Accepted to Phys. Fluids 10.1063/5.0184562

Machine Learning Method for Shale Gas Adsorption Capacity Prediction and Key Influencing Factors Evaluation

Yu Zhou(周玉),[1] *Bo Hui*(惠波),[2] *Jinwen Shi*(师进文),[1] *Huaqiang Shi*(石华强),[2] *Dengwei Jing*(敬登伟)[1,a)]

[1]*State Key Laboratory of Multiphase Flow in Power Engineering & International Research Center for Renewable Energy, School of Energy and Power Engeering, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China*

[2]*Oil and Gas Technology Institute, PetroChina Changqing Oilfield Company, Xi'an, Shaanxi 710018, China*

[a)]Corresponding author: Tel.:+86-29-82668769, Email: dwjing@xjtu.edu.cn

Abstract: Shale gas plays a pivotal role in the global energy landscape, emphasizing the need for accurate shale gas-in-place (GIP) prediction to facilitate effective production planning. Adsorbed gas in shale, the primary form of gas storage under reservoir conditions, is a critical aspect of this prediction. In this study, a machine learning Gaussian Process Regression (GPR) model for methane adsorption prediction was established and validated using published experimental data. Five typical variables, i.e. total organic carbon (TOC), clay minerals, temperature, pressure, and moisture were considered which were derived from the Marine shale of the Longmaxi Formation in the Sichuan Basin through correlation analysis. The performance of the GPR model was compared with the widely used Extreme Gradient Boosting (XGBoost) model. It turned out that our GPR model had better accuracy for predicting methane adsorption in shale with an average relative error of less than 3%. Furthermore, a variance-based sensitivity analysis method in conjunction with kernel density estimation theory was employed to conduct a global sensitivity analysis, quantifying the nonlinear influence of each variable methane adsorption. The findings indicate that TOC is the most significant factor affecting methane adsorption, while clay minerals have a limited direct impact but can enhance their influence through interactions with other influencing factors. Finally, based on the GPR model, a GIP prediction method was proposed that eliminates the need for calculating the density of the adsorbed phase. These findings are expected to extend the shale gas reserve assessment methodologies and offer valuable insights for further exploring the adsorption mechanisms of shale gas.

Accepted to Phys. Fluids 10.1063/5.0184562

Keywords: Shale gas; shale gas-in-place; Gaussian process regression; Adsorption; Global sensitivity analysis;

1 Introduction

Shale gas, a methane-rich unconventional natural gas, significantly enhances global energy reserves.[1–3] In the U.S., the shale gas revolution, driven by long horizontal drilling and multistage hydraulic fracturing, has effectively addressed the energy gap.[4,5] China is rapidly progressing in shale gas exploration, notably achieving commercial exploitation of the Longmaxi Formation shale in the Sichuan Basin.[6] At present, shale gas is widely considered one of the promising alternatives to alleviate the global shortage of fossil energy.[7] Despite its pivotal importance, uncertainties persist in evaluating recoverable resources.[8] The primary source of uncertainty lies in gas adsorption assessment, directly impacting total GIP estimation and optimal production strategy formulation.

Within a shale gas reservoir, the predominant constituents comprise either free gas or adsorbed gas phases.[9,10] The adsorbed gas accounts for 43.4% to 61.1% of the total shale gas in the Longmaxi Formation of the Sichuan Basin, China, highlighting the crucial importance of accurately predicting adsorption capacity.[8] The adsorption capacity is influenced by shale reservoir conditions and properties, including pressure, temperature, moisture, mineral composition, and others.[11–15] To explore the factors affecting shale adsorption, researchers commonly employ isothermal adsorption experiments from various shale samples. Specifically, classical models such as Langmuir and Dubinin-Astakhov models will generally be used firstly to determine maximum adsorption capacity.[16,17] Subsequently, linear regressions between maximum adsorption capacity and shale properties such as TOC and mineral composition will be established. A successful linear regression suggests an influential role of the property in methane adsorption, while an unsuccessful regression implies a negligible effect.[13,16,18–20] However, this approach often fails to provide a quantitative characterization of the multifaceted factors that significantly affects the methane adsorption properties, thereby constraining the comprehensive understanding of shale gas adsorption mechanisms. Additionally, these factors may exhibit nonlinear effects on methane adsorption.

Shale gas formations, in contrast to coal bed methane, typically reside at greater

depths, ranging from 1000m to 3000m, and exhibit elevated pressure and temperature conditions, reaching up to 360K and 30Mpa.[21,22] To investigate shale gas adsorption mechanisms, extensive experiments have been conducted under varying pressure and temperature conditions. Most researchers [8,21–24] used classical models to estimate shale gas adsorption. These classical models fall into two categories: molecular layer models (including monolayer and multilayer adsorption models) and pore-filling models. While these traditional models effectively describe methane adsorption, they primarily focus on absolute adsorption amounts. However, experimental measurements yield the Gibbs excess adsorption amount, which significantly differs from absolute adsorption at high pressures but approximates it at low pressures.[25] To address this, the Gibbs equation is employed for the conversion of excess adsorption into absolute adsorption, necessitating knowledge of adsorbed phase density or volume. Unfortunately, direct measurement or even calculation of these quantities is currently unfeasible. The adsorbed phase density is often treated as a fitting parameter or estimated using empirical methods.[26] In certain instances, the fitted density value of the adsorbed phase may surpass the density of liquid methane at its boiling point, which is physically unreasonable.[27] Empirical methods often make the assumption that the volume of the adsorbed phase remains invariant, regardless of the adsorption quantity, or depends solely on temperature, a premise that may lack a robust physical foundation.[8] Consequently, the use of classical models to estimate methane adsorption may lead to a loss of physical accuracy due to the treatment of adsorbed phase density. Moreover, classical models exhibit inherent shortcomings, such as relevant foundational assumptions and inability to transcend the limitations associated with various shale types. Meanwhile, parameters for most traditional adsorption models are typically derived by fitting data from isothermal adsorption experiments, a process that is both time-consuming and labor-intensive. Consequently, these limitations diminish the classical models' applicability in extensive numerical frameworks used for predicting and optimizing shale gas production curves.

Recently, machine learning approaches have been considered as excellent alternatives for classical models when it comes to complex systems.[28–32] There are only a few studies employing machine learning for the prediction methane adsorption amount. Meng et al. [33]gathered published data from shale gas adsorption experiments to construct several common black-box machine learning models, which include extreme gradient boosting (XGBoost), artificial neural network (ANN), random forest (RF), and support vector machine (SVM). They utilized pressure, temperature, TOC,

and moisture as input variables, with excess adsorption amount of shale gas as the output. Multiple evaluation metrics were employed to assess the training and predictive performance of these models. The results demonstrated the superior predictive capabilities of all four machine learning models, with XGBoost exhibiting the best overall performance. In contrast, Nait Amar et al. [34] investigated two white-box machine learning techniques for methane adsorption prediction. They reported that the Gene Expression Programming technique outperformed the Group Method of Data Handling technique, achieving a correlation coefficient value of 0.9837 across all data points. However, it is worth noting that both works employed modeling data from different formations, which disregarded the inherent heterogeneity between formations, potentially limiting the generalization capabilities of machine learning models. While machine learning offers a potent tool for predicting methane adsorption amounts, the application of these models necessitates attention to detail, particularly concerning the data source and formation-specific characteristics.

In this study, a published databank from the Longmaxi Formation in the Sichuan Basin comprising 988 experimental adsorption datasets were utilized to estimate methane adsorption quantities. Employing a data-driven approach, we developed a high precision Gaussian Process Regression (GPR) machine learning model that obviates the need for calculating adsorption phase density or volume, enabling the prediction of methane excess adsorption. Evaluation metrics were applied to assess the training and prediction performance of the GPR model and compared with the widely used XGBoost model. Furthermore, a global sensitivity analysis using the GPR model was conducted to quantitatively evaluate the nonlinear impact of input parameters on methane adsorption. Finally, a GIP estimation model will be developed, which is more suitable for engineering and science needs, considering both shale reservoir conditions and properties. These findings are expected to extend the shale gas reserve assessment methodologies and offer valuable insights for further exploring the adsorption mechanisms of shale gas.

2 Data gathering and preprocessing

2.1 Data gathering

In this study, machine learning models will be employed to predict methane adsorption. The predictive performance of any model is inherently contingent on the

quality of the training dataset. Here a reliable databank consisting of 988 experimental measurements from published literature will be employed which comprises data points collected from the Longmaxi shale formation in the Sichuan Basin, China, as outlined in **Table 1** (the complete list of all the data in literature can be found in Supporting information). In these literatures, the adsorption amount and clay mineral content were determined by volumetric method and X-ray diffraction experiment, respectively. TOC content was determined by the carbon/sulfur analyzer. **Fig. 1** depicts the statistical properties of each parameter. The boxplot can provide useful statistical formation about minimum, lower quartile (Q_1), median (Q_2), mean, upper quartile (Q_3), and maximum values. Q1, marking the 25th percentile of the data, is the median of the lower half of the dataset, indicating that 25% of the data points are below this first quartile. Meanwhile, Q3, representing the 75th percentile of the data, is the median of the upper half of the dataset, signifying that 75% of the data points are below Q3, with 25% above.

Notably, the shale samples in the databank exhibit substantial heterogeneity, exemplified by the clay mineral content, which ranges from a minimum of 13.3% to a maximum of 64.3%, representing a nearly fivefold difference. Meanwhile, the maximum pressure within the databank reaches 50.84Mpa, aligning closely with actual shale gas reservoir formation pressures.

Table 1 Overview of Data Sources (detailed data are shown in Supporting information**)**

TOC (%)	Total clays (%)	Temperature (K)	Pressure (Mpa)	Moisture (%)	Gas adsorption (mmol/g)	references
2.93~3.66	27.2~31.3	303.15	0.88~14.98	3.34~3.96	0.0308~0.1591	Hu et al.[35]
2.59~4.79	13.3~31.1	318.15	0.44~9.11	0.47~2.05	0.0048~0.1903	Wang and Zhang[36]
0.45~4.13	21.0~58.0	293.15~373.15	0.34~11.36	1.65~1.96	0.0049~0.1500	Ji et al.[37]
2.80~4.33	24.9~34.8	318.15	0.40~15.64	0.64~0.82	0.0012~0.0653	Wang et al.[38]
1.21~3.24	20.6~64.3	308.15	1.88~20.00	0.21~3.23	0.0037~0.0815	Gao and Xiong[39]
3.66	16.0	303.15~353.15	0.10~22.25	1.04~4.22	0.0039~0.1177	Han et al.[40]
0.46~4.20	27.2~51.4	333.15	0.54~50.84	1.13~1.23	0.0006~0.0645	Qian et al.[41]
0.96~9.40	17.2~45.5	312.15	1.19~24.26	0.46~4.50	0.0046~0.1278	Yang et al.[27]

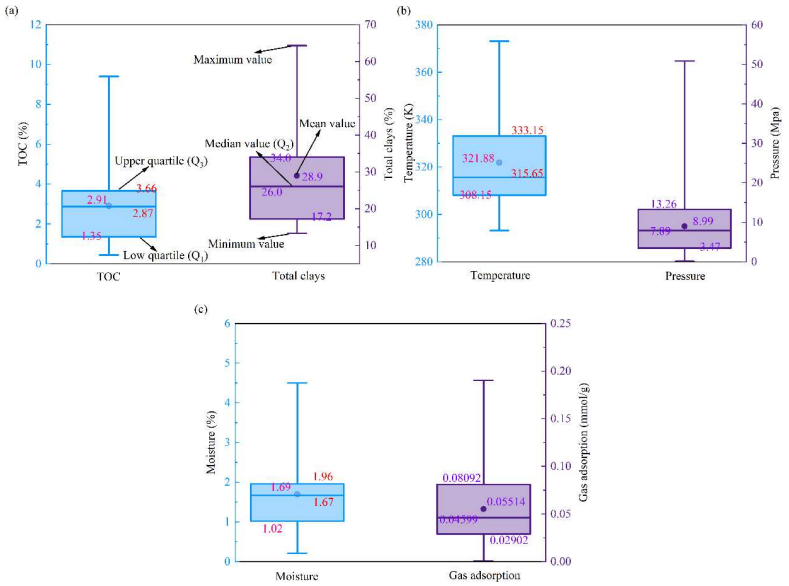

Fig. 1 Box plot for each parameter

2.2 Feature engineering

For machine learning algorithms, the selection of input parameters significantly impacts performance of the model. The adsorption capacity of methane in shale primarily relies on shale gas reservoir conditions and shale properties. In this study, five variables were collected, including pressure, temperature, moisture, TOC, and total clay minerals. However, potential high correlations among these variables could introduce redundancy and multicollinearity issues, increasing model complexity while diminishing accuracy and generalization capabilities. Therefore, we employed the Pearson correlation coefficient to identify and mitigate redundancy, ensuring data independence.

Pearson correlation coefficient is a classical correlation coefficient mainly used to characterize linear correlation, and its value is between -1 and 1. As the absolute magnitude of the Pearson correlation coefficient approaches 1, it indicates a high degree of correlation between the two variables. Specifically, the categorization of Pearson correlation (r) values is as follows: when r falls within the range of 0 to 0.35, it signifies weak correlations; within the range of 0.36 to 0.67, it denotes modest correlations; and when r is within the interval of 0.68 to 1.00, it signifies strong correlations.[42] Pearson

correlation coefficient is defined as follows:[43]

$$r\left(x_{oi},x_{oj}\right)=\frac{\sum_{k=1}^{n}\left(x_{oi,k}-x_{oi,ave}\right)\left(x_{oj,k}-x_{oj,ave}\right)}{\sqrt{\sum_{k=1}^{n}\left(x_{oi,k}-x_{oi,ave}\right)^{2}}\sqrt{\sum_{k=1}^{n}\left(x_{oj,k}-x_{oj,ave}\right)^{2}}}\ (i=1,2,3,4,5;\ j=1,2,3,4,5;\ n=988) \quad (1)$$

where x_{oi} and x_{oj} represent the ith and jth input parameters, respectively, $x_{oi,k}$ denotes the kth value of ith input parameter, and $x_{oi,ave}$ represents the average value of ith input. As indicated in **Fig. 2**, Pearson correlation coefficient among characteristics are all weak. TOC and total clay minerals have the largest Pearson correlation coefficient, which is only moderately correlated. It indicates that there may be no multicollinearity relationship among the five collected variables.

Although numerous variables affect methane adsorption in shale, not all of them are essential as inputs for machine learning models. Theoretically, incorporating more parameters could enhance the accuracy of predictions, but practically, capturing all variables is challenging. Different parameters require distinct instruments for measurement in the laboratory, each incurring varying costs. Meng et al.[33] and Menad Nait Amar et al. [34]have demonstrated the efficacy of using temperature, pressure, Total Organic Carbon (TOC), and water content as machine learning inputs for predicting methane adsorption. Notably, clay minerals, due to their large specific surface area, can serve as adsorption sites for methane.[44] Moreover, clay minerals do not exhibit a clear linear relationship with the other four input variables. Therefore, pressure, temperature, TOC, total clay minerals, and moisture are considered as input parameters to accurately predict methane adsorption amount.

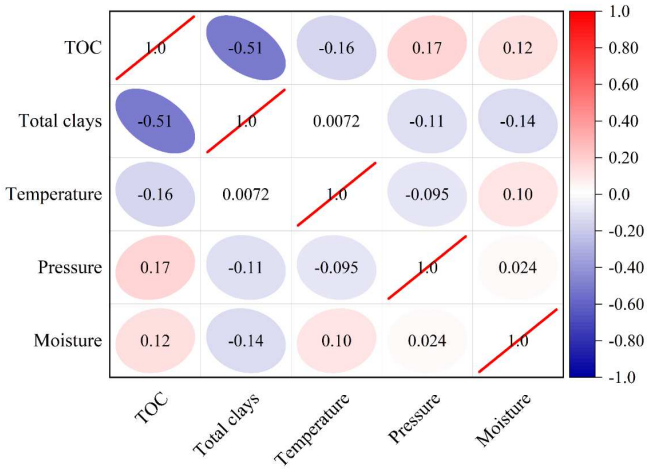

Fig. 2 Pearson correlation matrix of various variables

Furthermore, in order to eliminate the influence of dimension, Z-score normalization is applied to the experimental data:

$$x_i = \frac{x_{oi} - \mu_i}{\sigma_i} \tag{2}$$

$$y_a = \frac{y_{oa} - \mu_{oa}}{\sigma_{oa}} \tag{3}$$

where, the subscript i=1, 2, 3, 4, and 5 correspond to pressure, temperature, TOC, clay minerals and water content respectively; x_i represents the standardized value of these five input variables; x_{oi} represents the values of the five input variables before normalization; σ_i is the standard deviation of the input variable i; μ_i is the mean of the input variable i; y_a and y_{oa} represent methane adsorption before and after standardization, respectively. μ_{oa} and σ_{oa} represent the mean and standard deviation of methane adsorption, respectively.

Prior to model training, the databank underwent random partitioning into training and testing subsets, allocated at proportions of 80% and 20%, respectively, to mitigate overfitting.[45] The training subset was employed for model training, while the test subset served as a means to assess the model's generalization capacity.

Accepted to Phys. Fluids 10.1063/5.0184562

3 Methodology

3.1 Gaussian Process Regression (GPR)

GPR is a non-parametric Bayesian regression methodology, which is suitable for nonlinear regression problems with small samples and multiple dimensions.[46,47] In our work, a GPR model is developed to predict normalized methane adsorption on shales which can be expressed as,

$$y_a(\boldsymbol{x}) = f_a(\boldsymbol{x}) + \varepsilon_a \tag{4}$$

$$\boldsymbol{x} = [x_1, x_2, x_3, x_4, x_5] \tag{5}$$

where $\boldsymbol{x}$ represents the input vector; $y_a(\boldsymbol{x})$ and $f_a(\boldsymbol{x})$ correspond to the experimental output value of $\boldsymbol{x}$ vector and the predicted value of Gaussian process regression, respectively. ε_a is the Gaussian noise of y_a. Generally, it is assumed that ε_a conforms to a Gaussian distribution:

$$\varepsilon_a \sim N(0,\ \sigma_{an}^2) \tag{6}$$

where N is a normal distribution. Grounded in the tenets of Gaussian process regression, the Gaussian process prior distribution of $f_a(\boldsymbol{x})$ is given, which is expressed as:

$$f_a(x) \sim GP(0,\ \boldsymbol{K}_a) \tag{7}$$

Combing **Eqs.(4)-(7)**, the y_a prior distribution can be expressed as:

$$y_a(\boldsymbol{x}) \sim GP(0,\ \boldsymbol{K}_a + \sigma_{an}^2 \boldsymbol{I}_n) \tag{8}$$

where σ_{an} is the variance of noise ε_a; I_n is an identity matrix of order n; n represents the sample size of the training subsets; $\boldsymbol{K}_a$ is the covariance matrix, which is composed of covariance kernel functions. There are many possible options for the covariance function. A popular kernel is the Matern 5/2 covariance kernel (k_a) because of its stronger generalization ability and its wide application in practice.[46]

$$\boldsymbol{K}_a = \begin{bmatrix} k_a(\boldsymbol{x}_1, \boldsymbol{x}_1) & \cdots & k_a(\boldsymbol{x}_1, \boldsymbol{x}_n) \\ \vdots & k_a(\boldsymbol{x}_i, \boldsymbol{x}_j) & \vdots \\ k_a(\boldsymbol{x}_n, \boldsymbol{x}_1) & \cdots & k_a(\boldsymbol{x}_n, \boldsymbol{x}_n) \end{bmatrix} \tag{9}$$

$$k_a(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sigma_a^2 \left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right) \exp\left[-\frac{\sqrt{5}r}{\sigma_l}\right] \tag{10}$$

$$r=\sqrt{\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right)^T\left(\boldsymbol{x}_i-\boldsymbol{x}_j\right)} \tag{11}$$

where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ correspond to the i and j input vectors in the training subsets, respectively. σ_a^2 is the signal variance of the Matern 5/2 kernel function $k_a(\boldsymbol{x}_i, \boldsymbol{x}_j)$, and σ_l is the length scale of the ground variable of $k_a(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Therefore, with $\theta=(\sigma_{an}, \sigma_a, \sigma_l)$ as the hyperparameter in the Gaussian process regression, θ is optimized by obtaining θ partial derivative of the negative log-likelihood function $L(\theta)$ of the training sample, and then θ partial derivative according to the conjugate gradient method to obtain the minimum value $\theta_{\min}$, denoted by:

$$L(\theta)=-\frac{1}{2}\boldsymbol{Y}_{\mathrm{aT}}^{\mathrm{T}}\boldsymbol{Y}_{\mathrm{aT}}\boldsymbol{M}_a^{-1}-\frac{1}{2}\log\left|\boldsymbol{M}_a\right|+\frac{n}{2}\log 2\pi \tag{12}$$

$$\frac{\partial L(\theta)}{\partial\theta_i}=\frac{1}{2}\operatorname{trace}\left[\left(\beta\beta^{\mathrm{T}}\text{-}M^{-1}\right)\frac{\partial M}{\partial\theta_i}\right] \tag{13}$$

$$\boldsymbol{M}_a=\boldsymbol{K}_a+\sigma_{an}{}^2 I_{\mathrm{n}} \tag{14}$$

$$\beta=\boldsymbol{M}_a{}^{-1}\boldsymbol{Y}_{\mathrm{aT}} \tag{15}$$

$$\boldsymbol{Y}_{\mathrm{aT}}=\left[y_a\left(\boldsymbol{x}^{(1)}\right), y_a\left(\boldsymbol{x}^{(2)}\right), y_a\left(\boldsymbol{x}^{(3)}\right), \ldots, y_a\left(\boldsymbol{x}^{(n)}\right)\right]^{\mathrm{T}} \tag{16}$$

where M_{a} and β are intermediate variables; $\boldsymbol{Y}_{\mathrm{aT}}$ is the normalized methane adsorption vector corresponding to the training subsets; θ_1 is equal to $\sigma_{an}{}^2$. For the test input $\boldsymbol{x}^*$, which is the normalized TOC, total clay minerals, temperature, pressure and moisture, the joint distribution of the output value $f_a(\boldsymbol{x}^*)$ and $\boldsymbol{Y}_{\mathrm{aT}}$ also conforms to the Gaussian distribution, which are expressed as:

$$\begin{bmatrix}\boldsymbol{Y}_{\mathrm{aT}}\\ f_a\left(\boldsymbol{x}^*\right)\end{bmatrix} \sim N\left(0,\begin{bmatrix}\boldsymbol{K}_a\left(\boldsymbol{X},\boldsymbol{X}\right)+\sigma_{an}^2\boldsymbol{I}_n & \boldsymbol{K}_a\left(\boldsymbol{X},\boldsymbol{x}^*\right)\\ \boldsymbol{K}_a\left(\boldsymbol{x}^*,\boldsymbol{X}\right) & \boldsymbol{K}_a\left(\boldsymbol{x}^*,\boldsymbol{x}^*\right)\end{bmatrix}\right) \tag{17}$$

where $\boldsymbol{X}$ is the input corresponding to the training subsets. The probability distribution for predicted $f_a(\boldsymbol{x}^*)$ value is expressed by:

$$P\left(f_a\left(\boldsymbol{x}^*\right)\middle|\boldsymbol{Y}_{aT}\right)=N\left(\overline{f_a\left(\boldsymbol{x}^*\right)}, c_r\right) \tag{18}$$

where $\overline{f_a\left(\boldsymbol{x}^*\right)}$ and c_{r} are the predicted mean and variance of normalized methane adsorption, respectively. $\overline{f_a\left(\boldsymbol{x}^*\right)}$ is given as:

$$\overline{f_a\left(\boldsymbol{x}^*\right)}=\boldsymbol{K}_a\left(\boldsymbol{x}^*,\boldsymbol{X}\right)\left[\boldsymbol{K}_a\left(\boldsymbol{X},\boldsymbol{X}\right)+\sigma_{an}^2\boldsymbol{I}_{\mathrm{n}}\right]^{-1}\boldsymbol{Y}_{\mathrm{aT}} \tag{19}$$

$$c_r=\boldsymbol{K}_a\left(\boldsymbol{x}^*,\boldsymbol{x}^*\right)-\boldsymbol{K}_a\left(\boldsymbol{x}^*,\boldsymbol{X}\right)^T\times\left[\boldsymbol{K}_a\left(\boldsymbol{X},\boldsymbol{X}\right)+\sigma_{an}^2\boldsymbol{I}_{\mathrm{n}}\right]^{-1}\boldsymbol{K}_a\left(\boldsymbol{X},\boldsymbol{x}^*\right) \tag{20}$$

Therefore, for the test input $\boldsymbol{x}^*$, its prediction value $\overline{f_a\left(\boldsymbol{x}^*\right)}$ which can be obtained by **Eq.(19)**.

3.2 Extreme gradient boosting (XGBoost)

XGBoost, proposed by Chen and Guestrin,[48] is an integrated gradient algorithm based on decision trees. In contrast to traditional Gradient Boosting Decision Trees, XGBoost employs Newton's method to solve the loss function's extreme values and expands the Taylor series of the loss function to the second order. Additionally, regularization terms are incorporated into the loss function to mitigate overfitting, resulting in high-precision prediction and classification with improved computational efficiency. XGBoost operates on the principle of combining a set of weak learners through integration techniques to iteratively create a strong learner. During training, an initial learner is generated by fitting the entire training set, akin to the aforementioned weak learner. Subsequently, additional learners are sequentially added to fit the residuals of the preceding learners, enhancing learning efficiency. This process is iteratively repeated until predefined training criteria are met. Ultimately, the sum of predicted values from each learner is used as the final prediction. XGBoost algorithm specific derivation process, reference.[49,50]

3.3 Evaluation metrics systems

To statistically assess the model's performance and accuracy, two metric systems are employed: Root Mean Squared Error and Mean Absolute Percentage Error, defined as follows:

1. Root Mean Squared Error:

$$RMES = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\overline{f_a\left(\boldsymbol{x}^*\right)}_{i,pred} - y_{i,real}\right)^2} \tag{21}$$

2. Mean Absolute Percentage Error:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{i,real} - \overline{f_a\left(\boldsymbol{x}^*\right)}_{i,pred}}{y_{i,real}}\right| \times 100\% \tag{22}$$

where $\overline{f_a\left(\boldsymbol{x}^*\right)}_{i,pred}$ and $y_{i,real}$ are predicted and measured values of ith data, respectively.

Accepted to Phys. Fluids 10.1063/5.0184562

4. Results and discussion

4.1 Machine learning models training and prediction

To achieve accurate predictions of methane adsorption capacity in shale, optimization of GPR and XGBoost models is performed using a training set comprising 790 data samples. The remaining 198 data samples are designated for testing. Each data sample consists of five input variables (pressure, temperature, TOC, total clay minerals, and moisture) and one output variable (methane adsorption amount). To ensure consistency, the adsorption data with different units are standardized. Cross-verification of the two algorithms enhances their generalization capabilities, resulting in the determination of optimal hyperparameters (shown in **Table 2**). For the XGBoost algorithm, hyperparameter ranges align with commonly used values in prior research.[33,50] GPR algorithm hyperparameters are directly calculated from the training subset through the conjugate gradient method, as previously mentioned.

Table 2 Optimized hyperparameters of the GPR and XGBoost model

Model	Hyperparameter	value
GPR	the variance of noise(σ_{an})	0.0267
	the signal variance(σ_a)	4.9053
	the length scale of kernel(σ_l)	1.4407
XGBoost	Num estimators	50.0
	max depth	5.0
	learning rate	0.5
	min child weight	1.0
	col sample by tree	1.0

In the realm of machine learning algorithm optimization, the principle of 'no free lunch' prevails, highlighting that the suitability of each algorithm is contingent upon the unique characteristics of the data sets in question. In other words, not all machine learning algorithms can accurately predict methane adsorption amount. In assessing the accuracy of machine learning predictions, standard statistical indices such as Root Mean Square Error (*RMSE*) and Mean Absolute Percentage Error (*MAPE*) are routinely employed. These indices are instrumental in evaluating the performance of GPR and XGBoost models, and the results are shown in **Table 3**.

Accepted to Phys. Fluids 10.1063/5.0184562

Table 3 Evaluation metrics of the GPR and XGBoost models

	Training data		Testing data	
	RMES (mmol/g)	*MAPE* (%)	*RMES* (mmol/g)	*MAPE* (%)
GPR	0.0003	0.3622	0.0015	1.5116
XGBoost	0.0012	2.4140	0.0038	4.7839

As shown by this table, both the GPR and XGBoost algorithms demonstrate accurate predictions of methane adsorption quantities, with *MAPE* values below 5% for both the training and test subsets. However, a comparative analysis shows that the GPR model consistently exhibits lower *RMSE* and *MAPE* values than the XGBoost model in both the training and test subsets. In general, a smaller *RMES* value signifies better predictive performance. For the training set, the *RMES* values corresponding to GPR and XGBoost algorithms are small, which indicates that both algorithms can capture the structure and rules of the data. The *MAPE* of the XGBoost algorithm (4.7839) is nearly three times as high as that of the GPR model (1.5116) for the test subset. This discrepancy suggests that the GPR algorithm excels in robustness compared to the XGBoost algorithm for predicting methane adsorption quantities, possibly due to its suitability for small-sample data.

Most evaluation index systems can only reflect the overall effect of machine learning training and testing, and the *RMSE* and *MAPE* used here are the same. In order to obtain the training and prediction effect of each data point, the cross graph of GPR and XGboost algorithms is plotted, as shown in **Fig. 3**. In each plot, the X-axis represents experimentally measured adsorption quantities, while the Y-axis displays adsorption quantities predicted by the models. The line with a slope of 1 represents 100% agreement. If the data points are closer to the slope 1, it means that the training and prediction effect of the algorithm is better. It is obvious that the training and prediction results of GPR algorithm are better than XGboost in both training and test subsets. Therefore, **Fig. 3** further demonstrates the high capability of GPR algorithm for predicting methane adsorption, especially when encountering unsee data.

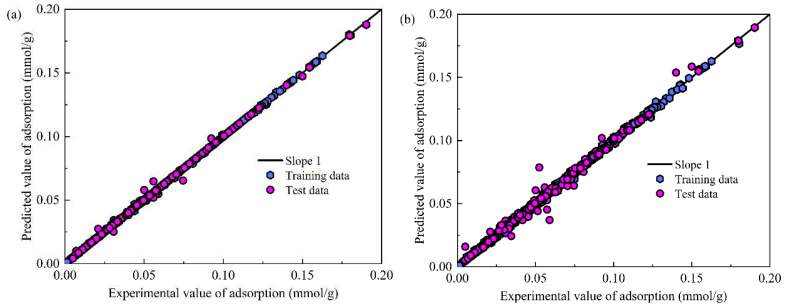

Fig. 3 Comparisons between model results and measured ones: (a) GPR; (b) XGBoost

Fig. 4 illustrates the performance of the GPR model in predicting methane adsorption capacity, indicating a high level of agreement with experimental values for both training and testing subsets. Thus, once the reservoir pressure, temperature, TOC, clay minerals, and moisture are known, the established GPR model can be used to accurately predict the methane adsorption capacity. Although both GPR and XGBoost models can accurately predict methane adsorption, the GPR model is more recommended. This preference for the GPR algorithm is rooted in its superior predictive performance, as previously noted. Additionally, GPR exhibits a smaller hyperparameter set, aligning with the Occam's razor principle advocating for simplicity when accuracy is comparable.

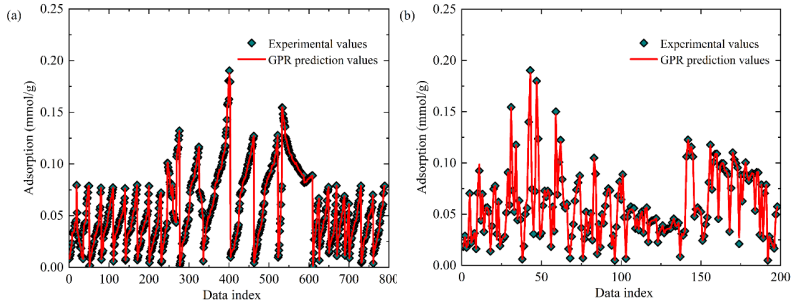

Fig. 4 Performance of the GPR models (a) Training data (b) Testing data

4.2 Global sensitivity analysis

As a further step, a quantitative investigation of the influence of input parameters on methane adsorption has been performed using variance-based sensitivity analysis

(VBSA) with the trained GPR model. In contrast to local sensitivity analysis methods, such as single-factor analysis, VBSA is a comprehensive method for conducting global sensitivity analysis that provides first-order and total-effect indices, facilitating a quantitative assessment of variable interactions. Unlike conventional global sensitivity analysis methods like experimental design, VBSA employs a random algorithm, such as the Monte Carlo method, to sample and analyze the entire parameter variation range. As a result, VBSA is particularly well-suited for nonlinear models, especially in cases where no explicit expression is available. These advantages have led to successful applications of VBSA in diverse fields, including petroleum engineering, earth science, and the chemical industry.[51–53]

When employing the VBSA method for sensitivity analysis, it is imperative to begin by defining the range and probability distribution of the variables. Utilizing kernel density estimation (KDE) theory, the probability density distribution functions for various characteristics are derived from the collected dataset. KDE theory revolves around the application of a smooth peak function, known as the "kernel," to approximate the distribution of sample points, thereby simulating the actual sample distribution. For an in-depth exploration of KDE theory, one can refer to the work of Parzen[54]. The probability density functions for each input parameter, obtained through KDE theory, are depicted in **Fig. 5**. Overall, these derived probability density functions closely align with the actual distribution of the sample data. This observation suggests that the VBSA method can serve as an effective quantitative tool for assessing the impact of each characteristic on methane adsorption. Furthermore, in conjunction with the acquired probability density function, an input vector, denoted as $\boldsymbol{x}$, is generated through sampling using the Monte Carlo method. Subsequently, the trained GPR model is employed to predict the corresponding output, $\overline{f_a\left(\boldsymbol{x}^*\right)}$.

After multiple sampling, sensitivity indices are calculated using the VBSA method based on the GPR model's prediction outcomes. The computation of sensitivity indices is closely tied to the sample size; a larger sample size yields more accurate results but increases computational demands. Hence, a convergence test was conducted to determine an optimal number of samples. **Fig. 6** displays the calculated first-order and total sensitivity indices at different sampling frequencies. It is evident from **Fig. 6** that once the sampling frequency surpasses 1.3×10^6 and is further increased, the changes in the first-order and total sensitivity indices become negligible, signifying that the convergence criterion is met at a sampling frequency of 1.3×10^6.

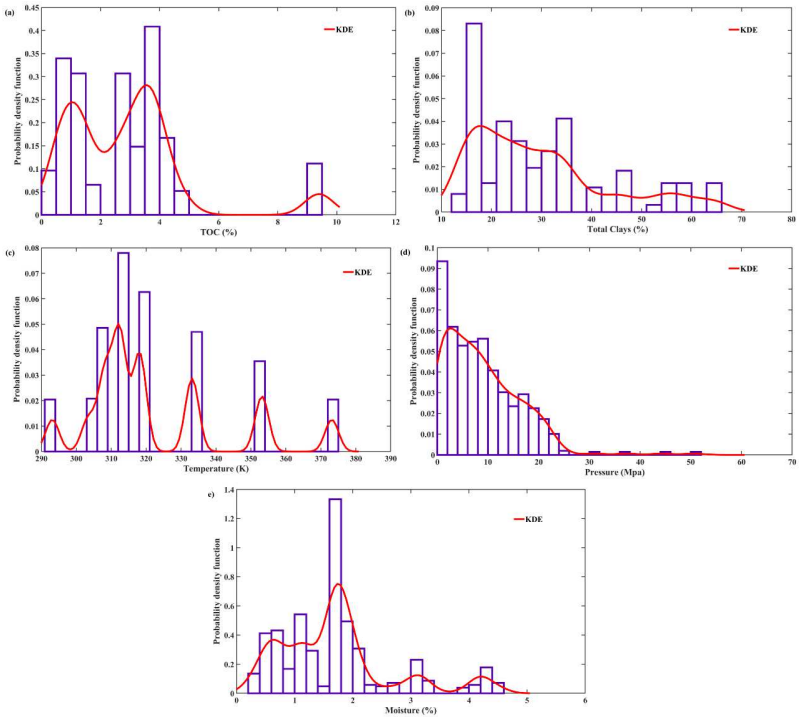

Fig. 5 Probability density distribution of different variables：

（a）TOC；（b）Total Clays；（c）Temperature；（d）Pressure；（e）Moisture

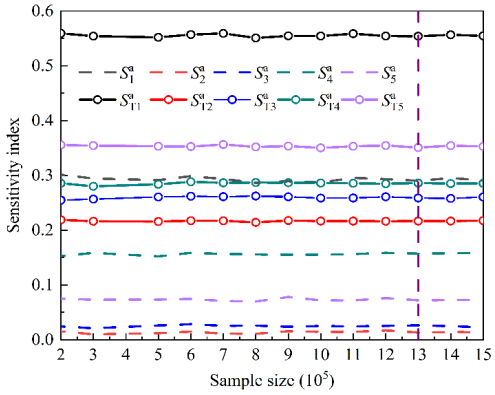

Fig. 6 Estimates of the first-order and total-effect indices of different variables as a function of sample size

Fig. 7 shows the first-order and total-effect indices for each input parameter at the

sample size of 1.3×10^6. The first-order index quantifies the isolated influence of an individual parameter on the GPR model output, while the total-effect index assesses the cumulative impact of an individual parameter along with other parameters on the GPR model output. The disparity between these indices reflects the impact of parameter interactions on the model output, referred to as "different parameters with similar effects". Therefore, the total-effect index can be emphatically analyzed, mainly because the total-effect index includes the first-order index and the higher-order index.

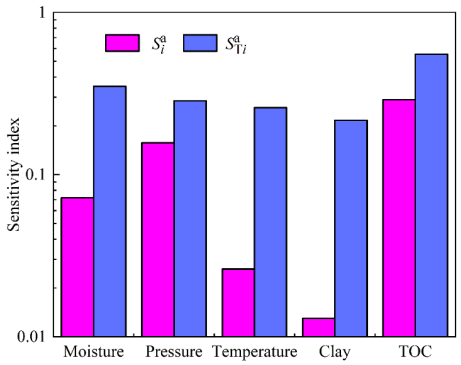

Fig. 7 First-order and total-effect indices of different variables

According to **Fig. 7**, TOC is the most important parameter affecting shale gas adsorption, followed by moisture, pressure, temperature and clay minerals from the total-effect index. These results align with prior research, emphasizing the dominant role of TOC in governing shale methane adsorption.[55] It is assumed that the abundant nano-scale pores (with pore widths <2 nm) found in shale samples offer a substantial quantity of adsorption sites for methane adsorption.[25] The adsorption affinity of non-polar methane gas is observed to be notably higher for hydrophobic organic matter as compared to hydrophilic inorganic minerals. The impact of water on adsorption primarily stems from its competitive adsorption with methane, particularly on hydrophilic clay minerals.[27,56] Moreover, at higher water content levels, water molecules obstruct the pore space, leading to a substantial reduction in adsorption capacity.[57] Temperature and pressure are two indexes that have not been quantitatively evaluated in previous studies. Nevertheless, the VBSA method has quantitatively evaluated their impact on methane adsorption, revealing that temperature and pressure exert a substantial influence on methane adsorption. It suggests a potential variation in the adsorption mechanism between deep and shallow shale gas reservoirs.

The impact of clay minerals on methane adsorption has been a subject of

controversy.[27,55,58] Hu and Mischo[55] analyzed four shale samples from the Longmaxi Formation, finding no straightforward linear relationship between clay mineral content and maximum adsorption capacity. In contrast, Bi et al.[58] conducted experiments that indicated a linear correlation between clay minerals and maximum adsorption capacity. **Fig. 7** illustrates that clay minerals exhibit a low first-order index but a high total-order index, implying their influence on methane adsorption capacity. This influence arises from their substantial specific surface area, providing ample adsorption sites for methane. It is worth mentioning that the influence of clay minerals on methane adsorption is mainly through the interaction with other influencing factors. For example, clay minerals have strong water absorption characteristics and by adsorbing some water, it can reduce the influence of water on methane adsorption. Therefore, for shales in a dry state with a low TOC content, clay minerals usually have a strong adsorption capacity for methane.[59] In traditional methods like linear regression, the interaction between clay minerals and other factors affecting methane adsorption is often overlooked, potentially resulting in a significant underestimation of the impact of clay minerals on methane adsorption. Meanwhile, Assessing the influence of clay minerals on methane adsorption from a nonlinear perspective aligns better with the physical significance of the phenomenon.

4.3 Accurate shale gas-in-place prediction from GPR

One of the primary objectives in estimating methane adsorption is to predict the GIP, which can be expressed as: [60]

$$GIP = n_{free} + n_a \tag{23}$$

where n_{free} and n_a are the amount of free gas amount and absolute adsorption gas amount, respectively. As the data employed in the GPR model represents the excess adsorption amount directly measured in the laboratory, it is necessary to transform the absolute adsorption amount using the Gibbs equation, which is expressed as:

$$n_a = n_e + V_a \rho_g \tag{24}$$

where n_e is the excess adsorption amount, mmol/g; V_a is the volume of the adsorbed phase, cm^3/g. Combining **Eq.(23)**, the GIP can be calculated as shown:

$$GIP = n_{free} + n_e + V_a \rho_g = n_e + \left(1 - S_w\right) V_p \rho_g \tag{25}$$

where V_p is the pore volume; S_w represents the water saturation, which can be calculated using the following formula:[27]

$$S_w = m\rho_{grain}\frac{1-\varphi}{\varphi} \tag{26}$$

where m is the moisture, %; φ is the porosity; ρ_{grain} is the grain density, g/cm^3. Therefore, Combining **Eq.(19)** and **Eq.(26)**, **Eq. (25)** can be modified to the following expression:

$$GIP = n_{free} + n_e + V_a\rho_g = \overline{f_a\left(\boldsymbol{x}^*\right)} + \left(1 - m\rho_{grain}\frac{1-\varphi}{\varphi}\right)V_p\rho_g \tag{27}$$

The method for calculating GIP, which omits the need for considering adsorbed phase density or volume, distinguishes itself from prior approaches [61,62] and enhances the physical interpretability of GIP calculations. Simultaneously, it is imperative to emphasize that the precision of GIP estimation through this methodology predominantly hinges on the accuracy of the projected excess adsorption, as opposed to the absolute adsorption. Some scholars [8] propose that it's possible to calculate GIP content solely through laboratory measurements of excess adsorption without relying on an adsorption model. However, in practical applications, conducting adsorption experiments for sampling is a time-consuming process. The use of the GPR model for predicting methane excess adsorption requires obtaining parameters such as TOC, clay minerals, temperature, pressure, and moisture, which can be directly derived from logging curves. Then, combined with the pore structure parameters, GIP can be predicted. This approach significantly reduces the required time for GIP estimation.

The Eq.(27) underscores that the precision for predicting shale gas-in-place (GIP) hinges largely on the accuracy of GPR in forecasting methane's excess adsorption in shale under varying temperatures and pressures. Consequently, the accuracy of GPR model to predict methane's excess adsorption directly correlates with reliable GIP estimations. Fig. 8 illustrates the GPR model's efficacy in predicting methane's excess adsorption in shale under various temperatures and pressures. The average error between the predicted excess adsorption value of GPR model and the experimental value is 3.51%, less than 5%. This capability suggests that the GPR model provides a more accurate prediction of methane's excess adsorption, which in turn implies a high level of accuracy in estimating shale gas-in-place (GIP) with GPR model. It is worth mentioning that when excess adsorption is predicted by the GPR model from Fig. 8, the input TOC content is 3.66%, the clay mineral content is 16.00%, and the moisture is about 4.22%.

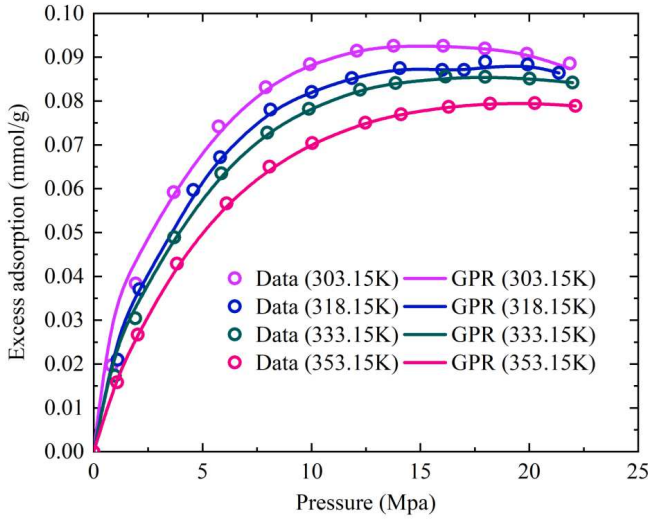

Fig. 8 Comparison between the measured methane excess adsorption and the predictions of GPR model under different temperatures and pressures (the data from Han et al.[40])

4.4 Geological application of GIP estimation

The Longmaxi shale in the southern Sichuan Basin is regarded as one of the most promising shale gas reservoirs in China and has successfully undergone commercial exploitation. Consequently, accurately estimating the GIP of the Longmaxi shale is of paramount importance. The shale of Longmaxi Formation has a very large span and is usually a zoned block when it is developed. We take the shale of Longmaxi Formation in Fuling Block as an example to show the reliability and accuracy of the GPR method in predicting GIP. The shale in the Longmaxi Formation in the Fuling area is typically found at depths ranging from 2000 to 3000 meters, with a nominal surface temperature of 15 ℃, a geothermal gradient of 27.3℃/km, and a pressure gradient of 15 Mpa/km. The shale rock exhibits an average porosity of 4.5% and a skeletal density of 2.62 g/cm³. Additionally, it possesses an average TOC content of 2.66% and total clays comprising 32.7% of its composition. The average water saturation in this shale formation is 32.3% [63].

It states that based on the geological data and burial depth provided, predictions have been made regarding the excess adsorbed gas and GIP for the Longmaxi shale

Formation in the Fuling area. **Fig. 9** displays the GIP prediction based on the GPR model. As burial depth increases, excess adsorption initially increases, peaks at a depth of 1200 meters, and then decreases. However, the GIP content consistently rises with increasing burial depth. This phenomenon is primarily attributed to the fact that shale reservoirs exhibit absolute adsorption rather than excess adsorption. Typically, the absolute adsorption amount in middle to deep shale formations is significantly greater than the excess adsorption amount. Moreover, as burial depth increases, free gas content increases accordingly. These findings suggest the great potential for substantial reserves in the middle to deep shale sections of the Longmaxi Formation.

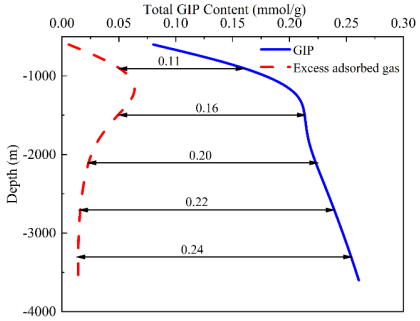

Fig. 9 Relationship between total GIP content and depth

Moreover, in order to quantitatively analyze the prediction accuracy of this method in Fuling block of Longmaxi Formation. The Well Jiaoye 1 in the Fuling shale gas field approximated the average total gas content of the shale reservoir (2377-2415m High-quality shale gas layer) as 6.03m^3/t (0.269mmol/g) using logging curves.[63] The GIP calculated via the GPR model, is 0.231mmol/g, resulting in a 14% calculation error. This margin is considered acceptable in engineering contexts, given that the GPR model's input parameters for simplicity, such as Total Organic Carbon (TOC) and clay minerals, represent average values across the Fuling block. Therefore, it is recommended to use GPR model to predict GIP.

5 Conclusions

In this study, the GPR machine learning algorithm is established for predicting the adsorption capacity of the Longmaxi shale formation in the Sichuan Basin. Subsequently, global sensitivity analysis for each input factor is conducted using a variance-based method combined with kernel density estimation theory. Lastly, leveraging the GPR algorithm, a GIP prediction method is proposed that eliminates the

need to calculate the density of the adsorbed phase. The key conclusions are as follows:

(1) Complex nonlinear relationships exist among TOC, clay minerals, temperature, pressure, and moisture, all of which can serve as direct machine learning inputs.

(2) For the same shale reservoir, both the GPR and XGBoost algorithms effectively predict methane's excess adsorption in shale. Statistically, the GPR algorithm is better suited for this prediction.

(3) TOC is the most influential factor of methane adsorption in shale. The effect of clays minerals on methane adsorption is mainly through interaction with other influencing factors.

(4) Utilizing the GPR model, this method accurately predicts the real GIP reserves in deep in-situ temperature and pressure shale formations.

Abbreviations

GIP:	Gas-in-place
GPR:	Gaussian Process Regression
XGBoost:	Extreme Gradient Boosting
TOC:	Total Organic Carbon

CRediT authorship contribution statement

Yu Zhou: Conceptualization, Methodology, Software, Writing – original draft. **Bo Hui**: Formal analysis, Visualization, Data curation. **Jinwen Shi**: Investigation, Data curation, Software. **Huaqiang Shi**: Validation, Supervision. **Dengwei Jing**: Formal analysis, Writing – Review & editing, Supervision.

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

Acknowledgement

The authors gratefully acknowledge the financial support of the National Natural Science Foundation of China (nos. 52025061 and 51888103). This work was also supported by the Innovative Talent Promotion Plan of Shaanxi Province-Scientific and Technological Innovation Team (no. 2023-CX-TD-25) and the Zhuhai Innovation and Entrepreneurship Team Project (2120004000225) "Key Technologies and Industrialization of Solar Powered Multi-Energy Conversion and Complementary Integrated Electricity, Heating and Hydrogen Energy System".

Supplementary data

The data that support the findings of this study are available from Supporting information.

References

[1] L. Chen, L. Zuo, Z. Jiang, S. Jiang, K. Liu, J. Tan, and L. Zhang, "Mechanisms of shale gas adsorption: Evidence from thermodynamics and kinetics study of methane adsorption on shale," Chem Eng J **361**, 559–570 (2019).

[2] W. Zhang, W. Chen, T. Wang, and Y. Yuan, "A self-similarity mathematical model of carbon isotopic flow fractionation during shale gas desorption," Phys. Fluids **31**(11), 112005 (2019).

[3] J. Gu, G. Liu, F. Gao, Y. Hu, and D. Ye, "Multiple seepage model and gas production simulation of shale with power law fracture distribution," Phys. Fluids **35**(2), 022010 (2023).

[4] R.S. Middleton, R. Gupta, J.D. Hyman, and H.S. Viswanathan, "The shale gas revolution: Barriers, sustainability, and emerging opportunities," Appl. Energy **199**, 88–95 (2017).

[5] Y. Xia, J. Goral, H. Huang, I. Miskovic, P. Meakin, and M. Deo, "Many-body dissipative particle dynamics modeling of fluid flow in fine-grained nanoporous shales," Phys. Fluids **29**(5), 056601 (2017).

[6] X. Ma, and J. Xie, "The progress and prospects of shale gas exploration and development in southern Sichuan Basin, SW China," Pet. Explor. Dev. **45**(1), 172–182 (2018).

[7] G. Sheng, H. Deng, H. Zhao, Z. Rui, T. Hu, J. Liu, and M. Lu, "Life cycle integrated flow simulation for hydraulic fracturing horizontal wells in shale gas reservoirs," Phys. Fluids **35**(10), 103318 (2023).

[8] X. Tang, N. Ripepi, N.P. Stadie, L. Yu, and M.R. Hall, "A dual-site Langmuir equation

589 for accurate estimation of high pressure deep shale gas resources," Fuel **185**, 10–17
590 (2016).
591 [9] H. Nie, L. Pei I, D. Wei, D. Jianghui, S. Chuanxiang, L. Mi, W. Jin, D. Wei, L. Peixian
592 Zhang and Donghui, and S. Haikun, "Enrichment characteristics and exploration
593 directions of deep shale gas of Ordovician-Silurian in the Sichuan Basin and its
594 surrounding areas, China," China Pet. Explor. Dev. **49**(4), 744–757 (2022).
595 [10] L. Germanou, M.T. Ho, Y. Zhang, and L. Wu, "Shale gas permeability upscaling from
596 the pore-scale," Phys. Fluids **32**(10), 102012 (2020).
597 [11] D.J.K. Ross, and R. Marc Bustin, "The importance of shale composition and pore
598 structure upon gas storage potential of shale gas reservoirs," Mar. Pet. Geol. **26**(6), 916–
599 927 (2009).
600 [12] J. Zou, R. Rezaee, Q. Xie, L. You, K. Liu, and A. Saeedi, "Investigation of moisture
601 effect on methane adsorption capacity of shale samples," Fuel **232**, 323–332 (2018).
602 [13] S. Zhou, H. Xue, Y. Ning, W. Guo, and Q. Zhang, "Experimental study of
603 supercritical methane adsorption in Longmaxi shale: Insights into the density of
604 adsorbed methane," Fuel **211**, 140–148 (2018).
605 [14] W. Shen, X. Li, T. Ma, J. Cai, X. Lu, and S. Zhou, "High-pressure methane adsorption
606 behavior on deep shales: Experiments and modeling," Phys. Fluids **33**(6), 063103
607 (2021).
608 [15] J. Deng, Q. Zhang, L. Zhang, Z. Lyu, Y. Rong, and H. Song, "Investigation on the
609 adsorption properties and adsorption layer thickness during CH4 flow driven by
610 pressure gradient in nano-slits," Phys. Fluids **35**(1), 016104 (2023).
611 [16] J. Zou, R. Rezaee, Q. Xie, and L. You, "Characterization of the combined effect of
612 high temperature and moisture on methane adsorption in shale gas reservoirs," J. Pet.
613 Sci. Eng. **182**, 106353 (2019).
614 [17] F. Yang, C. Xie, S. Xu, Z. Ning, and B.M. Krooss, "Supercritical Methane Sorption
615 on Organic-Rich Shales over a Wide Temperature Range," Energy Fuels **31**(12),
616 13427–13438 (2017).
617 [18] M. Gasparik, P. Bertier, Y. Gensterblum, A. Ghanizadeh, B.M. Krooss, and R. Littke,
618 "Geological controls on the methane storage capacity in organic-rich shales," Int. J.
619 Coal Geol. **123**, 34–51 (2014).
620 [19] B. Yang, Y. Kang, L. You, X. Li, and Q. Chen, "Measurement of the surface diffusion
621 coefficient for adsorbed gas in the fine mesopores and micropores of shale organic
622 matter," Fuel **181**, 793–804 (2016).
623 [20] Y. Ma, J. Pan, D. Fu, Y. Wang, Y. Wang, D. Chang, L. Cheng, and W. Fu, "The organic
624 geochemistry, pore structure and methane adsorption/storage capacity of lacustrine
625 shales from the Cretaceous Madongshan Formation, Liupanshan Basin, China," J. Nat.
626 Gas Sci. Eng. **96**, 104287 (2021).
627 [21] X. Tang, N. Ripepi, K. Luxbacher, and E. Pitcher, "Adsorption Models for Methane
628 in Shales: Review, Comparison, and Application," Energy Fuels **31**(10), 10787–10801
629 (2017).
630 [22] J. Li, Z. Chen, K. Wu, K. Wang, J. Luo, D. Feng, S. Qu, and X. Li, "A multi-site
631 model to determine supercritical methane adsorption in energetically heterogeneous
632 shales," Chem. Eng. J. **349**, 438–455 (2018).

[23] S. Zhou, D. Zhang, H. Wang, and X. Li, "A modified BET equation to investigate supercritical methane adsorption mechanisms in shale," Mar. Pet. Geol. **105**, 284–292 (2019).
[24] J. Li, B. Li, C. Ren, Y. Zhang, and B. Wang, "An adsorption model for evaluating methane adsorption capacity in shale under various pressures and moisture," J. Nat. Gas Sci. Eng. **81**, 103426 (2020).
[25] H. Tian, T. Li, T. Zhang, and X. Xiao, "Characterization of methane adsorption on overmature Lower Silurian-Upper Ordovician shales in Sichuan Basin, southwest China: Experimental results and geological implications," Int. J. Coal Geol. **156**, 36–49 (2016).
[26] W. Ren, G. Li, S. Tian, M. Sheng, and L. Geng, "Adsorption and Surface Diffusion of Supercritical Methane in Shale," Ind. Eng. Chem. Res. **56**(12), 3446–3455 (2017).
[27] F. Yang, C. Xie, Z. Ning, and B.M. Krooss, "High-pressure methane sorption on dry and moisture-equilibrated shales," Energy Fuels **31**(1), 482–492 (2017).
[28] X. Zheng, and X. Deng, "State-of-Health Prediction for Lithium-Ion Batteries with Multiple Gaussian Process Regression Model," IEEE Access **7**, 150383–150394 (2019).
[29] D. Kang, X. Wang, X. Zheng, and Y.P. Zhao, "Predicting the components and types of kerogen in shale by combining machine learning with NMR spectra," Fuel **290**, 120006 (2021).
[30] R. Ershadnia, M.A. Amooie, R. Shams, S. Hajirezaie, Y. Liu, S. Jamshidi, and M.R. Soltanian, "Non-Newtonian fluid flow dynamics in rotating annular media: Physics-based and data-driven modeling," J. Pet. Sci. Eng. **185**, 106641 (2020).
[31] A. Bemani, A. Baghban, A.H. Mohammadi, and P.Ø. Andersen, "Estimation of adsorption capacity of CO2, CH4, and their binary mixtures in Quidam shale using LSSVM: Application in CO2 enhanced shale gas recovery and CO2 storage," J. Nat. Gas Sci. Eng. **76**, 103204 (2020).
[32] H. Sun, L. Zhou, D. Fan, L. Zhang, Y. Yang, K. Zhang, and J. Yao, "Permeability prediction of considering organic matter distribution based on deep learning," Phys. Fluids **35**(3), 032014 (2023).
[33] M. Meng, R. Zhong, and Z. Wei, "Prediction of methane adsorption in shale: Classical models and machine learning based models," Fuel **278**, 118358 (2020).
[34] M. Nait Amar, A. Larestani, Q. Lv, T. Zhou, and A. Hemmati-Sarapardeh, "Modeling of methane adsorption capacity in shale gas formations using white-box supervised machine learning techniques," J. Pet. Sci. Eng. **208**, 109226 (2022).
[35] H. Hu, F. Hao, X. Guo, F. Dai, Y. Lu, and Y. Ma, "Investigation of methane sorption of overmature Wufeng-Longmaxi shale in the Jiaoshiba area, Eastern Sichuan Basin, China," Mar. Pet. Geol. **91**, 251–261 (2018).
[36] M. Wang, and D. Zhang, "Influences of moisture on adsorption and desorption of methane on gas shales," Energy Sources Part A, 1–19 (2020).
[37] W. Ji, Y. Song, Z. Jiang, L. Chen, Z. Li, X. Yang, and M. Meng, "Estimation of marine shale methane adsorption capacity based on experimental investigations of Lower Silurian Longmaxi formation in the Upper Yangtze Platform, south China," Mar. Pet. Geol. **68**, 94–106 (2015).
[38] M. Wang, Z. Lun, C. Zhao, H. Wang, C. Luo, X. Fu, C. Li, and D. Zhang, "Influences

677 of Primary Moisture on Methane Adsorption within Lower Silurian Longmaxi Shales
678 in the Sichuan Basin, China," Energy Fuels **34**(9), 10810–10824 (2020).
679 [39] Z. Gao, and S. Xiong, "Methane Adsorption Capacity Reduction Process of Water-
680 Bearing Shale Samples and Its Influencing Factors: One Example of Silurian Longmaxi
681 Formation Shale from the Southern Sichuan Basin in China," J. Earth. Sci. **32**(4), 946–
682 959 (2021).
683 [40] W. Han, A. Li, A. Memon, and M. Ma, "Synergetic Effect of Water, Temperature, and
684 Pressure on Methane Adsorption in Shale Gas Reservoirs," ACS Omega **6**(3), 2215–
685 2229 (2021).
686 [41] C. Qian, X. Li, Q. Zhang, Y. Li, W. Shen, H. Xing, P. Shu, L. Han, Y. Cui, and Y.
687 Huang, "Methane Adsorption Characteristics Under In Situ Reservoir Conditions of the
688 Wufeng–Longmaxi Shale in Southern Sichuan Basin, China: Implications for Gas
689 Content Evaluation," Nat. Resour. Res. **32**(3), 1111–1133 (2023).
690 [42] R. Taylor, "Interpretation of the Correlation Coefficient: A Basic Review," J. Diagn.
691 Med. Sonogr. **6**, 35–39 (1990).
692 [43] M.T. Puth, M. Neuhäuser, and G.D. Ruxton, "Effective use of Pearson's product-
693 moment correlation coefficient," Anim. Behav. **93**, 183–189 (2014).
694 [44] J. Tan, P. Weniger, B. Krooss, A. Merkel, B. Horsfield, J. Zhang, C.J. Boreham, G.
695 Van Graas, and B.A. Tocher, "Shale gas potential of the major marine shale formations
696 in the Upper Yangtze Platform, South China, Part II: Methane sorption capacity," Fuel
697 **129**, 204–218 (2014).
698 [45] D. Zhao, J. Hou, B. Wei, H. Liu, Q. Du, Y. Zhang, and Z. Sun, "Fast prediction method
699 of displacement front in heterogeneous porous media using deep learning and
700 orthogonal design," Phys. Fluids **35**(8), 083312 (2023).
701 [46] C.Edward. Rasmussen, and C.K.I. Williams, *Gaussian Processes for Machine*
702 *Learning* (MIT Press, 2006).
703 [47] B. Sun, R. Li, W. Cui, D. Fan, and Y. Shen, "Gaussian process regression for the side-
704 by-side foil pair," Phys. Fluids **35**(10), 107133 (2023).
705 [48] T. Chen, and C. Guestrin, in *Proceedings of the ACM SIGKDD International*
706 *Conference on Knowledge Discovery and Data Mining* (Association for Computing
707 Machinery, 2016), pp. 785–794.
708 [49] S. Pan, Z. Zheng, Z. Guo, and H. Luo, "An optimized XGBoost method for predicting
709 reservoir porosity using petrophysical logs," J. Pet. Sci. Eng. **208**, 109520 (2022).
710 [50] J. Zhang, Y. Sun, L. Shang, Q. Feng, L. Gong, and K. Wu, "A unified intelligent
711 model for estimating the (gas + n-alkane) interfacial tension based on the eXtreme
712 gradient boosting (XGBoost) trees," Fuel **282**, 118783 (2020).
713 [51] M.R. Andalibi, P. Bowen, A. Carino, and A. Testino, "Global uncertainty-sensitivity
714 analysis on mechanistic kinetic models: From model assessment to theory-driven
715 design of nanoparticles," Comput. Chem. Eng. **140**, 10697 (2020).
716 [52] W. Ren, and H.C. Lau, "Analytical modeling and probabilistic evaluation of gas
717 production from a hydraulically fractured shale reservoir using a quad-linear flow
718 model," J. Pet. Sci. Eng. **184**, 106516 (2020).
719 [53] N.A.S. Hamm, J.W. Hall, and M.G. Anderson, "Variance-based sensitivity analysis
720 of the probability of hydrologically induced slope instability," Comput. Geosci. **32**(6),

721 803–817 (2006).
722 [54] E. Parzen, "On Estimation of a Probability Density Function and Mode," Ann. Math.
723 Stat. **33**, 1065–1076 (1962).
724 [55] K. Hu, and H. Mischo, "High-Pressure Methane Adsorption and Desorption in Shales
725 from the Sichuan Basin, Southwestern China," Energy Fuels **34**(3), 2945–2957 (2020).
726 [56] R.T. MARTIN, in *Clays Clay Miner*, edited by E. INGERSON (Pergamon, 1962), pp.
727 28–70.
728 [57] J. Li, X. Li, X. Wang, Y. Li, K. Wu, J. Shi, L. Yang, D. Feng, T. Zhang, and P. Yu,
729 "Water distribution characteristic and effect on methane adsorption capacity in shale
730 clay," Int. J. Coal Geol. **159**, 135–154 (2016).
731 [58] H. Bi, Z. Jiang, J. Li, F. Xiong, P. Li, and L. Chen, "Ono-Kondo Model for
732 Supercritical Shale Gas Storage: A Case Study of Silurian Longmaxi Shale in Southeast
733 Chongqing, China," Energy Fuels **31**(3), 2755–2764 (2017).
734 [59] F. Yang, Z. Ning, R. Zhang, H. Zhao, and B.M. Krooss, "Investigations on the
735 methane sorption capacity of marine shales from Sichuan Basin, China," Int. J. Coal
736 Geol. **146**, 104–117 (2015).
737 [60] H. Nie, P. Li, W. Dang, J. Ding, C. Sun, M. Liu, J. Wang, W. Du, P. Zhang, D. Li, and
738 H. Su, "Enrichment characteristics and exploration directions of deep shale gas of
739 Ordovician–Silurian in the Sichuan Basin and its surrounding areas, China," Pet. Explor.
740 Dev. **49**(4), 744–757 (2022).
741 [61] L. Pan, X. Xiao, H. Tian, Q. Zhou, and P. Cheng, "Geological models of gas in place
742 of the Longmaxi shale in Southeast Chongqing, South China," Mar. Pet. Geol. **73**, 433–
743 444 (2016).
744 [62] F. Miao, D. Wu, X. Liu, X. Xiao, W. Zhai, and Y. Geng, "Methane adsorption on
745 shale under in situ conditions: Gas-in-place estimation considering in situ stress," Fuel
746 **308**, 121991 (2022).
747 [63] X. Guo, D. Hu, Z. Wei, Y. Li, and X. Wei, "Discovery and exploration of Fuling shale
748 gas field," China Petrol. Explorat **21**(3), 24–37 (2016).
749