost all correlations for prediction of crude oil properties, any error in estimation of bubble point pressure will propagate throughout estimates of other fluid properties such as oil formation volume factor, oil viscosity, oil density, etc. Hence, it is absolutely necessary that the prediction for bubble point pressure be as accurate as possible [5–7].

Although PVT sampling and experiments provide reliable results, they are relatively expensive, time consuming and also results are heavily dependent on the validity of the reservoir fluid samples, especially when the reservoir has depleted below the bubble point pressure [1,3,8]. Numerous numbers of equations of states (EOS) [9–12] have been published in the literature to model reservoir fluid phase behavior but none of them can be considered to be so accurate to predict all properties of reservoir fluids at all conditions [13–15]. Moreover, the equation of state requires the extensive knowledge of the detailed compositions of the reservoir fluids and the determination of such quantities is expensive and time consuming [16]. Furthermore, the accuracy of the EOS prediction depends heavily on the nature of the fluid, on the type of selected equation and on the operator-dependent tuning procedures. This method also involves several numerical computations [16,17].

To overcome the aforementioned problems, equations of state and several empirically derived correlations and eventually soft computing techniques have been developed based on available data of different regions of the world. A large number of correlations for estimation of bubble point pressure have been offered in the literature over the last few years to go with a handful of correlations published earlier [18–26]. Many of these new correlations are based on data from a single geographical area and cannot be used globally due to different characteristics of fluids in each area [16,24].

* Corresponding author at: Institut de Recherche en Génie Chimique et Pétrolier (IRGCP), Paris Cedex, France.
E-mail address: a.h.m@irgcp.frcp (A.H. Mohammadi).

0378-3820/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.fuproc.2013.06.007

Since the 1950s, many researchers have tried to develop correlations for predicting saturation pressure of reservoir fluids with different degrees of accuracy. The first work was done by Standing [18] on 22 hydrocarbon systems in California. Standing's correlation was based on the assumption that the bubble point pressure is a function of gas solubility, gas gravity, oil gravity, and reservoir temperature. Based on Henry's law, Lasater [20] developed a correlation for estimation of bubble point pressure on 158 samples from 137 reservoirs in Canada, the United States and South America. This correlation is based on systems essentially free from non-hydrocarbon components. Glaso [21] used a total of 45 oil samples from the North Sea to develop graphical form models for bubble point pressure, saturated oil formation volume factor, total oil formation volume factor and dead oil viscosity. This correlation provides corrections for the presence of non-hydrocarbon components such as H_2, N_2, H_2S and the paraffinicity of the oil. Al-Marhoun [22] used 160 data set points to develop correlations for estimating bubble point pressure and oil formation volume factor from 69 Middle Eastern hydrocarbon mixtures. In 1992, Dokla and Osman [27] presented correlations for estimating the bubble point pressure and the oil formation volume factor for UAE crude oil. In 1993, Petrosky and Farshad [28] developed correlations for the Gulf of Mexico crude oil samples using 90 PVT analysis data set. Velarde et al. [29] proposed a new correlation for bubble point pressure based on 2097 PVT data sets from 97 laboratory reports by introducing a new coefficient in the equation presented by Petrosky and Farshad [28]. In addition, other correlations were also proposed by researchers through different regression techniques. Valko and McCain [24] provided an in-depth analysis of various bubble point correlations.

Although, large numbers of correlations for estimation of bubble point pressure of reservoir oils have been offered in the literature over the last few years, many of these new correlations are based on data from a single geographical area. Most of these correlations were derived using petroleum service company laboratory fluid property data [24].

To overcome the shortcomings associated with the earlier correlation methods, researchers have utilized artificial intelligence based methods foremost of which is the classical artificial neural network (ANN) and its variants [30–36].

Gharbi and Elsharkawy [37] presented models for prediction of bubble point pressure as functions of solution gas–oil ratio, gas specific gravity, oil specific gravity, and temperature for Middle Eastern crude oils. Gharbi et al. [34] used universal multilayer perceptron based models for prediction of bubble point pressure using 5200 data points from all over the world. Al-Marhoun and Osman [32] presented correlations for bubble point pressure of Saudi Arabian crudes using ANN. This study shows a significant increase in accuracy over Al-Marhoun's previous correlations [22,38], also developed for Saudi crude. El-Sebakhy et al. [39] used support vector regression for generating correlations for bubble point pressure and saturated oil formation volume factor using three different published PVT databases. Moghadam et al. [40] used ANN for prediction of bubble point pressure based on 218 crude oil PVT data points from different locations in Iranian reservoirs.

The developed neural network correlations often do not perform to expectations and generally have several shortcomings that include; ANN is normally a black box modeling scheme that is based on the trial-and-error approach in which its architectural parameters have to be guessed in advance, such as, number and size of hidden layers, learning rate, training algorithm parameters, initial random weights and most importantly the tendency to be stocked at the local minima, and its inability to handle uncertainties [16,41]. The hidden layer is also disadvantageous since it does not reflect the physical phenomena of the process that is being simulated. Hence, while it shows precision within the training data set, it cannot predict accurately the 'unseen' data sets. This is indicated in McCain et al. [42].

The existent complexity, fuzziness and uncertainty existent in addition to nonlinear behavior of most reservoir parameters require a powerful tool to overcome these challenges [43]. The recent development and success of applying support vector machine modeling to solve various difficult engineering problems has drawn the attention to its potential applications in the petroleum industry [4,44–46]. In the last decades, various prediction models have been proposed that include artificial neural network models [47,48], adaptive neuro-fuzzy inference system (ANFIS) models [49,50] and support vector machine (SVM) [4,46]. This study presents a new model for the estimation of bubble point pressure for reservoir crude oils based on support vector machine modeling approach. The model was built and tested using a comprehensive data set collected from previously published literature. A quantitative analysis of the model was carried out to establish the adequacy and accuracy of the developed model. In addition, the model efficiency was tested against the measured saturation pressures, and the prediction made by available empirical correlations in the literature, the PR [9], and SRK-EOS [10]. Next, the model was used to simulate the results of gas injection or isothermal swelling test. Finally, outlier diagnosis was performed for detection of the probable doubtful experimental saturation pressure data.

2. Data acquisition

The reliability of the models for prediction of physical properties and phase behaviors of fluids generally depends on the comprehensiveness of the employed data set for their development [4,51]. For this purpose, a comprehensive data set was assembled from previously published literature [14,15,52–69]. A total of 130 experimental data points were obtained, each containing experimental values for temperature, crude oil compositions of C_1 through C_{7+}, non-hydrocarbon composition (N_2, CO_2, and H_2S), specific gravity of the heptane-plus fraction (SG_{C7+}), molecular weight of the heptane-plus fraction (MW_{C7+}), and the corresponding saturation pressure (P_b) measurement. The data set consists of data from a variety of crude oils of various composition ranges and from various geographical locations i.e. Middle East, North America and North Sea. Table 1 shows the range and the corresponding statistical parameters of the variables that have been used to develop and test the model. As can be seen, the data represent wide ranges of experimental conditions. Reservoir temperatures vary from 128 to 314 °F, specific gravities of the heptane-plus from 0.743 to 0.942, molecular weights of the heptane-plus from 134 to 324, and bubble point pressures from 313 to 6880 psia.

3. Model development

3.1. Background of the support vector machine

The support vector machine has been preliminary proposed for classification problems utilizing the hyper-planes to define decision boundaries between the experimental data points of different classes [70,71]. In this regression technique, the experimental data points from the input space are mapped into a high-dimensional or even

Table 1
Ranges of input/output variables used for developing and testing the model.

Parameter	Minimum	Maximum	Average
Temperature (°F)	128	324	177.3
Molecular weight of heptane-plus (MW_{C7+})	134	324	230.9
Specific gravity of heptane-plus (SG_{C7+})	0.743	0.942	0.861
Bubble point pressures (psi)	313	6880	2283.2
Methane	5.63	74.18	33.10
Ethane	0.84	12.45	7.35
Propane	0.43	11.87	6.33
Butanes	0.95	8.40	4.58
Pentanes	0.40	6.65	3.27
Hexanes	0	6.65	3.20
Heptane-plus	10.72	83.20	40.63
Nitrogen	0	1.67	0.36
Carbon dioxide	0	9.11	1.09
Hydrogen sulfide	0	3.68	0.14

Table 2
Compositions and saturation pressures for crude oils collected from literature.

No	N_2	CO_2	H_2S	C_1	C_2	C_3	C_4	C_5	C_6	C_{7+}	SG_{C7+}	MW_{C7+}	T	P_{sat}	Ref.
1	0.36	0.17	0	27.23	8.93	8.6	6.07	3.71	2.55	42.38	0.879	271	130	1365	[54]
2	0.29	0.48	0	28.36	8.29	7.38	5.06	3.42	4.41	42.31	0.88	252	133	1632	[54]
3	0.33	0.22	0	25.56	6.87	6.39	5.61	4.68	4.33	46.01	0.878	222	133	1595	[54]
4	0.35	0.47	0	26.52	7.71	6.05	4.03	4.08	4.22	46.57	0.892	253	135	1500	[54]
5	0.12	0.51	0	28.81	7.91	6.46	3.17	2.07	3.25	47.7	0.876	250	134	1615	[54]
6	0.16	0.3	0	24.66	7.51	6.92	4.76	2.65	3.43	49.61	0.875	239	134	1400	[54]
7	0.21	0.15	0	27.77	7.68	6.19	3.41	1.81	2.84	49.94	0.863	228	134	1590	[54]
8	0.11	0.28	0	27.53	6.9	5.87	4.42	2.47	3.52	48.9	0.879	245	135	1540	[54]
9	0.79	0.1	0	24.79	6.84	6.5	5.37	4.09	3.67	47.85	0.884	227	134	1399	[54]
10	0.036	0.17	0	31.22	6.5	4.75	2.53	1.71	2.64	50.45	0.889	264	135	1690	[54]
11	0	0.17	0	28.56	8.63	7.43	3.81	2.61	4.11	44.68	0.862	249	134	1548	[54]
12	0.45	0.08	0	29.35	7.85	7.28	6.03	4.81	4.53	39.62	0.863	227	134	1705	[54]
13	0.56	0.07	0	29.9	7.84	6.98	6.05	2.84	2.66	43.1	0.877	242	133	1653	[54]
14	0.03	0.3	0	27.8	8.26	6.73	3.72	4.08	4.2	44.88	0.88	254	134	1645	[54]
15	0.06	0.81	0	31.34	7.42	5.68	2.82	1.63	2.64	47.6	0.892	270	132	1751	[54]
16	0	0.15	0	19.5	8.17	8.43	6	2.92	3.61	51.22	0.849	225	132	1025	[54]
17	0	0	0	36.15	12.17	8.05	5.81	4.79	5.24	27.79	0.772	191	212	2238	[15,53]
18	0.48	0.41	0	33.25	7.55	6.21	4.11	2.2	2.88	42.91	0.88	236	136	1640	[54]
19	0.15	0.19	0	31.15	8.92	7.38	4.07	3.48	4.34	40.32	0.848	221	135	1650	[54]
20	0.5	0.87	0	29.85	7.33	5.98	3.48	1.96	4.28	46.2	0.882	271	134	1634	[54]
21	0.68	1.02	0.04	25.49	6.37	6.53	5.15	4.88	3.81	46.03	0.902	264	134	1533	[54]
22	0.22	1.37	0	27.92	6.51	5.16	2.52	1.58	2.59	52.13	0.893	255	135	1548	[54]
23	0.19	2.51	0	31.89	7.51	6.59	4.77	3.51	3.29	39.74	0.92	324	134	1825	[54]
24	0.18	2.46	0	30.26	7.3	6.36	4.61	3.54	3.22	42.08	0.914	299	134	1825	[54]
25	0.2	1.23	0.01	26.54	6.64	5.54	4.15	3.14	3.34	49.22	0.91	290	132	1597	[54]
26	0.44	0.83	0	27.75	7.54	6.9	4.61	3.1	3.11	45.72	0.902	272	133	1565	[54]
27	0.14	0.54	0	29.44	7.96	6.19	3.96	3.25	3.84	44.68	0.901	290	134	1580	[54]
28	0.15	0.39	0	27.55	7.86	6.3	3.59	2.58	3.68	47.9	0.896	152	132	1380	[54]
29	0.26	1.26	0	28.27	6	4.85	3.6	4	4.26	47.5	0.908	274	134	1680	[54]
30	0.2	0.8	0	31.42	7.6	7.23	4.88	1.97	0.99	44.91	0.889	251	132	1738	[54]
31	0.54	0.5	0	27.79	6.89	6.34	4.45	3.29	3.5	46.7	0.887	214	133	1655	[54]
32	0.41	0.65	0	30.21	6.89	6.6	5.88	4.62	4.27	40.47	0.875	247	135	1880	[54]
33	0.2	1.19	0	32.26	7.57	5.97	3.85	3.17	3.63	42.16	0.879	134	133	1788	[54]
34	0.1	1.16	0	32.93	7.78	6.06	3.73	2.92	3.7	41.62	0.879	279	133	1805	[54]
35	0.2	0.49	0	29.46	8.51	6.63	3.97	3.37	3.79	43.59	0.9	145.8	135	1500	[54]
36	0.15	0.48	0	28.51	8.17	6.36	4.15	3.73	3.87	44.59	0.897	141.9	135	1500	[54]
37	0.49	0.47	0	31.55	8.66	7.92	5.48	3.47	4.45	37.51	0.904	289	133	1715	[54]
38	0.15	0.65	0	30.48	7.27	5.15	2.04	1.96	2.76	49.54	0.866	239	130	1670	[54]
39	0.45	0.51	0	30.56	7.25	7.28	5.24	3.76	3.85	41.1	0.892	268	135	1735	[54]
40	0.23	1.06	0.01	24.75	6.78	6.8	6.01	4.35	3.89	46.12	0.905	256	134	1530	[54]
41	0.19	0.12	0	28.62	8.66	6.52	4.78	2.71	3.62	44.78	0.925	274	135	1766	[54]
42	0.21	0.72	0	33.12	7.44	6.07	3.22	1.79	3.05	44.38	0.9	271	140	1730	[54]
43	0.08	1.37	0.69	35.97	8.67	5.94	2.97	1.5	3.1	39.71	0.898	274	168	2505	[54]
44	0.13	1.45	0.75	36.02	8.83	6.13	4.1	2.8	2.49	37.3	0.896	230.1	166	2873	[54]
45	0	1.12	0.21	26.95	10.76	9.23	4.97	2.62	3.33	40.81	0.876	249	208	1877	[54]
46	0	1.25	0	33.35	9.24	6.07	2.4	1.62	2.78	43.29	0.851	252	230	2110	[54]
47	0	0.68	0	48.79	10.57	6.21	3.46	4.56	2.8	24.8	0.847	218	230	2730	[54]
48	0.05	0.85	0	41.05	11.07	7.07	4.82	2.89	3.19	29.01	0.848	198	241	3335	[54]
49	0.06	0.94	0	44.44	10.76	6.18	4.04	2.63	2.57	28.38	0.846	195	240	3630	[54]
50	0.04	0.78	0	40.91	10.72	6.64	4.44	2.62	2.51	31.34	0.845	202	241	3120	[54]
51	0.01	0.89	0.3	39.68	10.37	7.18	4.47	3.09	2.75	31.53	0.932	197	243	3061	[54]
52	0.06	0.85	0	40.7	10.54	6.55	4.41	2.72	2.81	31.36	0.847	195	243	3180	[54]
53	0.03	0.97	0	41.64	10.82	7.52	4.5	3	2.43	29.08	0.85	208	241	3069	[54]
54	0.06	1.1	0	43.13	10.46	7.24	4.57	2.82	2.34	28.29	0.884	214	241	3218	[54]
55	0.03	1.04	0.03	41.88	10.6	7.01	4.29	3.16	2.89	29.07	0.848	200	235	3234	[54]
56	0.02	0.99	0.04	38.79	10.07	6.86	4.18	2.92	2.77	33.38	0.849	210	240	3019	[54]
57	0.3	0.9	0	53.47	11.46	8.79	4.56	2.09	1.51	12.69	0.864	143	176	4460	[52]
58	0.11	2.35	0	35.21	6.72	6.24	5.07	5.23	4.1	34.97	0.841	213	250	2547	[52]
59	0.55	1.03	0	36.47	9.33	8.85	6	3.78	3.56	30.43	0.837	200	234	2746	[52]
60	1.64	0.08	0	28.4	7.16	10.48	8.4	3.82	4.05	35.97	0.843	252	131	1694	[52]
61	0.24	0.39	0	5.82	0.84	0.43	1.14	3.01	4.92	83.2	0.942	304	160	313	[54]
62	0.06	5.01	2.67	23.03	8.49	8.26	5.39	4.38	3.89	38.82	0.877	254	198	1731	[54]
63	0.1	1.32	0.34	8.86	4.13	5.41	3.91	3.45	3.81	68.67	0.934	243	178	677	[54]
64	0.16	0.91	0	36.47	9.67	6.95	5.37	2.85	4.33	33.29	0.852	218	220	2620	[53]
65	0.9	1.49	0	51.54	6.57	4.83	3.07	2.38	2.17	27.05	0.833	265	195	4566	[53]
66	0	0	0	57.53	10.16	5.83	3.28	2.71	1.4	19.11	0.81	203	212	5065	[53]
67	0.56	3.55	0	45.34	5.48	3.7	2.35	1.6	1.33	36.12	0.836	253	199	3885	[53]
68	1.64	0.08	0	28.4	7.16	10.48	8.4	3.82	4.05	35.97	0.843	252	131	1708	[60]
69	0.45	0.44	0	35.05	4.64	2.46	7.66	1.6	5.46	48.24	0.836	225	180	2520	[60]
70	0	0	0	42	8	6	2	1	4	37	0.823	255	155	3250	[60]
71	0.56	3.55	0	45.34	5.48	3.7	2.35	1.6	1.33	36.12	0.817	255	199	3885	[58]
72	0.4	1	0	45.4	4.2	0.89	1.08	0.94	1.01	45.08	0.888	250	160	3466	[58]
73	0.67	2.11	0	34.93	7	7.82	5.48	3.8	3.04	35.15	0.855	230	238	2724	[58]
74	0.34	0.84	0	49.23	6.32	4.46	3.04	2.26	2.06	31.45	0.865	230	200	3981	[58]
75	0.44	0.38	0	49.1	7.6	6.13	3.84	2.52	2	28	0.836	231	199	3739	[58]

(continued on next page)

Table 2 (*continued*)

No	N_2	CO_2	H_2S	C_1	C_2	C_3	C_4	C_5	C_6	C_{7+}	SG_{C7+}	MW_{C7+}	T	P_{sat}	Ref.
76	0.9	0.16	0	47.12	5.97	4.62	3.49	2.55	2.19	33	0.85	217	166	3394	[58]
77	0.36	1.06	0	50.5	4.54	0.9	1.15	0.89	1.6	39	0.901	291	156	3690	[58]
78	0.33	0.19	0	35.42	3.36	0.9	0.95	0.4	0.72	57.73	0.917	255	159	2307	[58]
79	0.41	0.44	0	40.48	7.74	8.2	5.45	3.64	2.83	31.42	0.845	210	208	2886	[58]
80	0	0	0	52	3.81	2.37	1.72	1.2	2.06	36.84	0.841	199	200	3839	[59]
81	0.45	0.44	0	35.05	4.64	2.46	1.66	1.6	5.46	48.24	0.9	225	180	2520	[52]
82	0.11	2.35	0	35.21	6.72	6.24	5.07	5.23	4.1	34.97	0.841	213	250	2547	[52]
83	0.55	1.03	0	36.47	9.93	8.85	6	3.78	3.56	30.43	0.837	200	234	2746	[52]
84	1.64	0.08	0	28.4	7.16	10.48	8.4	3.82	4.05	35.97	0.843	252	131	1694	[52]
85	0.25	2.19	1.16	16.33	6.29	7.48	6.09	4.36	3.58	52.27	0.88	249	215	1261	[55]
86	0.24	1.53	0.6	13.16	6.38	7.62	6.77	5.65	6.37	51.68	0.876	275	190	1140	[55]
87	0.17	0.65	1.93	12.59	6.05	6.51	4.26	4.52	1.14	62.18	0.877	230	239	1490	[55]
88	0.32	3.69	0.68	21.55	8.6	7.66	6.4	5.07	2.62	43.41	0.869	243	239	1591	[55]
89	0.43	3.47	3.68	19.49	8.28	6.85	4.3	4.18	2.42	46.9	0.876	246	230	993	[55]
90	0.21	0.34	0	20.04	7.93	8	6.6	5.87	5.08	45.93	0.861	230	235	900	[55]
91	0.77	1.99	1.4	17.38	6.42	7.62	5.62	4.53	5.14	49.13	0.891	267	234	1190	[55]
92	0.21	0.75	0.51	6.05	2.59	5.83	7.69	6.14	5.42	64.81	0.857	231	148	352	[56]
93	0.88	1.34	0	5.63	2.51	4.6	7.31	5.99	4.71	67.03	0.855	224	128	376	[56]
94	0.3	0.01	0	7.14	1.54	3.71	7.31	6.65	6.19	67.15	0.86	233	140	374	[56]
95	0.35	0.56	1.41	9.99	1.45	1.87	3.64	4.47	5.23	71.03	0.872	258	148	506	[56]
96	0.31	0.28	0.02	6.8	1.98	4.01	6.62	6.57	6.65	66.76	0.858	237	144	374	[56]
97	0.29	0.46	0.49	10.75	1.11	1.58	3.68	4.03	4.75	72.86	0.861	261	145	519	[56]
98	0.33	0.35	0	6.72	2.19	4.04	5.54	5.3	4.47	71.06	0.858	225	138	360	[56]
99	0.41	0.26	0	6.14	2.38	4.71	6.27	5.64	4.68	69.51	0.86	225	134	346	[56]
100	0.53	0.12	0	22.8	6.45	8.51	6.6	4.71	4.24	46.04	0.864	242	156	1325	[56]
101	0.78	0.1	0	20.64	5.8	7.05	6.9	5.95	5.11	47.67	0.857	237	157	1209	[56]
102	0.6	0.12	0	23.71	7.18	8.71	6.85	5.17	2.85	44.81	0.868	238	156	1386	[56]
103	1.13	0.13	0	25.45	6.99	9.6	7.87	5.93	1.88	41.02	0.851	237	157	1446	[56]
104	0.37	0.02	0	17.28	6.11	8.23	5.47	3.76	3.96	54.8	0.851	242	156	537	[56]
105	0.54	0.18	0	21.62	6.03	8.39	7.1	4.75	3.85	47.54	0.863	236	156	1257	[56]
106	1.39	0.28	0	21.32	6.21	8.11	6.37	3.84	5.45	47.03	0.87	257	156	1220	[56]
107	0.68	0.16	0	22.84	6.28	7.83	6.24	4.63	3.44	47.9	0.864	226	156	1407	[56]
108	0.39	0.14	0	21.4	6.4	7.43	5.06	4.34	4.45	50.39	0.842	245	149	1124	[56]
109	1.02	0.12	0	19.76	5.52	6.89	4.87	4.49	2.67	54.66	0.85	247	160.5	1152	[56]
110	0.3	0.9	0	53.47	11.46	8.79	4.56	2.09	1.51	16.92	0.836	173	176	4460	[52]
111	1.67	1.38	0	26.68	9.25	10.18	6.4	4.03	0	40.41	0.855	217	210	1954	[61]
112	0.65	0.02	0	45.02	12.45	8.93	6.03	3.02	1.44	22.44	0.81	184	140	3002	[62]
113	0.25	0.24	0	40.91	10.38	9.01	5.13	3.28	2.21	28.58	0.8	182	300	3043	[63]
114	0.24	0.27	0	66.83	8.28	5.15	3.31	2.04	1.85	12.03	0.8	182	215	4810	[64]
115	0	0	0	52	3.81	2.37	1.72	1.2	2.01	36.84	0.841	199	201	3828	[59]
116	0	0	0	46.78	8.77	7.44	4.01	2.56	4.02	26.4	0.766	158	212	2941	[15,53]
117	0	0	0	74.18	5.32	4.67	2.58	0.97	1.56	10.72	0.766	159	212	4753	[15,53]
118	0	0	0	73.36	5.35	4.71	2.62	1	1.62	11.18	0.767	161	212	4742	[15,53]
119	0	0.01	0	31	10.41	11.87	7.32	4.41	2.55	32.43	0.743	199	130	1620	[65]
120	1.02	0.4	0	54.62	11.47	7.33	4.01	1.96	1.42	17.77	0.808	218	164	3900	[65]
121	0.33	3.03	0	41.33	8.89	5.95	4.12	2.66	0	33.69	0.848	200	200	3200	[66]
122	0.71	3.84	1.17	59.49	5.55	4.06	3.06	1.58	1.27	19.27	0.78	297	158	6880	[67]
123	0	0	0	60.88	7.38	5.03	2.78	1.96	1.84	20.13	0.836	191	190	5000	[68]
124	0	9.11	0	45.58	5.11	3.03	2.18	1.92	1.54	31.53	0.894	270	190	3840	[68]
125	0.31	0.69	0	47.69	6.64	4.59	2.59	1.16	1.69	34.64	0.869	234	232	3800	[58]
126	0.03	8.39	0	47.43	10.29	6.12	4.2	2.89	2.05	18.61	0.83	180	295	4000	[58]
127	0.52	6.47	0	39.58	10.68	7.27	5.28	3.65	2.9	23.67	0.858	176	310	3627	[58]
128	0.34	7.1	0	48.43	9.24	5.84	4.39	3.21	2.28	19.17	0.805	183	314	4082	[58]
129	0.38	7.03	0	48.73	8.93	5.48	4.05	3	2.14	20.26	0.805	181	309	4156	[58]
130	1.67	2.18	0	60.51	7.52	4.74	4.12	2.97	0	16.29	0.789	181	246	4823	[69]

infinite dimensional feasible region (margin) using a definite function, in which a maximal separating plane is constructed [70–72]. The objective of a SVM calculation step is to find the optimum hyper-plane, from which the distance to all of the experimental data points is minimum.

For a set of data with $x = \{(x_1, y_1),\ldots,(x_n, y_n)\}$, the SVM method uses the following relations to express the separating plane in the input space of the problem [70–72]:

$$f(x) = w^T x + b \tag{1}$$

Subject to the following constraint when the data of two classes are separable:

$$\begin{cases} f(x_i) \geq 1 & \text{if} \quad y_i = +1 \\ f(x_i) \geq -1 & \text{if} \quad y_i = -1 \end{cases} \tag{2}$$

where, x denotes the input vector of parameters of the model; w represents the nonlinear function that maps the input space to a high-dimensional feature space and performs linear regression; y denotes the outputs; b is the bias term; and superscript T denotes the transpose matrix. Support vectors (SVs) are those points that satisfy the aforementioned constraints [46,70,71,73].

The distance between the plane crossing the data points standing for $f(x) = +1$ and $f(x) = -1$, is called the margin. The separating plane denotes the largest value of the margin. The best (optimal) separating plane can be obtained by maximization of the margin and minimization of the noise which introduce the slack variable as below [70–72]:

$$\min\left(\frac{1}{2}\left\|w\right\|^2\right) + C\sum_{i=1}^{N}\zeta_i \tag{3}$$

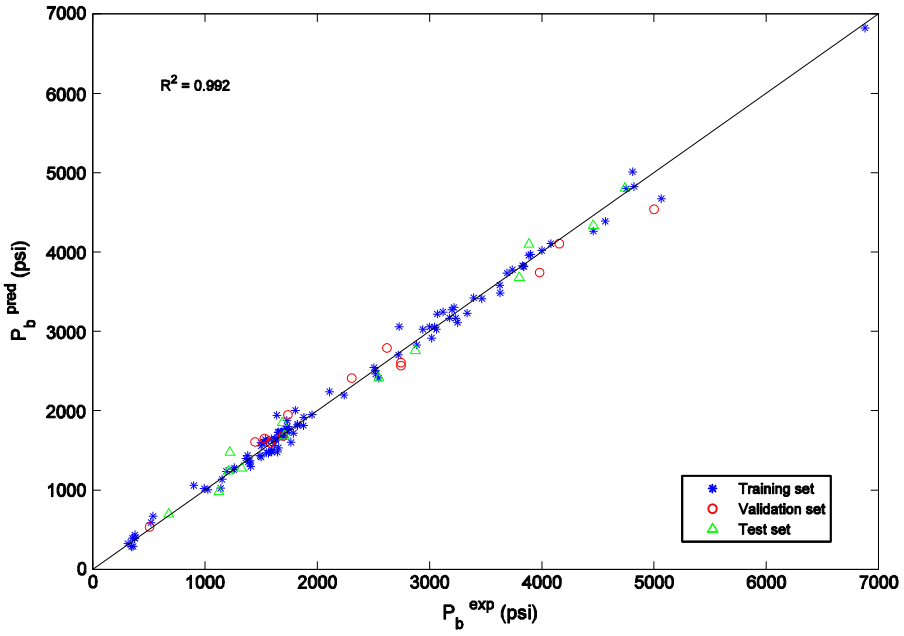

Fig. 1. Comparison between the results of the developed LSSVM [70] model and the applied data of saturation pressure.

In the preceding expression, C denotes a positive constant, and ξ (slack variable) stands for the distance between the data points that exist in the false class side and the margin of its virtual class.

Considering the abovementioned equations, it is possible to conclude that we have a typical problem of convex optimization [46,70,71,73]. This can be solved by using the Lagrange multiplier method that can result in the following equation [70–72]:

$$L(w,b;a)=\frac{1}{2}w^{T}w-\sum_{i=1}^{N}\alpha_i\left(y_i\left[w^{T}x_i+b\right]-1\right) \tag{4}$$

With the Lagrange multipliers $\alpha_i \geq 0$ the solution is defined through the Lagrangian saddle point [70,71]. For a given regression problem, the goal of the SVM technique is to find the optimal hyper-plane, from which the distance to all the data points is minimum, as mentioned earlier.

Due to the specific formulation of the SVM algorithm, sparse solutions can be found and both linear and nonlinear regressions can be pursued for solving the corresponding problems [46,70–73].

In 1999, Suykens and Vandewalle [70] proposed a modification description to the original SVM called least-square SVM (LSSVM). This method applies a set of linear equations using support vectors (SVs) instead of quadratic programming problems in order to facilitate the solution of the original SVM. The subsequent least-square SVM method benefits from advantages similar to those of SVM though it only requires solving a set of linear equations (linear programming), resulting in an easier-to-implement and faster alternative to the traditional SVM method [46,51,70,71,73,74].

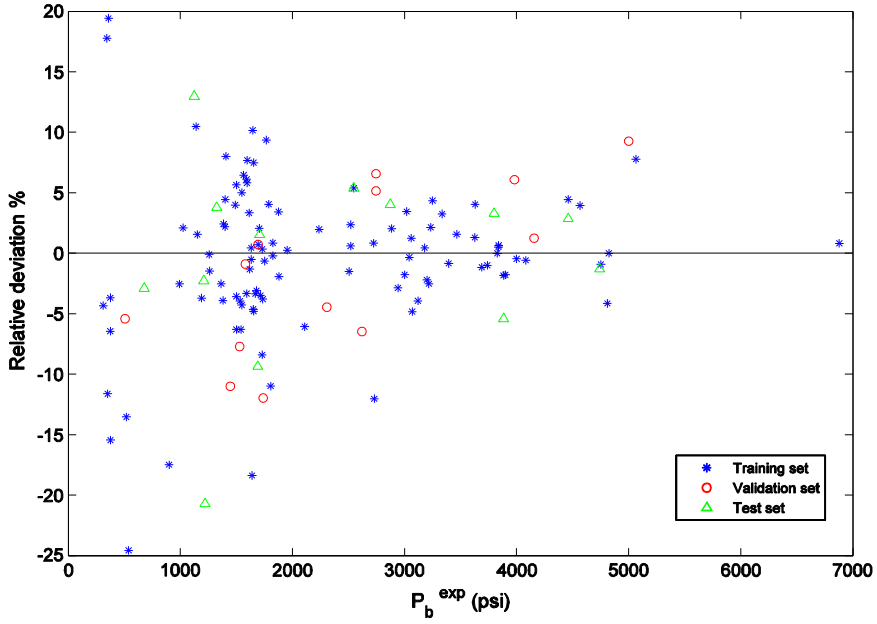

Fig. 2. Percent relative error distribution for saturation pressure estimation made by LSSVM [70] model.

Table 3
Statistical parameters of the obtained LSSVM model.

Statistical parameter	
Training set	
R^2 [a]	0.994
Average absolute relative deviation [b]	4.4
Standard deviation error [c]	100.19
Root mean square error [d]	99.71
N [e]	104
Validation set	
R^2	0.987
Average absolute relative deviation	5.9
Standard deviation error	192.40
Root mean square error	186.20
N	13
Test set	
R^2	0.990
Average absolute relative deviation	5.8
Standard deviation error	137.28
Root mean square error	131.90
N	13
Total	
R^2	0.992
Average absolute relative deviation	4.7
Standard deviation error	115.14
Root mean square error	114.72
N	130

[a] $R^2 = 1 - \dfrac{\sum_i^N (\text{Calc.}(i)/\text{Est.}(i) - \exp.(i))^2}{\sum_i^N (\text{Calc.}(i)/\text{Est.}(i) - \text{average}(\exp.(i)))^2}$

[b] $\%\text{AARD} = \frac{100}{N}\sum_i^N \dfrac{|\text{Calc.}(i)/\text{Est.}(i) - \exp.(i)|}{\exp.(i)}$

[c] $\text{Std} = \frac{1}{N}\sum_i^N \sqrt{(\text{Calc.}(i)/\text{Est.}(i) - \text{average}(\text{Calc.}(i)/\text{Est.}(i)))^2}$

[d] $\text{RMSE} = \left(\dfrac{\sum_{i=1}^N (\text{Calc}(i)/\text{Est.}(i) - \exp(i))^2}{N}\right)^{1/2}$

[e] Number of data points.

The regression error of the LSSVM [70] approach is defined as the difference between the predicted property values and experimental ones, which is considered as an addition to the constraint of the optimization problem. In the traditional SVM method, the value of the regression error is generally optimized during the calculations while in the LSSVM [70], the error is mathematically defined [51,70,71,73–75].

The cost function (penalized cost function) of the applied strategy is determined as below [4,51,70,71,73,74]:

$$Q_{\text{LSSVM}} = \frac{1}{2}w^T w + \gamma\sum_{k=1}^{N} e_k^2 \tag{5}$$

subject to:

$$y_k = w^T\varphi(x_k) + b + e_k, k = 1, 2, \ldots, N \tag{6}$$

In Eqs. (5) and (6) e_k stands for the regression error of N training objects (the least-squares error approach), and γ indicates the relative weight of the summation of the regression errors compared to the regression weight [51,70,71,73,74].

The weight coefficient (w) is generally written as follows [51,70,71,73,74]:

$$w = \sum_{k=1}^{N} \alpha_k x_k \tag{7}$$

in which:

$$\alpha_k = 2\gamma\, e_k \tag{8}$$

Assuming the linear regression between the independent and the dependent variables of the LSSVM [70] algorithm, Eq. (6) can be re-written as [51,70,71,73,74]:

$$y = \sum_{k=1}^{N} \alpha_k x_k^T x + b \tag{9}$$

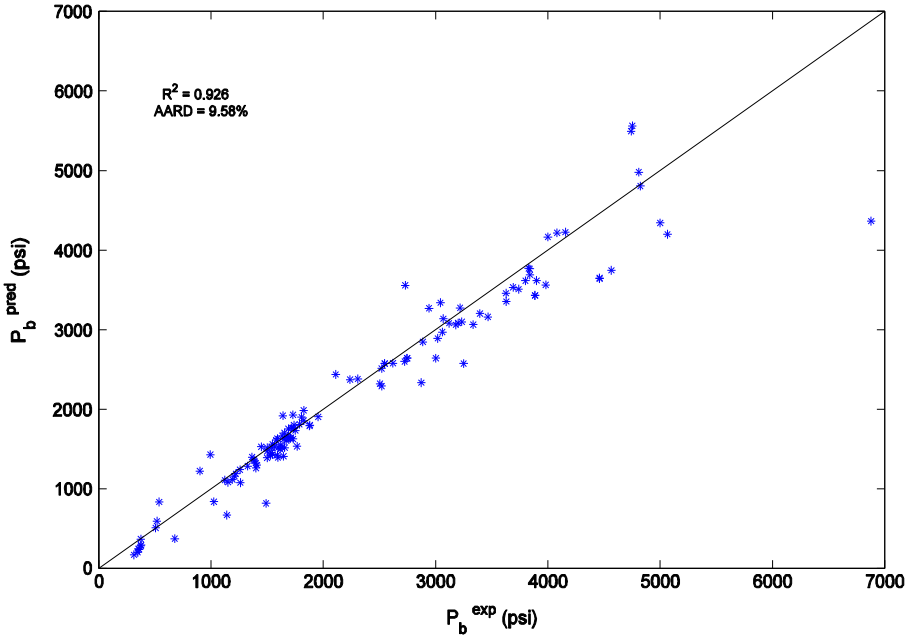

Fig. 3. Cross plot for saturation pressure estimation made by Bandyopadhyay and Sharma [6].

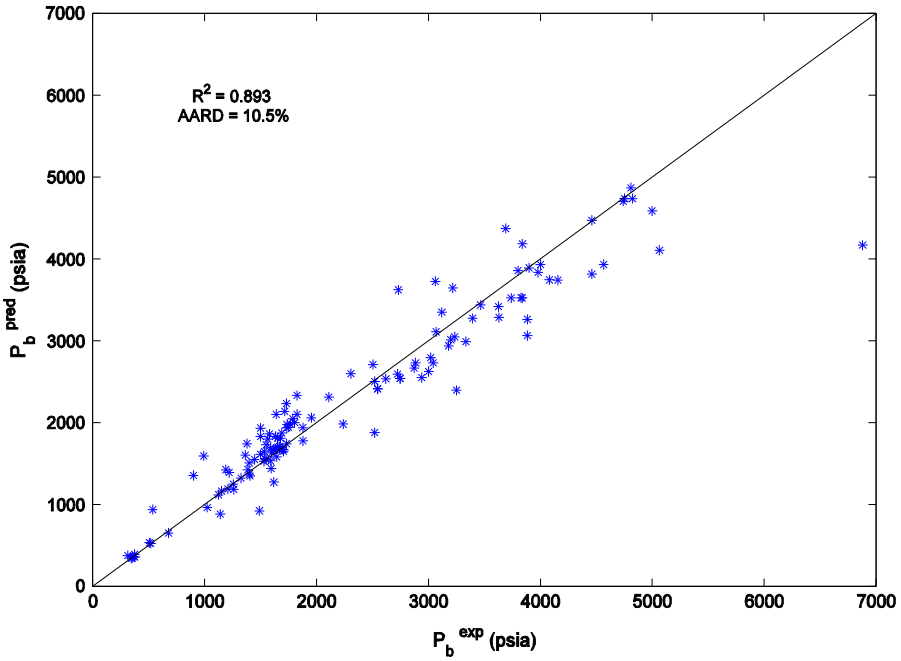

Fig. 4. Cross plot for saturation pressure estimation made by SRK [10].

The Lagrange multipliers [51,73,74,76,77] in Eq. (9) are calculated as:

$$\alpha_k = \frac{(y_k - b)}{x_k^T x + (2\gamma)^{-1}} \tag{10}$$

The aforementioned linear regression equation can be converted to nonlinear regression using Kernel function as follows [51,70,71,74]:

$$f(x) = \sum_{k=1}^{N} \alpha_k K(x, x_k) + b \tag{11}$$

where $K(x,x_k)$ is the Kernel function calculated from the inner product of the two vectors x and x_k in the feasible region built by the inner product of the vectors $\Phi(x)$ and $\Phi(x_k)$ as follows [4,51,70,71,73,74]:

$$K(x, x_k) = \Phi(x)^T . \Phi(x_k) \tag{12}$$

The most commonly used kernel function is the radial basis function (RBF) [4,44,51,70,78] that we used in this study:

$$K(x, x_k) = \exp\left(-\left|\left|x_k - x\right|\right|^2 / \sigma^2\right) \tag{13}$$

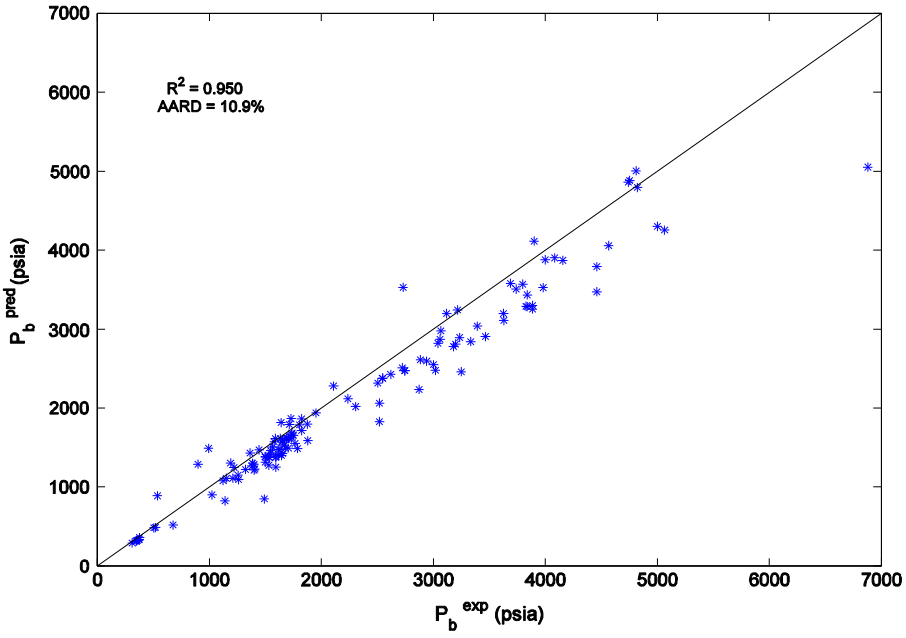

Fig. 5. Cross plot for saturation pressure estimation made by PR [9].

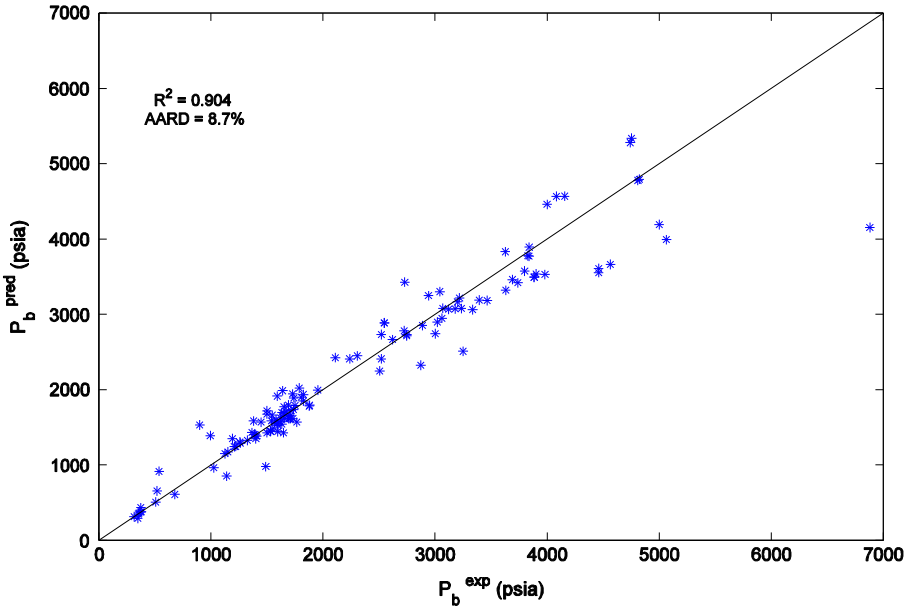

Fig. 6. Cross plot for saturation pressure estimation made by Elsharkawy [54].

where, σ^2 is squared bandwidth which is optimized by an external optimization algorithm during the calculations [4,44,46,73,78,79].

3.2. Computational procedure

A review on recent researches [6,54,80] shows that the bubble point pressure of reservoir crude oil is a function of reservoir temperature, hydrocarbon and non-hydrocarbon reservoir fluid compositions, and characteristics of the heptane-plus fraction as follows:

$$P_b = f(T, Z_i, SG_{C7+}, MW_{C7+}) \tag{14}$$

In the next step, the database is divided into three sub data sets including the "Training" set, the "Optimization" set, and the "Test" set. Generally, the "Training" set is used to generate or build the model structure, the "Optimization" set is applied for optimization of the model and adjusting the model parameters, and finally, the "Test (prediction)" set is used to investigate the prediction capability and validity of the proposed model [4,46]. To increase the model capability, it is better to divide the database randomly. For this purpose, 80%, 10%, and 10% of the main data sets are randomly selected for the "Training" set (104 data points), the "Optimization" set (13 data points), and the "Test" set (13 data points). In the distribution of the data through the three sub data sets, we generally perform many distributions to avoid the local accumulations of the data in the feasible region of the problem. As a result, the acceptable distribution is the one with homogeneous accumulations of the data on the domain of the three sub data set [51]. The two main parameters of this algorithm are σ^2 and γ, which are supposed to be optimized using an appropriate optimization strategy.

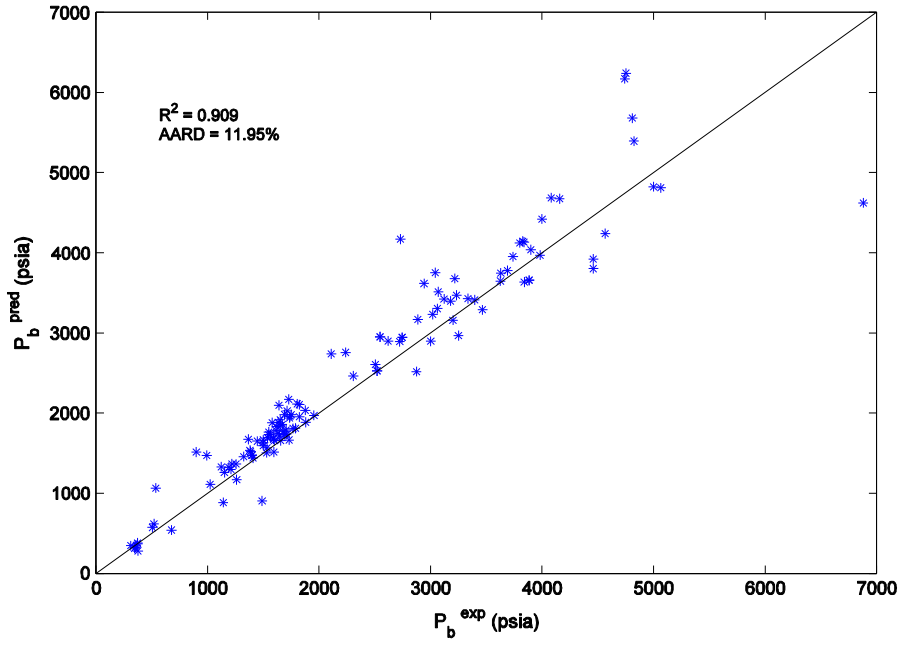

Fig. 7. Cross plot for saturation pressure estimation made by AlQuraishi [80].

Table 4
Accuracy of different models for predicting saturation pressure for different crudes from literature.

Black oils								
Model	Sample no.							
	85	86	87	88	89	90	91	AAD%
Experimental P_b (psia)	1261	1140	1490	1591	993	900	1190	
PR-EOS	1092	823	849	1615	1489	1290	1303	26.93
SRK-EOS	1185	882	922	1695	1593	1353	1426	29.13
Elsharkawy	1290	849	977	1913	1387	1531	1349	29.37
Bandyopadhyay	1076	672	819	1633	1429	1224	1117	27.06
AlQuraishi	1169	885	905	1676	1471	1514	1324	28.84
LSSVM (*this study*)	1280	1021	1431	1644	1018	1057	1234	**6.14**

Volatile oils										
Model	Sample no.									AAD%
	52	66	110	112	114	127	128	129	130	
Experimental P_b (psia)	4460	5065	4460	3002	4810	3627	4082	4156	4823	
PR-EOS	3791	4235	3473	2548	5007	3198	3902	3871	4513	11.36
SRK-EOS	4472	4105	3815	2624	4867	3418	3743	3740	4738	8.14
Elsharkawy	3608	3992	3554	2742	4778	3830	4567	4568	4794	10.88
Bandyopadhyay	3648	4199	3641	2642	4978	3457	4216	4224	4805	8.79
AlQuraishi	3801	4808	3921	2895	5678	3646	4685	4674	5393	10.35
LSSVM (*this study*)	4333	4671	4262	3055	5010	3580	4106	4105	4824	**2.68**

Heavy oils									
Model	Sample no.								AAD%
	92	93	94	95	96	97	98	99	
Experimental P_b (psia)	352	376	374	506	374	519	360	346	
PR-EOS	321	358	331	481	333	486	326	312	8.33
SRK-EOS	345	392	361	531	360	524	358	344	2.56
Elsharkawy	347	376	432	506	374	653	361	291	7.37
Bandyopadhyay	100	143	212	328	128	404	79	45	20.97
AlQuraishi	326	280	380	574	367	617	351	314	10.07
LSSVM (*this study*)	393	434	398	533	388	589	290	285	**11.64**

Synthetic oil					
Model	Sample no.				AAD%
	116	117	118	119	
Experimental P_b (psia)	2941	2238	4753	4742	5.62
PR-EOS	2594	2117	4883	4862	6.47
SRK-EOS	2550	1981	4737	4705	10.43
Elsharkawy	3248	2408	5339	5280	11.36
Bandyopadhyay	3268	2372	5559	5493	12.48
AlQuraishi	3617	2756	6239	6171	26.88
LSSVM (*this study*)	3025	2194	4797	4803	**1.76**

The mean square error (MSE) of the results of the LSSVM-algorithm [70] is defined as:

$$MSE = \frac{\sum_{i=1}^{n}\left(P_{pred_i} - P_{\exp_i}\right)}{n} \quad (15)$$

where P is the saturation pressure, subscripts *pred* and *exp* indicate the predicted and experimental saturation pressure values, respectively, and n is the number of samples from the initial population.

4. Results and discussion

There are generally two parameters in LSSVM [70] algorithm including σ^2 and γ, which are supposed to be optimized regarding the specified problem. The optimization procedure has been repeated several times as attempts to reach to the global optimum of the problem. In this work, we have applied the Coupled Simulated Annealing (CSA) optimization technique [81].

The values of the probable global optima of the problem including σ^2 and γ are 56.47648 and 531.03495, respectively.

To evaluate the performance of the proposed model, two types of analysis are performed. First, the accuracy of the new model in the prediction of crude oil saturation pressure is compared with the accuracy of PR [9] and SRK [10] equations of state and also all existing literature correlations to estimate saturation pressures for the experimentally measured samples collected from the literature. Second, the validity of the model is checked against EOS by studying the sensitivity of the proposed model to the most input variables. Next, the model is used to simulate the results of controlled condition laboratory tests such as isothermal swelling tests. Finally, outlier diagnoses were performed for detection of the probable doubtful saturation pressure data.

4.1. Model accuracy

The details of the experimental data used for developing the model are listed in Table 2. A comparison between the results of the proposed model and the corresponding experimental bubble point pressure values is shown in Fig. 1. A tight cloud of points about 45° line for training, validation and testing data set indicate the robustness of the proposed LSSVM [70] model. Fig. 2 shows the percent relative error distribution for all experimental data points. Furthermore, some important statistical parameters of the proposed model including squared

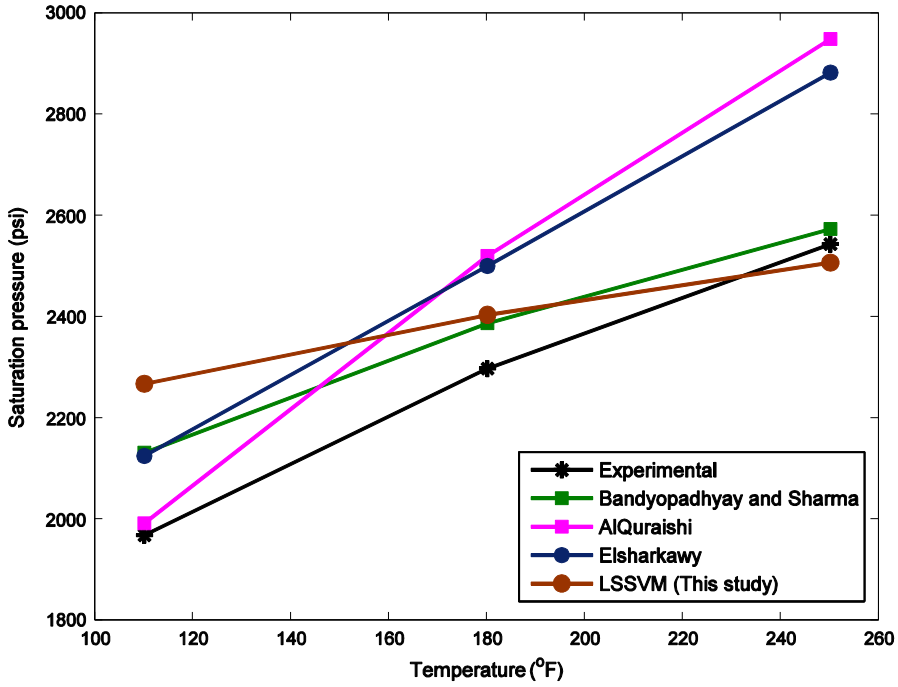

Fig. 8. Saturation pressure for sample no. 82 [52].

correlation coefficient (R^2), average absolute relative deviation, standard deviation error, and root mean square error are reported in Table 3. The results illustrate that an excellent agreement exists between the prediction of LSSVM [70] and the experimental data.

The efficiency of the model was also tested against PR-EOS [9], SRK-EOS [10], Elsharkawy [54], Bandyopadhyay and Sharma [6], and AlQuraishi [80] correlations. Cross plots between predictive correlations and experimental data are illustrated in Figs. 3–7. These crossplots show the degree of agreement between the experimentally measured data and the predicted values. As illustrated in Fig. 1, it evidences that the predictions of the bubble point pressures made by the proposed LSSVM [70] model yields the closest agreement with the experimental data among the selected correlations. The AlQuraish [80] correlation shows the largest number of scattered points in Fig. 7, which is reflected by the highest average absolute relative deviation obtained (AARD = 11.95%).

Table 4 reports the comparison of accuracy among different methods for predicting saturation pressure for four groups of crudes from the literature [52,53,55,56,62,64,69]. The first group are black oils from Moharam [55] represented by sample numbers 85–91. For such group, the proposed LSSVM [70] model has 6.14% AARD, SRK [10], PR [9], Elsharkawy [54], Bandyopadhyay and Sharma [6], and AlQuraishi [80] correlations have 29.13%, 26.93%, 29.37%, 27.06%, and 28.84%, respectively. The second group are volatile oils from various sources [52,53,55,56,62,64,69]. Both Bandyopadhyay and Sharma [6] and SRK [10] models have the same AARD (8%) and they are more accurate than the PR [9], Elsharkawy [54] and AlQuraishi [80] models (AARD of 11%). For this group the proposed LSSVM [70] model has the lowest value of AARD (2.7%). The third group contains heavy crude oils from Wu and Rosenegger [56]. The proposed model, SRK [10], PR [9], Elsharkawy [54], Bandyopadhyay and Sharma [6], and AlQuraishi [80] have AARD of 11.64%, 2.56% 8.33%, 7.37%, 20.97%, and 10.07%, respectively. The results show that PR-EOS [9] performed more satisfyingly compared with the other predictive models. Intelligent systems are strongly governed by the training data. It should be mentioned here that, just a few data (7 data points) belong to $P_b < 500$ psia. It was expected that for these data points the accuracy of the model decreases. Consequently, the error of the LSSVM [70] model increases in this area. Nonetheless, by providing more samples in the range of $P_b < 500$ psia, the accuracy of the LSSVM [70] model would be improved. The last group are synthetic hydrocarbon mixtures from Danesh et al. [15,53]. It could be easily observed that the LSSVM [70] model performs outstandingly better than other predictive models. As a whole, to make a judgment based on AARD reported in Table 4, it is clearly obvious that the proposed LSSVM [70] model performs significantly better in the estimation of saturation pressures for black, volatile, heavy crudes and synthetic mixtures.

4.2. Validity of the model

A validity analysis is performed to check whether the developed model is physically correct or not. Thus, the model responses to most important variables are studied and compared to those of the experimental values.

The saturation pressure versus temperature for crude oil sample (No. 82) reported by Coats and Smart [52] has been used to check validity of the model. Fig. 8 shows the comparison between LSSVM [70] to a change in temperature as compared with other predictive correlations. As expected, this figure shows that all the studied models predict an increase in the saturation pressure as a function of increasing temperature. Also, Fig. 8 illustrates that the model captured the trend of increasing the saturation pressure as a function of increasing the temperature. It is important to note that the measured saturation pressures at different temperatures for this sample (No. 82) were not used in developing the model.

The ability of the empirical model to predict saturation pressure during single contact experiments by rich gas is examined. The data used for isothermal swelling tests are from the data bank published by Jaubert et al. [82]. The molar composition for the crude oil and injected gas used for comparison in this section is tabulated in Tables 5–6.

The compositions of the hydrocarbon components reported in Table 5 were characterized into single component numbers C_1 to C_6

Table 5
Molar compositions and physical properties of studied oil sample from Jaubert et al. [82].

Compound	Mole percent	Properties of the cuts from C_6 to C_{20+}	
		Molar weight (g/mol)	Density at 15 °C (kg/m^3)
Hydrogen sulfide	0.383		
Nitrogen	0.450		
Carbon dioxide	2.070		
Methane	26.576		
Ethane	7.894		
Propane	6.730		
Cut C_4			
i-Butane	1.485		
n-Butane	3.899		
Cut C_5			
i-Pentanes	1.937		
n-Pentane	2.505		
Cut C_6			
i-Hexanes	1.783	86.2	679.4
n-Hexane	1.568		
Cut C_7			
i-Heptanes	0.531		
Benzene	0.538		
Cyclanes C_7	2.229	92.6	726.2
n-Heptane	1.013		
Cut C_8			
i-Octanes	1.215		
Toluene	1.004		
Cyclanes C_8	1.065	108.9	750.9
n-Octane	0.849		
Cut C_9			
i-Nonanes	0.866		
Aromatics C_9	1.043		
Cyclanes C_9	0.620	120.1	771.6
n-Nonane	0.522		
Cut C_{10}			
i-Decanes	1.267		
Aromatics C_{10}	0.433	137.9	787.8
n-Decane	0.333		
Undecanes (cut C_{11})	2.635	149.0	803.7
Dodecanes (cut C_{12})	2.285	163.0	815.4
Tridecanes (cut C_{13})	2.364	177.0	827.0
Tetradecanes (cut C_{14})	2.038	191.0	841.2
Pentadecanes (cut C_{15})	1.752	205.0	858.8
Hexadecanes (cut C_{16})	1.589	219.0	862.7
Heptadecanes (cut C_{17})	1.492	234.0	858.6
Octadecanes (cut C_{18})	1.263	248.0	864.8
Nonadecanes (cut C_{19})	0.812	263.0	877.1
Eicosane plus (C_{20+})	12.962	450.0	956.0

and the remaining components were lumped into a single C_{7+} parameter. The lumping was carried as follows [83]:

$$Z_{C_{7+}} = \sum_{i=7}^{n} Z_{C_i} \tag{16}$$

Table 6
Molar compositions of the injected gas for isothermal swelling tests from Jaubert et al. [82].

Compound	Mole percent
Hydrogen sulfide	0.000
Nitrogen	0.000
Carbon dioxide	0.000
Methane	88.000
Ethane	7.000
Propane	3.000
i-Butane	0.500
n-Butane	0.500
i-Pentanes	0.500
n-Pentane	0.500

$$\mathrm{MW}_{C_{7+}} = \sum_{i=7}^{n} Z_{C_i} \cdot \mathrm{MW}_i / Z_{C_{7+}} \tag{17}$$

$$\mathrm{SG}_{7+} = Z_{C_{7+}} \cdot \mathrm{MW}_{C_{7+}} / \sum_{i=7}^{n} \left(Z_{C_i} \cdot \mathrm{MW}_i / \mathrm{SG}_i \right) \tag{18}$$

Fig. 9 shows the comparison of bubble point pressures predicted by the proposed model versus the laboratory measured values when the crude oil sample was contacted by various amounts of injected gas. The model captured the correct physical trend of increasing saturation pressure as a function of the increasing amount of injected methane-rich gas. It can also be observed that the model predicts the saturation pressure accurately at lower values of injected gas, but at higher percentages (~35%) of injected gas, the model predictions deviate from the measured values. However, this indicates that the proposed model is adequate for the accurate prediction of crude oil saturation pressures as well as simulating a gas injection test.

4.3. Outlier diagnosis

The applied saturation pressure data contain definite uncertainties. These uncertainties affect, indeed, the prediction capability of the obtained correlation [45,84,85]. To this end, we have applied the method of Leverage Value Statistics [86,87]. The graphical detection of the suspended (doubtful) data or outliers is undertaken through sketching the Williams plot on the basis of the calculated H values [45,75,85]. A detailed description of the equations and computational procedure of this method can be found elsewhere [45,75,85]. The Williams plot has been sketched in Fig. 10 for the results using the LSSVM [70] model. The existence of the majority of data points in the ranges $0 \le H \le 0.3226$ and $-3 \le R \le 3$ reveals that the applied model is statistically correct and valid. Furthermore, it shows that the whole data except three in the data set are located within the applicability domains of the applied models. Therefore, there are only three points (numbers 47, 66, and 123) in the data set which are within this domain and consequently we can state it as probable doubtful datum. It may be possible to eliminate these probable outliers from the proposed model results and develop more accurate ones; however, our aim, herein, has been to investigate the ability of all of the investigated models to predict the whole saturation pressure values from a data set in the literature.

5. Conclusion

Presented here is a novel intelligent model for the prediction of saturation pressure of crude oils as a function of hydrocarbon and non-hydrocarbon reservoir fluid compositions, temperature and characteristics of the heptane-plus fraction. The proposed model shows better performance in predicting saturation pressure of crude oils with stable performance, and achieved the lowest AARD, lowest RMSE, and the highest correlation coefficient (R^2) compared to the well-established empirical formulas in the literature and most widely used EOS. In addition, the validity of the model was examined for the most important input variables and indicated that the model successfully captured the physical trend of changing the saturation pressure as a function of temperature. Finally, the Leverage mathematical strategy was used to evaluate the experimental data of crude oil saturation pressure. The results show that only three of the data points were found to be outliers (doubtful experimental data) while all of the investigated data were interpreted to be within the applicability of the proposed model.

Nomenclature

AARD	Average Absolute Relative Deviation, %
b	bias term
C	a positive constant

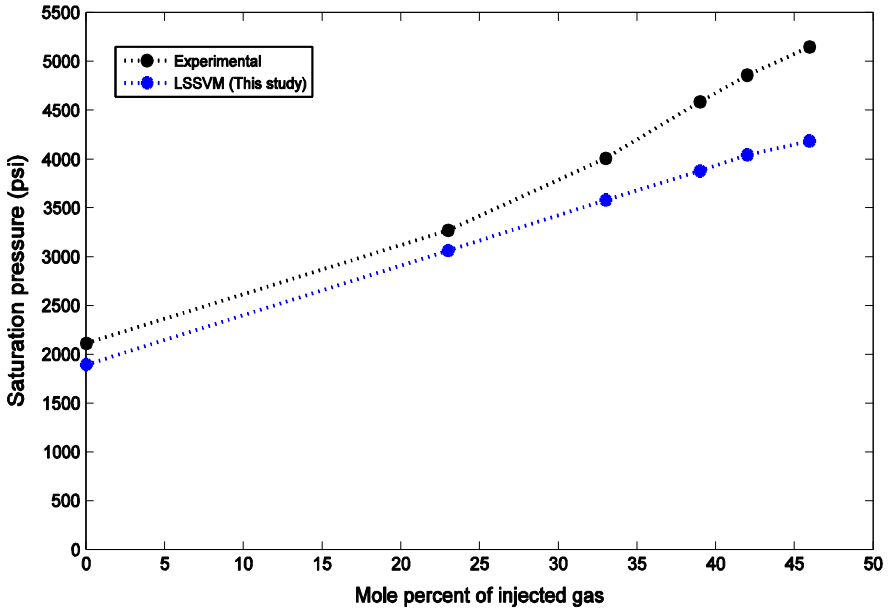

Fig. 9. Isothermal swelling test of black oil and rich gas at 230 °F.

e_k	regression error
$K(x,x_k)$	Kernel function
LSSVM	least-square support vector machine
MW_{C7+}	molecular weight of heptane-plus fraction in crude oil
P_b	bubble point pressure, psi
R^2	squared correlation coefficient
RBF	radial basis function
RMSE	root mean square error
SG_{C7+}	specific gravity of heptane-plus fraction in crude oil
T	reservoir temperature, °F
Z_i	mole percent of i component
α_i	Lagrange multipliers
γ	relative weight of the summation of the regression errors
σ^2	squared bandwidth
ξ	slack variable

Acknowledgment

A. Farasat would like to acknowledge the departments of research and technology of the National Iranian Oil Company and Iranian Offshore Oil Company for support throughout this research.

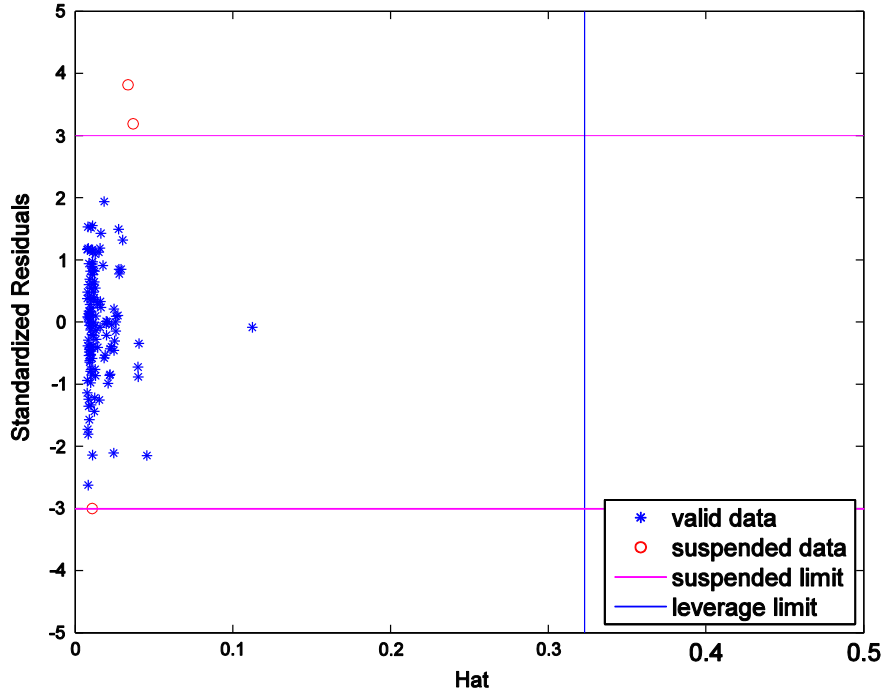

Fig. 10. Detection of the probable doubtful data and the applicability domain of the developed LSSVM [70] model.

References

[1] S.S. Ikiensikimama, J.A. Ajienka, Impact of PVT correlations development on hydrocarbon accounting: the case of the Niger Delta, Journal of Petroleum Science and Engineering 81 (2012) 80–85.
[2] A. Khoukhi, S. Albukhitan, PVT properties prediction using hybrid genetic-neuro-fuzzy systems, International Journal of Oil, Gas and Coal Technology 4 (2011) 47–63.
[3] E.A. El-Sebakhy, Data mining in forecasting PVT correlations of crude oil systems based on Type1 fuzzy logic inference systems, Computers & Geosciences 35 (2009) 1817–1826.
[4] S. Rafiee-Taghanaki, M. Arabloo, A. Chamkalani, M. Amani, M.H. Zargari, M.R. Adelzadeh, Implementation of SVM framework to estimate PVT properties of reservoir oil, Fluid Phase Equilibria 346 (2013) 25–32.
[5] M.B. Standing, Oil-system Correlation, McGraw-Hill Book Co., New York City, 1962.
[6] P. Bandyopadhyay, A. Sharma, Development of a new semi analytical model for prediction of bubble point pressure of crude oils, Journal of Petroleum Science and Engineering 78 (2011) 719–731.
[7] M. Asoodeh, P. Bagheripour, Estimation of bubble point pressure from PVT data using a power-law committee with intelligent systems, Journal of Petroleum Science and Engineering 90–91 (2012) 1–11.
[8] P.L. Moses, Engineering applications of phase behavior of crude oil and condensate systems, SPE Journal of Petroleum Technology 38 (1986) 715–723.
[9] D.Y. Peng, D.B. Robinson, A new two-constant equation of state, Industrial and Engineering Chemistry Fundamentals 15 (1976) 59–64.
[10] G. Soave, Equilibrium constants from a modified Redlich–Kwong equation of state, Chemical Engineering Science 27 (1972) 1197–1203.
[11] N.C. Patel, A.S. Teja, A new cubic equation of state for fluids and fluid mixtures, Chemical Engineering Science 37 (1982) 463–473.
[12] G. Schmidt, H. Wenzel, A modified van der Waals type equation of state, Chemical Engineering Science 35 (1980) 1503–1512.
[13] T.H. Ahmed, Comparative study of eight equations of state for predicting hydrocarbon volumetric phase behavior, SPE Reservoir Engineering 3 (1988) 337–348.
[14] R. Wu, L. Rosenegger, Comparison of PVT properties from equation of state analysis and PVT correlations for reservoir studies, Journal of Canadian Petroleum Technology 39 (2000).
[15] A. Danesh, D.H. Xu, A.C. Todd, Comparative study of cubic equations of state for predicting phase behaviour and volumetric properties of injection gas-reservoir oil systems, Fluid Phase Equilibria 63 (1991) 259–278.
[16] S.O. Olatunji, A. Selamat, A.A.A. Raheem, Predicting correlations properties of crude oil systems using type-2 fuzzy logic systems, Expert Systems with Applications 38 (2011) 10911–10922.
[17] E.S. Osman, M. Al-Marhoun, Artificial neural networks models for predicting PVT properties of oil field brines, Proceedings 14th SPE Middle East oil and gas show and conference, Bahrain, 2005.
[18] M. Standing, A pressure–volume–temperature correlation for mixtures of California oils and gases, Drilling and Production Practice (1947) 275–287.
[19] A.A. Al-Shammasi, Bubble point pressure and oil formation volume factor correlations, Middle East Oil Show and Conference, Bahrain, 1999.
[20] J. Lasater, Bubble point pressure correlation, Journal of Petroleum Technology 10 (1958) 65–67.
[21] O. Glaso, Generalized pressure–volume–temperature correlations, Journal of Petroleum Technology 32 (1980) 785–795.
[22] M. Al-Marhoun, PVT correlations for Middle East crude oils, Journal of Petroleum Technology 40 (1988) 650–666.
[23] A. Malallah, R. Gharbi, M. Algharaib, Accurate estimation of the world crude oil PVT properties using graphical alternating conditional expectation, Energy & Fuels 20 (2006) 688–698.
[24] P.P. Valkó, W.D. McCain Jr., Reservoir oil bubblepoint pressures revisited; solution gas–oil ratios and surface gas specific gravities, Journal of Petroleum Science and Engineering 37 (2003) 153–169.
[25] T. Kartoatmodjo, Z. Schmidt, Large data bank improves crude physical property correlations, Oil and Gas Journal 92 (1994) 51–55.
[26] M.H. Nikpoor, H.H. Khanamiri, New empirical correlations for predicting crude oil properties, Petroleum Science and Technology 29 (2011) 1649–1658.
[27] M.E. Dokla, M.E. Osman, Correlation of PVT properties for UAE crudes, SPE Formation Evaluation 7 (1992) 41–46.
[28] G.E. Petrosky Jr., F.F. Farshad, Pressure–volume–temperature correlations for Gulf of Mexico crude oils, SPE Annual Technical Conference and Exhibition, Houston, USA, 1993.
[29] J. Velarde, T.A. Blasingame, J.W.D. McCain, Correlation of black oil properties at pressures below bubble point pressure — a new approach, Annual Technical Meeting, Calgary, Alberta, 1997.
[30] A.M. Elsharkawy, Modeling the properties of crude oil and gas systems using RBF network, SPE Asia Pacific Oil and Gas Conference and Exhibition, Society of Petroleum Engineers Inc., Perth, Australia, 1998.
[31] N. Varotsis, V. Gaganis, J. Nighswander, P. Guieze, A novel non-iterative method for the prediction of the PVT behavior of reservoir fluids, SPE Annual Technical Conference and Exhibition, Society of Petroleum Engineers, Houston, Texas, 1999.
[32] M.A. Al-Marhoun, E.A. Osman, Using artificial neural networks to develop new PVT correlations for Saudi crude oils, International Petroleum Exhibition and Conference, Abu Dhabi, UAE, 2002.
[33] H.M. Goda, E.M.E.-M. Shokir, K.A. Fattah, M.H. Sayyouh, Prediction of the PVT data using neural network computing theory, Nigeria Annual International Conference and Exhibition, Khaled A. Fattah, Abuja, Nigeria, 2003.
[34] R.B. Gharbi, A.M. Elsharkawy, M. Karkoub, Universal neural-network-based model for estimating the PVT properties of crude oil systems, Energy & Fuels 13 (1999) 454–458.
[35] S. Dutta, J.P. Gupta, PVT correlations for Indian crude using artificial neural networks, Journal of Petroleum Science and Engineering 72 (2010) 93–109.
[36] S.O. Olatunji, A. Selamat, A.A. Abdul Raheem, S. Omatu, Modeling the correlations of crude oil properties based on sensitivity based linear learning method, Engineering Applications of Artificial Intelligence 24 (2011) 686–696.
[37] R.B.C. Gharbi, A.M. Elsharkawy, Neural network model for estimating the PVT properties of Middle East crude oils, SPE Reservoir Evaluation and Engineering 2 (1999) 255–265.
[38] M.A. Al-Marhoun, New correlations for formation volume factors of oil and gas mixtures, Journal of Canadian Petroleum Technology 31 (1992).
[39] E.A. El-Sebakhy, T. Sheltami, S.Y. Al-Bokhitan, Y. Shaaban, P.D. Raharja, Y. Khaeruzzaman, Support vector machines framework for predicting the PVT properties of crude-oil systems, SPE Middle East Oil and Gas Show and Conference, Bahrain, 2007.
[40] J.N. Moghadam, K. Salahshoor, R. Kharrat, Introducing a new method for predicting PVT properties of Iranian crude oils by applying artificial neural networks, Petroleum Science and Technology 29 (2011) 1066–1079.
[41] A. Eslamimanesh, F. Gharagheizi, M. Illbeigi, A.H. Mohammadi, A. Fazlali, D. Richon, Phase equilibrium modeling of clathrate hydrates of methane, carbon dioxide, nitrogen, and hydrogen + water soluble organic promoters using Support Vector Machine algorithm, Fluid Phase Equilibria 316 (2012) 34–45.
[42] W.D.J. McCain, R.B. Soto, P.P. Valko, T.A. Blasingame, Correlation of bubblepoint pressures for reservoir oils — a comparative study, SPE Eastern Regional Meeting, Pittsburgh, Pennsylvania, 1998.
[43] S. Nowroozi, M. Ranjbar, H. Hashemipour, M. Schaffie, Development of a neural fuzzy system for advanced prediction of dew point pressure in gas condensate reservoirs, Fuel Processing Technology 90 (2009) 452–457.
[44] A. Chamkalani, A.H. Mohammadi, A. Eslamimanesh, F. Gharagheizi, D. Richon, Diagnosis of asphaltene stability in crude oil through "two parameters" SVM model, Chemical Engineering Science 81 (2012) 202–208.
[45] A.H. Mohammadi, A. Eslamimanesh, F. Gharagheizi, D. Richon, A novel method for evaluation of asphaltene precipitation titration data, Chemical Engineering Science 78 (2012) 181–185.
[46] A. Shokrollahi, M. Arabloo, F. Gharagheizi, A.H. Mohammadi, Intelligent model for prediction of CO2 — reservoir oil minimum miscibility pressure, Fuel 112 (2013) 375–384.
[47] M.A. Ahmadi, M. Ebadi, A. Shokrollahi, S.M.J. Majidi, Evolving artificial neural network and imperialist competitive algorithm for prediction oil flow rate of the reservoir, Applied Soft Computing 13 (2013) 1085–1098.
[48] A. Chamkalani, M. Arabloo, R. Chamkalani, M.H. Zargari, M.R. Dehestani-Ardakani, M. Farzam, Soft computing method for prediction of CO2 corrosion in flow lines based on neural network approach, Chemical Engineering Communications 200 (2013) 731–747.
[49] F.-J. Chang, Y.-T. Chang, Adaptive neuro-fuzzy inference system for prediction of water level in reservoir, Advances in Water Resources 29 (2006) 1–10.
[50] A. Khajeh, H. Modarress, B. Rezaee, Application of adaptive neuro-fuzzy inference system for solubility prediction of carbon dioxide in polymers, Expert Systems with Applications 36 (2009) 5728–5732.
[51] A. Eslamimanesh, F. Gharagheizi, A. Mohammadi, H.D. Richon, Phase equilibrium modeling of structure H clathrate hydrates of methane + water "insoluble" hydrocarbon promoter using QSPR molecular approach, Journal of Chemical & Engineering Data 56 (2011) 3775–3793.
[52] K.H. Coats, G.T. Smart, Application of a regression-based EOS PVT program to laboratory data, SPE Reservoir Engineering 1 (1986) 277–299.
[53] A. Danesh, D.h. Xu, A.C. Todd, A grouping method to optimize oil description for compositional simulation of gas-injection processes, SPE Reservoir Engineering 7 (1992) 343–348.
[54] A.M. Elsharkawy, An empirical model for estimating the saturation pressures of crude oils, Journal of Petroleum Science and Engineering 38 (2003) 55–77.
[55] H.M. Moharam, M.A. Fahim, Prediction of viscosity of heavy petroleum fractions and crude oils using a corresponding states method, Industrial and Engineering Chemistry Research 34 (1995) 4140–4144.
[56] R. Wu, L. Rosenegger, Integrated oil PVT characterization — lessons from four case histories, Journal of Canadian Petroleum Technology 38 (1999).
[57] K.S. Pedersen, A.L. Blilie, K.K. Meisingset, PVT calculations on petroleum reservoir fluids using measured and estimated compositional data for the plus fraction, Industrial and Engineering Chemistry Research 31 (1992) 1378–1384.
[58] K.S. Pedersen, A. Fredenslund, P. Thomassen, Properties of Oil and Natural Gases, Gulf Publishing, Houston, 1989.
[59] A. Hoffman, J. Crump, C. Hocott, Equilibrium constants for a gas-condensate system, Journal of Petroleum Technology 5 (1953) 1–10.
[60] T. Ahmed, Hydrocarbon Phase Behavior, Gulf Publishing, Houston, 1989.
[61] Y.K. Li, L. Nghiem, A. Siu, Phase behaviour computations for reservoir fluids: effect of pseudo-components on phase diagrams and simulation results, Journal of Canadian Petroleum Technology 24 (1985).
[62] J.L. Vogel, L. Yarborough, The effect of nitrogen on the phase behavior and physical properties of reservoir fluids, SPE/DOE Enhanced Oil Recovery Symposium, Tulsa, Oklahoma, 1980.
[63] C. Williams, E. Zana, G. Humphrys, Use of the Peng–Robinson equation of state to predict hydrocarbon phase behavior and miscibility for fluid displacement, SPE/DOE on Enhanced Oil Recovery, Tulsa, OK, 1980.
[64] J. Drohm, W. Goldthorpe, R. Trengove, Enhancing the evaluation of PVT data, Offshore South East Asia Show, Singapore, 1988.

[65] K.C. Hong, Lumped-component characterization of crude oils for compositional simulation, Symp. on Enhanced Oil Recovery, Tulsa, OK, 1982.
[66] R. Agarwal, Y.K. Li, L. Nghiem, A regression technique with dynamic parameter selection for phase-behavior matching, SPE Reservoir Engineering 5 (1990) 115–120.
[67] W.G. Riemens, A.M. Schulte, L.N.J. Jong, Birba field PVT variations along the hydrocarbon column and confirmatory field tests, Journal of Petroleum Technology 40 (1988) 83–88.
[68] B. Jhaveri, G. Youngren, Three-parameter modification of the Peng–Robinson equation of state to improve volumetric predictions, SPE Reservoir Engineering 3 (1988) 1033–1040.
[69] R. Jacoby, V. Berry Jr., A method for predicting pressure maintenance performance for reservoirs producing volatile crude oil, Transactions of AIME 213 (1958) 59.
[70] J.A.K. Suykens, J. Vandewalle, Least squares support vector machine classifiers, Neural Processing Letters 9 (1999) 293–300.
[71] J.A.K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, Least squares support vector machines, Tutorial IJCNN, 2003.
[72] N. Cristianini, J. Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, Cambridge university press, 2000.
[73] F. Gharagheizi, A. Eslamimanesh, F. Farjood, A. Mohammadi, H.D. Richon, Solubility parameters of nonelectrolyte organic compounds: determination using quantitative structure–property relationship strategy, Industrial and Engineering Chemistry Research 50 (2011) 11382–11395.
[74] K. Pelckmans, J.A.K. Suykens, T. Van Gestel, J. De Brabanter, L. Lukas, B. Hamers, B. De Moor, J. Vandewalle, LS-SVMlab: a matlab/c toolbox for least squares support vector machines, Tutorial, KULeuven-ESAT, Leuven, Belgium, 2002.
[75] M. Fazavi, S.-M. Hosseini, M. Arabloo, A. Shokrollahi, M. Amani, Applying a smart technique for accurate determination of flowing oil/water pressure gradient in horizontal pipelines, Journal of Dispersion Science and Technology (2013), (in press).
[76] H. Liu, X. Yao, R. Zhang, M. Liu, Z. Hu, B. Fan, Accurate quantitative structure–property relationship model to predict the solubility of C60 in various solvents based on a novel approach using a least-squares support vector machine, The Journal of Physical Chemistry B 109 (2005) 20565–20571.
[77] D.T. Manallack, D.J. Livingstone, Neural networks in drug discovery: have they lived up to their promise? European Journal of Medicinal Chemistry 34 (1999) 195–208.
[78] F. Gharagheizi, A. Eslamimanesh, A.H. Mohammadi, D. Richon, QSPR approach for determination of parachor of non-electrolyte organic compounds, Chemical Engineering Science 66 (2011) 2959–2967.
[79] K. Price, R. Storn, Differential evolution a simple and efficient adaptive scheme for global optimization over continuous spaces, Technical Report TR-95-012, ICSI, 1998.
[80] A.A. AlQuraishi, Determination of crude oil saturation pressure using linear genetic programming, Energy & Fuels 23 (2009) 884–887.
[81] S. Xavier-de-Souza, J.A.K. Suykens, J. Vandewalle, D. Bollé, Coupled simulated annealing, IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics 40 (2010) 320–335.
[82] J.-N. Jaubert, L. Avaullee, J.-F. Souvay, A crude oil data bank containing more than 5000 PVT and gas injection data, Journal of Petroleum Science and Engineering 34 (2002) 65–107.
[83] A. Danesh, PVT and Phase Behaviour of Petroleum Reservoir Fluids, First ed. Elsevier Science B.V, Amsterdam, Netherlands, 1998.
[84] F. Gharagheizi, A. Eslamimanesh, M. Sattari, A.H. Mohammadi, D. Richon, Development of corresponding states model for estimation of the surface tension of chemical compounds, AICHE Journal 59 (2013) 613–621.
[85] A.H. Mohammadi, F. Gharagheizi, A. Eslamimanesh, D. Richon, Evaluation of experimental data for wax and diamondoids solubility in gaseous systems, Chemical Engineering Science 81 (2012) 1–7.
[86] P. Gramatica, Principles of QSAR models validation: internal and external, QSAR and Combinatorial Science 26 (2007) 694–701.
[87] C.R. Goodall, Computation Using the QR Decomposition, Elsevier, Amsterdam, North-Holland, 1993.