Development of a hydrate risk assessment tool based on machine learning for CO_2 storage in depleted gas reservoirs

Yamada, Kenta; Fernandes, Bruno Ramon Batista; Kalamkar, Atharva; Jeon, Jonghyeon; Delshad, Mojdeh; Farajzadeh, Rouhi; Sepehrnoori, Kamy

DOI
10.1016/j.fuel.2023.129670

Publication date
2023

Document Version
Final published version

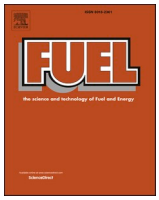
Published in
Fuel

Fuel 357 (2024) 129670

Contents lists available at ScienceDirect

Fuel

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



Development of a hydrate risk assessment tool based on machine learning for CO_2 storage in depleted gas reservoirs

Kenta Yamada [a], Bruno Ramon Batista Fernandes [b,*], Atharva Kalamkar [c], Jonghyeon Jeon [a], Mojdeh Delshad [a], Rouhi Farajzadeh [d,e], Kamy Sepehrnoori [a]

[a] Hildebrand Department of Petroleum and Geosystems Engineering, The University of Texas at Austin, Austin, TX, United States
[b] Center for Subsurface Energy and the Environment, The University of Texas at Austin, Austin, TX, United States
[c] Computational Engineering, The University of Texas at Austin, Austin, TX, United States
[d] Faculty of Civil Engineering and Geosciences, Delft University of Technology University, Delft, the Netherlands
[e] Shell Global Solutions International, The Hague, the Netherlands

ARTICLE INFO

Keywords:
Carbon capture and storage
Machine learning
Depleted gas reservoir
CO_2 hydrate
Reservoir simulation
Cold CO_2 injection

ABSTRACT

Depleted gas reservoirs are attractive sites for Carbon Capture and Storage (CCS) due to their huge storage capacities, proven seal integrity, existing infrastructure and subsurface data availability. However, CO_2 injection into depleted formations can potentially lead to hydrate formation near the wellbore due to Joule-Thomson cooling, which might cause injectivity issues. Some challenges encountered when modeling and simulating this process are the computational time caused by Newton's convergence issues and instability. The objective of this work is to propose a novel approach for hydrate risk assessment during CO_2 injection into depleted gas reservoirs using physics-based Machine Learning (ML) approach. First, the selection of input parameters for the ML models is performed based on sensitivity study results using an analytical solution for different operational and petrophysical values. Then the ML models are tuned and tested using datasets from numerical reservoir simulation results based on a wide range of input parameter values. To the best of our knowledge, this is the first time that an ML approach is used for risk assessment of CO_2 hydrate in its storage in depleted gas reservoirs. The ML models developed in this study presented an efficient performance to predict hydrate-forming events. The deep neural network model performed best with a 95% recall value and 84% precision value. These results show that the ML model can be further utilized for risk assessment in the screening stage, and the combination of screening by ML, followed by detailed analysis with numerical simulation in high-risk cases can be an efficient probing workflow for future CCS projects.

1. Introduction

To achieve the goal of the Paris Agreement [1], reducing greenhouse gas (GHG) emissions is inevitable. Carbon dioxide (CO_2) is the most common GHG, and reducing its emissions is critical [2]. Carbon capture and storage (CCS) is one of the methods that can reduce the excessive amount of CO_2 in the atmosphere.

Depleted gas reservoirs are suitable candidates for CO_2 storage because of storage capacities, proven seal integrity, existing infrastructure and subsurface data availability, making the capital-intensive project more feasible. However, because of the pressure difference between the injection well and the reservoir, injected CO_2 expands, resulting in significant reduction in temperature. This phenomenon is called Joule-Thomson (J-T) cooling effect. This cooling of CO_2 can be so high that solid hydrates can form in the presence of brine [3], impairing the injectivity of CO_2 into the well and reservoir. In addition, injectivity issues caused by hydrate formation could also lead to loss of well integrity [4]. Therefore, it is critical to know in advance if hydrate formation would happen under the operational conditions. Furthermore, the injection temperature will also impact the kinematic viscosity of the injected CO_2 which is inversely proportional to the injectivity [5]. Hence, a reduction in temperature will increase injectivity for the gaseous CO_2 whereas the injectivity will decrease if liquid CO_2 is injected into the reservoir. The injection of cold CO_2 can also have geomechanical impacts with changes in both thermal and tensile stresses which can lead to change in the formation fracturing pressure as well [6]. In the work from Khurshid and Fujii [6], the injection of cold

* Corresponding author.
E-mail address: brbfernandes@utexas.edu (B.R.B. Fernandes).

https://doi.org/10.1016/j.fuel.2023.129670
Received 18 April 2023; Received in revised form 15 July 2023; Accepted 28 August 2023
Available online 4 September 2023

Nomenclature

c	Specific heat capacity, J/(kg K)
H	Reservoir thickness, m
k	Absolute permeability, m^2
k_r	Relative permeability, dimensionless
M	Injection mass rate, kg/s
n	Relative permeability exponent, dimensionless
P	Pressure, Pa
r_e	Drainage radius, m
r_w	Well radius, m
S	Saturation, dimensionless
S_{wr}	Water residual saturation, dimensionless
S_{lr}	Liquid residual saturation, dimensionless
S_{gr}	Gas residual saturation, dimensionless
t	Time, sec
T	Temperature, K

Greek

α	Gas Joule-Thomson coefficient, K/Pa
ρ	Density, kg/m^3
ϕ	Porosity, dimensionless
μ	Gas viscosity, Pa.s

Subscript

g	Gas property
inj	Injection fluid property
l	Liquid CO_2 property
min	Minimum
r	Rock property
res	Reservoir property
w	Water property

Superscript

0	End-point relative permeability

CO_2 lead to a significant reduction in the formation fracturing pressure.

To investigate the risk of hydrate formation within a reservoir formation, it is necessary to have an approach capable of computing both the pressure and temperature fields within the reservoir formation over time. The injection of CO_2 in aquifers and oil and gas reservoirs has been extensively investigated in the literature [7], with approaches specially developed for modeling and simulating the complex isothermal compositional fluid flow of brine, oil, and gas in subsurface porous media promoted by CO_2 flooding [8], and further extended for conditions in which a fourth CO_2 rich liquid phase is formed [9,10]. However, the number of analytical and numerical approaches that can handle the heat and fluid flow in porous media is more limited [3,11–23]. Analytical and numerical solutions can both be used for the hydrate risk assessment.

One of the approaches for predicting the formation of hydrates is with phase diagrams. Estimating the hydrate phase diagrams of mixture of gases (i.e., CO_2 and methane) including the effect of salinity is quite challenging. Van der Waals and Platteeuw [24] used an adsorption model based on statistical thermodynamics to calculate the chemical potential of water in the hydrate phase which was further improved by many authors [25–28]. Coelho et al. [29] presented a compositional wellbore simulator capable of assessing the potential hydrate formation risk using the model from van der Waals and Platteeuw [24], but the formation of hydrates within the porous formation was not considered. Another way for modeling the formation of hydrates is with kinetic models for the formation and dissociation of hydrates. Since Kim et al. [30], many other studies were proposed for modeling hydrate growth/dissociation in water/gas systems through kinetics with different assumptions [31–33]. Kinetic models have been used for quantitatively model the formation of hydrates in wellbore [34–36] and reservoir formations [22,23]. For example, Janicki et al. [22] presented a kinetic based modeling approach for the methane hydrate dissociation coupled with heat and mass transfer for simulating the production of methane from hydrate deposits with an in-house simulator (UMSICHT HyReS) and CMG STARS [20]. It is not clear from their modeling and results whether Janicki et al. [22] took the J-T effect into consideration. While the formation of hydrates poses a risk for the injectivity of CO_2 during CCS, Ahmad et al. [23] suggested to store CO_2 in the hydrate form at depleted methane hydrate bearing formations. The authors presented a coupled mass, momentum, and heat transfer model along with a hydrate growth kinetic model. A term to account for the J-T effect was not considered in the energy balance presented and their results suggested that hydrate plugging of the formation was not an issue for the cases presented by the authors. While the phase diagram approach can only evaluate the formation of hydrates qualitatively, it is an adequate approach for hydrate risk assessment and has been used before [37]. Machine Learning (ML) algorithms are powerful tools capable of identifying and reproducing patterns in large datasets. This technology has already been applied in different scientific and engineering applications. Several studies show successful implementation of machine learning for CO_2 injection projects and prediction of CO_2 storage performance [38], prediction of fluid properties [39], and operational optimization [40]. Khanal et al. [41] developed a proxy saline aquifer model based on machine learning using a large set of simulated data. Their model could correctly forecast the CO_2 sequestered under various trapping mechanisms. Furthermore, Acharya and Bahadur [42] used machine-learning algorithms to predict hydrate-stable regions which would offer another option besides the approach by van der Waals and Platteeuw [24].

In this study, a machine learning model was developed for performing risk assessment of hydrate formation in the reservoir rock formation during the CO_2 storage processes in depleted gas reservoirs as a pre-screening tool. To train and test the machine learning tool, a dataset is required. Two approaches were used to create the hydrate risk assessment dataset based on the hydrate phase equilibrium diagram. The first approach considered the computation of the reservoir condition during the injection of CO_2 with an analytical solution, while the second approach considered a simplified numerical simulation model. The dataset is used for parameter importance analysis and calibration of the different machine learning models: random forest, gradient boosting, neural network, and deep neural network. To the best of our knowledge, this has been the first-time machine learning is used to develop a tool capable of assessing the risk of hydrate formation during CO_2 sequestration in depleted gas reservoirs from operation conditions and petrophysical and fluid properties. Furthermore, the tool developed provides the ranges of operational conditions with a risk of hydrate formation without the need for a complex numerical model set up. In this work, we assume the injected CO_2 is pure. It is important to mention that the presence of impurities in the injected CO_2 will affect the hydrate phase equilibrium [37]. Special attention should be given to the water content present in the injected CO_2 which can lead to plugging due to hydrate formation and corrosion [29,43].

2. Methods

2.1. Potential hydrate formation analysis

This study analyzes hydrate formation using the hydrate phase diagram which is a common approach because of its simplicity [44]. Fig. 1 shows the phase equilibrium diagrams for CO_2 and CH_4 hydrates (P-T

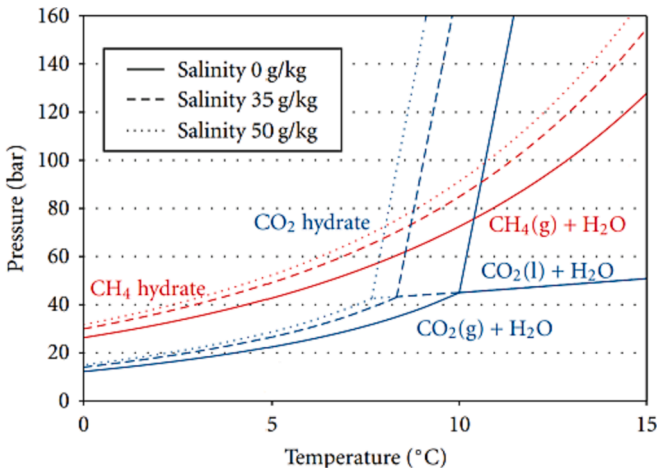


Fig. 1. Equilibrium curves of methane hydrate and carbon dioxide hydrates for different brine salinities [22].

diagrams) and the impact of the salinity [37]. These curves define the onset pressure and temperature conditions when hydrate forms, which allows a simple comparison of the hydrate curve and calculated pressure–temperature profile to analyze hydrate formation, as shown in Fig. 2. This method cannot evaluate the severity of hydrate formation, such as its accumulation and impacts on injectivity since it lacks hydrate formation/dissociation (kinetic) reactions [45] or other mechanisms to quantify the amount of hydrate. However, the main purpose of this study is to assess the hydrate formation risk through initial screening to be followed by a more comprehensive quantification analysis. Hence, the proposed method is fit-for-purpose.

In this study, we assume fresh water (zero salinity), corresponds to the worst-case scenario regarding the risk of hydrate formation. The CO_2 hydrate curve at zero salinity from Fig. 1 was digitized and fitted to

$$T_{hyd} = \begin{cases} 0.1696P^3 - 2.1877P^2 + 10.952P + 262.69 & (P < 4.5MPa) \\ 0.1714P + 282.39 & (P \geq 4.5MPa) \end{cases}, \quad (1)$$

where T_{hyd} is hydrate curve temperature in degree K and P is reservoir pressure in MPa. In addition, it is observed that the temperature front is considerably slower than the CO_2 front, to the point that the temperature changes are only observed within the CO_2-saturated area [3]. The temperature front retardation is caused by the specific heat capacity of the rock and residual water, as presented in [11]. These assumptions reduce the hydrate formation problem to computing pressure and temperature profiles during the CO_2 injection at specific operational and reservoir conditions.

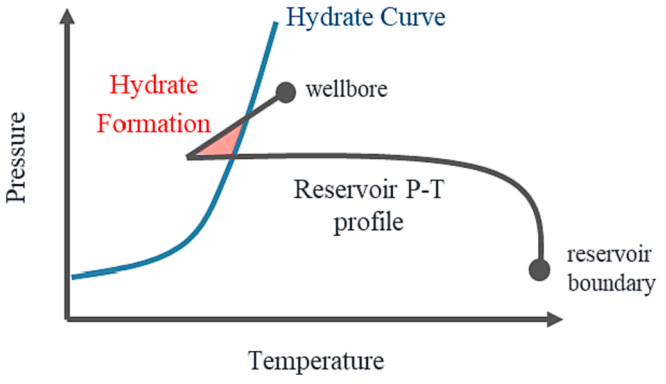


Fig. 2. Hydrate formation analysis comparing hydrate curve and reservoir pressure–temperature profile.

2.2. Temperature and pressure estimation

2.2.1. Analytical solution

The first approach to calculate temperature and pressure is based on an analytical solution [11] where the pressure diffusivity equation due to cold CO_2 injection is solved considering the J-T cooling effect. Assuming a steady-state pressure, single-phase flow, and constant thermodynamic properties, the minimum temperature at the temperature front at any given time can be computed as [11]

$$T_{min} = -\frac{\alpha M \mu_g}{4\pi k k_{rg} H \rho_g} \times ln\left[\left\{\frac{c_g}{\phi(1-S_w)\rho_g c_g + \phi S_w \rho_w c_w + (1-\phi)\rho_r c_r}\right\}\frac{Mt}{\pi H r_w^2} + 1\right] + min\left(T_{res}, T_{inj}\right) \quad (2)$$

where α is the Joule-Thomson coefficient, M is the mass injection rate, μ_g is the gas viscosity, k is the permeability, k_{rg} is the gas relative permeability, H is the formation thickness, ρ_g is the gas density, c_g is the gas specific heat capacity, ϕ is the porosity, S_w is the water saturation, ρ_w is the water density, c_w is the water specific heat capacity, ρ_r is the rock density, c_r is the rock specific heat capacity, t is the time, r_w is the well radius, T_{res} is the reservoir temperature, and T_{inj} is the bottomhole temperature of the injected fluid. Also, the distance from the well at which the minimum temperature is located is given by

$$r_{T_{min}} = r_w\sqrt{\left\{\frac{c_g}{\phi(1-S_w)\rho_g c_g + \phi S_w \rho_w c_w + (1-\phi)\rho_r c_r}\right\}\frac{Mt}{2\pi H r_w^2} + 1}. \quad (3)$$

It is important to emphasize that the water is assumed to be always immobile in the analytical solution, meaning that the water saturation in Eqs. (2) and (3) corresponds to the residual water saturation.

While the analytical solution is adequately verified with the fully coupled non-isothermal numerical simulator TOUGH2/EOS7C [3], Mathias et al. [11] implied the potential overestimate of temperature decrease caused by J-T cooling due to their assumption of constant thermodynamic properties.

As for pressure, while the original analytical solution assumes a steady-state pressure difference (pressure in any location within the reservoir minus wellbore pressure was constant over time), the actual pressure is needed for our approach. For this purpose, a drainage radius r_e is assumed where the pressure does not change from its initial value. Hence, the pressure at the location where the minimum temperature occurs is computed as

$$P_{T_{min}} = P_{res} + \frac{M\mu}{2\pi k k_r H \rho_f} \ln\left[\frac{r_e}{r_{T_{min}}}\right], \quad (4)$$

where P_{res} is the reservoir pressure.

2.2.2. Numerical solution

The second approach to obtain the temperature and pressure fields is using a numerical reservoir simulator. A significant advantage of numerical simulation is that it can consider more realistic reservoir properties and phase behavior than the analytical solution. This study uses the non-isothermal compositional reservoir simulator CMG GEM to simulate CO_2 injection in methane reservoirs. Although some researchers [22,46] conducted hydrate formation analysis using CMG-STARS because it can model hydrate kinetic reactions, STARS has limitations in modeling such a process. In STARS, the liquid CO_2 phase is modeled as a slightly compressible fluid. Also, the phase equilibrium between the vapor and liquid phase uses K-Values, unlike GEM, which combines stability analysis and flash calculation. Furthermore, including the liquid CO_2 phase in STARS can lead to convergence issues and inaccurate results when there is a wide range in pressure and

temperature. Therefore, GEM is selected when the calculated amount of formed or dissociated hydrates is not required. Phenomena such as geomechanics, geochemistry, and CO_2 dry-out can be modeled by GEM but were neglected in this study. Each of these phenomena can add significant complexity to both the simulation models and the machine learning and will be carefully investigated in future work.

2.3. Machine learning

This study aims for the development of ML model to classify CO_2 injection problems as "potential risk of hydrate" or "limited risk of hydrate" based on operational and reservoir conditions. This is a binary classification problem. First, the input parameters were carefully selected depending on their importance to the output to avoid degrading the ML performance due to overfitting. This was followed by the creation of the training/testing dataset. The quality and size of the dataset strongly impact the ML performance. In this study, we applied four different ML methods for binary classification problems: random forest, gradient boosting, neural network, and deep neural network, and compared their performance. Further details regarding the techniques can be found in ref. [47].

The ML performance is evaluated by the number of cases the model correctly or incorrectly predicts against the real values present in a test dataset. A confusion matrix summarizes them by categorizing all cases into four types, which are true-positive, true-negative, false-positive, and false-negative, as shown in Fig. 3.

This study focuses on two evaluation metrics to measure ML performance: the recall value and the precision value calculated by the equations below [48].

$$RECALL = \frac{TruePositive}{TruePositive + FalseNegative} \tag{5}$$

$$PRECISION = \frac{TruePositive}{TruePositive + FalsePositive} \tag{6}$$

In the recall value, the numerator is the number of correctly predicted potential hydrate cases, and the denominator is the total true potential hydrate cases, so it represents the fraction of actual potential hydrate cases that the ML model correctly screened. On the other hand, in the precision value, the numerator is the same as in the recall, but the denominator is the number of total predicted potential hydrate cases, so it measures the fraction of correct answers in the predicted potential hydrate cases. In this study, recall is more important than precision because this ML model is supposed to be used as a risk assessment screening tool. In the case of suspected potential hydrate risk the screening process should be followed by detailed analysis. Our main objective in ML training and testing is to obtain models with high recall while retaining a reasonable precision value.

The ML model classifies a case by calculating a probability value and using a threshold value to identify it as positive (1) or negative (0). Therefore, each ML model performance can be optimized by calibrating the threshold value. A lower threshold enables the prediction of a positive case (potential hydrate risk) even when the ML model predicts a low probability of a potential hydrate case, which results in a higher recall value. In contrast, a higher threshold tends to predict positive cases more strictly, resulting in a higher precision value. Considering that our target is obtaining a high recall while retaining reasonable precision, we optimized the threshold value in each ML model. Moreover, the trade-off relationship between precision and recall according to the threshold can be drawn in a Precision-Recall (P-R) curve, as shown in Fig. 4. The larger the area under the P-R curve is, the higher both precision and recall are. The Area Under Curve (AUC) is used to compare the performance of different ML models.

3. Results

3.1. Workflow

This study follows a general machine learning model development procedure described in Section 2.3, consisting of selecting key parameters, dataset development, machine learning model development, and performance check (Fig. 5).

3.2. Parameter selection

Since CO_2 hydrate forms depending on a variety of parameters, this step aims to rank the importance of these parameters and select the input parameters for the machine learning model. We perform sensitivity analysis using the analytical solution to evaluate the parameter importance. Although there are limitations with the analytical method as described in Section 2.2.1, it still provides qualitative values and is a preferred method because of its simplicity.

There are 20 input parameters in the analytical model for obtaining the temperature and pressure profiles. All parameters and their value settings are summarized in Table 1. Parameters 1 to 10 are used as impactful parameters, and their ranges are based on ongoing CCS projects [49]. Thermodynamic properties Nos. 11 to 14 are calculated based

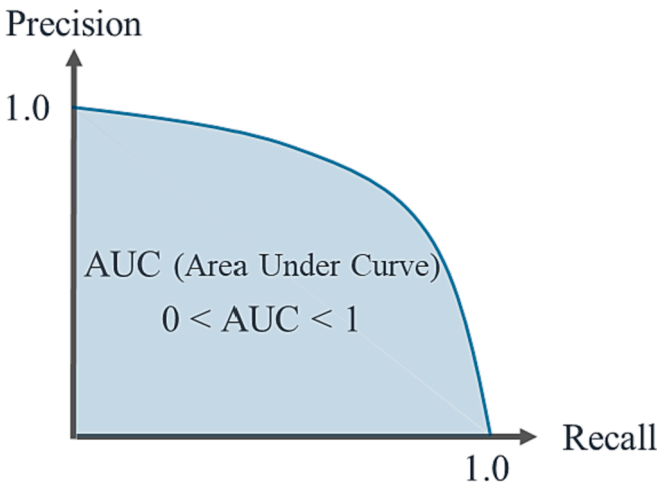


Fig. 4. Precision–Recall curve and Area Under Curve.

		Predicted	
		Negative (limited hydrate risk)	Positive (potential hydrate risk)
Actual	Negative (limited hydrate risk)	True Negative	False Positive
	Positive (potential hydrate risk)	False Negative	True Positive

Fig. 3. Description of the confusion matrix.

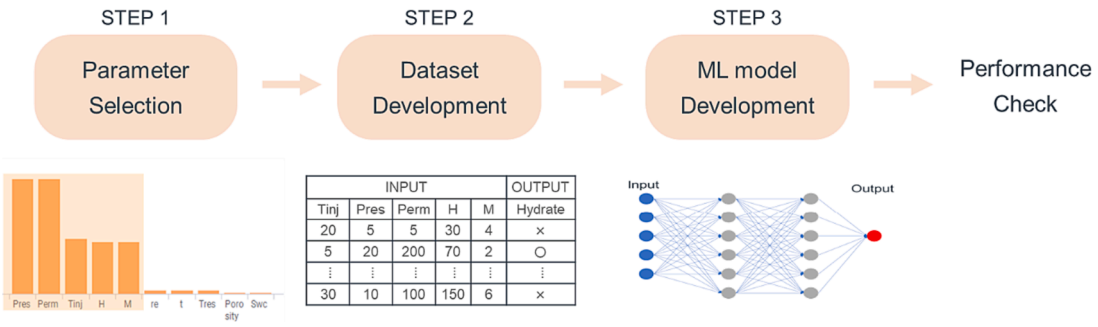


Fig. 5. Workflow to develop machine learning model.

Table 1
Properties used in the sensitivity analysis.

No.	Parameter	Range	No.	Parameter	Value
1	t(days)	1 – 3650	11	ρ_g(kg/m^3)	Calculated from correlations
2	M(kg/s)	1 – 50	12	μ_g(Pa • s)	
3	P_{res}(MPa)	1 – 20	13	c_g(J/kg/K)	
4	T_{inj}(°C)	10 – 40	14	α(K/Pa)	
5	T_{res}(°C)	60 – 120	15	r_w(m)	0.1
6	r_e(m)	10^3 – 10^6	16	k_{rg}(−)	1
7	H(m)	10 – 200	17	ρ_w(kg/m^3)	992
8	ϕ(−)	0.05 – 0.3	18	c_w(J/kg/K)	4037
9	k(mD)	5 – 1000	19	ρ_r(kg/m^3)	2600
10	S_w(−)	0.2 – 0.5	20	c_r(J/kg/K)	1000

on the bottomhole injection temperature and initial reservoir pressure. Dranchuk and Abou-Kassem's equation [50] is used to calculate the compressibility factor, the Jossi et al. model [51] is used for viscosity, and the Joule-Thomson coefficient is calculated by its definition. In addition, parameters 15 to 20 are fixed because they usually have less variations in temperature and pressure ranges considered here.

The dataset with 10 sensitivity parameters were randomly sampled from a uniform distribution within their value ranges. Additionally, a uniform distribution for the log of permeability was used. Then cases in which the combination of T_{inj} and P_{res} are within the hydrate region in the phase diagram were excluded since these conditions always lead to a potential risk of hydrate formation. The hydrate risk for these cases is not affected by the pressure drawdown and the J-T effect and including such cases in the ML training degrades the model performance by reducing its ability to capture the complex fluid flow and cooling behavior in favor of simply tagging if specific parameters are within the hydrate region. This also means that there is no need to use the ML model for performing a risk assessment of cases in which T_{inj} and P_{res} are within the hydrate region since these can be instantly assigned as potentially high-risk cases. Finally, twenty thousand cases were prepared and a sensitivity analysis was performed using an analytical solution.

The Kendall's correlation coefficient was used to rank the importance of the parameters as shown in Fig. 6. Reservoir pressure, rock permeability, bottomhole injection temperature, reservoir thickness, and CO_2 mass injection rate were observed to have high importance, unlike the other five parameters. Hydrate formation events were observed to often occur near the wellbore due to slow temperature front speed and the reservoir radius was only used to guarantee the imposed boundary condition of infinite reservoir. Time is also less important because hydrate forms during the early injection time since CO_2 injection into the initial (low) reservoir pressure has high J-T cooling effects [37]. As for the initial water saturation, despite of Creusen's [37] observation of its importance for the near wellbore conditions such as the multiphase flow, salt precipitation, and drying-out effect, our study assumes a simpler reservoir model and these effects are not investigated in this study. However, the increase in salinity and the drying-out effect both mitigate the hydrate risk leading our approach to consider the worst scenario for hydrate risk (zero salinity and no water dry-out). One may notice that porosity and reservoir temperature also have low importance. However, the initial reservoir temperature will have an impact on the properties of the gas in-place in the numerical model which is not considered by the analytical solution. We also kept the porosity since rock compressibility will be considered for the numerical model. Therefore, the reservoir radius, time, and initial water saturation (equivalent here to the residual water saturation) were removed from the input list for the ML models based on the results of numerical simulations.

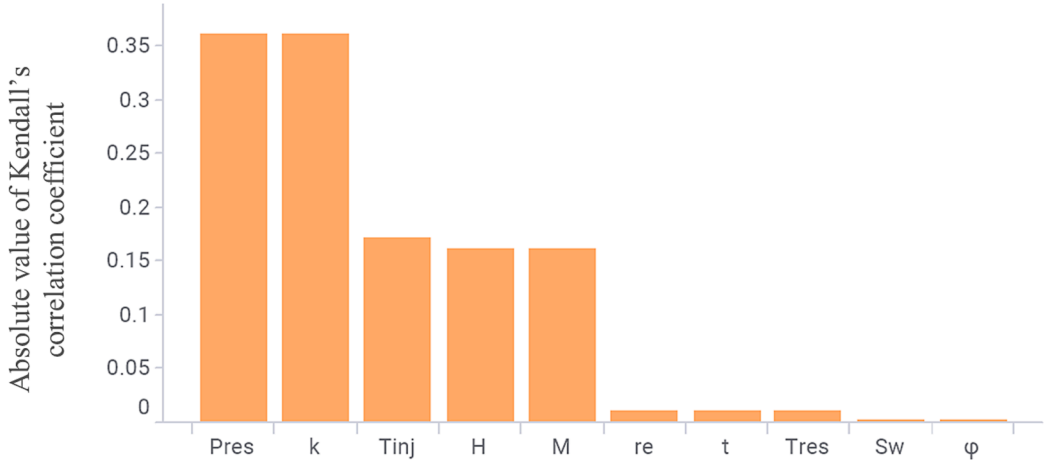


Fig. 6. Parameter importance ranking – absolute value of Kendall's coefficient to potential hydrate formation occurrence.

3.3. Dataset development

Numerical simulations using CMG GEM [21] were performed to develop an extensive dataset for the machine learning model. The hydrate formation equation, Eq. (1), is implemented in CMG-CMOST [52] to identify potential hydrate formation risk cases if any of the reservoir gridblock has the pressure and temperature conditions within the hydrate region. The reservoir model is assumed to be homogeneous depleted gas reservoir with infinite boundary (r_e), modeled through a pore volume multiplier at the outermost gridblock, as shown in Fig. 7. One should keep in mind that assuming a homogeneous reservoir will limit the accuracy of the model. However, the effects of heterogeneity are quite complex to include. For instance, the risk of hydrate may decrease in a layered reservoir because the flow will be diverted to the layer with the highest permeability. On the other hand, the heterogeneity can create flow barriers leading to an increase in the injection pressure and more J-T cooling. The model is initially saturated with methane and water is at the residual saturation (immobile water). Pure CO_2 is injected for a period of 1 year since hydrates are formed at an early time of injection. We performed a total of 17,697 sensitivity simulations varying seven parameters to create the dataset in a similar manner to before. The parameter ranges and basic properties are summarized in Table 2.

The numerical simulations resulted in 844 cases of potential hydrate formation risk and 16,784 cases of limited hydrate formation risk. Also, 69 cases were terminated and discarded due to exceeding the simulation time limit (20 times larger than usual) or abnormal temperature (the simulator will stop the run if temperatures lower than 0 °C are reached). The results are summarized in Fig. 8 as a plot of the initial reservoir pressure versus the injection temperature. The red region in Fig. 8 is limited by the CO_2 hydrate PT diagram at zero salinity presented in Fig. 1 at zero salinity. As mentioned before, a combination of T_{inj} and P_{res} within this region is immediately assumed to be a potential hydrate risk and the ML models are not used to predict cases within it. Such cases are not included in the training/testing dataset in favor of better prediction related to fluid flow and J-T cooling, as discussed before. The key findings from Fig. 8 are:

- Potential risk of hydrate formation increases at low bottomhole injection temperature and low initial reservoir pressure.
- The size of the dots represents the temperature reduction caused by the J-T cooling effect, and it is larger in the cases with gaseous CO_2 due to its higher J-T coefficient. This may cause hydrate formation even when the injection temperature is greater than 25 °C.
- All simulation terminations occur when P_{res} and T_{inj} are in the vapor region. These cases require small time-steps due to large temperature change by J-T cooling effect and CO_2 phase transitions. Also, some cases resulted in temperatures less than 0 °C (outside the modeling capability of GEM.

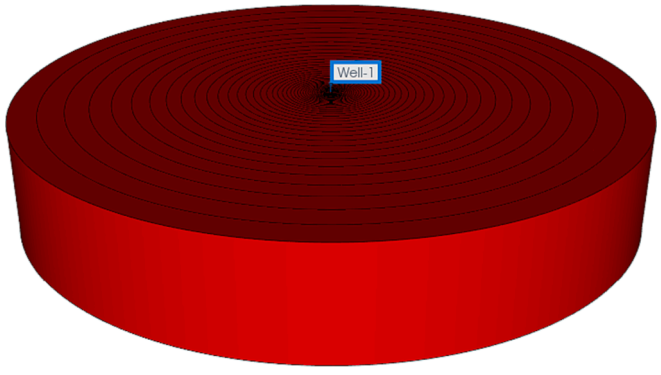


Fig. 7. Homogeneous radial reservoir model used in this study. The injection well is in the center.

Table 2
Properties used in the sensitivity analysis.

No.	Sensitivity parameter	Range	Fixed parameters	Value
1	M(kg/s)	6.5 – 50	Grid number (r,θ,z)	(100, 1, 1)
2	P_{res}(MPa)	1 – 20	Well radius, r_w (m)	0.1
3	T_{inj}(°C)	5 – 40	Residual water saturation,S_{wr}	0.2
4	T_{res}(°C)	60 – 120	Initial gas saturation	CH_4
5	H(m)	10 – 200	Injection fluid	Pure CO_2
6	ϕ(−)	0.05 – 0.3	Relative permeability (water-liquid)	$S_{wr} = 0.2$, $S_{lr} = 0.2$, $k^0_{rw} = 0.2$, $k^0_{rl} = 0.9$, $n_w = 2$, $n_l = 2$
7	k(mD)	5 – 1000	Relative permeability (liquid–gas)	$S_{lr} = 0.4$, $S_{gr} = 0.05$, $k^0_{rl} = 0.9$, $k^0_{rg} = 1.0$, $n_l = 2$, $n_g = 2$

Fig. 9 shows the ranking of parameter importance to hydrate formation from the numerical solution based on Kendall's correlation coefficient (similar to Fig. 6).

The order of parameter importance is (1) injection temperature, (2) reservoir pressure, (3) permeability, (4) reservoir thickness, (5) mass flow rate, and reservoir temperature and porosity are less important. These results are similar to the analytical solution. Furthermore, Fig. 10 shows the frequency of hydrate formation versus the parameter values in the top 5 important parameters. Each parameter has a clear relationship with the hydrate formation, but injection temperature and reservoir pressure have a strong relation containing clear thresholds at 15 °C and 4 MPa, respectively. These results can lead to two recommendations for hydrate formation risk mitigation. The first one is that a bottomhole injection temperature higher than 15 °C significantly reduces the probability of hydrate formation, consistent with the general rule of thumb presented by Le Goff et al. [49]. The second one is that at an initial reservoir pressure higher than 4 MPa, the probability of hydrate formation is reduced since the CO_2 J-T coefficient decreases as the pressure increases [37].

3.4. Machine learning model development

Four different types of ML models are tested and compared in this study. The free ML model library (scikit-learn) is used to develop the ML models. The seven input parameters used in the numerical simulation are normalized, and the output is either 0 (hydrate does not form) or 1 (hydrate does form). Here, 80% of the results from the numerical simulations are used for the training process of the ML model, while the other 20% are used for the ML model performance test. Hyper parameters in each model are optimized by using a grid search algorithm 3-fold cross-validation [53] within the training dataset, and tuned parameters that were changed from the default value in the library are summarized in Table 3. The development was conducted using Python 3.9.16 and scikit-learn 1.2.2 on a Mac device with a Core i7-9750H processor and 16 GB memory.

3.5. Machine learning model performance

Fig. 11 shows the Precision-Recall (P-R) curve for the four ML models. The AUC (Area Under Curve) in each model is summarized in Table 4. Since a larger AUC suggests a better model performance, the models can be ordered according to best to worst performance as deep neural network, neural network, random forest, and gradient boosting.

In addition, we maximize ML performance by optimizing the

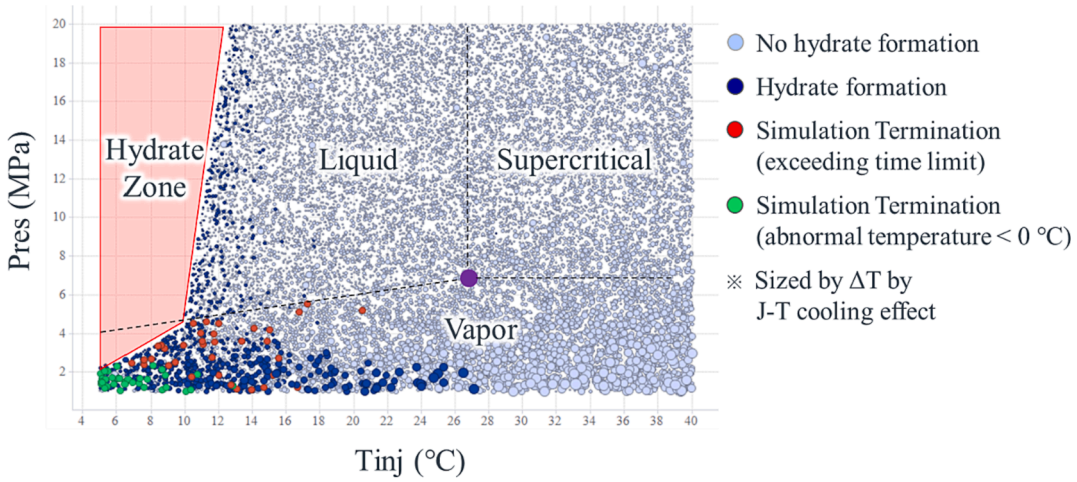


Fig. 8. Potential hydrate formation map for reservoir pressure vs. injection temperature (T_{inj}).

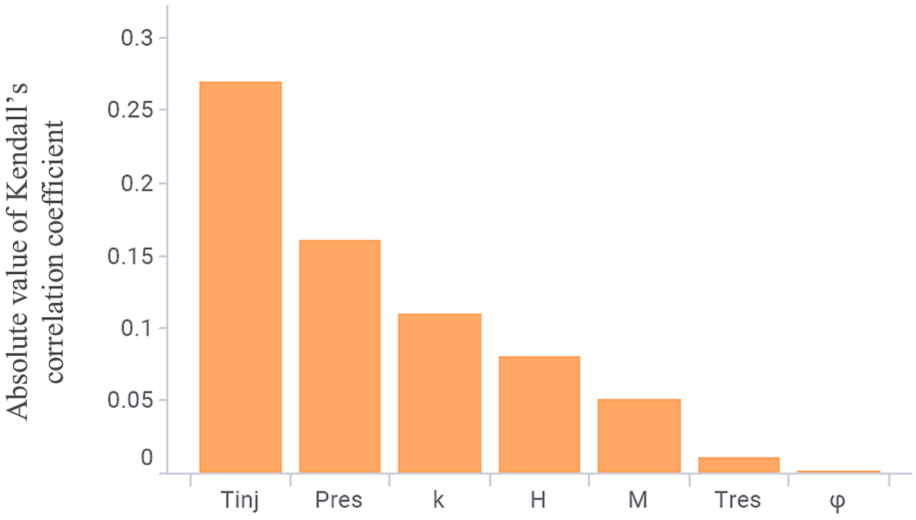


Fig. 9. Parameter importance ranking from simulation results - absolute value of Kendall's coefficient to potential hydrate formation occurrence.

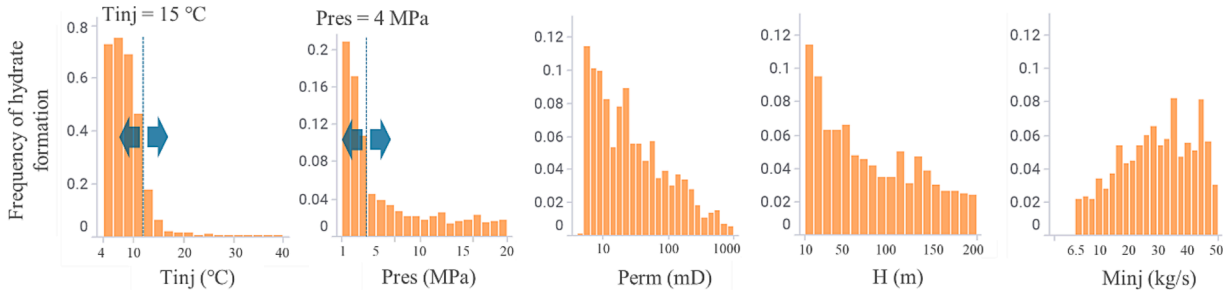


Fig. 10. Frequency of potential hydrate formation vs. input parameter values.

threshold in each model. Since our target in ML performance is a high recall value with retaining a reasonable precision value, the threshold is tuned to obtain a recall value of 95%. Tuned threshold values and resultant precision values are summarized in Table 4. Deep neural network and neural network model show great performance 84% and 76% of precision value, respectively, while retaining more than 95% of recall value. The third highest precision value was obtained by the random forest and gradient boosting with both presenting a value around 60%. These results suggest that neural network has a better prediction capability for complex phenomena such as the potential hydrate formation considered in this study than a simpler machine learning algorithm, and the deep neural network adds further capability to improve their performance.

Fig. 12 shows the confusion matrix from the testing dataset obtained by the deep neural network model with an optimal threshold of 0.299. It shows that 160 out of 168 true hydrate cases are correctly estimated, and the resultant recall value reaches 95.2%. Fig. 13 shows the hydrate prediction by deep neural network model in the initial reservoir pressure and injection temperature map. Another point to emphasize is that unlike the numerical simulations, the ML does not require long calculations

Table 3
Key hyperparameter settings in ML models.

No.	Model	Key hyperparameters
1	random forest	n_estimators: 275, class_weight: balanced
2	gradient boosting	learning_rate: 0.05, n_estimators: 300, subsample: 0.9, min_samples_split: 10, min_samples_leaf: 2, max_depth; 7, random_state: 42, max_features: sqrt,
3	neural network	Hidden_layer_sizes: 100, batch_size: 128, learning_rate: adaptive, learning_rate_init: 0.01, max_iter: 1000, beta_1: 0.8, beta_2: 0.9, n_iter_no_change: 25
4	deep neural network	Hidden_layer_sizes: 200,100, batch_size: 128, learning_rate: adaptive, learning_rate_init: 0.01, max_iter: 1000, beta_1: 0.8, beta_2: 0.9, n_iter_no_change: 25

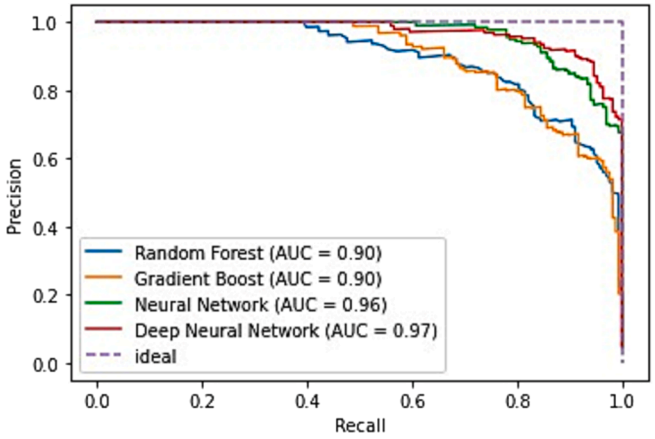


Fig. 11. Precision-Recall curve in the performance test.

Table 4
ML performance comparison in four different methods.

Model	AUC	Optimized Performance (Recall > 95%)	
		Precision (%)	Threshold
Deep Neural Network	0.97	84	0.299
Neural Network	0.96	76	0.054
Random Forest	0.90	60	0.139
Gradient Boost	0.90	60	0.019

under any condition, and can provide answers in less than one second. These results indicate that this ML model can correctly predict 95% of the potential hydrate formation cases with just seven input parameters without numerical issues (i.e. convergence issues, large material balance error, simulation failure), which provides an efficient hydrate risk screening tool.

For the testing data, out of 16% of the hydrate cases predicted by the deep neural network were false positives. However, considering the amount of time and effort that needs to be dedicated to factors such as model setup and CPU time, the ML quick hydrate risk screening model will efficiently support the site selection process for future CCS projects.

4. Conclusions

This study presented the development of an efficient screening tool to assess the risk of hydrate formation during CO_2 injection into depleted gas reservoirs using a machine learning model.

Hydrate formation analysis using a hydrate phase diagram is a practical approach to determine whether the operation and reservoir conditions will result in potential hydrate formation. It simplifies hydrate formation problems to a comparison of the hydrate curve and pressure–temperature profile due to CO_2 injection.

Here, an analytical solution was first used to generate a dataset utilized for the parameter importance analysis. This was followed by conducting non-isothermal numerical simulations to generate a training/testing dataset for the ML models through 17,697 simulations. Four ML models were generated: random forest, gradient boosting, neural network, and deep neural network. The key insights observed in this study are:

- The analytical solution was useful for calculating the pressure and temperature profile during CO_2 injection, but it considers several simplifying assumptions. Numerical simulation might show more realistic results, but it requires more time and effort, and sometimes there are numerical issues near the CO_2 phase transition.
- Numerical simulations suggest that bottomhole injection temperature and reservoir pressure have the greatest impact on potential hydrate formation events.

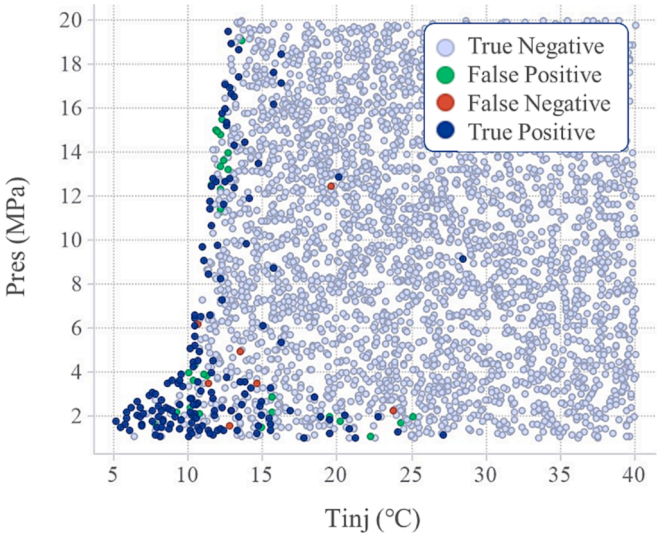


Fig. 13. Potential hydrate formation risk map calculated by deep neural network model for reservoir pressure vs. injection temperature (T_{inj}).

		Predicted	
		Negative (limited hydrate risk)	Positive (potential hydrate risk)
Actual	Negative (limited hydrate risk)	3325	30
	Positive (potential hydrate risk)	8	160

Fig. 12. Confusion matrix of test performance by deep neural network model.

- The ML model developed in this study has seven input parameters selected based on the results of sensitivity studies, and the model is trained and tested using numerical simulation results.
- The deep neural network model with a tuned threshold shows the best performance, with 95% of recall value without 84% of precision value. This means that the ML model can instantly predict 95% of the hydrate case without numerical issues with just seven input parameters.

The ML model should be improved for early risk assessment in the screening stage. The combination of screening by ML followed by detailed analysis with numerical simulation can be an efficient workflow for future CCS projects.

CRediT authorship contribution statement

Kenta Yamada: Investigation, Methodology, Software, Visualization, Writing – original draft. **Bruno Ramon Batista Fernandes:** . **Atharva Kalamkar:** Methodology, Software. **Jonghyeon Jeon:** Writing – original draft. **Mojdeh Delshad:** Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Rouhi Farajzadeh:** Supervision, Writing – review & editing. **Kamy Sepehrnoori:** Funding acquisition, Supervision, Writing – review & 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

Data will be made available on request.

Acknowledgement

The authors acknowledge the sponsors of the RS-JIP from the Center for Subsurface Energy and the Environment at The University of Texas at Austin for supporting this research. The authors also thank the Computer Modeling Group for providing access to the CMG-GEM and CMG-CMOST software. Kenta Yamada would like to acknowledge ITOCHU Oil Exploration Co., Ltd. Rouhi Farajzadeh would like to acknowledge Shell Global Solutions International for their permission to publish this work.

References

[1] UNFCCC. Paris Agreement to the United Nations Framework Convention on Climate Change. 2015.
[2] Ipcc. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Geneva, Switzerland: IPCC; 2014.
[3] Oldenburg CM. Joule-Thomson cooling due to CO_2 injection into natural gas reservoirs. Energ Conver Manage 2007;48:1808–15. https://doi.org/10.1016/j.enconman.2007.01.010.
[4] Porse SL, Wade S, Hovorka SD. Can We Treat CO2 Well Blowouts like Routine Plumbing Problems? A Study of the Incidence, Impact, and Perception of Loss of Well Control. Energy Procedia 2014;63:7149–61. https://doi.org/10.1016/j.egypro.2014.11.751.
[5] Sokama-Neuyam YA, Aggrey WN, Boakye P, Sarkodie K, Oduro-Kwarteng S, Ursin JR. The effect of temperature on CO_2 injectivity in sandstone reservoirs. Scientific African 2022;15:e01066.
[6] Khurshid I, Fujii Y. Geomechanical analysis of formation deformation and permeability enhancement due to low-temperature CO_2 injection in subsurface oil reservoirs. J Petrol Explor Prod Technol 2021;11:1915–23. https://doi.org/10.1007/s13202-021-01133-1.
[7] Ajayi T, Gomes JS, Bera A. A review of CO_2 storage in geological formations emphasizing modeling, monitoring and capacity estimation approaches. Pet Sci 2019;16:1028–63. https://doi.org/10.1007/s12182-019-0340-8.
[8] Young LC. Compositional Reservoir Simulation: A Review. SPE J 2022;27:2746–92. https://doi.org/10.2118/208610-PA.
[9] Fernandes BRB, Marcondes F, Sepehrnoori K. A new four-phase adaptive implicit method for compositional reservoir simulation. J Comput Phys 2021;435:110263. https://doi.org/10.1016/j.jcp.2021.110263.
[10] Fernandes BRB, Marcondes F, Sepehrnoori K. Development of a Fully Implicit Approach with Intensive Variables for Compositional Reservoir Simulation. J Pet Sci Eng 2018;169:317–36. https://doi.org/10.1016/j.petrol.2018.05.039.
[11] Mathias SA, Gluyas JG, Oldenburg CM, Tsang C-F. Analytical solution for Joule-Thomson cooling during CO2 geo-sequestration in depleted oil and gas reservoirs. Int J Greenhouse Gas Control 2010;4:806–10. https://doi.org/10.1016/j.ijggc.2010.05.008.
[12] Jang M, Chun TS, An J. An Analytical Heat Transfer Model in Oil Reservoir during Long-Term Production. Energies 2022;15:2544. https://doi.org/10.3390/en15072544.
[13] Delshad M, Thomas SGG, Wheeler MFF. Parallel Numerical Reservoir Simulations of Nonisothermal Compositional Flow and Chemistry. SPE J 2011;16:239–48. https://doi.org/10.2118/118847-PA.
[14] Varavei A, Sepehrnoori K. An EOS-Based Compositional Thermal Reservoir Simulator. All Days, The Woodlands, Texas: SPE; 2009, p. SPE-119154-MS. https://doi.org/10.2118/119154-MS.
[15] Marcondes F, Varavei A, Sepehrnoori K. An Eos-Based Numerical Simulation of Thermal Recovery Process Using Unstructured Meshes. Braz J Chem Eng 2015;32:247–58. https://doi.org/10.1590/0104-6632.20150321s00003293.
[16] Ishimoto K, Pope GA, Sepehrnoori K. An equation-of-state steam simulator. In Situ 1987;11:1–37.
[17] Zaydullin R, Voskov DV, James SC, Henley H, Lucia A. Fully compositional and thermal reservoir simulation. Comput Chem Eng 2014;63:51–65. https://doi.org/10.1016/j.compchemeng.2013.12.008.
[18] Agarwal A, Tchelepi HA. Adaptive Implicit Method for Thermal Compositional Reservoir Simulation. All Days, Calgary, Alberta, Canada: SPE; 2008, p. SPE-117438-MS. https://doi.org/10.2118/117438-MS.
[19] Liu H, Chen Z, Guo X, Shen L. Development of a Scalable Thermal Reservoir Simulator on Distributed-Memory Parallel Computers. Fluids 2021;6:395. https://doi.org/10.3390/fluids6110395.
[20] CMG. STARS User Guide. Calgary, Alberta, Canada: CMG; 2022.
[21] CMG. GEM User Guide. Calgary, Alberta, Canada: CMG; 2022.
[22] Janicki G, Schlüter S, Hennig T, Lyko H, Deerberg G. Simulation of Methane Recovery from Gas Hydrates Combined with Storing Carbon Dioxide as Hydrates. Journal of Geological Research 2011;2011:1–15. https://doi.org/10.1155/2011/462156.
[23] Ahmad S, Li Y, Li X, Xia W, Chen Z, Ullah N. Numerical analysis of CO_2 hydrate growth in a depleted natural gas hydrate formation with free water. Greenhouse Gas Sci Technol 2019;9:1181–201. https://doi.org/10.1002/ghg.1924.
[24] van der Waals JH, Platteeuw JC. Clathrate Solutions. In: Prigogine I, editor. Advances in Chemical Physics. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2007. p. 1–57. https://doi.org/10.1002/9780470143483.ch1.
[25] Parrish WR, Prausnitz JM. Dissociation Pressures of Gas Hydrates Formed by Gas Mixtures. Ind Eng Chem Proc Des Dev 1972;11:26–35. https://doi.org/10.1021/i260041a006.
[26] Kihara T. Virial Coefficients and Models of Molecules in Gases. Rev Mod Phys 1953;25:831–43. https://doi.org/10.1103/RevModPhys.25.831.
[27] Kihara T. Virial Coefficients and Models of Molecules in Gases. B Rev Mod Phys 1955;27:412–23. https://doi.org/10.1103/RevModPhys.27.412.
[28] Munck J, Skjold-Jørgensen S, Rasmussen P. Computations of the formation of gas hydrates. Chem Eng Sci 1988;43:2661–72. https://doi.org/10.1016/0009-2509(88)80010-1.
[29] Coelho FMC, Sepehrnoori K, Ezekoye OA. A Coupled Hydrate and Compositional Wellbore Simulator: Understanding Hydrate Inhibition from Associated Brines in Oil and Gas Production. SPE Prod Oper 2021;36:858–72. https://doi.org/10.2118/206716-PA.
[30] Kim HC, Bishnoi PR, Heidemann RA, Rizvi SSH. Kinetics of methane hydrate decomposition. Chem Eng Sci 1987;42:1645–53. https://doi.org/10.1016/0009-2509(87)80169-0.
[31] Clarke M, Bishnoi PR. Determination of the activation energy and intrinsic rate constant of methane gas hydrate decomposition. Can J Chem Eng 2001;79:143–7. https://doi.org/10.1002/cjce.5450790122.
[32] Clarke MA, Bishnoi PR. Determination of the intrinsic rate constant and activation energy of CO_2 gas hydrate decomposition using in-situ particle size analysis. Chem Eng Sci 2004;59:2983–93. https://doi.org/10.1016/j.ces.2004.04.030.
[33] Englezos P, Kalogerakis N, Dholabhai PD, Bishnoi PR. Kinetics of formation of methane and ethane gas hydrates. Chem Eng Sci 1987;42:2647–58. https://doi.org/10.1016/0009-2509(87)87015-X.
[34] Zerpa LE, Sloan ED, Sum AK, Koh CA. Overview of CSMHyK: A transient hydrate formation model. J Pet Sci Eng 2012;98–99:122–9. https://doi.org/10.1016/j.petrol.2012.08.017.
[35] Zerpa LE, Rao I, Aman ZM, Danielson TJ, Koh CA, Sloan ED, et al. Multiphase flow modeling of gas hydrates with a simple hydrodynamic slug flow model. Chem Eng Sci 2013;99:298–304. https://doi.org/10.1016/j.ces.2013.06.016.
[36] Wang Y, Koh CA, Dapena JA, Zerpa LE. A transient simulation model to predict hydrate formation rate in both oil- and water-dominated systems in pipelines. J Nat Gas Sci Eng 2018;58:126–34. https://doi.org/10.1016/j.jngse.2018.08.010.
[37] Creusen MCM. Near wellbore effects induced by CO_2 injection and the influence on injectivity in depleted gas reservoirs. Delft University of Technology; 2018. Master Thesis.
[38] Kim Y, Jang H, Kim J, Lee J. Prediction of storage efficiency on CO_2 sequestration in deep saline aquifers using artificial neural network. Appl Energy 2017;185:916–28. https://doi.org/10.1016/j.apenergy.2016.10.012.

[39] Ahmadi MA, Kashiwao T, Rozyn J, Bahadori A. Accurate prediction of properties of carbon dioxide for carbon capture and sequestration operations. Pet Sci Technol 2016;34:97–103. https://doi.org/10.1080/10916466.2015.1107847.
[40] You J, Ampomah W, Sun Q. Co-optimizing water-alternating-carbon dioxide injection projects using a machine learning assisted computational framework. Appl Energy 2020;279:115695. https://doi.org/10.1016/j.apenergy.2020.115695.
[41] Khanal A, Shahriar MF. Physics-Based Proxy Modeling of CO_2 Sequestration in Deep Saline Aquifers. Energies 2022;15:4350. https://doi.org/10.3390/en15124350.
[42] Acharya PV, Bahadur V. Thermodynamic features-driven machine learning-based predictions of clathrate hydrate equilibria in the presence of electrolytes. Fluid Phase Equilib 2021;530:112894. https://doi.org/10.1016/j.fluid.2020.112894.
[43] Cole IS, Corrigan P, Sim S, Birbilis N. Corrosion of pipelines used for CO_2 transport in CCS: Is it a real problem? Int J Greenhouse Gas Control 2011;5:749–56. https://doi.org/10.1016/j.ijggc.2011.05.010.
[44] Subsea Engineering Handbook. Elsevier; 2019. https://doi.org/10.1016/C2016-0-03767-1.
[45] Sun Q, Kang YT. Review on CO_2 hydrate formation/dissociation and its cold energy application. Renew Sustain Energy Rev 2016;62:478–94. https://doi.org/10.1016/j.rser.2016.04.062.
[46] Lin T-K, Dahyar M, Lee M-J, Hsieh B-Z. Study of the formation mechanisms of CO_2 hydrates from matching the experimental data with a porous media setting by multiphase flow-geochemical-thermal reservoir simulator. J Taiwan Inst Chem Eng 2020;114:115–24. https://doi.org/10.1016/j.jtice.2020.09.015.
[47] Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN COMPUT SCI 2021;2:160. https://doi.org/10.1007/s42979-021-00592-x.
[48] Zhou Z-H. Machine Learning. Singapore: Springer Singapore; 2021. https://doi.org/10.1007/978-981-15-1967-3.
[49] Le Goff T-H, Lagarde F, Thibeau S, Brisset A. CO_2 Injection in Depleted Reservoirs: Analysis. Modelling SSRN Journal 2021. https://doi.org/10.2139/ssrn.3815663.
[50] Dranchuk PM, Abou-Kassem H. Calculation of Z Factors For Natural Gases Using Equations of State. J Can Pet Technol 1975;14. https://doi.org/10.2118/75-03-03.
[51] Jossi JA, Stiel LI, Thodos G. The viscosity of pure substances in the dense gaseous and liquid phases. AIChE J 1962;8:59–63. https://doi.org/10.1002/aic.690080116.
[52] CMG. CMOST-AI User Guide. Calgary, Alberta, Canada: CMG; 2023.
[53] Yang L, Shami A. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 2020;415:295–316. https://doi.org/10.1016/j.neucom.2020.07.061.