Renewable and Sustainable Energy Reviews 173 (2023) 113103

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

A reliable model to predict the methane-hydrate equilibrium: An updated database and machine learning approach

Mostafa Hosseini [a], Yuri Leonenko [a,b,*]

[a] Department of Earth and Environmental Sciences, University of Waterloo, Waterloo, ON, N2L 3G1, Canada
[b] Department of Geography and Environmental Management, University of Waterloo, N2L 3G1, Canada

ARTICLE INFO

Keywords:
Methane
Hydrate
Machine learning
Temperature
Modeling
Extremely randomized trees

ABSTRACT

Gas hydrates are a type of crystalline compounds that consists of water and small gas molecules. A wide range of applications of gas hydrates in storing natural gas in the form of artificially created solid hydrates, known as solidified natural gas technology, gas separation processes, and seawater desalination technology, has attracted great interest in scientific and practical studies. Gas hydrate formation may also cause deleterious effects such as blockage of gas pipelines. Therefore, accurate prediction of equilibrium conditions for gas hydrates is of great interest. The purpose of this study was to propose machine learning based models to predict methane-hydrate formation temperature for a wide range of brines. In this regard, firstly, a comprehensive database including 987 data samples covering 15 different brines was gathered from the literature. After data cleaning and preparation, three different models of multilayer perceptron, decision tree, and extremely randomized trees were used and tested. The results showed that the extremely randomized trees is capable of predicting methane-hydrate formation temperature with good accuracy. The root mean squared error for this model for the testing dataset was acquired as 0.6248, which shows its great accuracy. The findings of this study can be used as a reliable tool to predict the methane-hydrate formation PT curve in the pure water, single-salt brines, and multi-salt brines.

1. Introduction

Gas hydrates are a type of crystalline compounds that consists of water and small gas molecules known as the host and guest molecules, respectively. These compounds are formed at low temperatures and high pressures [1]. CH_4 (methane), C_2H_6 (ethane), N_2 (nitrogen), CO_2 (carbon dioxide), H_2 (hydrogen), and H_2S (hydrogen sulfide) can be mentioned as guest molecules in Gas hydrates [2]. The formation of gas hydrates occurs when guest molecules are trapped inside the cage structures created by host molecules [1]. Crystal structures of gas hydrates can be classified into three classes sI, sII, and sH, each of which is composed of different polyhedral combinations, where each small gas molecule affects the gas hydrate equilibrium and structure [1–3].

Natural gas hydrates can be mentioned as one of the energy sources because they have been discovered in many offshore and permafrost regions [4,5]. Natural gas is the cleanest fossil fuel which plays a crucial role in supplying the world's energy needs. Natural gas contains approximately 90% and in most cases much more methane. Among fossil fuels, the use of natural gas as the primary energy will continue to grow until 2040. High calorific value, low carbon emissions, and the improvement of electricity generation efficiency, compared to coal and gasoline, increases the use of natural gas, especially for electricity generation applications. More clearly, the use of natural gas for electricity generation reduces CO2 emissions by about 50% and 33%, respectively, compared to the use of coal and oil [6]. Thus, efficient options for natural gas storage and transportation are critical around the world [7]. The benefits that come with storing natural gas in the form of artificially created solid hydrates, known as solidified natural gas (SNG), provide a unique opportunity to store and ship it around the world. The SNG technology provide us with various advantages such as high volume storage capacity, environmental friendliness, high safety, cost-effectiveness, and large-scale natural gas storage [8].

It is noteworthy to mention that the SNG technology focuses on the use of methane gas due to the availability of this gas. Also, a significant portion of research over the past three decades has focused on methane (natural gas) storage in hydrates [6]. Moreover, as salts have a significant effect on both the thermodynamics and kinetics of methane hydrates, much attention has recently been directed to the use of seawater in addition to pure water for the formation of methane hydrates [8]. In

* Corresponding author. Department of Earth and Environmental Sciences, University of Waterloo, N2L 3G1, Canada.
E-mail address: leonenko@uwaterloo.ca (Y. Leonenko).

https://doi.org/10.1016/j.rser.2022.113103
Received 21 July 2022; Received in revised form 26 November 2022; Accepted 1 December 2022
Available online 11 December 2022

Nomenclature

Parameters

AARD	Average Absolute Relative Deviation
ARD	Absolute Relative Deviation
C	Salt concentration, %wt
max_depth	maximum depth of a tree
max_features	maximum number of features
Mean	Mean value
n_estimator	Number of trees in an ensemble method
P	Pressure, MPa
R^2	Coefficient of determination
Res_i	Residual
RMSE	Root Mean Squared Error
T	Temperature, K
t	Target value
x	Ion's mole fraction
y	Model's Output

Abbreviations

ANN	Artificial Neural Network
BOP	Blow Out Preventer
CV	Cross-Validation
DOI	Digital Object Identifier
DT	Decision Tree
ET	Extremely randomized Trees
GBR	Gradient boosting regression
k-NN	k -Nearest Neighbor
LSSVM	Least Squares Support Vector Machine
MLP	Multilayer Perceptron
RF	Random Forest
SVR	Support Vector Regression

addition, the use of seawater can improve the methane hydrates formation economics [9].

Other applications of gas hydrates include their use in gas separation processes using selective partitioning of the target gas between the hydrate crystal phase and gas phase [10–12]. On the other hand, the formation of natural gas hydrates in an offshore wellbore and in oil and gas pipelines within cold regions causes flow-assurance problems [13,14]. Therefore, accurate prediction of equilibrium conditions for gas hydrates is necessary for a variety of reasons. Thus far, temperatures and pressures of gas hydrate equilibrium have been obtained from experiments [1], thermodynamics models [15–19], empirical correlations [20–27], and artificial intelligence techniques [28,29].

In some cases, experimental measurements of gas hydrate equilibrium conditions are costly and time-consuming. The van der Waals and Platteeuw (vdW-P) model, which is based on classical statistical thermodynamics, is one of the most common thermodynamic models for predicting the equilibrium conditions of gas hydrates [30]. However, the vdW-P model also has its limitations [30,31]. For instance, calculating the Langmuir adsorption constants is complex [32,33]. In addition, the vdW-P model has a high error in calculating the gas-hydrate equilibrium at low and high temperatures [34]. Apart from the vdW-P model, other thermodynamic models based on the equations of state have been developed [35,36]. However, estimating the intermolecular interaction parameters of equations of state is also complex [37]. In addition, these thermodynamic models do not have acceptable predictions at low temperatures and high pressures [38]. Although empirical correlations are regularly updated and are straightforward to use, employment of them for mixed gas hydrates over a wide range of temperatures and pressures leads to considerable errors and thus limits their applications in industrial problems [27,30,34].

In recent years, a number of researchers have examined the application of machine learning in the development of prediction models because these techniques have shown promising prediction accuracy [39–42]. Recently, the application of machine learning techniques in predicting gas hydrate formation conditions has received more attention. Soroush et al. (2015) proposed a model based on the artificial neural network (ANN) method to estimate the hydrate formation temperature for different gas mixtures. They utilized 279 experimental data points for construction of their network. They were able to show that the predictions of the proposed ANN model were reasonably consistent with the experimental data [28]. Mesbah et al. (2017) developed a model based on the (least square support vector machine) LSSVM algorithm to predict hydrate formation temperature for a wide range of natural gas mixtures using 279 experimental data points. Their model presented a high accuracy with a squared correlation coefficient (R^2) of 0.9918 [43]. In the following, Amar (2021) modeled the hydrate formation temperature of natural gases using gene expression programming (GEP) model. A database containing 279 experimental data points was employed. The developed GEP model could accurately estimate the hydrate formation temperature an average absolute relative error (AARE) of 0.1397% [44]. The employed database was similar to that of Mesbah et al. (2017) [43]. More recently, Xu et al. [45] used five different machine learning predictive tools, including multiple linear regression, support vector regression, k -nearest neighbor, gradient boosting regression, and random forest, for predicting CH4-hydrate formation condition in different brines. They used 702 data samples in this regard and achieve excellent results. Among their models, gradient boosting regression performed better than the other models.

The purpose of this study is to propose machine learning based models to predict CH_4-hydrate formation temperature for a wide range of brines. In this regard, firstly, a comprehensive database was gathered from the open literature. The database includes 987 data samples for pure water, 15 different single-salt brines, and 6 different multi-salt systems. Three different predictive models of Multilayer Perceptron (MLP) ANN, Decision Tree (DT), and Extremely Randomized Trees (ET) were implemented to accurately predict the methane-hydrate formation PT data. The performance of the developed models was also evaluated by comparison with the literature. The structure of this study is as follows. A brief explanation of the predictive models is provided in section 2. Database development and preparation are presented in section 3. Section 4 is devoted to the result and discussion encompassing model tuning and fitting, accuracy assessment, and comparing the results with the literature. Finally, the concluding remarks are presented in section 5. It is worthwhile mentioning that the prepared database as well as the outputs of the developed models are provided as supplementary material to this study.

2. Methods

In this study, three different predictive models of Multilayer Perceptron (MLP) ANN, Decision Tree (DT), and Extremely Randomized Trees (ET) are considered. The Sickit-Learn library [46] developed for Python is used for model implementation. For the MLP, the function sklearn. neural_network.MLPRegressor () is used. MLP is a black-box model and the fitting process is not explicitly shown. The structure of MLP includes one input layer, one or more hidden layer(s), and one output layer. The number of neurons in the input and output layers corresponds to the number of input and output features. The number of neurons in the hidden layer(s) is arbitrary. For the DT, the function of sklearn. tree.DecisionTreeRegressor () is used. This predictive tool is of non-parametric models. DT works through developing simple decision rules based on the input variables. As DT can be visualized, it can be

Table 1
The sources of the gathered database with the number of data samples for each brine.

Index	Brine	Index	Year	Reference	N
Pure Water					
1	–	1	1983	de Roo et al. [48]	9
		2	1991	Adisasmito et al. [66]	11
		3	1995	Maekawa et al. [61]	3
		4	2001	Jager and Sloan [22]	12
		5	2006	Maekawa and Imai [60]	28
		6	2017	Hu et al. [64]	6
		7	2017	Kamari et al. [55]	17
		8	2021	Bhawangirkar and Sangwai [68]	13
Single Salt					
1	$CaBr_2$	1	2009	Mohammadi et al. [63]	10
		2	2017	Hu et al. [53]	6
		3	2021	Bhawangirkar and Sangwai [68]	41
2	$CaCl_2$	1	2003	Kharrat and Dalmazzone [57]	26
		2	2006	Atik et al. [67]	6
		3	2008	Mohammadi et al. [62]	12
		4	2018	Hu et al. [54]	20
		5	2019	Aregbe et al. [65]	21
		6	2019	Du et al. [50]	14
		7	2020	Li et al. [59]	5
3	K_2CO_3	1	2009	Mohammadi et al. [63]	10
4	KBr	1	2009	Mohammadi et al. [63]	12
5	KCl	1	2008	Mohammadi et al. [62]	13
		2	2009	Haghighi et al. [51]	3
		3	2016	Cha et al. [69]	15
		4	2017	Hu et al. [58]	27
		5	2019	Aregbe et al. [65]	23
		6	2020	Li et al. [59]	7
6	KHCOO	1	2017	Kamari et al. [55]	15
7	$MgBr_2$	1	2021	Bhawangirkar and Sangwai [68]	40
8	$MgCl_2$	1	1998	Kang et al. [56]	22
		2	2006	Atik et al. [67]	21
		3	2009	Haghighi et al. [51]	4
		4	2009	Mohammadi et al. [63]	10
		5	2017	Chong et al. [70]	6
		6	2019	Du et al. [50]	5
		7	2020	Li et al. [59]	5
9	Na_2SO_4	1	2006	Maekawa and Imai [60]	13
10	NaBr	1	2006	Maekawa and Imai [60]	27
		2	2009	Mohammadi et al. [63]	13
11	NaCl	1	1983	de Roo et al. [48]	23
		2	1991	Dholabhai et al. [49]	6
		3	1995	Maekawa et al. [61]	50
		4	2001	Jager and Sloan [22]	42
		5	2006	Maekawa and Imai [60]	30
		6	2008	Mohammadi et al. [62]	5
		7	2009	Haghighi et al. [51]	6
		8	2016	Cha et al. [69]	16
		9	2017	Hu et al. [64]	28
		10	2019	Aregbe et al. [65]	36
		11	2019	Du et al. [50]	13
		12	2020	Li et al. [59]	7
12	NaHCOO	1	2017	Kamari et al. [55]	15
13	NaI	1	2006	Maekawa and Imai [60]	9
14	NH_4Cl	1	2016	Cha et al. [69]	21
15	$ZnBr_2$	1	2021	Bhawangirkar and Sangwai [68]	38
Multi Salt					
1	$CaCl_2$+$MgCl_2$	1	2017	Hu et al. [52]	3
2	NaCl + $CaCl_2$	1	1991	Dholabhai et al. [49]	24
		2	2017	Hu et al. [52]	15
3	NaCl + $CaCl_2$+$MgCl_2$	1	2017	Hu et al. [52]	3
4	NaCl + KCl	1	1991	Dholabhai et al. [49]	31
		2	2017	Hu et al. [52]	6
5	NaCl + KCl + $CaCl_2$+$MgCl_2$	1	2017	Hu et al. [52]	6
6	NaCl + $MgCl_2$	1	2017	Hu et al. [52]	3

Table 2
The ranges for brine concentrations based on both salt and ion types.

Brine	*C*, %wt	Ion	*x*, mole fraction
$CaBr_2$	0.1–32	Na^+	0.0066–0.1167
$CaCl_2$	1.5–35.5	K^+	0.0064–0.1388
K_2CO_3	5–10	NH_4^+	0.0174–0.0361
KBr	5–10	Zn^+	0.0001–0.0139
KCl	2.6–40	Ca^{++}	0.0001–0.0820
KHCOO	5–20	Mg^{++}	0.0001–0.0323
$MgBr_2$	0.1–15	Cl^-	0.0038–0.1640
$MgCl_2$	1–15	Br^-	0.0002–0.0814
Na_2SO_4	10–10	I^-	0.0132–0.0132
NaBr	5–20	$HCOO^-$	0.0111–0.0621
NaCl	2.1–30	CO_3^{--}	0.0068–0.0143
NaHCOO	5–20	SO_4^{--}	0.0139–0.0139
NaI	10–10		
NH_4Cl	5–10		
$ZnBr_2$	0.1–15		

considered as a with-box model. However, if the DT is too deep, it will lose its simplicity to be visually understood and interpreted. For the ET, the function of sklearn. ensemble.ExtraTreesRegressor () is used. ET is an ensemble method that benefits from averaging several DTs to provide a superior result and minimize the over-fitting probability. More details about the used predictive methods are available in the literature [46, 47].

3. Database development

3.1. Literature survey

A comprehensive literature review was performed to gather all the data available on the CH4-hydrate phase equilibria. In this regard, different studies, including those that had a database [22,45,48–70] were recognized containing relevant data. The gathered database includes 987 data samples for pure water, 15 different single-salt brines, and 6 different multi-salt systems. Covering 15 different salts, it is the largest database gathered on the topic of interest so far. A comprehensive introduction to the gathered data is provided. The details of the gathered data are elaborated in Table 1. This table lists the publications and corresponding data. As one can see, NaCl has been the most studied brine in the literature as 12 different studies reported data on CH_4-hydrate phase equilibria for this system. Table 2 displays the range of salt concentrations in brines expressed as weight percentages of salt as well as the range of mole fractions of ions in salts for the database compiled from the literature. Among the studied brines, KCl has the broadest salt concentration (C) in the range of 2.6–40 %wt..

All the single-salt systems have the same trend regarding the pressure (P) and salt concentration (C) effects on the equilibrium temperature (T). The P-T diagrams for the NaBr (5%, 10% and 20%) system is depicted in Fig. 1 (a) as an example. This figure shows that in a constant P, an increase in the C results in a decrease in the equilibrium T. On the other hand, for a fixed C, there is a monotonic increase in T by increasing P. The same goes true for the multi-salts systems (see Fig. 1 (b) for NaCl + $CaCl_2$). This figure suggests that the concentration of different ions has a significant influence on the phase equilibria.

3.2. Data quality check

It is always probable that incorrect data creep into the reported database for several reasons such as errors in the devices or human error in recording. Without a correct and clean database, the developed predictive models would not be accurate and general enough. Hence, all the data were checked for any inconsistency. In data reported by Dholabhai et al. [49], there is an inconsistent T reported as 72.22 K. Considering the T values reported for other data samples and the authors' statement for the tested T range of 264–284 K, it seems that there has been an error

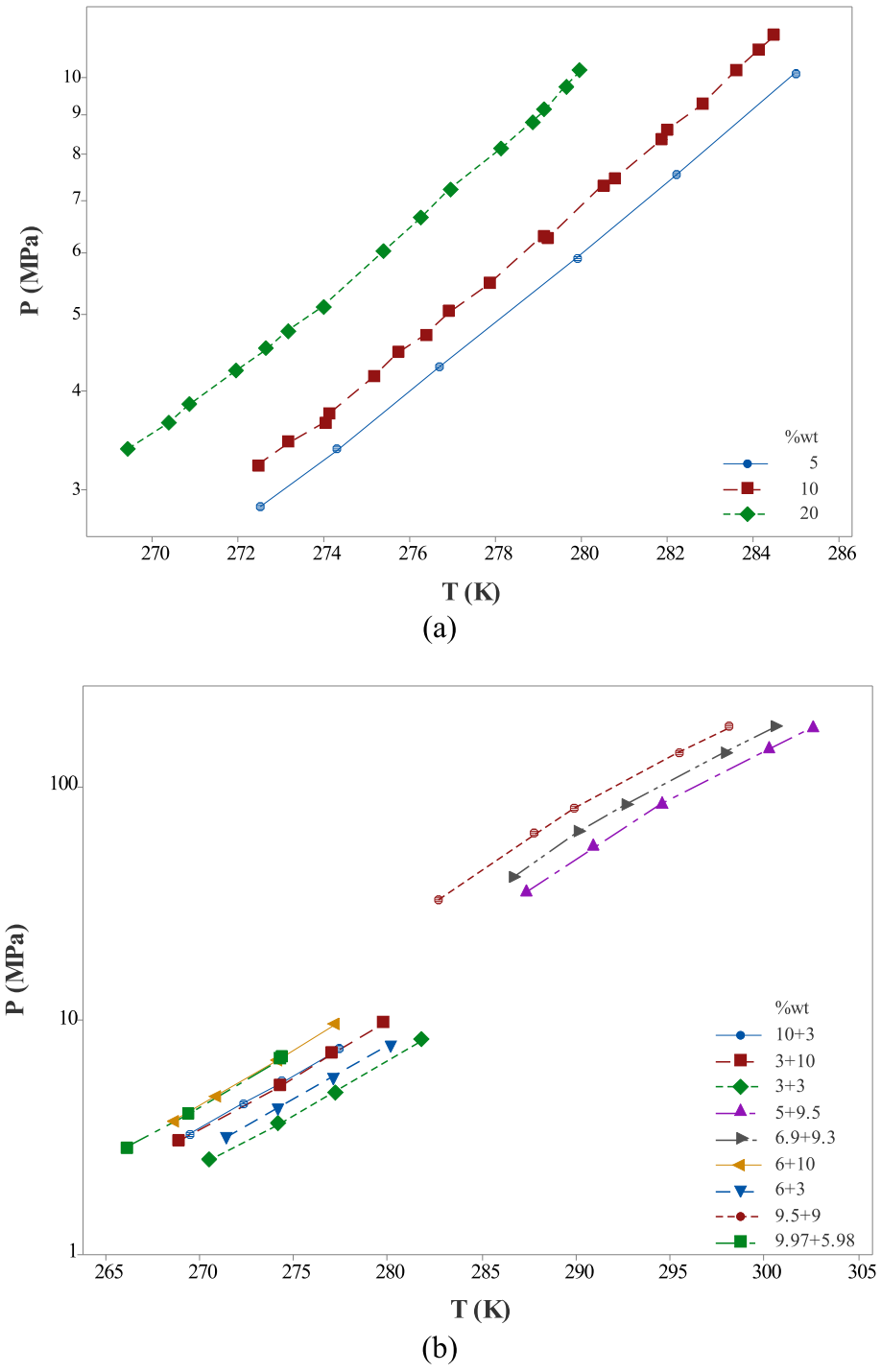


Fig. 1. The P-T diagram for (a) NaBr and (b) NaCl + $CaCl_2$ systems.

in recording, and the correct value might have been 272.22 K.

Additionally, the database was checked for duplicate samples. The presence of duplicates has a deleterious effect on the machine learning model's training as it cannot learn the true pattern. Table 3 lists the present duplicates in the gathered database. As can be seen, 5 and 5 cases were detected for pure water and NaCl systems, respectively. For the pure water, 4 cases of duplicates were reported in different studies. On the other hand, all the duplicates for NaCl brine were reported in the same study. As the duplicate values are very close to each other, the discrepancy might be due to the different methods of measurement or repeating the measurements. In this study, one of the duplicates was excluded from the database.

3.3. Feature selection for modeling

Almost all the experimental works in the literature have reported C in %wt. However, it is opted to use ion's mole fraction (x) for the modeling purpose for several reasons. First of all, using C, it is not possible to correctly address the multi-salts systems. As was illustrated in Fig. 1 (b), the effects of the presence of different salts cannot be adequately shown by C. Secondly, using ions concentration, there would be 12 characteristic features for brines, which is lower than the number of different salts, which is 15. Using a fewer number of features not only decreases the processing load but also lowers the overfitting probability. More importantly, using the ions concentration, it is possible to address

Table 3
The present duplicates in the gathered database.

Year	Author	Source	Brine	%wt of salt	P (MPa)	T (K)
1983	de Roo et al. [48]	Table 1	Pure Water	0	3.34	275.41
1983	de Roo et al. [48]	Table 1	Pure Water	0	3.34	275.95
2006	Maekawa and Imai [60]	Figure 4	Pure Water	0	4.19	277.88
2021	Bhawangirkar and Sangwai [68]	Table S1	Pure Water	0	4.19	277.8
2006	Maekawa and Imai [60]	Figure 4	Pure Water	0	5.72	280.9
2021	Bhawangirkar and Sangwai [68]	Table S1	Pure Water	0	5.72	281
2017	Kamari et al. [55]	Table1 1	Pure Water	0	7.25	283.3
2021	Bhawangirkar and Sangwai [68]	Table S1	Pure Water	0	7.25	283.1
1991	Adisasmito et al. [66]	Table II	Pure Water	0	8.55	284.7
2017	Kamari et al. [55]	Table1 1	Pure Water	0	8.55	284.8
2017	Hu et al. [64]	Table S1	NaCl	12	37.8	290.41
2017	Hu et al. [64]	Table S1	NaCl	12	37.8	290.44
2017	Hu et al. [64]	Table S1	NaCl	12	62	293.55
2017	Hu et al. [64]	Table S1	NaCl	12	62	293.75
1995	Maekawa et al. [61]	Figure 2	NaCl	20	6.85	270.25
1995	Maekawa et al. [61]	Figure 2	NaCl	20	6.85	270.60
2017	Hu et al. [64]	Table S1	NaCl	23	32	278.26
2017	Hu et al. [64]	Table S1	NaCl	23	32	278.35
2017	Hu et al. [64]	Table S1	NaCl	26	62	278.75
2017	Hu et al. [64]	Table S1	NaCl	26	62	278.85

the salts not present in the reported experimental data samples. Among the listed ions in Table 2, there are 6 cations and 6 anions. Thus, 36 different single-salt brines can be theoretically developed. As is listed in the same table, experimental data are available for only 15 single-salt systems. We will explore the capability of the developed predictive models for the single-salt systems the data for which are not provided directly. In addition to brine characteristics, P will also be used as the operational parameter to predict the equilibrium T for the CH_4-hydrate.

3.4. Data splitting

The gathered database includes 977 data samples after removing the duplicates. Data splitting in the ratio of 80/20 for training/testing was considered. Before splitting, 40 data samples belonging to $MgBr_2$ were set aside for additional testing. As is shown in Table 1, the experimental data for this brine were reported recently by Bhawangirkar and Sangwai [68]. Although there is no other reference reporting experimental data for this salt, data samples are available containing both Mg^{++} and Br^{-} ions for other salts. As additional testing, if the developed model provides an acceptable prediction for $MgBr_2$, it would be able to be used for any hypothetical salt the ions of which are available in the database. The remaining 937 data samples were split into training and testing subsets in the ratio of 80/20. A 5-fold Cross-Validation (CV) was used to validate the developed models.

3.5. Database supplementary file

The gathered and processed data are provided as supplementary material to this study in a Microsoft Excel worksheet. To eliminate any doubts and misunderstandings about the origin of the data, details of the Digital Object Identifier (DOI), publication year, Authors' names, and the source of data are listed for each data sample. The brine characteristics are mentioned in both C and x for the comprising ions. This database would be available to the readers for future studies.

4. Result and discussion

4.1. Model development

In this section, the steps of model development including hyperparameter optimization and model fitting are introduced. The Sickit-Learn library [46] developed for Python is used in this regard. The road map is as follows. Three different predictive models of MLP, DT, and ET are considered to develop. In the first step, the hyperparameters for each model are optimized. It is worthwhile mentioning that only the main ones are subjected to optimization. Then, the models are fitted several times due to the stochastic nature of the models. The best one for each model type is saved for the accuracy analysis. The accuracy will be checked using both graphical and statistical methods for the training, testing, and additional testing datasets.

4.1.1. Hyperparameter optimization

Hyperparameters denote the external shape of a model and should be determined before the training step. On the other hand, the model's parameters are determined during the training step. The best set of hyperparameters guarantees the best performance of the model for the given dataset. It goes without saying that just like all other actions, hyperparameter optimization is done just using the training dataset. Table 4 shows the determined hyperparameters as well as the searching space for each model. As the search space for the MLP was discrete, an exhaustive search was performed for networks with one hidden layer with neurons in the range of (1)-(10) and networks with two hidden layers with neurons in the range of (1,1)-(10,10). Among the 110 assessed cases, the MLP with two hidden layers with (10,10) neurons were determined to be the best structure. On the other hand, a randomized search with 20 iterations was used for DT and ET methods. The best 5-fold CV Root Mean Squared Error (RMSE) values are listed in the table. It is worthwhile mentioning the reported error values are for the standardized T.

4.1.2. Model fitting

The scaled values were used for MLP, DT, and ET models. Fig. 2 depicts the box-plot for the 5-fold CV RMSE acquired from 20 trials for MLP, DT, and ET models. As is obvious, ET yielded the best results with a mean value equal to 0.19. Regarding the accuracy, after ET are MLP and DT models with respective mean values of 0.23 and 0.29. All the RMSE values for ET are very close to each other. The closer error values to each other for different random states, the more stable the model. As each trial uses a different random state, the yielded results are not the same.

Table 4
The determined hyperparameters as well as the searching space for each model.

	Hidden_layer-size	max_depth	max_features	n_estimators	Best CV RMSE
Search Interval	(1)—(10, 10)	5–50	0.1–1	5–50	–
MLP	(10, 10)	–	–	–	0.1993
DT	–	43	0.7112	–	0.2735
ET	–	21	0.7580	43	0.1842

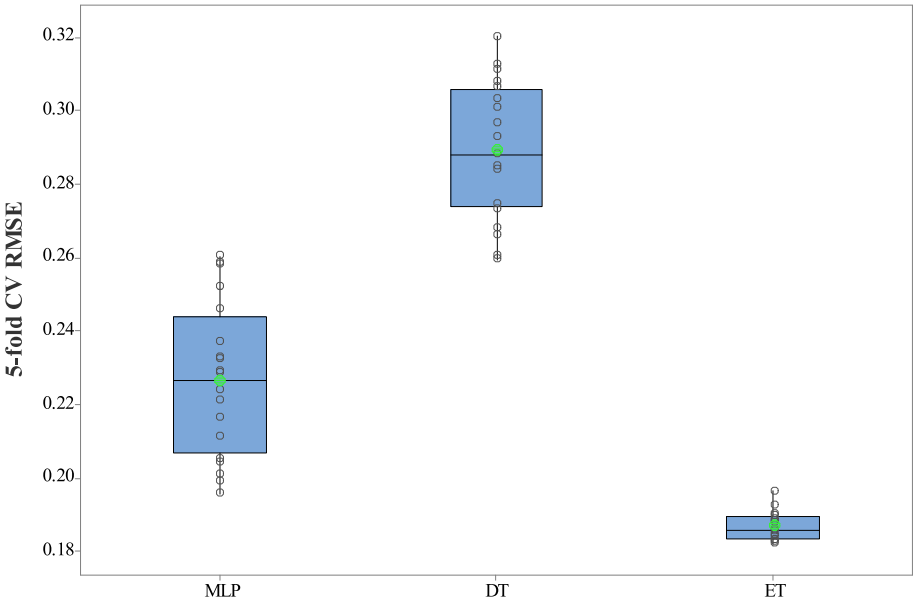


Fig. 2. The box-plot for the 5-fold CV RMSE acquired from 20 trials for MLP, DT, and ET models.

Table 5
The statistical parameters denoting the accuracy of the developed models in this study.

Model	Dataset	R2	RMSE	ARD (%)	AARD (%)	N
MLP	Training	0.9825	0.9978	0.0031	0.1943	749
	Testing	0.9837	0.9815	−0.0058	0.2263	188
DT	Training	1.0000	0.0000	0.0000	0.0000	749
	Testing	0.9621	1.4968	0.0431	0.3792	188
ET	Training	1.0000	0.0380	0.0000	0.0068	749
	Testing	**0.9934**	**0.6248**	**0.0026**	**0.1436**	188

This shows the stability of this model.

4.2. Accuracy assessment

Different statistical parameters can be used to show the error of predictions. In this study, Residual (Res_i), Coefficient of determination (R^2), RMSE, Absolute Relative Error (ARD), and Average Absolute Relative Error (AARD). The formulations of the aforementioned parameters are as follows:

$$Res_i = t_i - y_i \tag{1}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (Res_i)^2}{\sum_{i=1}^{n} (\bar{t} - t_i)^2} \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Res_i)^2} \tag{3}$$

$$ARD = \frac{100}{n} \times \sum_{i=1}^{n} \frac{Res_i}{t_i} \tag{4}$$

$$AARD = \frac{100}{n} \times \sum_{i=1}^{n} \frac{|Res_i|}{t_i} \tag{5}$$

where y and t are the model's output and the target value, respectively.

Table 5 lists the statistical parameters denoting the accuracy of the developed models in this study. The statistical parameters are calculated for the original T, not the scaled ones. The models must be compared using the testing dataset. In this table, the best values are bolded. As is shown, ET has yielded the best results by an RMSE equal to 0.6248 for the testing dataset. Except for ARD, the ET has exhibited the best results for all other parameters. The other important point is the difference between the training and testing errors. For MLP, the errors for the training and testing datasets are close to each other. On the other hand, DT shows a noticeable difference, which is indicative of overtraining. In other words, the developed DT model has failed in showing good performance as it did for the training data. This problem is addressed in the ET model, which is comprised of several individual DTs. By taking the average of individual estimators, not only has the ET model provided excellent results for the training dataset but also has shown very good performance for the testing dataset.

Fig. 3 illustrates the comparison between predicted and experimental T for the testing dataset. Subplots (a) and (b) shows the cross and residual plots, respectively. In both subplots, the x-axis depicts the experimental T. For a perfect prediction, all the data samples lie on the 45° line and zero line for the cross and residual plots, respectively. As one can see, the ET has provided the best results. Both statistical and graphical error analyses corroborate that the proposed ET model is well capable of predicting unseen data samples during the training phase.

4.3. Additional testing

As was mentioned in section 3.4, the data samples belonging to the $MgBr_2$ system were set aside for additional testing before splitting data for training and testing. The reason for selecting this specific system was the fact that there were no instances of this brine in multi-salts systems and also there were several occurrences of its comprising ions in other salts. Thus, it was expected that a well-trained model should be able to provide an adequate result for this system. Table 6 lists the statistical parameters. The best values are bolded. This table suggests that only MLP and ET have provided good results for $MgBr_2$. Fig. 4 shows both cross and residual plots. The residual plot shows that although having overall acceptable predictions, MLP is deficient for exhibiting a trend in residuals. For the first and second halves of the T, MLP tends to under and over-predict, respectively. All in all, the ET provided the best results for the additional testing as it did for training and testing datasets.

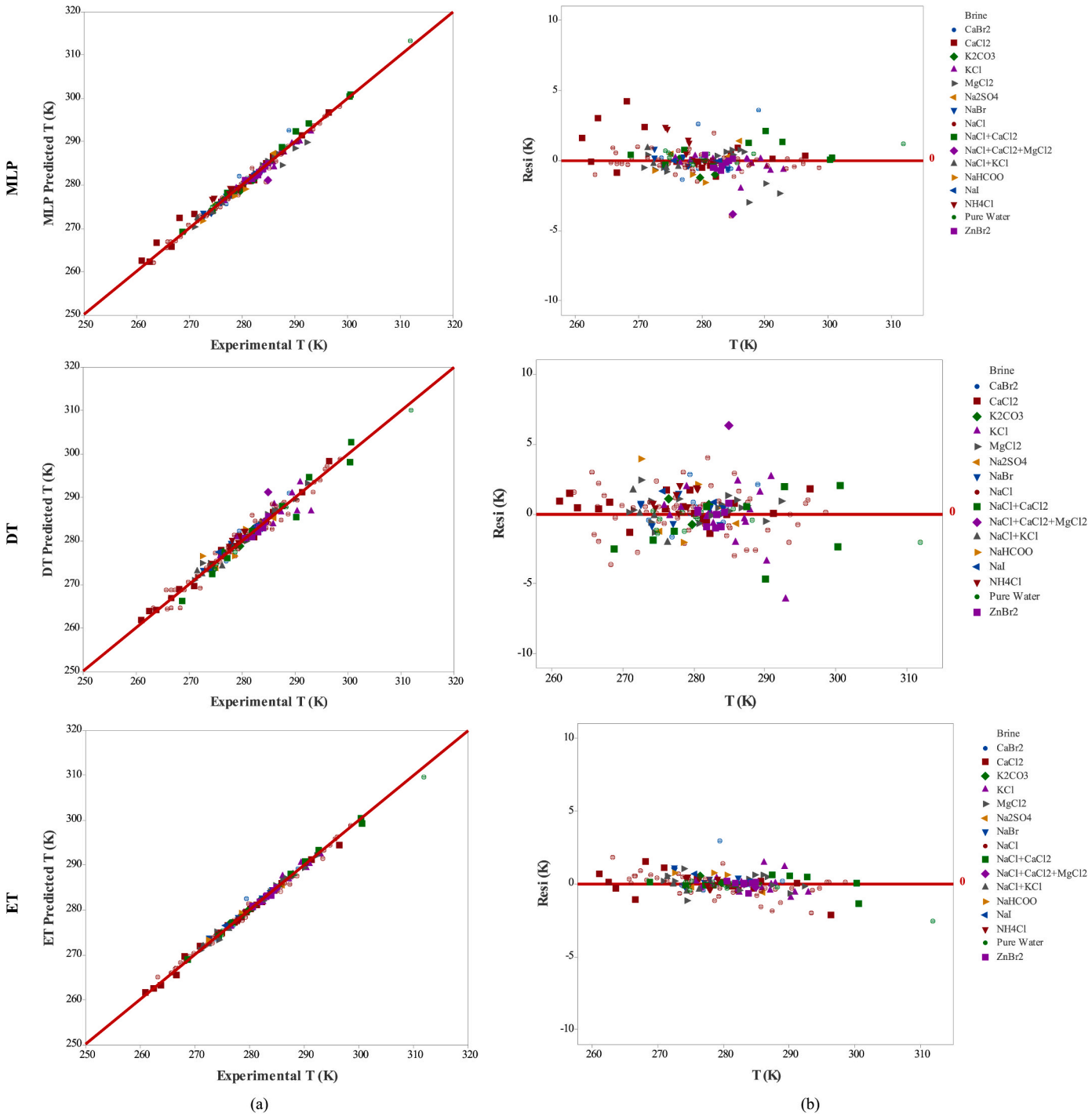


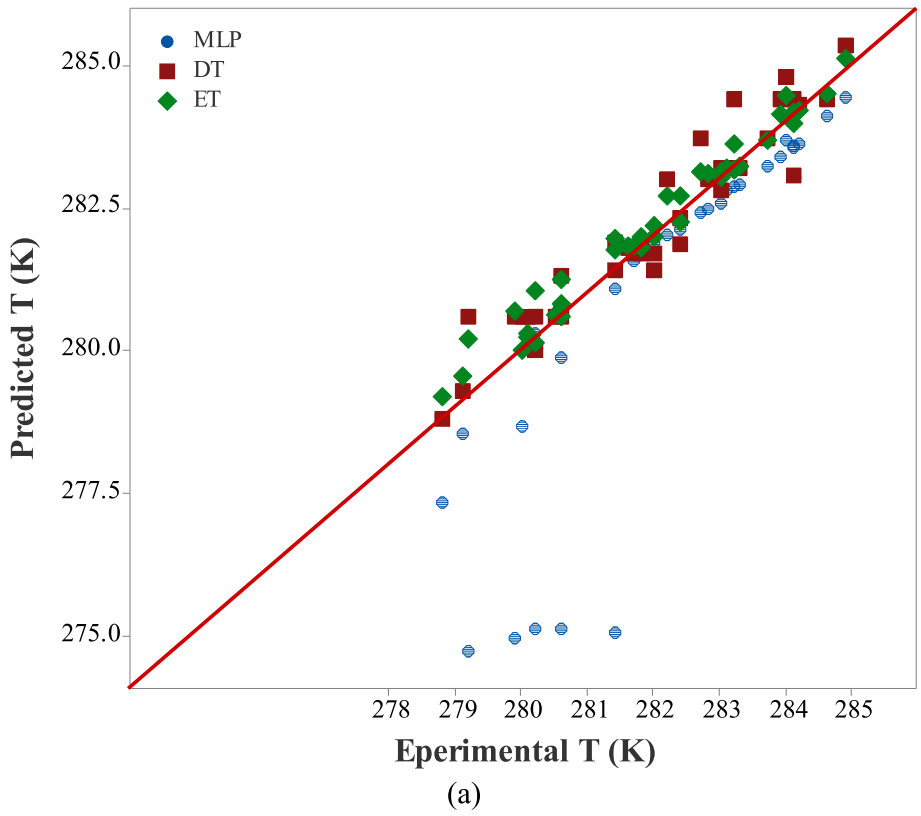
Fig. 3. Comparison of predicted and experimental T for the testing dataset. (a) cross-plot and (b) Residual plot.

Table 6
Statistical error analysis of different proposed models for the MgBr2 system.

Model	R2	RMSE	MPE (%)	MAPE (%)	N
MLP	−0.4026	1.9308	**−0.3321**	0.3488	40
DT	0.8997	0.5163	0.0726	0.1324	40
ET	**0.9534**	**0.3520**	0.0788	**0.0906**	40

4.4. Comparison with literature

The results acquired in this study are compared with the models published recently in 2021 by Xu et al. [45]. Table 7 shows the comparison between the models developed in this study and literature models regarding the RMSE for the testing dataset. The best values are bolded. As one can see the mean RMSE for the developed ET and GBT proposed by Xu et al. [45] are close to each other with a slight superiority of the latter. However, the models developed in this study cover a much wider range of different salts. To be more specific, 15 different salts are used in this study compared with 4 distinct salts used by Xu et al. [45] Regarding the different ions, this study addresses 6 cations and 6 anions. On the other hand, the study of Xu et al. [45] includes only 4 cations and 1 anion. The comparison reveals that the developed ET model has provided a comparable accuracy with the GBR model proposed by Xu et al. [45] but for a wider range of brines. Most of the time,



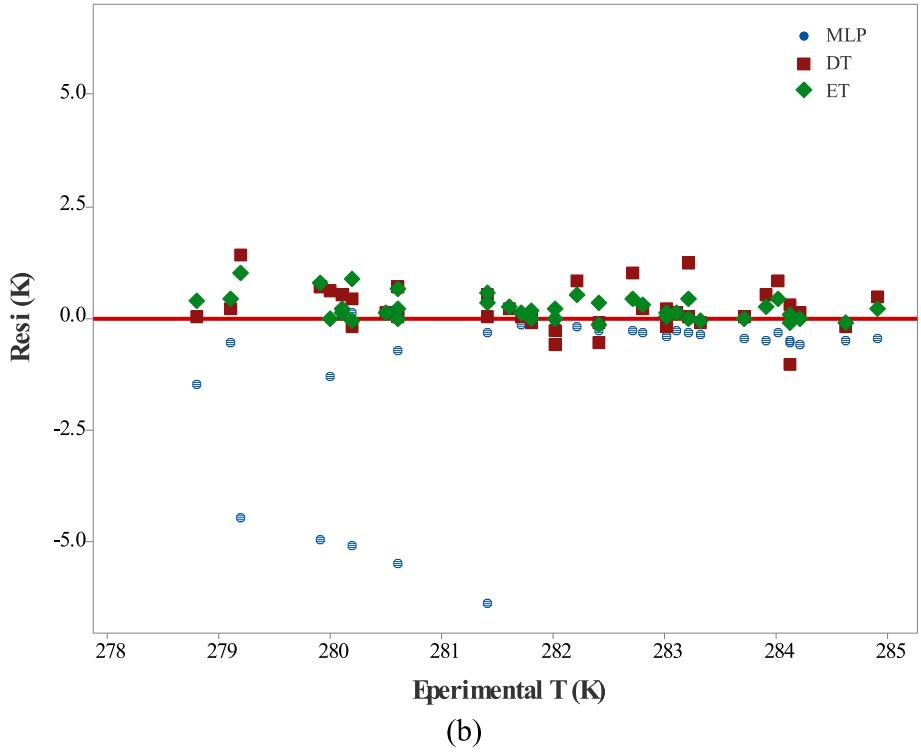


Fig. 4. Comparison among different models for predicting equilibrium T for MgBr2 system (a) cross-plot and (b) residual plot.

the prediction models based on machine learning methods do not provide an explicit equation for predicting equilibrium. However, implementing a web app or web-based user interface allows us to illustrate the application of these models.

5. Conclusion

The purpose of this study was to develop machine learning-based models to accurately predict the methane-hydrate formation temperature in presence of different brines. In this regard, firstly, a comprehensive database was gathered from the open literature. It covered a

Table 7
Comparison between the models developed in this study and literature models regarding the RMSE for the testing dataset.

Brine	Xu et al. [45]					This Study		
	MLR	k-NN	SVM	RF	GBR	MLP	DT	ET
$CaBr_2$	4.54	4.73	2.6	1.08	1.05	1.33	1.14	**0.83**
$CaCl_2$	–	–	–	–	–	1.47	0.99	**0.77**
$CaCl_2$+$MgCl_2$	17.94	12.53	1.02	4.3	**0.42**	–	–	–
K_2CO_3	–	–	–	–	–	0.94	0.84	**0.32**
KCl	4.22	1.67	1.5	1.47	0.66	0.63	2.08	**0.60**
KCl + $CaCl_2$	4.81	1.9	1.46	0.26	**0.06**	–	–	–
$MgBr_2$	–	–	–	–	–	1.93	0.52	**0.35**
$MgCl_2$	4.54	3.71	3.32	1.18	0.71	1.08	0.90	**0.49**
NH_4Cl	–	–	–	–	–	1.35	1.28	**0.31**
Na_2SO_4	–	–	–	–	–	0.84	0.89	**0.37**
NaBr	–	–	–	–	–	**0.45**	0.80	0.55
NaCl	4.24	1.66	2.44	0.81	0.7	0.78	1.53	**0.66**
NaCl + $CaCl_2$	3.98	2.19	1.88	0.92	**0.57**	0.99	2.30	**0.56**
NaCl + $CaCl_2$+$MgCl_2$	6.62	15.57	**0.02**	2.87	0.96	3.82	6.30	0.27
NaCl + KCl	4.85	2.29	2.94	0.58	0.91	0.57	1.33	**0.41**
NaCl + KCl + $CaCl_2$+$MgCl_2$	4.98	2.23	0.84	0.57	**0.1**	–	–	–
NaHCOO	–	–	–	–	–	1.16	2.83	**0.69**
NaI	–	–	–	–	–	**0.03**	1.23	0.50
Pure Water	4.73	**0.23**	2.73	0.38	0.59	0.49	0.92	0.78
$ZnBr_2$	–	–	–	–	–	0.41	0.66	**0.26**

wide range of pressure and temperature and 15 different brines. Data were checked for unreliable data and duplicates. It was opted to use ions' mole fractions to represent the brines so that the developed models would be useable for any possible combination of ions present in the database. Before data splitting into training and testing, $MgBr_2$ samples were set aside for additional testing. The remaining data were split in the ratio of 80/20 for respective training and testing. Three different models of MLP, DT, and ET were used. Firstly, the hyperparameters were optimized using an exhaustive search for the MLP model and a randomized search for DT and ET. Then the models were fitted using the training dataset. The ET model showed the best performance with an RMSE equal to 0.6248 K for the testing dataset. The order of model based on their accuracy was as ET > MLP > DT for the testing data set. Additionally, ET showed the best stability during fitting for 20 different random states. For the additional testing, ET resulted in an RMSE equal to 0.3520, which shows the capability of this model to be used for any brine that can be developed with ions available in the database. All the models also showed acceptable performance for the additional testing. The models were compared with the most recent ones proposed in the literature. The comparison showed a comparable level of accuracy with the developed models covering a much wider range of brines and ions. The findings of the present study can be used as a powerful tool for predicting the methane-hydrate formation conditions in the pure water, single-salt brines, and multi-salt brines.

Credit author statement

Mostafa Hosseini: Conceptualization, Methodology, Investigation, Validation, Writing – original draft. **Yuri Leonenko**: Supervision, Conceptualization, Methodology, Funding acquisition, Writing-Reviewing and Editing,

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors are unable or have chosen not to specify which data has been used.

Acknowledgments

Financial support for this work provided by Natural Sciences and Engineering Research Council of Canada (NSERC).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rser.2022.113103.

References

[1] Sloan Jr ED, Koh CA. Clathrate hydrates of natural gases. CRC press; 2007.
[2] Sloan ED. Clathrate hydrates: the other common solid water phase. Ind Eng Chem Res 2000;39:3123–9.
[3] Jeffrey GA. Hydrate inclusion compounds. J Incl Phenom 1984;1:211–22.
[4] Koh CA, Sum AK, Sloan ED. State of the art: natural gas hydrates as a natural resource. J Nat Gas Sci Eng 2012;8:132–8.
[5] Chong ZR, Yang SHB, Babu P, Linga P, Li X-S. Review of natural gas hydrates as an energy resource: prospects and challenges. Appl Energy 2016;162:1633–52.
[6] Veluswamy HP, Kumar A, Seo Y, Lee JD, Linga P. A review of solidified natural gas (SNG) technology for gas storage via clathrate hydrates. Appl Energy 2018;216:262–85.
[7] Qin Y, Shang L, Lv Z, He J, Yang X, Zhang Z. Methane hydrate formation in porous media: overview and perspectives. J Energy Chem 2022;74:454–80. https://doi.org/10.1016/j.jechem.2022.07.019.
[8] Bhattacharjee G, Veluswamy HP, Kumar A, Linga P. Stability analysis of methane hydrates for gas storage application. Chem Eng J 2021;415:128927.
[9] Omran A, Nesterenko N, Paecklar AA, Barrier N, Valtchev V. Toward economical seawater-based methane hydrate formation at ambient temperature: a combined experimental and computational study. ACS Sustain Chem Eng 2022;10:11617–26. https://doi.org/10.1021/acssuschemeng.2c03530.
[10] Eslamimanesh A, Mohammadi AH, Richon D, Naidoo P, Ramjugernath D. Application of gas hydrate formation in separation processes: a review of experimental studies. J Chem Thermodyn 2012;46:62–71.
[11] Shahnazar S, Hasan N. Gas hydrate formation condition: review on experimental and modeling approaches. Fluid Phase Equil 2014;379:72–85.
[12] Xu C-G, Li X-S. Research progress of hydrate-based CO 2 separation and capture from gas mixtures. RSC Adv 2014;4:18301–16.
[13] Zhang J, Wang Z, Liu S, Zhang W, Yu J, Sun B. Prediction of hydrate deposition in pipelines to improve gas transportation efficiency and safety. Appl Energy 2019;253:113521.
[14] Zhang J, Wang Z, Sun B, Sun X, Liao Y. An integrated prediction model of hydrate blockage formation in deep-water gas wells. Int J Heat Mass Tran 2019;140:187–202.
[15] Javanmardi J, Moshfeghian M, Maddox RN. An accurate model for prediction of gas hydrate formation conditions in mixtures of aqueous electrolyte solutions and alcohol. Can J Chem Eng 2001;79:367–73.
[16] Nasrifar K, Moshfeghian M. A model for prediction of gas hydrate formation conditions in aqueous solutions containing electrolytes and/or alcohol. J Chem Thermodyn 2001;33:999–1014.

[17] Eslamimanesh A, Mohammadi AH, Richon D. Thermodynamic model for predicting phase equilibria of simple clathrate hydrates of refrigerants. Chem Eng Sci 2011;66:5439–45.
[18] Moradi G, Khosravani E. Application of PRSV2 equation of state to predict hydrate formation temperature in the presence of inhibitors. Fluid Phase Equil 2012;333:18–26.
[19] Moradi G, Khosravani E. Modeling of hydrate formation conditions for CH4, C2H6, C3H8, N2, CO2 and their mixtures using the PRSV2 equation of state and obtaining the Kihara potential parameters for these components. Fluid Phase Equil 2013;338:179–87.
[20] Hammerschmidt EG. Formation of gas hydrates in natural gas transmission lines. Ind Eng Chem 1934;26:851–5.
[21] Motiee M. Estimate possibility of hydrates. Hydrocarb Process 1991;70:98–9.
[22] Jager MD, Sloan ED. The effect of pressure on methane hydration in pure water and sodium chloride solutions. Fluid Phase Equil 2001;185:89–99.
[23] Towler BF, Mokhatab SBT-HP. Quickly estimate hydrate formation conditions in natural gases: using this simple technique to predict unfavorable conditions can save time and cost, vol. 84; 200561+.
[24] Bahadori A, Vuthaluru HB. A novel correlation for estimation of hydrate forming condition of natural gases. J Nat Gas Chem 2009;18:453–7.
[25] Ghiasi MM. Initial estimation of hydrate formation temperature of sweet natural gases based on new empirical correlation. J Nat Gas Chem 2012;21:508–12.
[26] Mohamadi-Baghmolaei M, Hajizadeh A, Azin R, Izadpanah AA. Assessing thermodynamic models and introducing novel method for prediction of methane hydrate formation. J Pet Explor Prod Technol 2018;8:1401–12.
[27] Kummamuru NB, Perreault P, Lenaerts S. A new generalized empirical correlation for predicting methane hydrate equilibrium conditions in pure water. Ind Eng Chem Res 2021;60:3474–83.
[28] Soroush E, Mesbah M, Shokrollahi A, Rozyn J, Lee M, Kashiwao T, et al. Evolving a robust modeling tool for prediction of natural gas hydrate formation conditions. J Unconv Oil Gas Resour 2015;12:45–55.
[29] Zhong H-Q, Yao Q-L, Wang Y, He Y-F, Li Z-H. A graphical alternating conditional expectation to predict hydrate phase equilibrium conditions for sweet and sour natural gases. Math Probl Eng 2019;2019.
[30] Medeiros F de A, Segtovich ISV, Tavares FW, Sum AK. Sixty years of the van der Waals and Platteeuw model for clathrate hydrates—a critical review from its statistical thermodynamic basis to its extensions and applications. Chem Rev 2020;120:13349–81.
[31] Khan MN, Warrier P, Peters CJ, Koh CA. Advancements in hydrate phase equilibria and modeling of gas hydrates systems. Fluid Phase Equil 2018;463:48–61.
[32] Parrish WR, Prausnitz JM. Dissociation pressures of gas hydrates formed by gas mixtures. Ind Eng Chem Process Des Dev 1972;11:26–35. https://doi.org/10.1021/i260041a006.
[33] Chen G-J, Guo T-M. A new approach to gas hydrate modelling. Chem Eng J 1998;71:145–51.
[34] Chen X, Li H. New pragmatic strategies for optimizing Kihara potential parameters used in van der Waals-Platteeuw hydrate model. Chem Eng Sci 2022;248:117213.
[35] Mahabadian MA, Chapoy A, Burgass R, Tohidi B. Development of a multiphase flash in presence of hydrates: experimental measurements and validation with the CPA equation of state. Fluid Phase Equil 2016;414:117–32.
[36] Sirino TH, Neto MAM, Bertoldi D, Morales REM, Sum AK. Multiphase flash calculations for gas hydrates systems. Fluid Phase Equil 2018;475:45–63.
[37] MohamadiBaghmolaei M, Mahmoudy M, Jafari D, MohamadiBaghmolaei R, Tabkhi F. Assessing and optimization of pipeline system performance using intelligent systems. J Nat Gas Sci Eng 2014;18:64–76.
[38] Garapati N, Anderson BJ. Statistical thermodynamics model and empirical correlations for predicting mixed hydrate phase equilibria. Fluid Phase Equil 2014;373:20–8.
[39] Hosseini M, Rahimi R, Ghaedi M. Hydrogen sulfide solubility in different ionic liquids: an updated database and intelligent modeling. J Mol Liq 2020;317:113984.
[40] Hemmati-Sarapardeh A, Varamesh A, Husein MM, Karan K. On the evaluation of the viscosity of nanofluid systems: modeling and data assessment. Renew Sustain Energy Rev 2018;81:313–29.
[41] Esmaeili S, Sarma H, Harding T, Maini B. A data-driven model for predicting the effect of temperature on oil-water relative permeability. Fuel 2019;236:264–77.
[42] Zarei MM, Hosseini M, Mohammadi AH, Moosavi A. Model development for estimating calcium sulfate dihydrate, hemihydrate, and anhydrite solubilities in multicomponent acid and salt containing aqueous solutions over wide temperature ranges. J Mol Liq 2021;328:115473.
[43] Mesbah M, Soroush E, Rezakazemi M. Development of a least squares support vector machine model for prediction of natural gas hydrate formation temperature. Chin J Chem Eng 2017;25:1238–48.
[44] Amar MN. Prediction of hydrate formation temperature using gene expression programming. J Nat Gas Sci Eng 2021;89:103879.
[45] Xu H, Jiao Z, Zhang Z, Huffman M, Wang Q. Prediction of methane hydrate formation conditions in salt water using machine learning algorithms. Comput Chem Eng 2021;151:107358.
[46] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–30.
[47] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn 2006;63:3–42.
[48] De Roo JL, Peters CJ, Lichtenthaler RN, Diepen GAM. Occurrence of methane hydrate in saturated and unsaturated solutions of sodium chloride and water in dependence of temperature and pressure. AIChE J 1983;29:651–7.
[49] Dholabhai PD, Englezos P, Kalogerakis N, Bishnoi PR. Equilibrium conditions for methane hydrate formation in aqueous mixed electrolyte solutions. Can J Chem Eng 1991;69:800–5.
[50] Du J, Wang X, Liu H, Guo P, Wang Z, Fan S. Experiments and prediction of phase equilibrium conditions for methane hydrate formation in the NaCl, CaCl2, MgCl2 electrolyte solutions. Fluid Phase Equil 2019;479:1–8.
[51] Haghighi H, Chapoy A, Tohidi B. Methane and water phase equilibria in the presence of single and mixed electrolyte solutions using the cubic-plus-association equation of state. Oil Gas Sci Technol l'IFP 2009;64:141–54.
[52] Hu Y, Lee BR, Sum AK. Phase equilibrium data of methane hydrates in mixed brine solutions. J Nat Gas Sci Eng 2017;46:750–5.
[53] Hu Y, Makogon TY, Karanjkar P, Lee K-H, Lee BR, Sum AK. Gas hydrates phase equilibrium with CaBr2 and CaBr2+ MEG at ultra-high pressures. J Nat Gas Eng 2017;2:42–9.
[54] Hu Y, Makogon TY, Karanjkar P, Lee K-H, Lee BR, Sum AK. Gas hydrates phase equilibria for structure I and II hydrates with chloride salts at high salt concentrations and up to 200 MPa. J Chem Thermodyn 2018;117:27–32.
[55] Kamari A, Hashemi H, Babaee S, Mohammadi AH, Ramjugernath D. Phase stability conditions of carbon dioxide and methane clathrate hydrates in the presence of KBr, CaBr2, MgCl2, HCOONa, and HCOOK aqueous solutions: experimental measurements and thermodynamic modelling. J Chem Thermodyn 2017;115:307–17.
[56] Kang S-P, Chun M-K, Lee H. Phase equilibria of methane and carbon dioxide hydrates in the aqueous MgCl2 solutions. Fluid Phase Equil 1998;147:229–38.
[57] Kharrat M, Dalmazzone D. Experimental determination of stability conditions of methane hydrate in aqueous calcium chloride solutions using high pressure differential scanning calorimetry. J Chem Thermodyn 2003;35:1489–505.
[58] Hu Y, Lee K-H, Lee BR, Sum AK. Gas hydrate formation from high concentration KCl brines at ultra-high pressures. J Ind Eng Chem 2017;50:142–6.
[59] Li S, Wang J, Lv X, Ge K, Jiang Z, Li Y. Experimental measurement and thermodynamic modeling of methane hydrate phase equilibria in the presence of chloride salts. Chem Eng J 2020;395:125126.
[60] Maekawa T, Imai N. Equilibrium conditions of methane and ethane hydrates in aqueous electrolyte solutions. Ann N Y Acad Sci 2000;912:932–9.
[61] Maekawa T, Itoh S, Sakata S, Igari S, Imai N. Pressure and temperature conditions for methane hydrate dissociation in sodium chloride solutions. Geochem J 1995;29:325–9.
[62] Mohammadi AH, Afzal W, Richon D. Gas hydrates of methane, ethane, propane, and carbon dioxide in the presence of single NaCl, KCl, and CaCl2 aqueous solutions: experimental measurements and predictions of dissociation conditions. J Chem Thermodyn 2008;40:1693–7.
[63] Mohammadi AH, Kraouti I, Richon D. Methane hydrate phase equilibrium in the presence of NaBr, KBr, CaBr2, K2CO3, and MgCl2 aqueous solutions: experimental measurements and predictions of dissociation conditions. J Chem Thermodyn 2009;41:779–82.
[64] Hu Y, Makogon TY, Karanjkar P, Lee K-H, Lee BR, Sum AK. Gas hydrates phase equilibria and formation from high concentration NaCl brines up to 200 MPa. J Chem Eng Data 2017;62:1910–8.
[65] Aregbe AG, Sun B, Chen L. Methane hydrate dissociation conditions in high-concentration NaCl/KCl/CaCl2 aqueous solution: experiment and correlation. J Chem Eng Data 2019;64:2929–39.
[66] Adisasmito S, Frank III RJ, Sloan Jr ED. Hydrates of carbon dioxide and methane mixtures. J Chem Eng Data 1991;36:68–71.
[67] Atik Z, Windmeier C, Oellrich LR. Experimental gas hydrate dissociation pressures for pure methane in aqueous solutions of MgCl2 and CaCl2 and for a (methane+ethane) gas mixture in an aqueous solution of (NaCl+ MgCl2). J Chem Eng Data 2006;51:1862–7.
[68] Bhawangirkar DR, Sangwai JS. Phase equilibrium of methane hydrates in the presence of MgBr2, CaBr2, and ZnBr2 aqueous solutions. J Chem Eng Data 2021;66:2519–30.
[69] Cha M, Hu Y, Sum AK. Methane hydrate phase equilibria for systems containing NaCl, KCl, and NH4Cl. Fluid Phase Equil 2016;413:2–9.
[70] Chong ZR, Koh JW, Linga P. Effect of KCl and MgCl2 on the kinetics of methane hydrate formation and dissociation in sandy sediments. Energy 2017;137:518–29.