

A generalized machine learning-assisted phase-equilibrium calculation model for shale reservoirs

Fangxuan Chen^a, Sheng Luo^b, Shihao Wang^b, Hadi Nasrabadi^{a,*}

^a Harold Vance Department of Petroleum Engineering, Texas A&M University, United States

^b Chevron Company United States

ARTICLE INFO

Article history:

Received 4 November 2021

Revised 14 February 2022

Accepted 20 February 2022

Available online 2 March 2022

Keywords:

Proxy phase-equilibrium calculator

Shale reservoirs

Multi-layer perceptron

Physics-informed neural network

Machine learning

ABSTRACT

In compositional reservoir simulation, a significant portion of the CPU time is consumed in phase equilibrium calculations. Previous studies have incorporated the machine learning (ML) technique to accelerate and stabilize the phase equilibrium calculations. However, there are two main limitations: 1) previous work mainly focuses on conventional reservoirs, which cannot be extended to unconventional reservoirs; 2) previous studies are limited to fluid compositions with specific hydrocarbon components that narrows their application. In this paper, we propose a novel ML-assisted framework for phase equilibrium calculations in shale reservoirs. A general set of pseudo-components is considered to allow users to customize the composition of hydrocarbon mixtures. A pore size-dependent EOS is applied to simulate the fluid phase behavior in nano-scale conditions. In the stability test, the multilayer perceptron (MLP) is trained to predict the fluid phase state: single-phase or two-phase. For the fluid labeled as two-phase condition, the phase-split computation is performed to obtain the equilibrium ratio. Instead of using the initial estimate from the stability test, the MLP and the physics-informed neural network (PINN) are applied to obtain the initial estimates for the minimization program. The results show that, with the assistance of ML technique, we are able to reduce the computation time needed for the nano-scale phase equilibrium calculations by more than two orders of magnitude while maintaining 97% accuracy. Compared with MLP, PINN can accurately predict the equilibrium ratios with a limited range of input variables but require more training time. The progress of this study present a ML-assisted framework for phase equilibrium calculations and the generalized proxy phase-equilibrium calculator can be compiled into reservoir simulator to accelerate flash calculation.

© 2022 Elsevier B.V. All rights reserved.

1. Introduction

The widespread application of horizontal wells and multi-stage hydraulic fracturing has facilitated the development of shale reservoirs [1,2]. One important difference between conventional and shale reservoirs is their pore-size distribution. Distinguished from conventional reservoirs, which are dominated by macro-scale pores ($d > 50\text{nm}$), shale reservoirs have a large number of mesopores ($2\text{nm} < d < 50\text{nm}$) and micropores ($d < 2\text{nm}$) [3]. The volume of mesopores and micropores can account for 40% of the total pore volume in shale reservoirs [4]. Due to the confinement effect, fluid-phase behavior in mesopores and micropores deviates significantly from bulk fluidphase behavior. As a result, it is necessary to find a fast and accurate way to describe the fluid-phase behavior of shale reservoirs.

1.1. Phase behavior investigation methods

Methods commonly used to investigate the phase behavior of shale reservoirs include experimental studies, molecular simulation, and equation of state (EOS) modeling [5]. The experimental methods for phase-behavior analysis include adsorption-desorption experiments [6–8], differential scanning calorimetry (DSC) techniques [9–13], diffusion experiments [14], and lab-on-a-chip technology [15–17]. Unlike experimental methods that investigate phase behavior by measurement and observation, molecular simulation provides an atomic-level perspective to reveal the positions of fluid molecules [18–21]. Although experimental methods and molecular simulation can provide relatively accurate analyses of phase behavior, the two methods are laborious and time consuming. Equation of state (EOS) modeling is efficient and can be applied to phase-equilibrium calculations in reservoir simulations. The PengRobinson equation of state [22] (PR EOS) is a commonly used EOS in petroleum engineering. However, it becomes inaccurate

* Corresponding author.

E-mail address: hadi.nasrabadi@tamu.edu (H. Nasrabadi).

Nomenclature

ANN	Artificial neural network
BP	Backpropagation
DSC	Differential scanning calorimetry
EOS	Equation of state
LSSVM	Least-squares support vector machine
LSTM	Long short-term memory
MAE	Mean absolute error
ML	Machine learning
MLP	Multi-layer perceptron
MSE	Mean square error
ODE	Ordinary differential equation
PAH	Polycyclic aromatic hydrocarbon
PCNN	Physics-constrained neural network
PGNN	Physics-guided neural network
PINN	Physics-informed neural network
PR EOS	Peng-Robinson equation of state
PR-C EOS	Peng Robinson-confined equation of state
ReLU	Rectified linear unit
RVM	Relevance vector machine
SVM	Support vector machine
VLE	Vapor-liquid equilibrium
A	Helmholtz free energy
a_p	Confinement-modified energy parameter
BIP	Binary interaction parameter
b_p	Confinement-modified volume parameter
d	Diameter
E_{conf}	Configurational energy
K	Equilibrium ratio
k	Number of folds in cross validation
k_B	Boltzmann constant
\mathcal{L}	Loss function
MW	Molecular weight
N	Number of molecules
N_c	Coordination number
n	Number of moles
nc	Number of components
ns	Number of samples
P	Pressure
P_c	Critical pressure
q	Intramolecular contribution
Q	Canonical partition function
r_p	Pore radius
T	Temperature
T_c	Critical temperature
V	Volume
V_f	Free volume
w	Acentric factor
VSP	Volume shift parameter
λ	Translational contribution
δ_p	Square well width
ε_p	Square well depth
κ	Normalized fluid-pore surface affinity factor
μ	Chemical potential
$\rho_{max,i}$	Packing density
σ	Effective parameter

rate when pore diameter is reduced to a nanometer. For applicability with shale reservoirs, researchers have primarily improved PR EOS in the following three areas: 1) incorporating capillary pressure to model the confinement effect [23–27]; 2) combining the effect of capillary pressure and adsorption on phase equilibrium [28,29]; 3) tuning critical properties of pure components to nano-

scale conditions [30–32]; and 4) using the generalized van der Waals theory [33,34] to extend the application of PR EOS [35,36]. Although the modified PR EOS shows a higher accuracy, the required CPU time is increased due to the complexity of the modified PR EOS, particularly with the extended PR EOS (PR-C EOS) [36–38]. A significant portion of the total CPU time in a compositional reservoir simulator is spent on phase-equilibrium calculations [39], which need to be calculated repeatedly. As a result, the acceleration of phase-equilibrium calculations can significantly reduce the CPU time needed in reservoir simulations.

1.2. Machine-learning application in vapor-liquid equilibria studies

In recent years, machine learning (ML) techniques have been utilized in petroleum engineering to facilitate production forecasting [40–43], parameter optimization [44–47], compositional simulation [39,48–50], and image recognition [51–53]. Specifically, many researchers have applied ML techniques in the stabilization and acceleration of vapor-liquid equilibria (VLE) calculations [54–57]. Gaganis and Varotsis [39,58] used ML to speed up the compositional reservoir simulation. To determine phase stability, the Support Vector Machines (SVM) [59] algorithm was used. Rather than collecting data points uniformly, data points around the phase boundary were more likely to be selected to train the ML model. Data points with temperatures higher than cricondenbar or pressures higher than cricondenbar were ignored, as fluid under those conditions can be directly determined. For phase-split computation, Gaganis and Varotsis utilized a feed-forward artificial neural network (ANN) to estimate the equilibrium ratio, which is an important coefficient in the RachfordRice equation [60].

Kashinath et al. [61] improved upon Gaganis and Varotsis work by applying the Relevance Vector Machine (RVM) [62] to predict fluid-phase stability. Compared to SVM, RVM set a discriminant threshold for the classification results. The predicted result was accepted only when the posterior probability was higher than the threshold, therefore reducing the misclassification rate. Groven et al. [49] used ANN to predict the VLE properties of polycyclic aromatic hydrocarbons (PAHs). The predicted results were consistent with the experimental results. Mesbah et al. [63] applied the Least-Squares Support Vector Machine (LSSVM) [64] algorithm to predict the saturation pressure of a (CO₂)-hydrocarbon binary mixture. The predicted results were in good agreement with the results calculated by EOS. Wang et al. [65] proposed a proxy phase-equilibrium calculator that could predict the phase condition for a fluid system and provide the initial guess for phase-split computation if the fluid system was in a two-phase condition. Different from the previous work, they incorporated the effect of capillary pressure into the proxy phase-equilibrium calculator to better predict the fluid-phase behavior in shale reservoirs.

Based on our analysis of the previous work on ML-assisted VLE calculation, there are two main limits. First, most of the proposed ML models are only applicable to conventional reservoirs. Although Wang et al. [65] incorporated the effect of capillary pressure in their VLE calculations, this method is not accurate when the pore diameter is less than 10 nm [23]. In their work, the pore diameter is set to a range of 60–200 nm, which does not cover the majority of pore sizes in shale reservoirs [66]. Second, the previous ML models were trained by a specific set of components, which means the models are only accurate for a particular type of hydrocarbon mixture. When given a new hydrocarbon mixture, users must regenerate the data sets and then train a new ML model, which is time consuming and laborious. As a result, a generalized ML model that is suitable for all possible hydrocarbon mixtures is required for broader application.

In this paper, we propose a generalized proxy-phase equilibrium calculator for shale reservoirs, which have resolved the two

concerns mentioned above. The PR-C EOS [35] is applied to perform VLE calculations in nano-scale porous media. This method has a high accuracy at small pore diameter ranges but requires a long running time due to its complexity. The introduction of the ML technique can significantly cut down the running time without losing significant accuracy. The combination of the PR-C EOS and ML techniques allowed our proxy phase-equilibrium calculator to conduct VLE calculations in a fast and accurate way. In addition, a representative set of pseudo-components was considered in this generalized proxy phase-equilibrium calculator. Unlike previous ML models in which the number of components and the types of components were fixed, the general set of pseudo-components allows users to customize the composition of a hydrocarbon mixture, avoiding concerns of generating different ML models for various fluid compositions. In the stability test [67], the fluid phase behavior was predicted by the MLP with an accuracy of about 97%. In the phase-split computation [68], the PINN was applied to provide initial guesses for the Helmholtz free energy minimization algorithm by incorporating the underlying scientific principles into neural networks, which ensured the rationality of the output and reduced the number of iterations. In total, the generalized proxy phase-equilibrium calculator cut down CPU time by 99% compared to the traditional phase-equilibrium calculation method.

The paper is organized as follows. Section 2 details describes the PR-C EOS and the algorithm of phase-equilibrium calculation. Section 3 shows the general set of pseudo components, required parameters of PR-C EOS and the generation of data set. Section 4 elaborates the optimized parameters and structures of the neural networks in the ML models. Section 5 presents the analysis of the results and Section 6 is the summary and conclusions of our work.

2. Forward modeling

2.1. Extended PR EOS (PR-C EOS)

Travalloni et al. [35] proposed an extended PR EOS (PR-C EOS) to investigate fluid-phase equilibrium in nano-scale conditions. The fluidpore wall interaction is described by a square-well potential with the parameters of square-well depth (ε_p) and square-well width (δ_p), as shown in Fig. 1. The PR-C EOS reverts to PR EOS when the pore diameter reaches macro-scale conditions.

Based on the generalized van der Waals theory [33,34], the canonical partition function (Q) was calculated as follows:

$$Q(T, V, N_1, N_2, \dots, N_{nc}) = \prod_{i=1}^{nc} \left(\frac{q_i^{N_i}}{\lambda_i^{3N_i} N_i!} \right) V_f^N \exp \left(\int_{\infty}^T \frac{E_{conf}}{k_B T^2} dT \right) \quad (1)$$

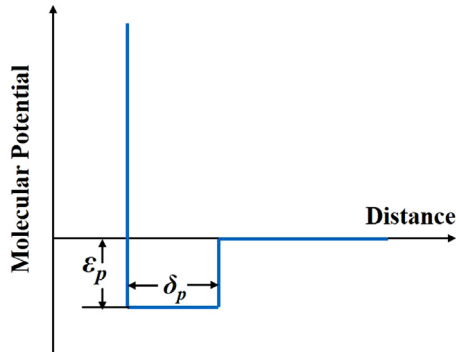


Fig. 1. Square-well potential model for the fluidpore wall interaction.

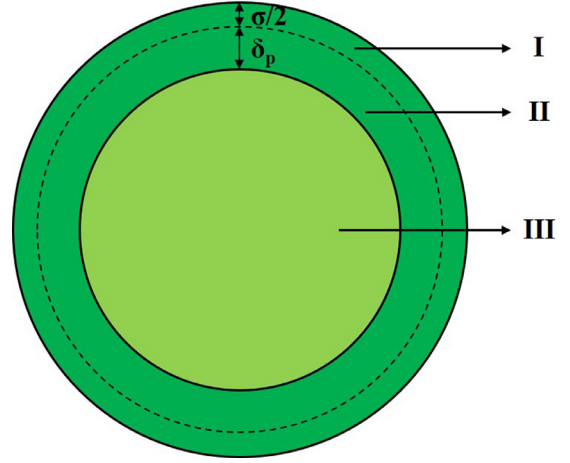


Fig. 2. Schematic figure of a cylindrical pore. The dark green regions (Regions I and II) are the surface-adsorbed regions, while the light green region (Region III) is the confined core region. Region II represents the surface-adsorbed region that is accessible to the mass center of the fluid molecules. F_{pr} is defined as the volume ratio of Region II over Regions II and III. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

where T denotes the temperature, V represents the total volume of the system, N is the total number of molecules, nc denotes the number of components in the system, q denotes intramolecular contribution, λ is translational contribution, V_f denotes the free volume, and E_{conf} represents the configurational energy. Detailed equations are listed in Appendix A.

The PR-C EOS divides a pore into two parts: the surface adsorbed region (Region I and II) and the confined core region (Region III), as shown in Fig. 2. The fluid molecules in the confined core region are only affected by fluid-fluid intermolecular interactions, while the fluid molecules in the surface adsorbed region are subject to both fluid-fluid and fluidpore wall interactions. Based on this assumption, the configurational energy (E_{conf}) is modeled as follows:

$$E_{conf} = - \sum_{i=1}^{nc} \sum_{j=1}^{nc} \left(\frac{N_j}{2} N_{c,ij} \varepsilon_{ij} \right) - \sum_{i=1}^{nc} (N_i F_{p,i} \varepsilon_{p,i}) \quad (2)$$

where $N_{c,ij}$ is the coordination number, ε_{ij} denotes the interaction energy between components i and j , $\varepsilon_{p,i}$ denotes the interaction energy between component i and the pore wall, and $F_{p,i}$ represents the percentage of molecules of component i in the surface-adsorbed region. The first part of Eq. (2) represents the fluid-fluid interactions, while the second part represents the fluidpore wall interactions. The term $F_{p,i}$ is modeled by the empirical expression as shown below:

$$F_{p,i} = F_{pr,i} + (1 - F_{pr,i}) \left(1 - \exp \left(- \frac{\varepsilon_{p,i}}{k_B T} \right) \right) \left(1 - \frac{x_i \rho}{\rho_{max,i}} \right)^{\theta_i} \quad (3)$$

$$F_{pr,i} = \frac{(r_p - \sigma_i/2)^2 - (r_p - \sigma_i/2 - \delta_{p,i})^2}{(r_p - \sigma_i/2)^2} \quad (4)$$

The term $F_{pr,i}$ is the volume fraction of the surface-adsorbed region accessible to the mass center of the fluid molecules in a pore (calculated as the volume ratio of Region II over Regions II and III), which is affected by the value of $\delta_{p,i}$. The term $\delta_{p,i}$ is the square-well width of the fluidpore wall interaction potential of component i . When the distance between a fluid molecule and a pore wall is less than $\delta_{p,i}$, the fluid molecule is subject to fluidpore wall interaction. In this paper, the square-well width (δ_p) of different types of hydrocarbons was set as $0.5\sigma_i$, which is widely used for square-well potential models in the existing literature [69–71].

The Helmholtz free energy (A) and chemical potential of component i (μ_i) were obtained using the following thermodynamic relations [72]:

$$A(T, V, N_1, N_2, \dots, N_{nc}) = -k_B T \ln_Q(T, V, N_1, N_2, \dots, N_{nc}) \quad (5)$$

$$\mu_i = \left(\frac{\partial A}{\partial N_i} \right)_{T, V, N_{j \neq i}} \quad (6)$$

Herein, the PR-C EOS model was used in data generation in the present study. The volume shift parameter (VSP) was incorporated into the PR-C EOS to calibrate the prediction of liquid density [73], and the binary interaction parameter (BIP) was considered to account for the deviation from ideal conditions [74]. In summary, the input parameters of the forward model were pressure (P), temperature (T), pore radius (r_p), number of moles (n), critical pressure (P_c), critical temperature (T_c), molecular weight (MW), acentric factor (w), volume shift parameter (VSP), square-well depth (ε_p) and binary interaction parameter (BIP). The Helmholtz free energy (A) and chemical potential (μ) were then calculated using Eqs. (5) and (6) based on the input parameters.

2.2. Stability test

The basic principle of the stability test using the PR-C EOS is the equality of the chemical potential of component i in the vapor and liquid phases, as shown in the equation below:

$$\mu_i^V = \mu_i^L \quad (7)$$

where μ_i^V and μ_i^L represent the chemical potential of component i in vapor and liquid phases, respectively. We first assumed that the fluid was in a single-phase condition and calculated the chemical potential of each component (μ_i^I) using Eq. (6). We then searched for every possible composition with the same chemical potential (μ_i^{II}) as μ_i^I . The Newton-Raphson iteration was used in the search task. If the new phase composition (n_i^{II}) was the same as the initial single-phase composition (n_i^I), the solution was considered trivial, and the fluid was considered in single-phase condition; otherwise, the fluid was considered in two-phase condition. If the fluid was in two-phase condition, the phase-split computation was conducted to obtain the compositions of the vapor and liquid phases. The flowchart of the stability test is shown in Fig. 3.

2.3. Phase-split computation

The phase-split computation was performed by directly minimizing the total Helmholtz free energy (A_{total}), which was calculated using the equation below:

$$A_{total} = A^I + A^{II} \quad (8)$$

where A^I and A^{II} represent the Helmholtz free energy of phases I and II , which can be calculated by Eq. (5). The variables of the Helmholtz free energy minimization algorithm are the equilibrium ratios of all pseudo-components (K_i), which can be used to calculate the number of moles of each component (n_i) in the vapor and liquid phases. The initial estimate of the equilibrium ratio was calculated based on the n_i^I and n_i^{II} obtained from the stability test. Since the fluid is determined to be in two-phase condition, the total Helmholtz free energy is minimized by changing the values of K_i . The lowest total Helmholtz free energy corresponds to the desired value of K_i .

3. Dataset generation

3.1. General set of pseudo-components

In the generalized proxy phase-equilibrium calculator, instead of using a fixed set of real components, a general set of pseudo-

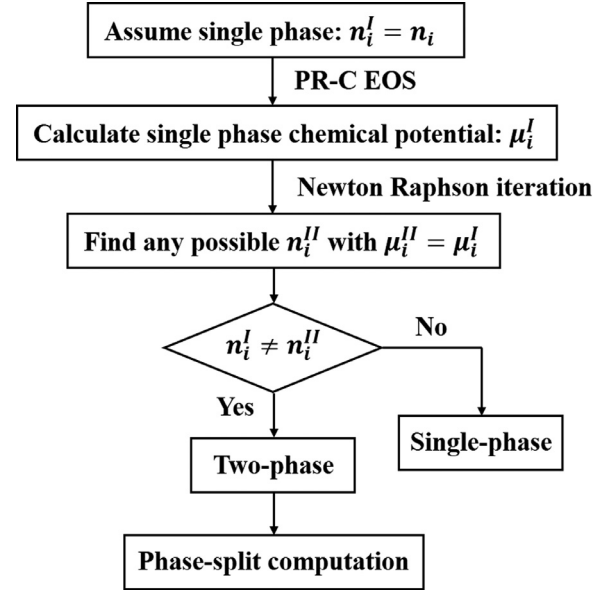


Fig. 3. Flowchart of stability test using PR-C EOS. The fluid is first assumed to be in single-phase condition, and the chemical potential of each component (μ_i^I) is calculated using Eq. (6). Then, every possible composition with the same chemical potential (μ_i^{II}) as μ_i^I is searched using the Newton-Raphson iteration. If the new phase composition (n_i^{II}) is the same as the initial single-phase composition (n_i^I), the solution is trivial, and the fluid is in single-phase condition. Otherwise, the fluid is in two-phase condition. If the fluid is in two-phase condition, the phase-split computation is conducted to obtain the compositions of the vapor and liquid phases.

components with PR-C EOS parameters is considered. Based on the literature review of compositional analysis [75–77], we selected six pseudo-components to represent the composition of various types of reservoir fluids, including $C_1 - N_2$, $C_2 - CO_2 - C_3$, $C_4 - 6$, $C_7 - 12$, C_{13-24} , C_{25+} . When the fluid composition is given, it can be categorized as one of the six pseudo-components, and the PR EOS parameters of each pseudo-component (n_i , $P_{c,i}$, $T_{c,i}$, MW_i , w_i , VSP_i , and $BIP_{i,j}$) can be obtained. If these PR EOS parameters are considered to be the input variables for the ML model, the ML model can be suitable for any fluid composition, which extends the application of our approach from fixed-fluid composition to generalized conditions.

3.2. Data range and acquisition

The input parameters of the generalized proxy phase-equilibrium calculator included P , T , r_p , n , P_c , T_c , MW , w , VSP , ε_p and BIP . The range of P and T followed the reference [52], and the range of r_p was based on the pore-size distribution of shale reservoirs [66], as shown in Table 1. The lower bound of r_p was set at 2 nm because the pore diameter should be larger than the effective diameter (σ) of the fluid molecules, as required by the PR-C EOS.

The values of the PR EOS parameters (n , P_c , T_c , MW , w , VSP , and BIP) were determined by applying the mixing rule based on the mole fraction of each real component in a pseudo-component, as

Table 1
Range of parameters of reservoir conditions.

Parameter	Unit	Minimum	Maximum
Pressure (P)	MPa	1	80
Temperature (T)	°C	40	100
Pore radius (r_p)	nm	2	50

Table 2
Range of parameters of the extended PR EOS.

Parameter	Unit	Minimum	Maximum
Critical pressure (P_c)	Pa	736,768	5,455,021
Critical temperature (T_c)	K	175	954
Molecular weight (MW)	g/mol	16	533
Acentric factor (w)	dimensionless	0.0084	1.3189
Volume shift parameter (VSP)	dimensionless	-0.1678	0.3535
Binary interaction parameter (BIP)	dimensionless	0	0.1303
Square-well depth (ε_p)	K	228	2788

shown in the equation below:

$$S_{pseudo} = \sum x_{real,i} S_{real,i} \quad (9)$$

where S denotes one of the PR EOS parameters and $x_{real,i}$ is the mole fraction of the real component i in a pseudo-component. Each pseudo component was categorized under several real components for the simplicity of data generation, as shown in Table B.1 in Appendix B. The values of the PR EOS parameters of the real components are listed in Tables B.2 and B.3. To clarify Eq. (9), an example of calculating the critical pressure of $C_2 - CO_2 - C_3$ is shown below:

$$P_{C_2-CO_2-C_3} = P_{C_2} \cdot x_{C_2} + P_{C_2CO_2} \cdot x_{CO_2} + P_{C_2C_3} \cdot x_{C_3} \quad (10)$$

$$x_{C_2} + x_{CO_2} + x_{C_3} = 1 \quad (11)$$

where $P_{C_2-CO_2-C_3}$, P_{C_2} , $P_{C_2CO_2}$, and $P_{C_2C_3}$ denote the critical pressure of $C_2 - CO_2 - C_3$, C_2 , CO_2 , and C_3 respectively and x_{C_2} , x_{CO_2} , and x_{C_3} represent the mole fraction of C_2 , CO_2 , and C_3 in $C_2 - CO_2 - C_3$.

The PR-C EOS parameter (ε_p) was calculated based on the equation [78] shown below

$$\kappa = \frac{\varepsilon_p \delta_p}{MW} \quad (12)$$

where κ is the normalized fluid-pore surface affinity factor, which can be obtained from the reference [78]. The values of ε_p of real components are listed in Table B.4. The value of ε_p of a pseudo-component was calculated using the same method as shown in Eq. (9).

In summary, for a sixpseudo-component system, each parameter of n , P_c , T_c , MW , w , VSP , and ε_p contained six values. The matrix of BIP had a symmetric distribution, and the values of the diagonal elements in the matrix were zeros. As a result, for a sixpseudo-component system, 36 elements in the BIP matrix could be described by 15 different values. Together with the parameters of the reservoir conditions (P , T , r_p), the total number of input parameters for the generalized model was 60. The ranges of different parameters of the extended PR EOS are summarized in Table 2.

3.3. Data samples

The samples were generated using the Latin hypercube sampling technique [79] to ensure randomness in a multidimensional distribution. A total of 123,050 samples were generated for the training and testing of the ML model.

4. ML models for phase-equilibrium calculation

The phase-equilibrium calculation consisted of two parts: the stability test and the phase split computation. In the stability test, the MLP was applied to predict the fluid-phase state (single-phase or two-phase) due to its high flexibility and efficiency. In the phase-split computation, the MLP and PINN models were trained to obtain the initial estimates of equilibrium ratios, rather than using the initial estimate from the stability test.

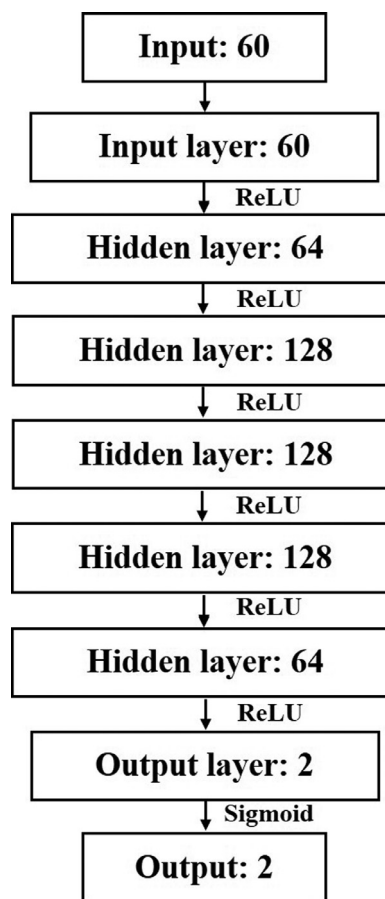


Fig. 4. Structure of the MLP of stability test. The numbers represent the number of nodes in each layer.

4.1. Stability test

The function of the ML model of stability test was to determine the fluid phase state: single-phase or two-phase. MLP is applied in the ML model to predict the phase state. The input variables of the MLP were the parameters of the reservoir conditions and EOS, a total number of 60, as mentioned in Section 3.2. The output variable was the label of the phase-state condition, including single-phase and two-phase conditions. The detailed structure of the MLP is shown in Fig. 4, which includes an input layer, five hidden layers, and an output layer. The activation function of the hidden layers was ReLU [80], whereas the activation function of the output layer was sigmoid. The loss function was set as binary cross-entropy because the stability test is a binary classification problem. The Adam optimization algorithm [81] was used in the MLP, which is computationally efficient and requires less memory. During the training process, the model is optimized by continuously changing the weight and bias values of each node. All input variables were scaled based on the normalization equation shown below:

$$y_{scaled} = (y - y_{min}) / (y_{max} - y_{min}) \quad (13)$$

where y denotes a variable among the 60 input variables.

A total of 123,050 samples were used for the stability test. The k -fold cross-validation technique [82,83] was applied in the training process. In the present study, k was set as 10, which means the samples were divided into 10 parts. Each time, one part was selected as the test set and the remaining nine parts were used for the training set. The training process was performed 10 times until each part had been selected as the test set. The accuracy ob-

Table 3
The optimized parameters of MLP.

Parameters of MLP	Optimal results
Number of hidden layers	5
Number of nodes	64, 128, 128, 128, 64
Batch size	64
Number of epochs	80
Learning rate of Adam	0.001

tained using k -fold cross-validation was more robust and trustworthy compared to the method of splitting data directly, as each sample was considered for training and testing.

To increase the accuracy of our ML model, several parameters of the MLP were optimized separately. First, the number of hidden layers were optimized to provide a proper MLP structure. The results showed that the structure with five hidden layers had the highest accuracy. Less layers may not be able to fully interpret the complex relationship between the input and output variables, while too many layers may lead to overfitting. Next, different combinations of nodes in each layer tested, and the results showed that this specific set of nodes (64, 128, 128, 128, and 64) had the best performance. By balancing accuracy and CPU time, the optimized batch size and the optimized number of epochs were 64 and 80, respectively. The optimal learning rate of Adam was 0.001. The optimized parameters of the MLP are summarized in Table 3.

4.2. Phase-split computation

The function of the ML model of phase-split computation was to obtain the initial guess of the equilibrium ratio for the Helmholtz free energyminimization algorithm. A good initial guess was shown to largely reduce the number of iterations and, therefore, lower the CPU time. In the forward modeling, the initial estimate of the equilibrium ratio was obtained from the stability test. In the generalized proxy phase-equilibrium calculator, the MLP and PINN models of phase-split computation were trained to obtain the equilibrium ratio as the initial estimate for the Helmholtz free energyminimization algorithm.

4.2.1. MLP neural network of phase split computation

The input variables were the parameters of the reservoir conditions and EOS, whereas the output variables were the equilibrium ratios of the six pseudo-components. The structure of the MLP neural network of the phase-split computation is shown in Fig. 5, which is similar to the MLP neural network used for the stability test. The activation functions of the hidden layers are ReLU whereas the activation function of the output layer is sigmoid since the value of the output is in the range of 0 and 1. The optimization algorithm used for this model was also Adam. Both the input and output variables were normalized using Eq. (13). The loss function was the mean square error (MSE), which was calculated based on the equation shown below:

$$MSE = \frac{1}{ns \cdot nc} \sum_{j=1}^{ns} \sum_{i=1}^{nc} (\hat{K}_{i,j} - K_{i,j})^2 \quad (14)$$

where ns represents the number of samples and $\hat{K}_{i,j}$ is the predicted equilibrium ratio of component i in sample j .

The total number of samples for phase-split computation was 52,102. The MLP neural network of phase-split computation was optimized following the same method for the optimization of the MLP neural network for the stability test. The optimized results are summarized in Table 4.

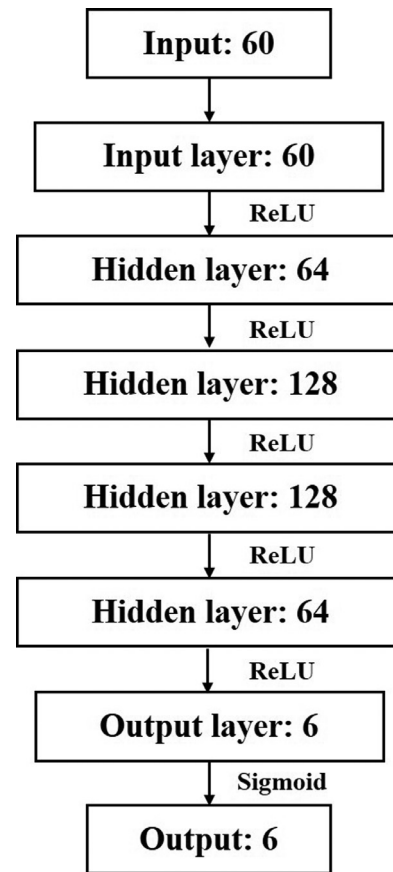


Fig. 5. Structure of the MLP of phase split computation. The numbers represent the number of nodes in each layer.

Table 4
The optimized parameters of MLP.

Parameters of MLP	Optimal results
Number of hidden layers	4
Number of nodes	64, 128, 128, 64
Batch size	64
Number of epochs	40
Learning rate of Adam	0.001

4.2.2. PINN of phase-split computation

Physics-based models are typically limited by the assumptions of ideal conditions, which may not fully interpret the complex relationships in engineering systems. Black-box ML models often encounter problems in that the predictions are inconsistent with basic physical laws. In addition, ML models require a substantial amount of data for training and fail to predict accurately when the values of the input variables are beyond the range of the training set. To avoid the limitations and leverage the advantages of the physics-based and ML models, many hybrid physics-data models have been proposed by researchers to incorporate physical mechanisms into ML models, including the physics-informed neural network (PINN), physics-constrained neural network (PCNN), and physics-guided neural network (PGNN).

The PINN was initially designed to solve partial differential equations or to determine potential partial differential equations to govern a system [84]. Recently, PINNs have been applied in petroleum engineering studies. Klie et al. [85] applied a PINN in the prediction of oil and gas production. Given that oil and gas rates are functions of time, the ordinary differential equation (ODE) set with respect to time was incorporated to train the ML model

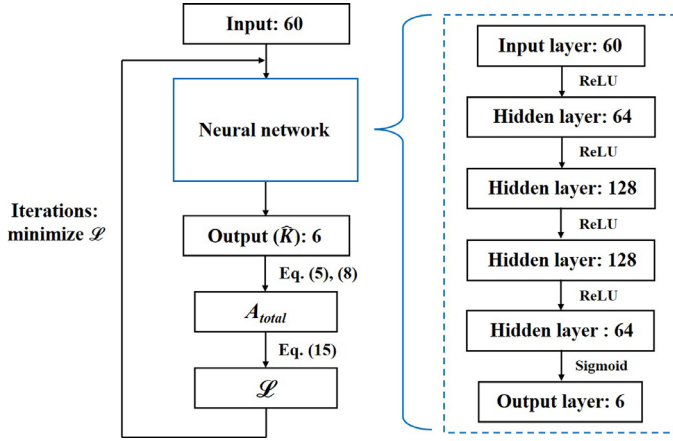


Fig. 6. Flow chart of the PINN of phase-split computation. The structure of the neural network is shown in the blue dashed box. The input variables are the parameters of reservoir conditions and EOS, whereas the output parameters of the neural network are the equilibrium ratios. A_{total} is calculated based on Eqs. (5) and (8) using \hat{K} predicted by the neural network. During the iterations, \mathcal{L} is continuously minimized and therefore, the neural network is trained to give more accurate predictions of \hat{K} . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

in a specific direction. Sarma et al. [86] applied the PINN in their steam-injection model to reduce operational costs and increase the production rate. The equations of mass conservation, energy conservation, and Darcy's law were incorporated into the PINN to ensure the rationality and accuracy of the output. In this paper, PINN is applied in the phase-split computation to obtain the initial guess for the Helmholtz free energy-minimization algorithm.

In the PINN, the input variables are the parameters of reservoir conditions and EOS whereas the output parameters are the equilibrium ratios. The flow chart of the PINN of phase split computation is shown in Fig. 6. The structure of the neural network in the PINN was the same as the neural network of the MLP for the phase-split computation, as shown in the blue dashed box in Fig. 6. Different from the MLP model that directly minimizes the deviations between the real values and the estimated values of the equilibrium ratios, the equilibrium ratios (\hat{K}) predicted by the neural network are used to calculate the total Helmholtz free energy (A_{total}) based on Eqs. (5) and (8). During the iterations, A_{total} is continuously minimized and therefore, the neural network is trained to give more accurate predictions of \hat{K} . It should be noted that, during the training process, the real values of the equilibrium ratios are not needed, which means that the training of PINN does not require a data set. What we need to prepare is only the different combination of the values of input variables. The optimization algorithm used was Adam, with a learning rate of 0.001. The loss function was calculated based on the equation shown below:

$$\mathcal{L} = \frac{1}{ns} \sum_{j=1}^{ns} (A_{total,j} / 10^7) \quad (15)$$

5. Results and discussion

In this work, all the ML models were trained on Keras [87] with the CPU of Intel Core i7-1065G7 and an installed memory of 16 GB. A comparison between the forward modeling and the proxy phase-equilibrium calculator is shown below.

5.1. Phase-equilibrium calculator using PR EOS

Herein, we first compare the average simulation time for the phase-equilibrium calculation using PR EOS and PR-C EOS, as

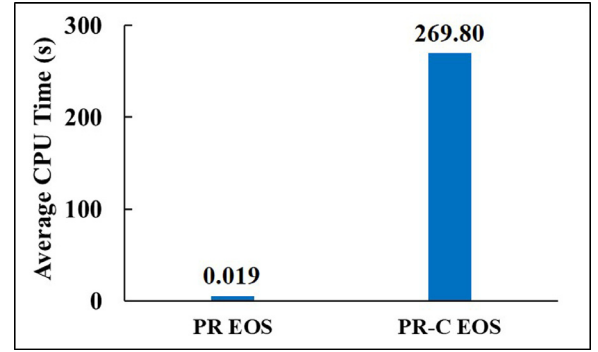


Fig. 7. The average CPU time for phase-equilibrium calculation for one sample using PR EOS and PR-C EOS.

shown in Fig. 7. In the previous flash calculation studies [88], the computation time per flash using cubic EOS is about 10 microseconds, which is much faster compared with our simulator. However, the aim of this section is not to optimize and accelerate traditional phase-equilibrium calculator, but to show that the phase-equilibrium calculation using PR-C EOS is much more complex compared with the phase-equilibrium calculation using PR EOS. The reasons are two folded: 1) the interaction between the fluid and pore wall is incorporated in PR-C EOS. As shown in Section 2.1, more parameters are introduced in PR-C EOS when the effect of fluid-pore wall interaction is taken into consideration. Compared with PR EOS, these extra computations require more CPU time; 2) the algorithm of solving the equilibrium of chemical potential (Eq. (7)) is more difficult with PR-C EOS. In the phase-equilibrium calculation using PR EOS, the method of tangent plane distance (TPD) is often applied to find the solutions. However, in the phase-equilibrium calculation using PR-C EOS, the TPD method is not applicable. The fluid is considered to be in single-phase condition only when all possible roots are checked to be trivial roots, which is much more time-consuming. As a result, the application of ML technique provided a more significant speedup in the phase-equilibrium calculations using PR-C EOS.

5.2. Stability test

Based on the optimized MLP-based stability test, the average accuracy of the test set using k -fold cross validation technique ($k = 10$) was 96.96%. We randomly selected 500 samples from the data set and performed the stability test by applying the algorithm in forward modeling, as well as the MLP-based stability test. The average CPU time for performing the stability test for one sample using the two methods is shown in Fig. 8. The results indicated

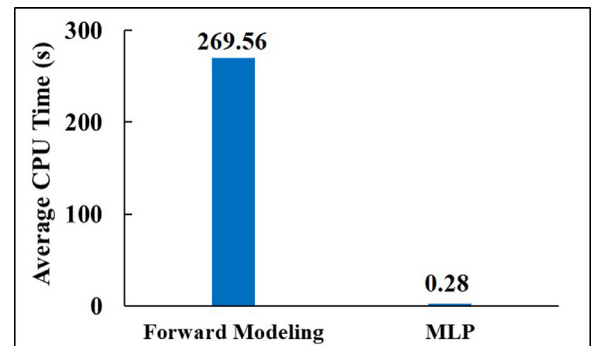


Fig. 8. Average CPU time for performing the stability test for one sample using forward modeling and the MLP.

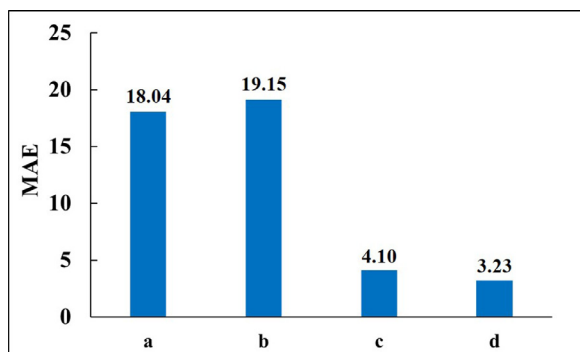


Fig. 9. Mean absolute error (MAE) of equilibrium ratios obtained by the initial estimate from the stability test and ML models: (a) the initial estimate obtained from the stability test, (b) MLP neural network trained with 1000 samples, (c) MLP neural network trained with 52,102 samples, (d) PINN trained with 1000 samples. The MAEs are calculated using Eq. (16).

that the utilization of the ML technique can significantly improve the computational efficiency by 963 times compared to the traditional method.

5.3. Phase-split computation

For the phase split computation, the PINN model was trained using 1000 input combinations, while the MLP models were trained with 1000 and 52,102 samples, respectively.

The mean absolute error (MAE) of the predicted equilibrium ratios was calculated based on the equation shown below:

$$MAE = \frac{1}{ns \cdot nc} \sum_{j=1}^{ns} \sum_{i=1}^{nc} \left| \frac{\hat{K}_{i,j} - K_{i,j}}{K_{i,j}} \right| \quad (16)$$

The comparisons between the deviations of the predicted equilibrium ratios by the initial estimate from the stability test and the ML models are shown in Fig. 9. The results show that the deviations of the equilibrium ratios obtained by MLP with 52,102 samples and PINN with 1000 input combinations are 77.26% and 82.12% lower than the deviations obtained from the stability test, respectively. The predicted results from the MLP with 1000 samples are worse than the initial estimate from the stability test, indicating that the training of MLP requires a large data set to achieve high accuracy. The results show that, by incorporating the PR-C EOS into the training of neural network, PINN can accurately predict the equilibrium ratios with a limited range of input variables. However, for the MLP neural network, a huge data set is required to achieve a high accuracy. As a result, for the condition that a huge data set is not available, PINN can be applied to avoid the concern of generating the data set since PINN directly minimizes the loss function.

Although the deviations of equilibrium ratios obtained by the ML models are significantly reduced compared with the initial estimate obtained from the stability test, the minimization program cannot be completely replaced by the ML models. The reason is that the equilibrium ratios will be used to determine the composition of each phase. A small deviation of equilibrium ratios can have significant effect on fluid composition. In addition, the error

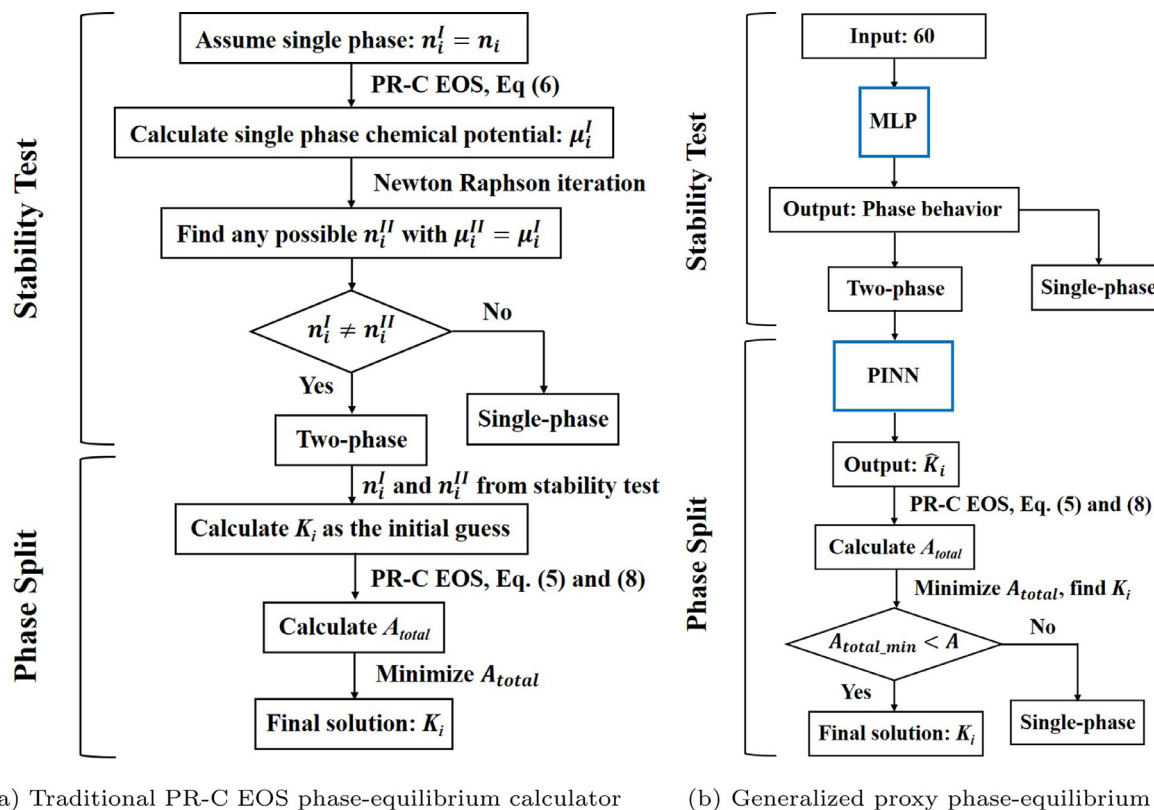


Fig. 10. Flow chart of the traditional PR-C EOS phase-equilibrium calculator and the generalized proxy phase-equilibrium calculator. In the traditional PR-C EOS phase-equilibrium calculator, the stability test is conducted by NewtonRaphson iterations to find possible solutions to the equality of chemical potentials, and the initial guess of the equilibrium ratio for the Helmholtz free energy minimization algorithm is obtained from the result in stability test. In the generalized proxy phase-equilibrium calculator, the phase behavior is directly predicted by the ML model, and the initial guess for the Helmholtz free energy minimization algorithm is obtained from the PINN of the phase-split computation.

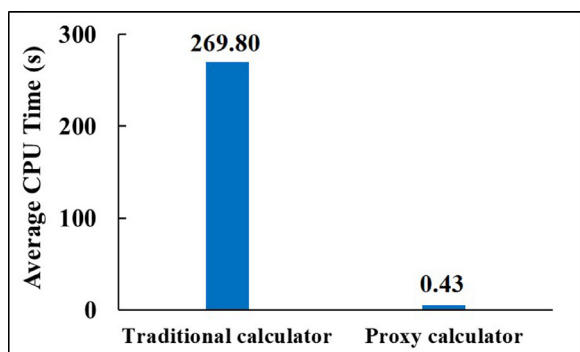


Fig. 11. Average CPU time of the phase-equilibrium calculation for one sample performed by the traditional PR-C EOS phase-equilibrium calculator and the generalized proxy phase-equilibrium calculator.

caused by the deviations will accumulate and magnify in the reservoir simulation process, which leads to a large discrepancy.

5.4. Phase-equilibrium calculation

Herein, the stability test and phase-split computation were integrated as phase-equilibrium calculations. The algorithms of the traditional PR-C EOS phase-equilibrium calculator and the generalized proxy phase-equilibrium calculator are shown in Fig. 10. In the traditional PR-C EOS phase-equilibrium calculator, the stability test is conducted by Newton-Raphson iterations to find possible solutions to the equality of chemical potentials. In the generalized proxy phase-equilibrium calculator, the results of the stability test are directly obtained from the ML model. For the phase-split computation, the initial guess of the equilibrium ratio for the Helmholtz free energy minimization algorithm is obtained from the result in stability test in the traditional PR-C EOS phase-equilibrium calculator, whereas the initial guess is obtained from the PINN of the phase-split computation in the generalized proxy phase-equilibrium calculator. In the generalized proxy phase-equilibrium calculator, we have included an additional phase-stability check: if the minimized A_{total} is higher than A_{single} , the fluid is considered in single-phase condition; otherwise, the system is in two-phase condition. Different from the traditional PR-C EOS phase-equilibrium calculation, here, the result of stability test is directly obtained from the ML prediction result, which has an accuracy of about 97%. The additional phase-stability check further inspects the accuracy of the ML-based phase stability analysis, and therefore increases the accuracy and the reliability of the phase equilibrium calculations.

We randomly selected 100 samples from the data set and performed the phase-equilibrium calculation using the traditional PR-C EOS phase-equilibrium calculator and the generalized proxy phase-equilibrium calculator. The average CPU time of the phase-equilibrium calculation for one sample is shown in Fig. 11. Compared to the traditional method, the average CPU time was reduced by 627 times with the assistance of the ML technique.

6. Conclusion

In the present work, we proposed a generalized ML-assisted phase-equilibrium calculation model to perform stability tests and phase-split computations for shale reservoirs. Compared to the previous work of ML-assisted phase-equilibrium calculation, our model has three main innovations as follows: 1) the extended PR EOS was applied to investigate the fluid-phase behavior in shale reservoirs; 2) a general set of pseudo-components was selected to allow users to customize the composition of the hydrocarbon mixture; 3) the physical equations were incorporated into the training

of the ML model to achieve high accuracy. For the stability test, the MLP model was trained based on the input variables of the reservoir conditions and the EOS parameters to determine whether the fluid was in single-phase or two-phase condition. If the fluid was in two-phase condition, the PINN model was used to obtain the initial guess of the equilibrium ratio for the Helmholtz free energy minimization algorithm instead of the estimate in stability test. The generalized proxy phase-equilibrium calculator was shown to be suitable for any fluid composition and could be implemented into reservoir simulators to perform phase-equilibrium calculations for shale reservoirs in a fast and accurate manner. In summary, we have the following findings and contributions:

(1) The extended PR EOS was applied to investigate the fluid-phase behavior in shale reservoirs, which ensures the accuracy of the generalized proxy phase-equilibrium calculator under nano-scale conditions.

(2) Compared with previous works in which the number of components and the types of components were fixed, the general set of pseudo-components allows users to customize the composition of a hydrocarbon mixture, avoiding concerns of generating different ML models for various fluid compositions.

(3) By incorporating the PR-C EOS into the training of neural network, PINN can accurately predict the equilibrium ratios with a limited range of input variables, which indicates that, for the condition that a huge data set is not available, PINN can be applied to avoid the concern of generating the data set.

(4) For the stability test, the accuracy of the MLP model was 96.96%, and the average CPU time was reduced by 963 times compared to the traditional method using Newton-Raphson iterations.

(5) For the phase-split computation, the deviations of the equilibrium ratios obtained by the PINN were 82.12% lower than the deviations obtained from the stability test. The MLP model had similar achievements when given a large data set.

(6) In total, the average CPU time for one phase-equilibrium calculation was reduced by 627 times using the generalized proxy phase-equilibrium calculator.

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.

CRediT authorship contribution statement

Fangxuan Chen: Conceptualization, Methodology, Visualization, Software, Writing – original draft. **Sheng Luo:** Conceptualization, Methodology, Software, Writing – review & editing. **Shihao Wang:** Conceptualization, Methodology, Writing – review & editing. **Hadi Nasrabadi:** Conceptualization, Supervision, Writing – review & editing.

Acknowledgement

We appreciate the support from the Texas A&M High Performance Research Computing Facility in performing the required phase-equilibrium calculations to generate the data set.

Appendix A

The free volume (V_f) is calculated based on the equations shown below:

$$V_f = V - \sum_{i=1}^{nc} \frac{N_i}{\rho_{\max,i}} \quad (\text{A.1})$$

Table B.1

The lumped real components of 6 pseudo components.

Pseudo Components	Real Components
$C_1 -$	C_1
N_2	N_2
$C_2 -$	C_2
$CO_2 -$	CO_2
C_3	C_3
C_{4-6}	C_4
	C_5
	C_6
C_{7-12}	C_7
	C_9
	C_{12}
C_{13-24}	C_{13}
	C_{19}
	C_{24}
C_{25+}	C_{25}
	C_{35}
	C_{45}

Table B.2

The PR EOS parameters of all real components.

Real Components	P_c (Pa)	T_c (K)	MW(g/mol)	w	VSP
C_1	4,600,155	190.6	16	0.008	-0.1595
N_2	3,394,388	126.2	28	0.040	-0.1927
C_2	4,883,865	305.4	30	0.098	-0.1134
CO_2	7,376,460	304.2	44	0.225	-0.0817
C_3	4,245,518	369.8	44	0.152	-0.0863
C_4	3,799,688	425.2	58	0.193	-0.0675
C_5	3,374,123	469.6	72	0.251	-0.0390
C_6	2,968,823	507.4	86	0.296	-0.0080
C_7	2,735,775	540.2	100	0.351	0.0033
C_9	2,310,210	594.6	128	0.444	0.0408
C_{12}	2,191,660	663.9	161	0.522	0.0695
C_{13}	2,070,070	682.4	175	0.560	0.0775
C_{19}	1,525,955	771.0	163	0.790	0.1225
C_{24}	1,254,404	823.2	324	0.940	0.1835
C_{25}	1,199,688	832.7	337	0.965	0.1990
C_{35}	902,806	905.9	445	1.179	0.2883
C_{45}	722,346	957.8	539	1.330	0.3583

$$\rho_{\max,i}\sigma_i^3 = 1.158 - 0.479\exp\left(0.621\left(0.5 - \frac{r_p}{\sigma_i}\right)\right) + 0.595\exp\left(4.014\left(0.5 - \frac{r_p}{\sigma_i}\right)\right) \quad (A.2)$$

$$\sigma_i = \sqrt[3]{\frac{1.158b_i}{N_{av}}} \quad (A.3)$$

The parameters $\rho_{\max,i}$ and σ_i represent the packing density of pure component i and the effective diameter of component i , respectively.

The parameter θ_i in Eq. (3) is calculated as follows:

$$\theta_i = \frac{r_p}{\delta_{p,i} + \sigma_i/2} \quad (A.4)$$

The confinement-modified energy parameter ($a_{p,ij}$) between components i and j and the confinement-modified volume parameter ($b_{p,i}$) of component i were calculated as follows:

$$a_{p,ij} = \sqrt{a_i a_j} \left(1 - \frac{2}{5} \frac{\sigma_{ij}}{r_p}\right) \quad (A.5)$$

$$b_{p,i} = \frac{N_{av}}{\rho_{\max,i}} \quad (A.6)$$

Table B.3

The values of BIP of all real components.

Real Components	C ₁ (Pa)	N ₂	C ₂	CO ₂	C ₃	C ₄	C ₅	C ₆	
C ₁	0.000000	0.025000	0.002689	0.105000	0.008537	0.014749	0.020640	0.028330	
N ₂	0.025000	0.000000	0.010000	0.000000	0.090000	0.095000	0.110000	0.110000	
C ₂	0.002689	0.010000	0.000000	0.130000	0.001662	0.004914	0.008578	0.013840	
CO ₂	0.105000	0.000000	0.130000	0.000000	0.125000	0.115000	0.115000	0.115000	
C ₃	0.008537	0.090000	0.001662	0.125000	0.000000	0.000866	0.002712	0.005978	
C ₄	0.014749	0.095000	0.004914	0.115000	0.000866	0.000000	0.000515	0.002305	
C ₅	0.020640	0.110000	0.008578	0.115000	0.002712	0.000515	0.000000	0.000643	
C ₆	0.028330	0.110000	0.013840	0.115000	0.005978	0.002305	0.000643	0.000000	
C ₇	0.035185	0.110000	0.018844	0.115000	0.009436	0.004615	0.002055	0.000400	
C ₉	0.047012	0.110000	0.027939	0.115000	0.016229	0.00968	0.005761	0.002566	
C ₁₂	0.054287	0.110000	0.033750	0.115000	0.020795	0.013300	0.008640	0.004592	
C ₁₃	0.058520	0.110000	0.037195	0.115000	0.023560	0.015555	0.010480	0.005967	
C ₁₉	0.081500	0.110000	0.056500	0.115000	0.039669	0.029200	0.022170	0.015390	
C ₂₄	0.094117	0.110000	0.067493	0.115000	0.049172	0.037552	0.029600	0.021720	
C ₂₅	0.096000	0.110000	0.069170	0.115000	0.050638	0.038853	0.030740	0.022730	
C ₃₅	0.116688	0.110000	0.087486	0.115000	0.066883	0.053464	0.044000	0.034450	
C ₄₅	0.132875	0.110000	0.102154	0.115000	0.080148	0.065610	0.055240	0.044580	
Real Components	C ₇	C ₉	C ₁₂	C ₁₃	C ₁₉	C ₂₄	C ₂₅	C ₃₅	C ₄₅
C ₁	0.035185	0.047012	0.054287	0.058520	0.081500	0.094117	0.096000	0.116688	0.132875
N ₂	0.110000	0.110000	0.110000	0.110000	0.110000	0.110000	0.110000	0.110000	0.110000
C ₂	0.018844	0.027939	0.033750	0.037195	0.056500	0.067493	0.069170	0.087486	0.102154
CO ₂	0.115000	0.115000	0.115000	0.115000	0.115000	0.115000	0.115000	0.115000	0.115000
C ₃	0.009436	0.016229	0.020795	0.023560	0.039669	0.049172	0.050638	0.066883	0.080148
C ₄	0.004615	0.009680	0.013300	0.015555	0.029200	0.037552	0.038853	0.053464	0.065610
C ₅	0.002055	0.005761	0.008640	0.010480	0.022170	0.029600	0.030740	0.044000	0.055240
C ₆	0.000400	0.002566	0.004592	0.005967	0.015390	0.021720	0.022730	0.034450	0.044580
C ₇	0.000000	0.000942	0.002289	0.003289	0.010895	0.016334	0.017220	0.027663	0.036900
C ₉	0.000942	0.000000	0.000295	0.000713	0.005470	0.009517	0.010200	0.018615	0.026410
C ₁₂	0.002289	0.000295	0.000000	0.000091	0.003234	0.006486	0.007055	0.014300	0.021260
C ₁₃	0.003289	0.000713	0.000091	0.000000	0.002243	0.005050	0.005554	0.012143	0.018627
C ₁₉	0.010895	0.005470	0.003234	0.002243	0.000000	0.000566	0.000744	0.004000	0.008064
C ₂₄	0.016334	0.009517	0.006486	0.005050	0.000566	0.000000	0.000012	0.001560	0.004378
C ₂₅	0.017220	0.010200	0.007055	0.005554	0.000744	0.000012	0.000000	0.001297	0.003931
C ₃₅	0.027663	0.018615	0.014300	0.012143	0.004000	0.001560	0.001297	0.000000	0.000716
C ₄₅	0.036900	0.026410	0.021260	0.018627	0.008064	0.004378	0.003931	0.000716	0.000000

Table B.4The values of ε_p of real components.

Real Components	$\varepsilon_p/k_B(K)$
C ₁	228.57
N ₂	444.71
C ₂	490.48
CO ₂	1202.76
C ₃	1215.24
C ₄	1842.76
C ₅	1913.58
C ₆	1975.45
C ₇	2030.54
C ₉	2125.72
C ₁₂	2242.81
C ₁₃	2276.95
C ₁₉	2448.06
C ₂₄	2561.16
C ₂₅	2581.55
C ₃₅	2756.77
C ₄₅	2896.25

Applying the mixing rule based on the mole fraction of each component in a mixture, the confinement-modified energy parameter (a_p) and the confinement-modified volume parameter (b_p) of the fluid mixture were calculated as follows:

$$a_p = \sum_{i=1}^{nc} \sum_{j=1}^{nc} (x_i x_j a_{p,ij}) \quad (\text{A.7})$$

$$b_p = \sum_{i=1}^{nc} x_i b_{p,i} \quad (\text{A.8})$$

The parameters of the PR EOS used in this paper were determined by the equations shown below:

$$m_i = \begin{cases} 0.3746 + 1.5423w_i - 0.2699w_i^2 & 0 < w_i \leq 0.5 \\ 0.3796 + 1.4850w_i - 0.1644w_i^2 + 0.01677w_i^3 & w_i > 0.5 \end{cases} \quad (\text{A.9})$$

$$a_i = \frac{0.4572R^2T_{c,i}^2}{P_{c,i}} \left(1 + m_i \left(1 - \sqrt{\frac{T}{T_{c,i}}} \right) \right)^2 \quad (\text{A.10})$$

$$b_i = \frac{0.07780RT_{c,i}}{P_{c,i}} \quad (\text{A.11})$$

References

- [1] H. Zhang, J. Sheng, Optimization of horizontal well fracturing in shale gas reservoir based on stimulated reservoir volume, *J. Pet. Sci. Eng.* 190 (2020) 107059.
- [2] N. Nagel, M. Sanchez-Nagel, F. Zhang, X. Garcia, B. Lee, Coupled numerical evaluations of the geomechanical interactions between a hydraulic fracture stimulation and a natural fracture system in shale formations, *Rock Mech. Rock Eng.* 46 (3) (2013) 581–609.
- [3] K.S. Sing, Reporting physisorption data for gas/solid systems with special reference to the determination of surface area and porosity (recommendations 1984), *Pure Appl. Chem.* 57 (4) (1985) 603–619.
- [4] R.F. Sigal, Pore-size distributions for organic-shale-reservoir rocks from nuclear-magnetic-resonance spectra combined with adsorption measurements, *SPE J.* 20 (04) (2015) 824–830.
- [5] X. Liu, D. Zhang, A review of phase behavior simulation of hydrocarbons in confined space: implications for shale oil and shale gas, *J. Nat. Gas Sci. Eng.* 68 (2019) 102901.
- [6] K. Morishige, H. Fujii, M. Uga, D. Kinukawa, Capillary critical point of argon, nitrogen, oxygen, ethylene, and carbon dioxide in MCM-41, *Langmuir* 13 (13) (1997) 3494–3498.
- [7] P. Russo, M.R. Carrott, P. Carrott, Trends in the condensation/evaporation and adsorption enthalpies of volatile organic compounds on mesoporous silica materials, *Microporous Mesoporous Mater.* 151 (2012) 223–230.
- [8] S. Kittaka, M. Morimura, S. Ishimaru, A. Morino, K. Ueda, Effect of confinement on the fluid properties of ammonia in mesopores of MCM-41 and SBA-15, *Langmuir* 25 (3) (2009) 1718–1724.
- [9] S. Luo, H. Nasrabadi, J.L. Lutkenhaus, Effect of confinement on the bubble points of hydrocarbons in nanoporous media, *AIChE J.* 62 (5) (2016) 1772–1780.
- [10] S. Luo, J.L. Lutkenhaus, H. Nasrabadi, Confinement-induced supercriticality and phase equilibria of hydrocarbons in nanopores, *Langmuir* 32 (44) (2016) 11506–11513.
- [11] S. Luo, J.L. Lutkenhaus, H. Nasrabadi, Use of differential scanning calorimetry to study phase behavior of hydrocarbon mixtures in nano-scale porous media, *J. Pet. Sci. Eng.* 163 (2018) 731–738.
- [12] X. Qiu, S.P. Tan, M. Dejam, H. Adidharma, Isochoric measurement of the evaporation point of pure fluids in bulk and nanoporous media using differential scanning calorimetry, *PCCP* 22 (13) (2020) 7048–7057.
- [13] X. Qiu, S.P. Tan, M. Dejam, H. Adidharma, Experimental study on the criticality of a methane/ethane mixture confined in nanoporous media, *Langmuir* 35 (36) (2019) 11635–11642.
- [14] P. Zeigermann, M. Dvoyashkin, R. Valiullin, J. Kärger, Assessing the pore critical point of the confined fluid by diffusion measurement (2009).
- [15] M. Alfi, H. Nasrabadi, D. Banerjee, Experimental investigation of confinement effect on phase behavior of hexane, heptane and octane using lab-on-a-chip technology, *Fluid Phase Equilib.* 423 (2016) 25–33.
- [16] M. Alfi, H. Nasrabadi, D. Banerjee, Effect of confinement on bubble point temperature shift of hydrocarbon mixtures: experimental investigation using nanofluidic devices, in: *SPE Annual Technical Conference and Exhibition, OnePetro*, 2017.
- [17] Q. Yang, B. Jin, D. Banerjee, H. Nasrabadi, Direct visualization and molecular simulation of dewpoint pressure of a confined fluid in sub-10 nm slit pores, *Fuel* 235 (2019) 1216–1223.
- [18] Z. Sun, X. Li, W. Liu, T. Zhang, M. He, H. Nasrabadi, Molecular dynamics of methane flow behavior through realistic organic nanopores under geologic shale condition: pore size and kerogen types, *Chem. Eng. J.* 398 (2020) 124341.
- [19] B. Jin, R. Bi, H. Nasrabadi, Molecular simulation of the pore size distribution effect on phase behavior of methane confined in nanopores, *Fluid Phase Equilib.* 452 (2017) 94–102.
- [20] B. Jin, H. Nasrabadi, Phase behavior of multi-component hydrocarbon systems in nano-pores using gauge-GCMC molecular simulation, *Fluid Phase Equilib.* 425 (2016) 324–334.
- [21] R. Bi, H. Nasrabadi, Molecular simulation of the constant composition expansion experiment in shale multi-scale systems, *Fluid Phase Equilib.* 495 (2019) 59–68.
- [22] D.-Y. Peng, D.B. Robinson, A new two-constant equation of state, *Ind. Eng. Chem. Fundam.* 15 (1) (1976) 59–64.
- [23] B. Nojabaei, R.T. Johns, L. Chu, Effect of capillary pressure on phase behavior in tight rocks and shales, *SPE Reservoir Eval. Eng.* 16 (03) (2013) 281–289.
- [24] N. Siripatrachai, T. Ertekin, R. Johns, Compositional simulation of discrete fractures incorporating the effect of capillary pressure on phase behavior, in: *SPE Improved Oil Recovery Conference, OnePetro*, 2016.
- [25] L. Du, L. Chu, Understanding anomalous phase behavior in unconventional oil reservoirs, in: *SPE Canadian Unconventional Resources Conference, OnePetro*, 2012.
- [26] D.R. Sandoval, M.L. Michelsen, W. Yan, E.H. Stenby, VT-based phase envelope and flash calculations in the presence of capillary pressure, *Ind. Eng. Chem. Res.* 58 (13) (2019) 5291–5300.
- [27] D.R. Sandoval, W. Yan, M.L. Michelsen, E.H. Stenby, The phase envelope of multicomponent mixtures in the presence of a capillary pressure difference, *Ind. Eng. Chem. Res.* 55 (22) (2016) 6530–6538.
- [28] D.R. Sandoval Lemus, W. Yan, E.H. Stenby, Phase equilibrium in shale including porous media effects, in: *Abu Dhabi International Petroleum Exhibition & Conference, OnePetro*, 2019.
- [29] D.R. Sandoval, W. Yan, M.L. Michelsen, E.H. Stenby, Influence of adsorption and capillary pressure on phase equilibria inside shale reservoirs, *Energy Fuels* 32 (3) (2018) 2819–2833.
- [30] L. Jin, Y. Ma, A. Jamili, Investigating the effect of pore proximity on phase behavior and fluid properties in shale formations, in: *SPE Annual Technical Conference and Exhibition, OnePetro*, 2013.
- [31] X. Xiong, D. Devegowda, G. Michel, R.F. Sigal, F. Civan, A fully-coupled free and adsorptive phase transport model for shale gas reservoirs including non-darcy flow effects, in: *SPE Annual Technical Conference and Exhibition, OnePetro*, 2012.
- [32] N.S. Alharthy, T.N. Nguyen, T.W. Teklu, H. Kazemi, R.M. Graves, Multiphase compositional modeling in small-scale pores of unconventional shale reservoirs, in: *SPE Annual Technical Conference and Exhibition, OnePetro*, 2013.
- [33] S.I. Sandler, The generalized van der Waals partition function. I. Basic theory, *Fluid Phase Equilib.* 19 (3) (1985) 238–257.
- [34] K.-H. Lee, M. Lombardo, S. Sandler, The generalized van der Waals partition function. II. Application to the square-well fluid, *Fluid Phase Equilib.* 21 (3) (1985) 177–196.
- [35] L. Travalloni, M. Castier, F.W. Tavares, Phase equilibrium of fluids confined in porous media from an extended Peng–Robinson equation of state, *Fluid Phase Equilib.* 362 (2014) 335–341.
- [36] S. Luo, B. Jin, J.L. Lutkenhaus, H. Nasrabadi, A novel pore-size-dependent equation of state for modeling fluid phase behavior in nanopores, *Fluid Phase Equilib.* 498 (2019) 72–85.
- [37] L. Travalloni, M. Castier, F.W. Tavares, S.I. Sandler, Thermodynamic modeling of confined fluids using an extension of the generalized van der Waals theory, *Chem. Eng. Sci.* 65 (10) (2010) 3088–3099.

- [38] S. Luo, F. Chen, D. Zhou, H. Nasrabadi, Multiscale pressure/volume/temperature simulation of decreasing condensate/gas ratio at greater than dewpoint pressure in shale gas-condensate reservoirs, *SPE J.* (2021) 1–13.
- [39] V. Gaganis, N. Varotsis, Machine learning methods to speed up compositional reservoir simulation, in: *SPE Europec/EAGE Annual Conference*, OnePetro, 2012.
- [40] C. Temizel, C.H. Canbaz, O. Saracoglu, D. Putra, A. Baser, T. Erfando, S. Krishna, L. Saputelli, et al., Production forecasting in shale reservoirs using LSTM method in deep learning, in: *SPE/AAPG/SEG Unconventional Resources Technology Conference*, Unconventional Resources Technology Conference, 2020.
- [41] H.-H. Liu, J. Zhang, F. Liang, C. Temizel, M.A. Basri, R. Mesdour, Incorporation of physics into machine learning for production prediction from unconventional reservoirs: a brief review of the gray-box approach, *SPE Reservoir Eval. Eng.* (2021) 1–12.
- [42] M. Mehana, E. Guiltinan, V. Vesselinov, R. Middleton, J.D. Hyman, Q. Kang, H. Viswanathan, Machine-learning predictions of the shale wells performance, *J. Nat. Gas Sci. Eng.* 88 (2021) 103819.
- [43] Y. Wang, H. Liu, Y. Zhou, Development of a deep learning-based model for the entire production process of steam-assisted gravity drainage (SAGD), *Fuel* 287 (2021) 119565.
- [44] S. Gorell, J. Andrews, J. Browning, Rationalizing the concept of net-to-gross with respect to reservoir heterogeneity and flow behavior utilizing machine learning analyses, *SPE Eastern Regional Meeting*, OnePetro, 2019.
- [45] A. Almasov, M. Onur, A.C. Reynolds, Production optimization of the CO₂ huff-n-puff process in an unconventional reservoir using a machine learning based proxy, in: *SPE Improved Oil Recovery Conference*, OnePetro, 2020.
- [46] L. Cadei, G. Rossi, M. Montini, P. Fier, D. Milana, A. Corneo, L. Lancia, D. Lofreno, E. Purlalli, F. Carducci, et al., Machine learning advanced algorithm to enhance production optimization: an ann proxy modelling approach, in: *International Petroleum Technology Conference*, OnePetro, 2020.
- [47] I. Al Selaiti, C. Mata, L. Saputelli, D. Badmaev, Y. Alatrach, E. Rubio, R. Mohan, D. Quijada, Robust data driven well performance optimization assisted by machine learning techniques for natural flowing and gas-lift wells in Abu Dhabi, in: *SPE Annual Technical Conference and Exhibition*, OnePetro, 2020.
- [48] O. Sola-Aremu, An inferable machine learning approach to predicting PVT properties of niger delta crude oil using compositional data, in: *SPE Annual Technical Conference and Exhibition*, OnePetro, 2019.
- [49] S.D. Groven, C. Desgranges, J. Delhommele, Prediction of the boiling and critical points of polycyclic aromatic hydrocarbons via wang-landau simulations and machine learning, *Fluid Phase Equilib.* 484 (2019) 225–231.
- [50] K. Wang, J. Luo, Y. Wei, K. Wu, J. Li, Z. Chen, Practical application of machine learning on fast phase equilibrium calculations in compositional reservoir simulations, *J. Comput. Phys.* 401 (2020) 109013.
- [51] Y. Wu, S. Misra, C. Sondergeld, M. Curtis, J. Jernigen, Machine learning for locating organic matter and pores in scanning electron microscopy images of organic-rich shales, *Fuel* 253 (2019) 662–676.
- [52] R. Akkurt, T.T. Conroy, D. Psaila, A. Paxton, J. Low, P. Spaans, Accelerating and enhancing petrophysical analysis with machine learning: a case study of an automated system for well log outlier detection and reconstruction, *SPWLA 59th Annual Logging Symposium*, OnePetro, 2018.
- [53] Y. Wang, H. Liu, M. Guo, X. Shen, B. Han, Y. Zhou, Image recognition model based on deep learning for remaining oil recognition from visualization experiment, *Fuel* 291 (2021) 120216.
- [54] S. Mohanty, Estimation of vapour liquid equilibria for the system carbon dioxide–difluoromethane using artificial neural networks, *Int. J. Refrig.* 29 (2) (2006) 243–249.
- [55] R. Sharma, D. Singhal, R. Ghosh, A. Dwivedi, Potential applications of artificial neural networks to thermodynamics: vapor–liquid equilibrium predictions, *Comput. Chem. Eng.* 23 (3) (1999) 385–390.
- [56] R. Petersen, A. Fredenslund, P. Rasmussen, Artificial neural networks as a predictive tool for vapor–liquid equilibrium, *Comput. Chem. Eng.* 18 (1994) S63–S67.
- [57] V.D. Nguyen, R.R. Tan, Y. Brondial, T. Fuchino, Prediction of vapor–liquid equilibrium data for ternary systems using artificial neural networks, *Fluid Phase Equilib.* 254 (1–2) (2007) 188–197.
- [58] V. Gaganis, N. Varotsis, An integrated approach for rapid phase behavior calculations in compositional modeling, *J. Pet. Sci. Eng.* 118 (2014) 74–87.
- [59] C.J. Burges, A tutorial on support vector machines for pattern recognition, *Data Min. Knowl. Discov.* 2 (2) (1998) 121–167.
- [60] H.H. Rachford, J. Rice, Procedure for use of electronic digital computers in calculating flash vaporization hydrocarbon equilibrium, *J. Pet. Technol.* 4 (10) (1952) 19.
- [61] A. Kashinath, M.L. Szulczewski, A.H. Dogru, A fast algorithm for calculating isothermal phase behavior using machine learning, *Fluid Phase Equilib.* 465 (2018) 73–82.
- [62] M.E. Tipping, Sparse Bayesian learning and the relevance vector machine, *J. Mach. Learn. Res.* 1 (Jun) (2001) 211–244.
- [63] M. Mesbah, E. Soroush, V. Azari, M. Lee, A. Bahadori, S. Habibnia, Vapor liquid equilibrium prediction of carbon dioxide and hydrocarbon systems using LSSVM algorithm, *J. Supercrit. Fluids* 97 (2015) 256–267.
- [64] H.-Z. Huang, H.-K. Wang, Y.-F. Li, L. Zhang, Z. Liu, Support vector machine based estimation of remaining useful life: current research status and future trends, *J. Mech. Sci. Technol.* 29 (1) (2015) 151–163.
- [65] S. Wang, N. Sobacki, D. Ding, L. Zhu, Y.-S. Wu, Accelerating and stabilizing the vapor–liquid equilibrium (VLE) calculation in compositional simulation of unconventional reservoirs using deep learning based flash calculation, *Fuel* 253 (2019) 209–219.
- [66] C.R. Clarkson, N. Solano, R.M. Bustin, A. Bustin, G. Chalmers, L. He, Y.B. Melnichenko, A. Radliński, T.P. Blach, Pore structure characterization of North American shale gas reservoirs using USANS/SANS, gas adsorption, and mercury intrusion, *Fuel* 103 (2013) 606–616.
- [67] R. Bi, A. Firoozabadi, P.C. Myint, Efficient and robust phase-split computations in the internal energy, volume, and moles (UVN) space, *Fluid Phase Equilib.* 526 (2020) 112729.
- [68] R. Bi, A. Zidane, A. Firoozabadi, Efficient and robust stability analysis in the internal energy, volume, and moles (UVN) space, *Fluid Phase Equilib.* 512 (2020) 112468.
- [69] J.K. Singh, S.K. Kwak, Surface tension and vapor–liquid phase coexistence of confined square-well fluid, *J. Chem. Phys.* 126 (2) (2007) 024702.
- [70] G.A. Chapela, S.E. Martínez-Casas, C. Varea, Square well orthobaric densities via spinodal decomposition, *J. Chem. Phys.* 86 (10) (1987) 5683–5688.
- [71] D.D. Carley, Thermodynamic properties of a square-well fluid in the liquid and vapor regions, *J. Chem. Phys.* 78 (9) (1983) 5776–5781.
- [72] T.L. Hill, An Introduction to Statistical Thermodynamics, Courier Corporation, 1986.
- [73] R. Hoseini, R. Dawe, A parametric methodology in tuning the Peng–Robinson (PR) equation of state for gas condensate systems, *Pet. Sci. Technol.* 32 (6) (2014) 662–672.
- [74] X. Xu, H. Chen, C. Liu, C. Dang, Prediction of the binary interaction parameter of carbon dioxide/alkanes mixtures in the pseudocritical region, *ACS Omega* 4 (8) (2019) 13279–13294.
- [75] Ø. Fevang, K. Singh, C.H. Whitson, et al., Guidelines for choosing compositional and black-oil models for volatile oil and gas-condensate reservoirs, in: *SPE Annual Technical Conference and Exhibition*, Society of Petroleum Engineers, 2000.
- [76] Y.-K. Li, L.X. Nghiem, A. Siu, Phase behaviour computations for reservoir fluids: effect of pseudo-components on phase diagrams and simulation results, *J. Can. Pet. Technol.* 24 (06) (1985).
- [77] T. Newley, R. Merrill, Pseudocomponent selection for compositional simulation, *SPE Reservoir Eng.* 6 (04) (1991) 490–496.
- [78] S. Luo, J.L. Lutkenhaus, H. Nasrabadi, Multiscale fluid-phase-behavior simulation in shale reservoirs using a pore-size-dependent equation of state, *SPE Reservoir Eval. Eng.* 21 (04) (2018) 806–820.
- [79] M. Stein, Large sample properties of simulations using latin hypercube sampling, *Technometrics* 29 (2) (1987) 143–151.
- [80] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, *ICML*, 2010.
- [81] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
- [82] K. Abou-Moustafa, C. Szepesvári, An a priori exponential tail bound for k-folds cross-validation, *arXiv preprint arXiv:1706.05801* (2017).
- [83] B.G. Marcot, A.M. Hanea, What is an optimal value of k in k-fold cross-validation in discrete Bayesian network analysis? *Comput. Stat.* (2020) 1–23.
- [84] M. Raissi, P. Perdikaris, G. E. Karniadakis, Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations, *arXiv preprint arXiv:1711.10561* (2017).
- [85] H. Klie, B. Yan, A. Klie, Transfer learning for scalable optimization of unconventional field operations, in: *Unconventional Resources Technology Conference*, 20–22 July 2020, Unconventional Resources Technology Conference (URTEC), 2020, pp. 2315–2333.
- [86] P. Sarma, S. Kyriacou, M. Henning, P. Orland, G. Thakur, D. Sloss, Redistribution of steam injection in heavy oil reservoir management to improve EOR economics, powered by a unique integration of reservoir physics and machine learning, in: *SPE Latin America and Caribbean Petroleum Engineering Conference*, OnePetro, 2017.
- [87] F. Chollet, et al., Keras: the python deep learning library, *Astrophys. Source Code Library* (2018) ascl-1806.
- [88] M.L. Michelsen, W. Yan, E.H. Stenby, A comparative study of reduced-variables-based flash and conventional flash, *SPE J.* 18 (05) (2013) 952–959.