Petroleum Science 20 (2023) 1659–1675

Contents lists available at ScienceDirect

Petroleum Science

journal homepage: www.keaipublishing.com/en/journals/petroleum-science

Original Paper

Shale gas production evaluation framework based on data-driven models

You-Wei He [a, *], Zhi-Yue He [a], Yong Tang [a], Ying-Jie Xu [a], Ji-Chang Long [a], Kamy Sepehrnoori [b]

[a] State Key Laboratory of Oil and Gas Reservoir Geology and Exploitation, Southwest Petroleum University, Chengdu, 610500, Sichuan, China
[b] Department of Petroleum and Geosystems Engineering, The University of Texas at Austin, Austin, TX, 78712, USA

ARTICLE INFO

Article history:
Received 9 September 2022
Received in revised form
5 December 2022
Accepted 7 December 2022
Available online 9 December 2022

Edited by Yan-Hua Sun

Keywords:
Shale gas
Production evaluation
Production prediction
Data-driven models
Carbon neutrality

ABSTRACT

Increasing the production and utilization of shale gas is of great significance for building a clean and low-carbon energy system. Sharp decline of gas production has been widely observed in shale gas reservoirs. How to forecast shale gas production is still challenging due to complex fracture networks, dynamic fracture properties, frac hits, complicated multiphase flow, and multi-scale flow as well as data quality and uncertainty. This work develops an integrated framework for evaluating shale gas well production based on data-driven models. Firstly, a comprehensive dominated-factor system has been established, including geological, drilling, fracturing, and production factors. Data processing and visualization are required to ensure data quality and determine final data set. A shale gas production evaluation model is developed to evaluate shale gas production levels. Finally, the random forest algorithm is used to forecast shale gas production. The prediction accuracy of shale gas production level is higher than 95% based on the shale gas reservoirs in China. Forty-one wells are randomly selected to predict cumulative gas production using the optimal regression model. The proposed shale gas production evaluation framework overcomes too many assumptions of analytical or semi-analytical models and avoids huge computation cost and poor generalization for numerical modelling.

1. Introduction

Conventional natural gas resources are unable to meet the human-being's demand. Advances of drilling and fracturing technologies enables the commercial production of unconventional natural gas resources, and shale gas has become the significant source of global natural gas production due to its wide distribution and giant reserves (Fig. 1). According to the energy report in 2015 from U.S. Energy Information Administration (EIA), the global recoverable resources of shale gas are about 214.55 trillion cubic meters (tcm), of which the United States has 17.63 trillion cubic meter and China has 31.58 trillion cubic meters (EIA, 2015). The U.S. shale oil and gas production has increased significantly by applying multi-stage fractured horizontal wells (MFHW) technology. In 2016, the natural gas exports of U.S. exceeded imports for the first time, and the natural gas exports of U.S. even reached 5.90×10^8 cubic meters in 2020 (Zhao et al., 2021). The shale gas revolution by the U.S. energy industry has achieved major success and reshaped the world energy landscape. Increasing the utilization of shale gas is of great significance for ensuring energy security, improving the energy structure, and reducing the environmental pollution to build a safe, clean, low-carbon, and efficient energy system to achieve carbon neutrality.

The commercial gas production is hard to achieve due to extremely low porosity and permeability (reaching to nano-Darcy) of shale gas reservoirs. Stimulation of original shale gas formations is required to enhance the permeability of shale reservoirs. Hydraulically fracturing technology along with horizontal wells has been acknowledged as an effective stimulation means to achieve the above-mentioned goal by generating complex fracture networks composed of natural fractures and hydraulic fractures to provide high-permeability flow channels for shale gas (McClure et al., 2016; Xue et al., 2021a). Sharp decline of shale gas production has been widely observed in most of shale gas reservoirs, leading to low gas recovery factor. Thus, accurate evaluation and

* Corresponding author.
E-mail address: youweihe_cupb@163.com (Y.-W. He).

https://doi.org/10.1016/j.petsci.2022.12.003

Nomenclature

a	Average distance between sample and all other samples in the same cluster
b	Average distance between sample and all samples in the next nearest cluster
CSS	Cluster sum of square
$d_{\text{Euclidean}}(x, \mu)$	Euclidean distance
$d_{\text{Manhattan}}(x, \mu)$	Manhattan distance
$Interf_{\text{Frachits}}$	Total interference degree of gas production caused by frac hits
$Interf_{\text{Frachits-}i}$	Interference degree of gas production caused by any frac hits
m	Number of clusters
MSE	Mean squared error
n	Number of samples in the cluster
$Q_{\text{Frachits-}i}$	Impacted gas production by any frac hits
R^2	Fitting-quality coefficient
R_{Frachits}	Total recovering degree of gas production caused by frac hits
$R_{\text{Frachits-}i}$	Recovering degree of gas production caused by any frac hits
s	Silhouette coefficient
t_{all}	Total interference time of frac hits
t_i	Time of any frac hits
x_i	Sample points in cluster
x_{ij}	The i^{th} sample in the j^{th} cluster
y_i	Actual value of the sample label in the test set
$\hat{y}_i$	Predicted value of the sample label in the test set
$\bar{y}$	Average of sample labels in the test set
z	Number of samples in the test set
α_i	Weight of individual frac hits
μ	Cluster centroid
μ_j	Cluster centroid of j^{th} cluster

Fig. 1. Global shale gas distribution (From https://www.worldenergy.org/assets/downloads/Summary-report_Unconventional-gas-a-global-phenomenon-World-Energy-Resources-1.pdf).

forecasting of shale gas production becomes a crucial issue. However, it is difficult to predict the production of shale gas wells due to the complex fracture networks (Li et al., 2018), dynamic fracture properties, fracturing interference (He et al., 2020; Guo et al., 2021; Qin et al., 2022) in multi-well pads, complicated multiphase flow (gas, water, oil, and fracturing fluids) after fracturing (Wu et al., 2021) and multi-scale flow (Zeng et al., 2021; He et al., 2022; Clarkson et al., 2016) as well as data quality and uncertainty (Fig. 2). Therefore, how to efficiently and accurately evaluate and predict the shale gas production is significant to improve the production performance of shale gas resources.

There are three methods for shale gas production prediction, including analytical approaches, numerical simulation, and emerging data-driven models. The analytical methods mainly include the material balance equation (MBE) and empirical production decline model. Although the MBE approach can be applied to estimate the oil content, gas content, and gas-oil ratio in the unconventional reservoirs (Ojo and Osisanya, 2006), its accuracy is significantly reduced under complex geological conditions. Empirical decline model is the most commonly used method for evaluating shale gas production, originally proposed by Arps (1945) to predict oil and gas production. However, the accuracy of Arps' original model in production evaluation of unconventional reservoirs is very low (Duong, 2010). Thus, the empirical decline model

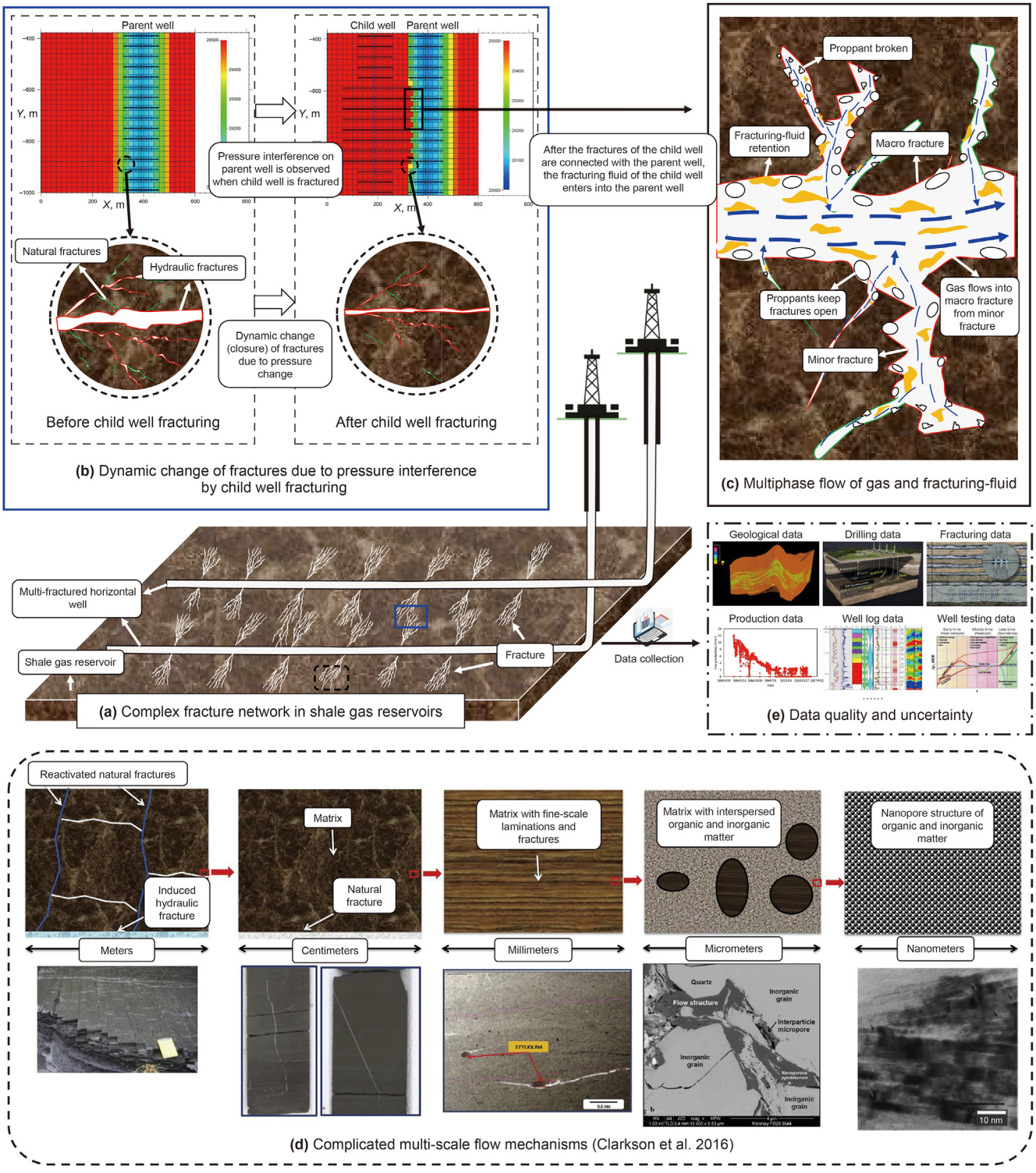


Fig. 2. Complicated static and dynamic mechanisms as well as data quality and uncertainty during the exploitation of shale gas reservoirs.

has been improved (Fetkovich, 1973; Carter, 1985; Palacio and Blasingame, 1993; Agarwal et al., 1999). Novel empirical decline models for tight/shale gas reservoirs have been developed since 2008, including four categories (i.e., related to Arps' exponential decline models (Ilk et al., 2008; Johnson et al., 2009; Mattar and Moghadam, 2009; Valko, 2009; Yu and Miocevic, 2013; Zhang et al., 2016), related to Arps' hyperbolic decline models (Fulford and Blasingame 2013; Maraggi et al., 2016), related to the rate decline models during the linear flow in fracture-dominated reservoirs (Josh and Lee, 2013; Ali et al., 2014; Wang et al., 2017a), and other methods (e.g., logistic growth curves (Clark et al., 2011), fractional decline curves (Zuo et al., 2016), and other complicated methods (Makinde and Lee, 2017; Mishra, 2012). Many parameters are required to be calculated by trial and error or should be adjusted

using the corresponding dimensionless curves for the above-mentioned models, causing them inconvenient for field application (Wang et al., 2020b). Although other improved empirical decline methods are emerging (Yuan et al., 2020), these models have common shortcomings in shale gas production prediction. For instance, the fluctuated gas production leads to inaccurate evaluation results, caused by large amounts of flowback water, frac-hits, shut-in, field tests, skin effect, workover, and other engineering operations (Wang et al., 2020b). Fig. 3 summarizes the main analytical models used for shale gas production prediction.

Recent advances in the numerical reservoir simulation techniques of unconventional reservoirs include two parts. The first one focuses on improving numerical modelling methods that account for complicated mechanisms such as the migration mechanism of natural gas (Miao et al., 2019; Ning et al., 2019), real gas effect (Wu et al., 2017), methane adsorption (Tian and Liu, 2020), phase behavior (Zhang et al., 2017), nonlinear flow (Liu et al., 2019), and stress sensitivity (Wang et al., 2018). However, the reservoir simulation requires numerical solution of partial differential equations, which yields heavy computational costs and makes it hard to characterize the dynamic properties of rock, fluids and fractures as well as stress. And it is quite hard to provide real-time reference for fracturing schemes adjustment, development plan decision, or enhanced oil recovery measures based on the numerical simulations. Also, the established geological model and numerical model can only be applied to the target reservoirs, which is limited for generalization.

The second methods intend to predict the shale gas production through numerical simulation (Wang et al., 2017b; Hu et al., 2017, 2020) by integrating multiple models (e.g., dual-continuum models (Ganzer, 2002; Azom and Javadpour, 2012; Jia et al., 2021a, 2021b), discrete fracture models (DFM) (Moinfar et al., 2011), and embedded discrete fracture models (EDFM) (Xu et al., 2018; Yu et al., 2018) into reservoir simulations to handle the irregular and complex natural and hydraulic fractures more accurately and efficiently. During the large-scale fracturing operations in unconventional gas reservoirs, the fracture networks near the parent well may dynamically change (open or close) since the stress will be changed. Even if the influence of formation pressure is not considered, the fracturing may cause that two wells are directly connected or communicated through natural fractures or hydraulic fractures (Sardinha et al., 2014; Guindon, 2015; Jacobs, 2017) (also known as frac hits) due to the small well spacing, which will also make the fracture networks dynamically change and form more complex fracture networks. During fracturing fluid flowback, the fracture conductivity may be reduced, and there is still fracturing fluid retention within the fractures even after flowback (Jia et al., 2018; Huang et al., 2020). Thus, two-phase flow of shale gas and fracturing fluids in the fractures are definitely existing and the impact of fracturing fluids cannot be ignored, which are hard to be reflected by the established numerical models (Wei et al., 2020). Furthermore, the dynamic changes of hydraulic fractures and natural fractures will also occur due to the change of stress during the production period. For instance, the hydraulic fractures will be partially closed due to proppant broken (Shi, 2021; Mao et al., 2020), and the natural fractures will also be partially closed due to the effect of stress. It may lead to the reopening of fractures and increase of fracture conductivity if some enhanced gas recovery (EGR) measures are performed, such as acidizing, refracturing, water injection, and gas injection. The predicted productivity will show a large error with the practical productivity if the originally designed fracturing parameters are used. Fig. 4 is a brief summary of numerical simulation methods for shale gas production prediction.

At present, production prediction of unconventional reservoirs is still challenging due to the complicated micro-flow mechanism of fluids in shale reservoirs and engineering problems. It is necessary to comprehensively consider the geological and engineering factors as well as multi-scale data for shale gas production evaluation. The emerging data-driven model provides a potential solution to the above technical challenges. Data-driven model mainly uses machine learning methods for modeling. Machine learning has been widely used to solve complicated problems in both engineering and science fields (Tontiwachwuthikul et al., 2020; Cao et al., 2022; Yan et al., 2022a, 2022b), including petrophysical modeling (Syed et al., 2020), recovery factor estimation of reservoirs (Makhotin et al., 2022), etc. Compared with traditional methods, data mining could identify intrinsic information and

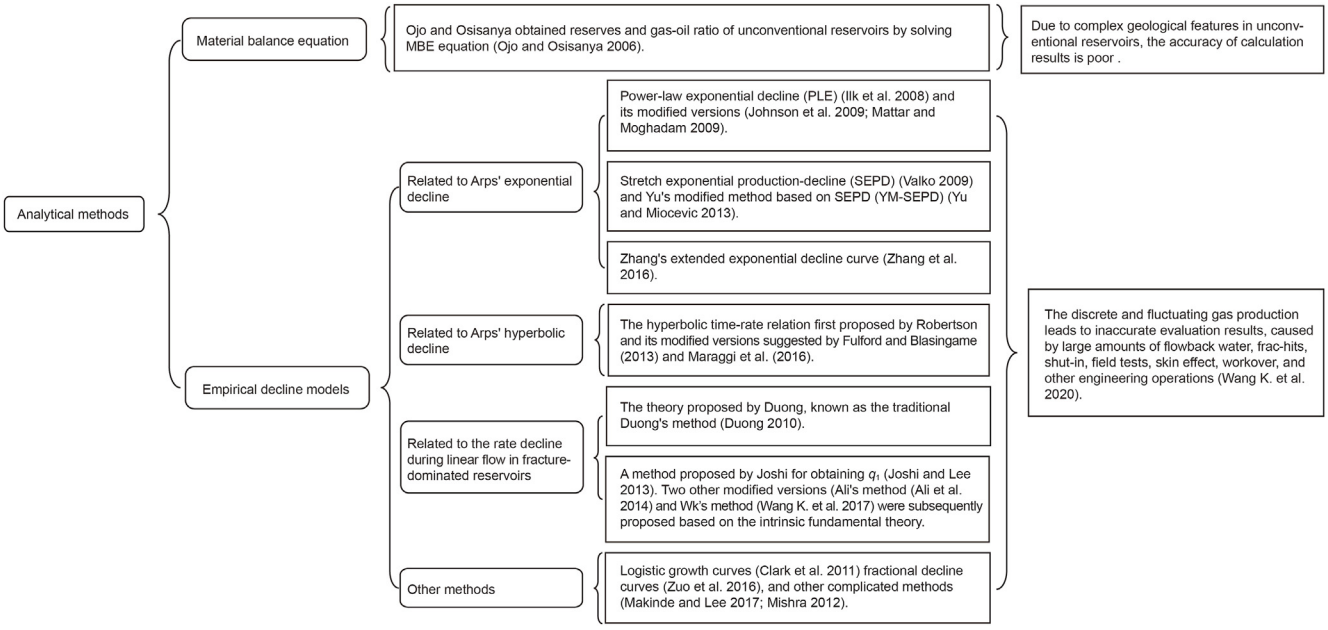


Fig. 3. Summary of analytical models for shale gas production prediction.

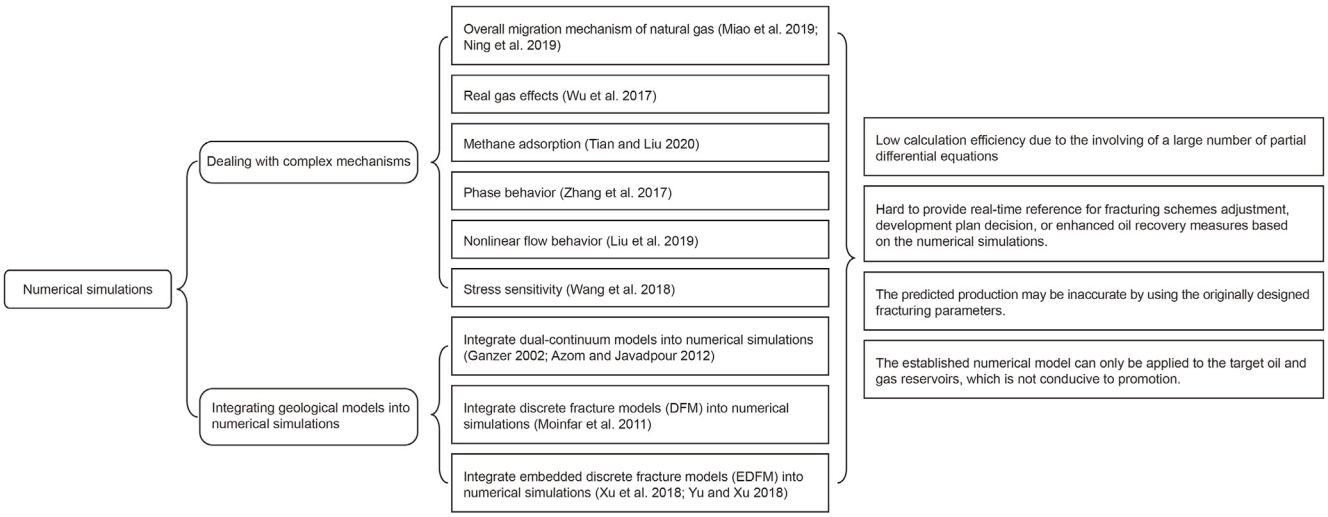


Fig. 4. Summary of numerical simulation methods for shale gas production prediction.

potential patterns that are difficult to visualize from existing data and provide new ideas for unsolved problems (Wang et al., 2021b; He et al., 2021). According to the data sources (using field data or simulation results), data-driven methods can be divided into two categories. Firstly, Wang et al. (2021b) tried to predict the production performance of unconventional reservoirs by combination of representative numerical model (MFHW in shale oil reservoirs) and deep belief network (DBN) models. Xue et al. (2021b) established a numerical model that can simulate three flow systems (i.e., shale matrix, stimulated reservoir volume (SRV), and hydraulic fracture) to obtain the modeling data set, and further developed a data-driven model by using the multi-objective random forest method to forecast the shale gas production. Secondly, Han and Kwon (2021) used the data from the Montney shale reservoir in Canada and deep neural network (DNN) models to build a data-driven model for predicting cumulative gas production. Vikara et al. (2020) used Marcellus shale parameters and gradient boosted regression tree (GBRT) algorithm to establish a data-driven model to evaluate shale gas productivity. Mehana et al. (2021) developed a data-driven model using Eagle Ford shale gas parameters and a new unsupervised machine learning method to quickly and accurately predict and update the estimated ultimate recovery (EUR) of unconventional gas wells. Liu et al. (2021) designed a deep-learning-based algorithm for EUR evaluation of shale gas wells based on EUR evaluation results of 282 wells in the WY shale gas field. Fig. 5 summarizes the data-driven models for predicting shale gas production.

However, the index system used in the data-driven models is often incomplete. And the data processing approaches in the data-driven models are different so that it is hard to determine whether the processed data set and the established model are optimal. A general and effective data processing method is significant and required.

To fill this gap, this work develops an integrated framework for shale gas production evaluation and prediction based on data-driven models (Fig. 6). Firstly, a comprehensive system of dominated factors was established, including four categories (i.e., geological factors, drilling factors, fracturing factors, and production factors) and a total of 35 dominated factors. Based on field data, the available factors can be determined based on the developed dominated factors. Data processing workflow is proposed to ensure the data quality, including missing value processing, correlation analysis, outlier analysis, data standardization, and PCA, etc. Results based on different data processing methods need to be compared to determine the optimal data processing method suitable for the basic data set. Data visualization is performed by considering two different situations to determine the final data set. The shale gas production evaluation model can be further developed to evaluate the shale gas production level. Finally, the random forest algorithm is used to establish the prediction model to quantitatively forecast the shale gas production. The proposed shale gas well production evaluation and prediction framework highlight the practical potential in shale gas production prediction.

2. Dominated factor system

The dominated factors of the shale gas production are complicated due to the complexities on shale gas evolution characteristics (free gas, absorbed gas), production and operations (multi-well pad with massively hydraulically fracturing, sharp decline of gas production, frac hits), and multi-scale flow mechanisms (nanopores, micro fractures, macro fractures, hydraulic fractures) in shale gas formations (Lawal et al., 2013; Huang et al., 2019; Wang et al., 2020a).

The dominated factors impacting shale gas production can be divided into four categories, including geological factors, drilling factors, fracturing factors, and production factors. It is inaccurate and unreasonable by only considering partial factors for estimating shale gas production. Thus, this work establishes a relatively completed dominated factor system of shale gas production by incorporating the full-cycle features (i.e., geological factors, drilling factors, fracturing factors, and production factors) and a total of 34 dominated factors. The dominated factor system for shale gas production developed in this work can be found in Fig. 7.

2.1. Geological factors

The geological factors directly and significantly affect the shale gas production since they represent hydrocarbon reserves and the ability of fluid flow in the porous media. The impact of geological factors on production is complicated so that the relationship among different geological factors needs to be identified. For instance, organic matter (OM) and inorganic minerals are not totally independent. Based on this principle, 15 geological factors are determined, including total organic carbon (TOC), vitrinite reflectance (R_o), total gas content, Poisson's ratio, minimum horizontal

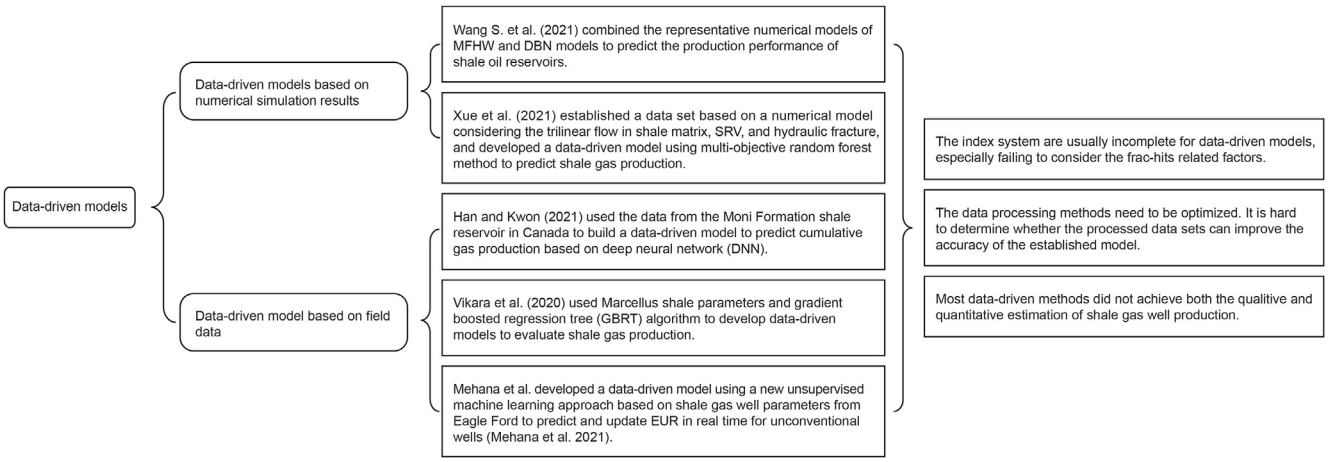


Fig. 5. Summary of data-driven models for shale gas production prediction.

principal stress, rock brittle index, porosity, permeability, reservoir thickness, filling degree, angle between the fracture and the maximum horizontal principal stress, pressure gradient and reservoir depth.

The TOC content affects both the gas saturation and the quality of shale gas reservoirs, and it plays a key role in assessing the production potential and the subsequent EGR techniques in shale gas reservoirs (Wang et al., 2019). The lower limit of TOC content for economic exploitation of shale gas reservoirs is about 2.5–3 wt% (Bowker, 2007), and the limit of TOC content available for shale gas reservoirs will become even lower with the technical advances.

There are two main mechanisms affecting shale gas production by TOC. On one hand, shale gas is *in-situ* generated and stored in the reservoirs. After shale gas is generated, there is almost no migration or the migration distance is extremely short. Thus, the hydrocarbon generation potential and shale gas production may be higher under higher TOC. For instance, the Marcellus Shale and Bakken Shale also show high TOC (greater than 7 wt%) in specific zones (Bhattacharya et al., 2016). On the other hand, TOC has strong effect on porosity and pore structure (Song et al., 2019). OM-hosted pores are the most prevalent pores in organic-rich shale (Wang et al., 2021c). Song et al. (2019) analyzed the samples from different shale gas reservoirs, and found that the presence of OM in shale strongly enhances the storage capacity by increasing the specific surface area and pore volume, which represents absorption storage capacity and free-gas storage capacity.

The thermal maturity and clay mineral content play an important role in evaluating the storage capacity of shale gas reservoirs. Shale samples with vitrinite reflectance (R_o) higher than 1% have a greater surface area, which indicates that more micropores are formed in the OM to further enhance shale storage capacity when the shale maturity reaches the oil window level (Li et al., 2019). Based on the Marcellus Shale and Mahantango Formation, although clay minerals have a bigger specific surface area, there is a negative correlation between the storage capacity of clay minerals and shale content (Ambrose et al., 2010), which may be attributed to the accessibility of the pore space. Marcellus shale is a pore system dominated by OM, and the OM content and shale storage capacity decrease with the increase of clay minerals. Therefore, clay minerals show a negative effect on shale gas reserves when shale is rich in OM.

Compared with TOC and organic maturity, the shale gas content can intuitively reflect the upper limit of shale gas production. Shale gas is composed of free gas, adsorbed gas and solution gas. The solution gas may be merged into adsorbed gas since solution gas content is relatively quite limited (Ambrose et al., 2010). There are still some difficulties in calculating the adsorbed gas content using the previous models. The first principle methods and empirical models are commonly applied to obtain the adsorbed gas content. However, the first principle method is hard to fully consider the complexity of the storage mechanism (Chen et al., 2017). Therefore, the empirical models are the preferred choice. The Langmuir model is the most commonly used method (Langmuir, 1918; Ross and Bustin, 2007), and adsorption experiments are regarded as an effective method to obtain Langmuir parameters.

However, these experiments are time-consuming since the adsorption process within the microscale and nanoscale shale pores is slow, and absorbed gas leakage during the coring process affects the accuracy of the adsorption experiments and causes testing errors (Chen et al., 2017). Besides, obtaining the shale samples is expensive. These challenges lead to the difficulty in determining the corresponding Langmuir parameters and the high uncertainty in the evaluation of the adsorbed gas content. Thus, desorption, isothermal adsorption, and logging interpretation methods are often used to measure the total gas content roughly. If the content of different types of shale gas can be calculated, the shale gas production can be evaluated more accurately.

Natural fracture is a key factor that affects the migration accumulation and production of shale gas. The natural fractures can increase the reservoir connectivity and the accumulation of free gas, improve the effective formation permeability, and finally increase the cumulative gas production. The shale reservoir characteristics in North America indicated that matrix permeability influences gas production potential significantly. Furthermore, the formation pressure coefficient is a key indicator reflecting the preservation conditions (Wang et al., 2016; Pang et al., 2018), and the high formation pressure coefficient is conducive to the shale gas enrichment. The formation pressure coefficient will increase with the increase of reservoir depth. Meanwhile, the rock near the fractures or faults may occur plastic deformation and enhance the sealing of the fractures or faults, inhibiting the escape of shale gas.

The large-scale natural fractures with good conductivity may destroy the formation sealing properties and be unfavorable for shale gas preservation. Filling degree and the angle between the fracture and the maximum horizontal principal stress are important factors for formation sealability. When the angle between the maximum horizontal principal stress and natural fracture is closer to 90°, the fracture generally remains closed, inhibiting the escape of shale gas (Patil et al., 2017).

Large-scale hydraulic fracturing is often used in shale gas

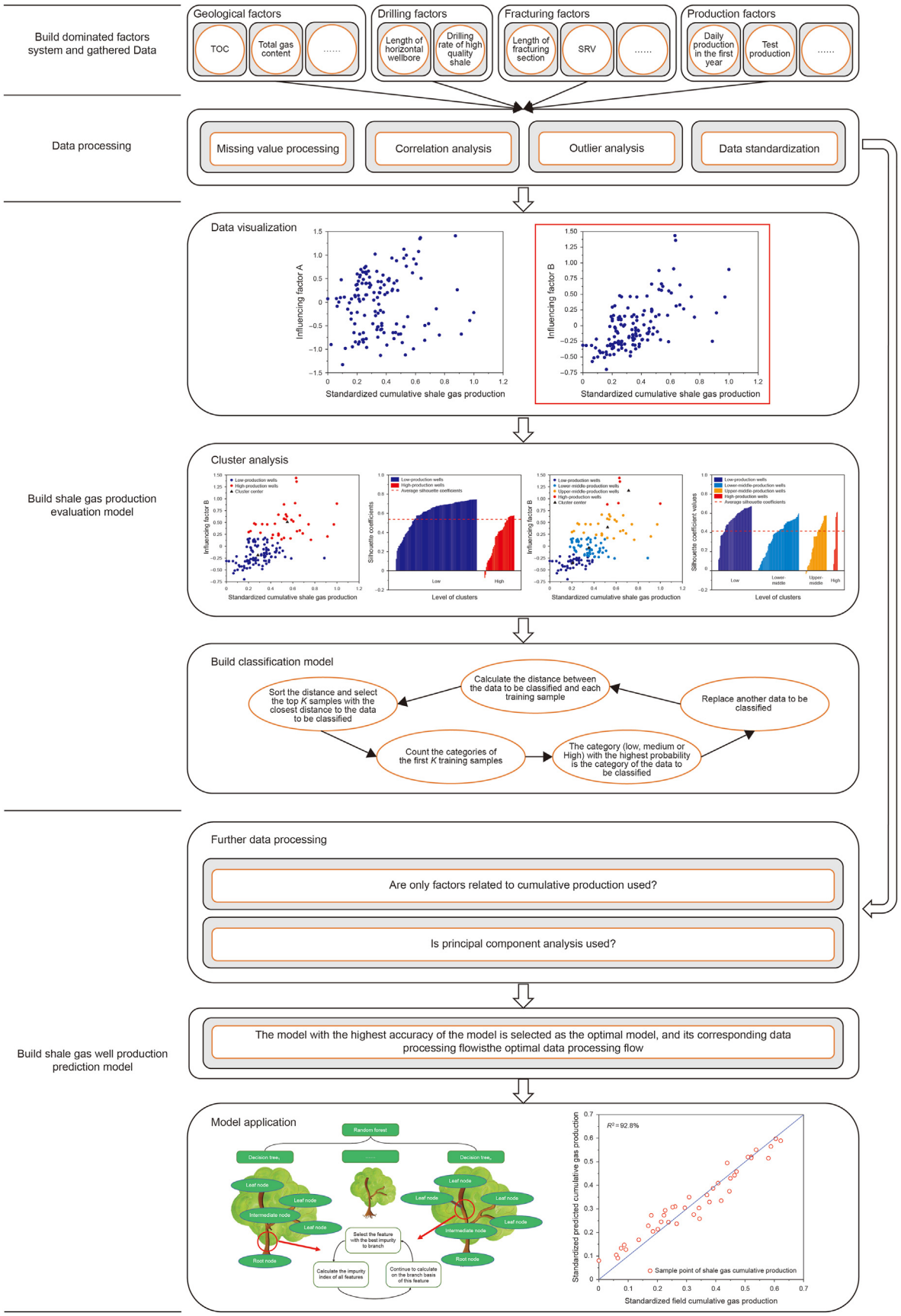


Fig. 6. Shale gas production evaluation framework based on data-driven models.

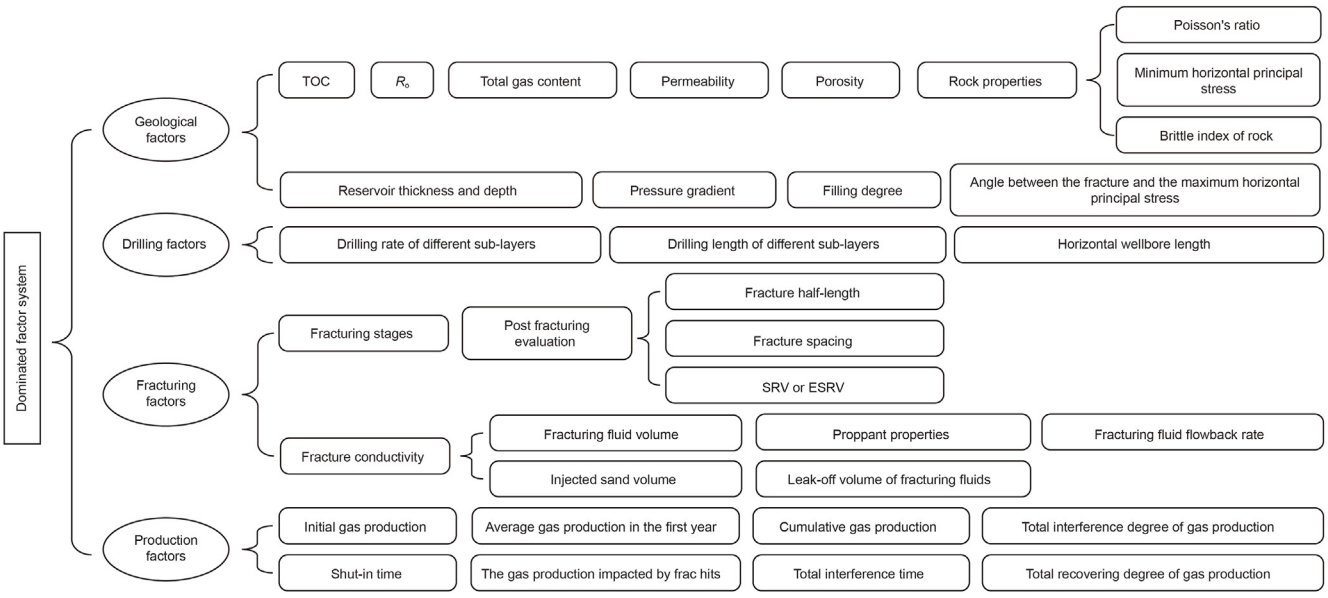


Fig. 7. The dominated factor system for shale gas production established in this work.

reservoirs. The final fracturing performance is directly related to rock properties such as rock brittleness and stress conditions (Grieser and Bray, 2007). Poisson's ratio, rock brittleness index and minimum horizontal principal stress can be used to quantify the above properties (Bhattacharya et al., 2019). The content of brittle minerals will affect the fracturing and the evolution of shale pores and natural fractures. There is a certain internal connection between the OM and brittle minerals. And the correlation coefficient between brittleness index and TOC is high (Yin et al., 2018), demonstrating that the brittle minerals significantly impact shale gas production.

The shale formation thickness will also affect the shale gas production. High shale gas production is easier to be achieved for thick shale reservoirs (Fan et al., 2020).

2.2. Drilling factors

The purpose of drilling is to form an effective channel between shale gas formations and wellbore. Baihly et al. (2015) analyzed the gas production of major shale gas reservoirs in the United States and indicated that drilling factors can affect the initial gas production of shale gas wells. The horizontal wellbore length, drilling encountered rate of different sub-layers, and drilling length of different sub-layers significantly affected the shale gas production.

As the horizontal wellbore length increases, the well-controlled reserves and the stable production period are also enhanced. Longer horizontal wellbore means bigger fracturing spacing and weaker fracture interference if the number of fracturing perforations is constant. There exists an optimal length of horizontal wellbore considering the drilling success rate and economic costs.

The reserves of different layers may be different for the same shale gas reservoir. Thus, the drilling encountered rate and drilling length of different sub-layers especially for gas-rich layers have significant impact on shale gas production, which are the vital factors to evaluate drilling quality and potential gas productivity.

2.3. Fracturing factors

The massively hydraulic fracturing is performed after drilling to generate the fracture networks to allow gas flowing to the wellbore from the tight formations. The properties of fracturing networks are tightly related to the shale gas production. Ten fracturing factors are proposed to represent the effect of fracturing on shale gas production, including fracturing stages, fracture half-length, fracture spacing, SRV, effective stimulated reservoir volume (ESRV), volume of fracturing fluids, volume of injected sand, fracturing fluids flowback rate, leak-off volume of fracturing fluids, and proppant properties.

The number of perforations can be used to approximatively characterize the fracturing scale. Generally, more fracturing fluids and sand need to be pumped for more perforations, and the final fracturing scale will be larger. The SRV is often used to estimate the range of the complex fracture networks (Cipolla and Wallace, 2014; Al-Rbeawi, 2020; Umar et al., 2021). SRV can be measured based on microseismic maps while it cannot accurately reflect the effective stimulated volume since the microseismic may not identify the natural fractures (Eisner and Staněk, 2018). Thus, a new factor, ESRV, is proposed to better describe the stimulated volume. Moos et al. (2011) tried to estimate ESRV of shale gas reservoirs by geomechanically modeling. The ultimate gas recovery, optimal fracture length and fracture spacing can be evaluated after the ESRV is determined. It can be also benefit of determining the percentage of the effective fractures (Umar et al., 2021). However, the data acquisition is limited and ESRV is hard to calculate so that SRV is still often used to evaluate fracturing performance and recoverable reserves without ESRV.

The hydraulic fracture half-length will affect the complexity of the fracture networks so that reasonable fracture half-length can enhance the gas production. Fracture spacing is also a significant parameter influencing shale gas production. Wide fracture spacing can only control partial shale reservoirs and resources. Tight fracture spacing will cause fracturing interference, which may cause difficulty in fracture initiation. Therefore, it is necessary to reasonably design reasonable fracture spacing to maximize gas productivity based on the specific formation properties of different reservoirs.

The impact of fracture conductivity on gas production is significant. The fracture conductivity is related to fracturing fluid volume, injected sand volume, fracturing fluid flowback rate, leak-off volume of fracturing fluids and proppant properties (e.g., sand-

liquid ratio, sand strength). Although fracture conductivity shows positive correlation to shale gas production, increasing fracture conductivity will not increase the gas production significantly when the conductivity is high enough. Thus, there exists optimal fracture conductivity.

2.4. Production factors

Well pad is applied in shale gas reservoirs to reduce the cost of drilling and fracturing. Generally, one well pad is composed of six to eight horizontal wells. Well spacing is about 300–500 m, causing well interference or even frac hits. Horizontal wells are connected through hydraulic fractures or natural fractures after frac hits, and fracturing fluids from the child wells flow into the parent well, increasing water production and significantly decreasing gas production of the parent well. And the dynamic stress and pore pressure may change the fracture networks, such as fracture closure. After drilling and fracturing operations, shale gas well can be opened for fracturing fluid flowback and production. Thus, it is hard to evaluate shale gas production potential only using absolute open flow rate, early gas production or average gas production due to complicated flow mechanisms and special features of shale gas reservoirs. The production time needs to be kept the same to equally evaluate the gas production. Cumulative gas production during the same production time was used as the measure of shale gas production. Multiple factors are also required to represent the gas production potential, such as initial gas production, average gas production in the first year.

Since well interference especially fracturing interference impacts the shale gas production a lot, the factors related to fracturing interference cannot be ignored (Guo et al., 2022; Zhang et al., 2022). The existing evaluation system hardly considers the fracturing interference for shale gas prediction. The gas production impacted by fracturing interference, impact degree, recovering degree of gas production, and interference time are determined to evaluate the effect of fracturing interference on shale gas production. However, these field data may not be used directly and need to be processed.

Firstly, if the well receives multiple frac hits from different child wells, new factor is required to reflect the impact of multiple frac hits. The weight of individual frac hits can be estimated based on the impacted gas production by different frac hits.

$$\alpha_i = \frac{Q_{\text{Frachits}-i}}{\sum_{i=1}^{n} Q_{\text{Frachits}-i}} \tag{1}$$

where α_i represents the weight of individual frac hits; and $Q_{\text{Frachits-}i}$ means the impacted gas production by any frac hits.

The total interference degree of gas production caused by frac hits can be determined by

$$Interf_{\text{Frachits}} = \sum_{i=1}^{n} \alpha_i \times Interf_{\text{Frachits}-i} \tag{2}$$

where $Interf_{\text{Frachits}}$ is the total interference degree of gas production caused by frac hits; and $Interf_{\text{Frachits-}i}$ denotes the interference degree of gas production caused by any frac hits.

The total recovering degree of gas production caused by frac hits can be characterized by

$$R_{\text{Frachits}} = \sum_{i=1}^{n} \alpha_i \times R_{\text{Frachits}-i} \tag{3}$$

where R_{Frachits} means the total recovering degree of gas production caused by frac hits; and $R_{\text{Frachits-}i}$ shows the recovering degree of gas production caused by any frac hits.

The total interference time of frac hits can be calculated by

$$t_{\text{all}} = \sum_{i=1}^{n} t_i \tag{4}$$

where t_{all} is the total interference time of frac hits; and t_i means the time of any frac hits.

Furthermore, shut in is also a key factor for shale gas production evaluation since shut in can be often observed for gas wells due to frac hits, low gas production, or other reasons. Except for the absolute shut-in time, the ratio of shut-in time to total production time may also be considered for different wells.

3. Shale gas production evaluation framework

An integrated and improved production evaluation method is developed for shale gas reservoirs, including the production level evaluation model and production regression prediction model.

3.1. Data processing

The available field data is limited compared to the comprehensive evaluation system so that the factors covering the full cycle of shale gas exploitation need to be obtained to ensure the evaluation accuracy. The practical factors can be determined using the field data of the A shale gas reservoir based on the developed factor system. Data processing is required to improve the data quality, such as the abnormal data and missing data.

Data processing methods mainly include missing value interpolation, data standardization, correlation analysis, outlier analysis and principal component analysis (PCA). Firstly, missing value interpolation is used to supplement the missing data. Secondly, data standardization is applied to avoid the ignorance of some parameters. Thirdly, spearman correlation analysis can be used to provide a basis for outlier analysis and subsequent processing of the modeling dataset. Finally, Mahalanobis distance method is used to remove the abnormal data to improve data quality. Other complicated data processing methods, such as PCA, are further used to figure out whether the use of complicated data processing methods can improve the modeling accuracy.

3.2. Shale gas production level evaluation model

The production evaluation is significant for making related measures to enhance shale gas recovery. The unsupervised learning algorithm (i.e., clustering algorithm) is used to develop the shale gas production evaluation model since the corresponding production level of individual well is not known before modeling. The adopted factor system is important for evaluation results.

Firstly, two factor systems are established, including the whole factors and related factors. Secondly, this work introduces a novel factor system for modeling according the above-mentioned two systems using the PCA. In this way, it can visualize the data, reduce the model uncertainty. Thirdly, data visualization for cumulative shale gas production is performed to show which factor system achieves better performance. Fourthly, Kmeans algorithm is used to cluster the data sets more suitable for modeling. Finally, KNN algorithm is applied to develop the model for predicting shale gas production levels based on the cluster analysis results. The principle of the Kmeans algorithm is shown in Fig. 8.

The clustering quality needs to be optimized by achieving minor difference within the cluster and large difference outside the

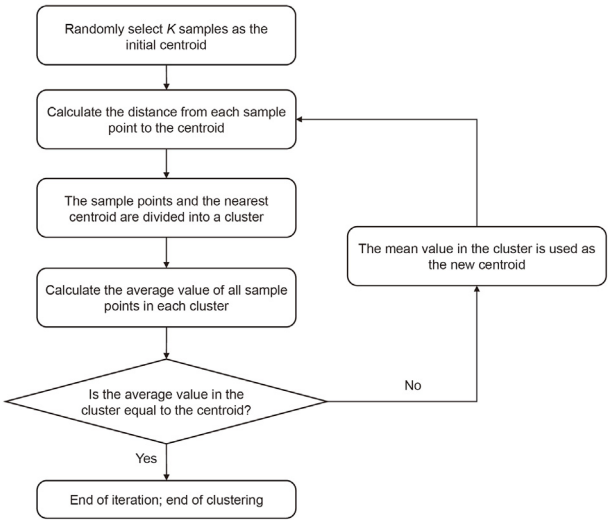


Fig. 8. The principle of the Kmeans algorithm.

cluster. The distance from the sample point to the centroid of its cluster can be used to evaluate the clustering quality, which can be calculated by the specific parameters. Euclidean distance (Danielsson, 1980) is equal to

$$d_{\text{Euclidean}}(x,\mu)=\sqrt{\sum_{i=1}^{n}(x_i-\mu)^2} \tag{5}$$

where $d_{\text{Euclidean}}(x, \mu)$ is the Euclidean distance; n is the number of samples in the cluster; x_i is the sample points in the cluster; and μ is the cluster centroid.

Manhattan distance (Ouarda and Souad, 2022) can be calculated by

$$d_{\text{Manhattan}}(x,\mu)=\sum_{i=1}^{n}(|x_i-\mu|) \tag{6}$$

where $d_{\text{Manhattan}}(x, \mu)$ is the Manhattan distance.

If the Euclidean distance is used, the cluster sum of squares (CSS) (McKenna et al., 2014) of the distances from all sample points in a cluster to the centroid can be expressed as

$$\text{CSS}=\sum_{j=1}^{m}\sum_{i=1}^{n}\left(x_{ij}-\mu_j\right)^2 \tag{7}$$

where m is the number of clusters; x_{ij} is the ith sample in the jth cluster; and μ_j is the cluster centroid of the jth cluster.

Generally, minor CSS indicates good clustering. However, CSS is hard to quantitatively evaluate the clustering quality since it will be impacted by many factors. The silhouette coefficient is preferred to assess the clustering (Wang et al., 2021a). The silhouette coefficient for an individual sample is expressed as

$$s=\frac{b-a}{\max(a,b)} \tag{8}$$

where s is the silhouette coefficient; a is the average distance between sample and other samples in the same cluster; b is the average distance between sample and all samples in the next nearest cluster.

The range of silhouette coefficient is between −1 and 1. Best clustering performance can be achieved when s is equal to 1. The number of clusters depends on the silhouette coefficient and practical purposes. After the production levels of shale gas wells are obtained by Kmeans algorithm, KKN approach can be used to predict the production level. Fig. 9 shows the principle of the KNN algorithm. The prediction results of production level can be helpful for making targeted EOR measures.

3.3. Shale gas production prediction model

The regression and classification ability of non-integrated models (e.g., linear regression and decision tree) are limited since minor data fluctuation may lead to large deviations. Random forest shows high accuracy and good stability, and the complexity of random forest is also lower compared with XGBoost and neural network. Therefore, the random forest is applied to develop the regression prediction model of shale gas well production. The schematic of the Random Forest approach can be found in Fig. 10.

Although there are some regression prediction models for production prediction, there lacks standardized data processing processes and methods, increasing the uncertainty of prediction results and reducing the reliability of results. Thus, this work also tries to response to these issues by performing different modelling.

Firstly, the data set obtained from the basic data processing steps (missing value interpolation, correlation analysis, outlier analysis and data standardization) is used as the benchmark for modeling, using the random forest. Secondly, the data set processed using the combination of the basic data processing steps and Spearman correlation analysis results and PCA is used for modeling and parameter adjustment. Finally, two modeling results are compared to determine the optimal model as the regression model to predict shale gas well production.

4. Application

The basic data are from the shale gas reservoirs in China. The developed framework is used to evaluate and predict the shale gas production.

4.1. Field data processing

Firstly, the cumulative gas production for 6 years are collected as the evaluation factor of shale gas production. Secondly, the

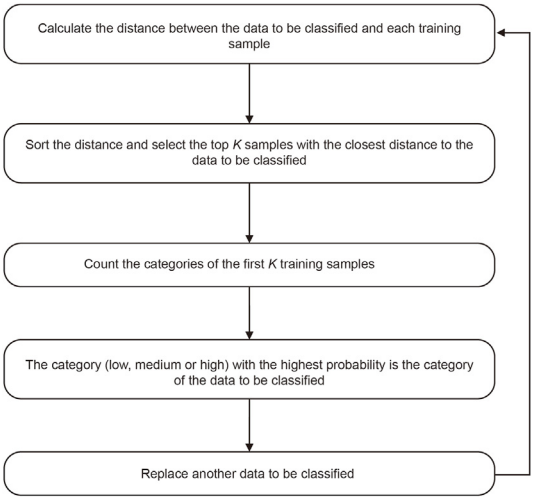


Fig. 9. The principle of the KNN algorithm.

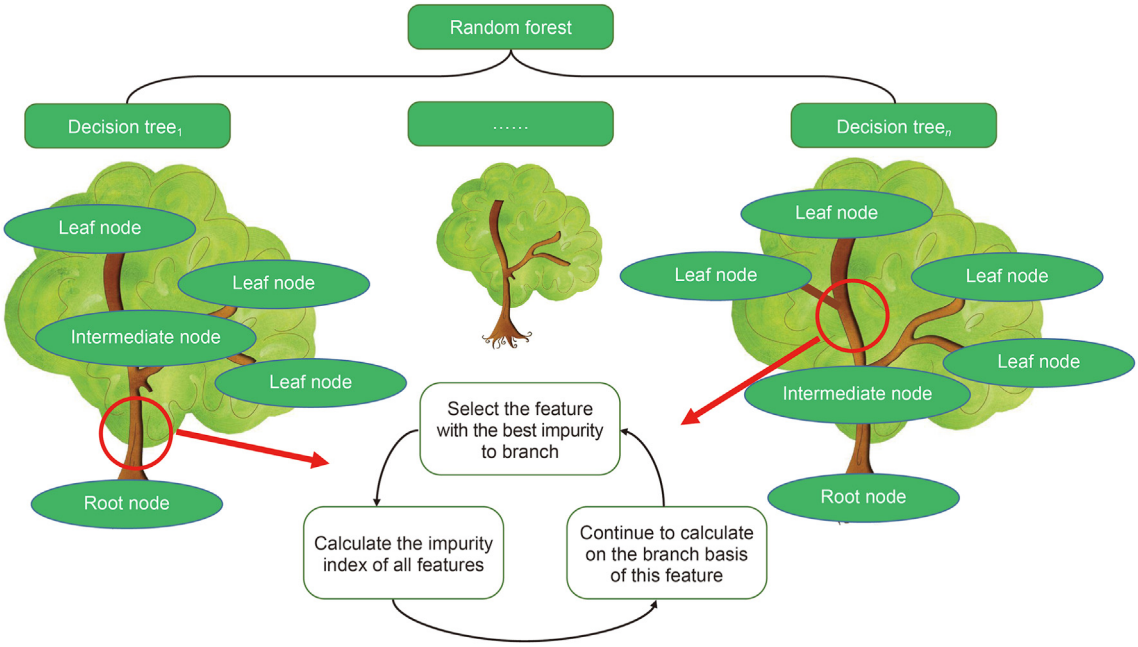


Fig. 10. Schematic of the Random Forest approach.

dominated factors are determined based on the developed factor system and field data. Thirdly, data processing is conducted to check and improve data quality. The results of Spearman correlation analysis before outlier analysis are shown in Fig. 11 and the correlation results should be checked. For instance, the shut-in time and ratio of shut-in time to total production time may only affect the cumulative production, and they have no correlation with other factors. Thus, the shut-in time and ratio of shut-in time to total production time need to be removed from the relevant factors when using the Mahalanobis distance method. Furthermore, TOC is negatively correlated with porosity and brittleness index, which is inconsistent with the Yin et al., (2018). Thus, TOC needs to be

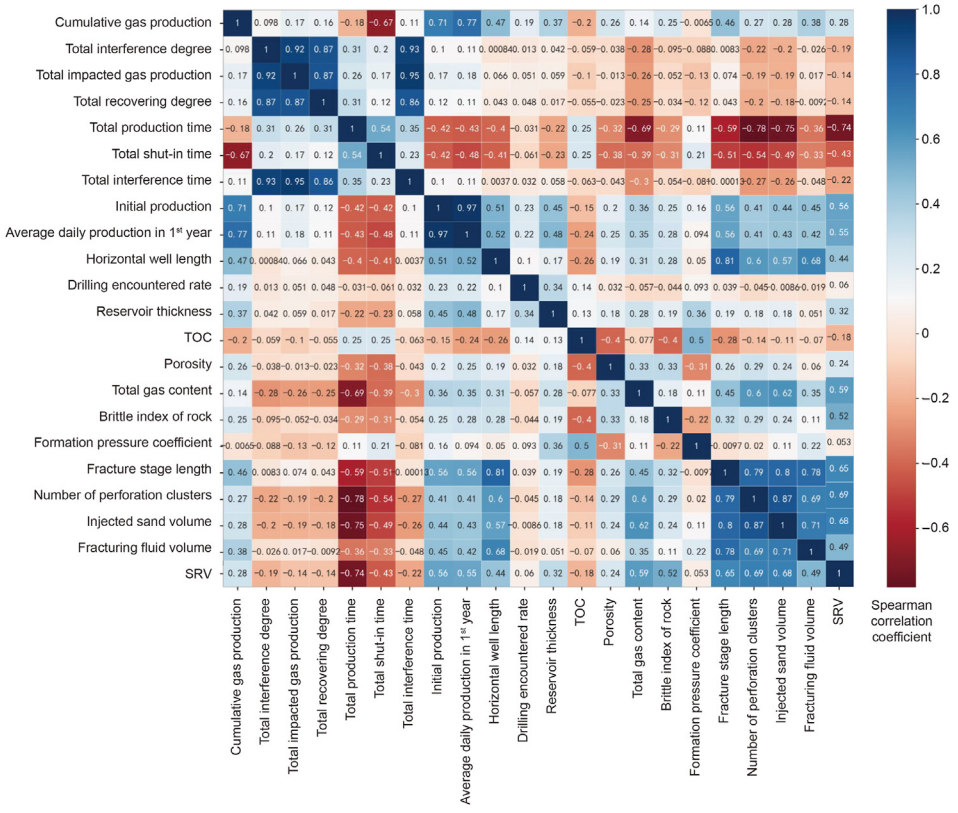


Fig. 11. Results of Spearman correlation analysis.

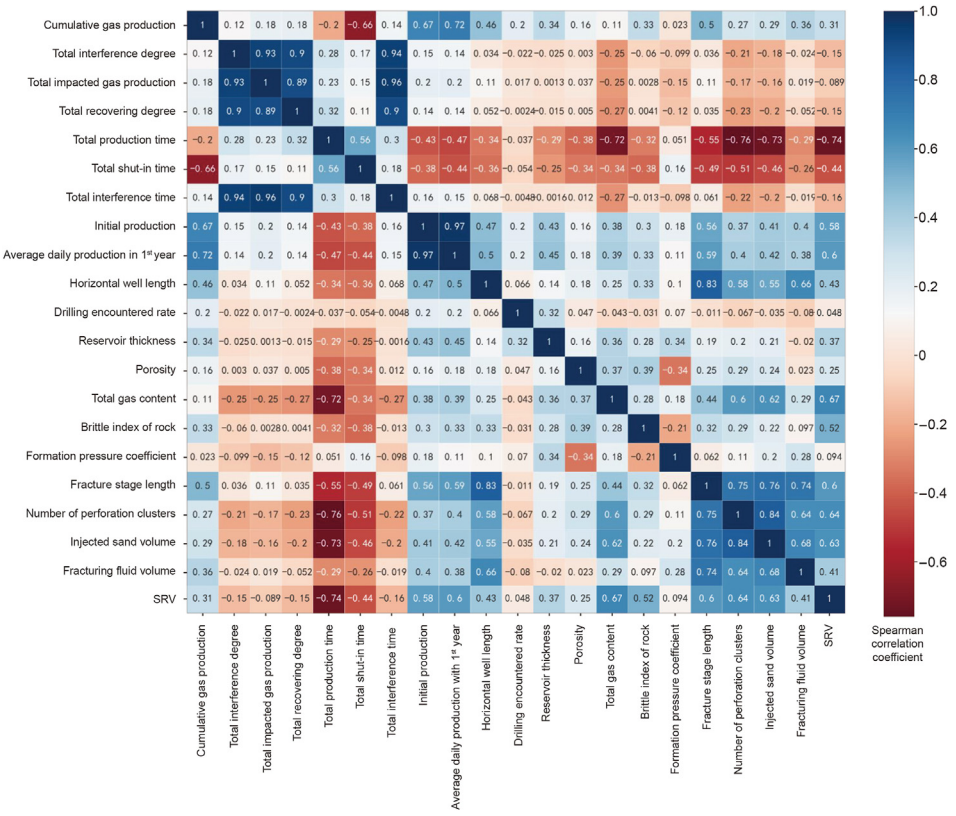


Fig. 12. Spearman correlation analysis after deleting outliers.

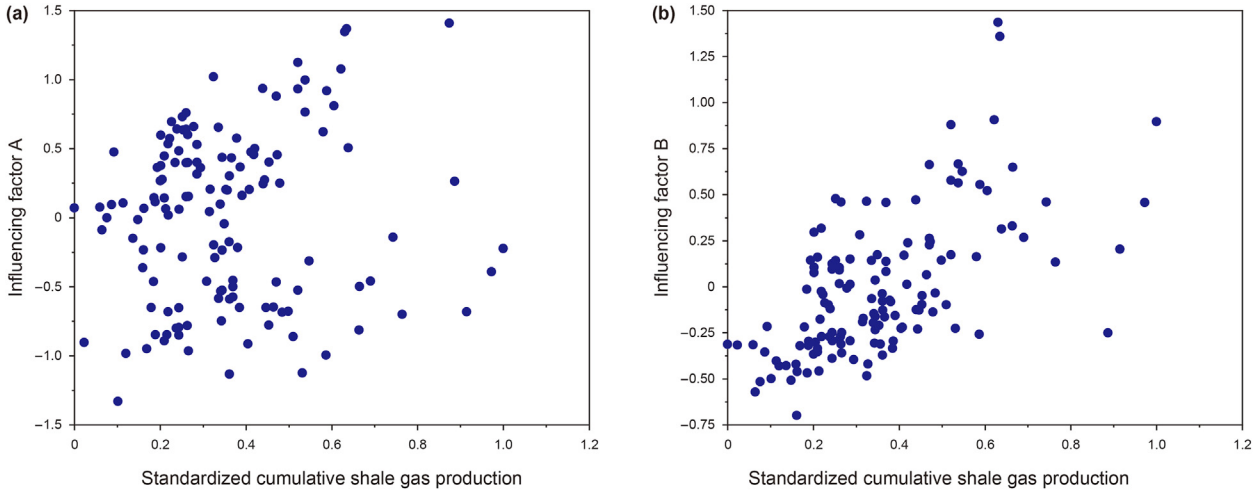


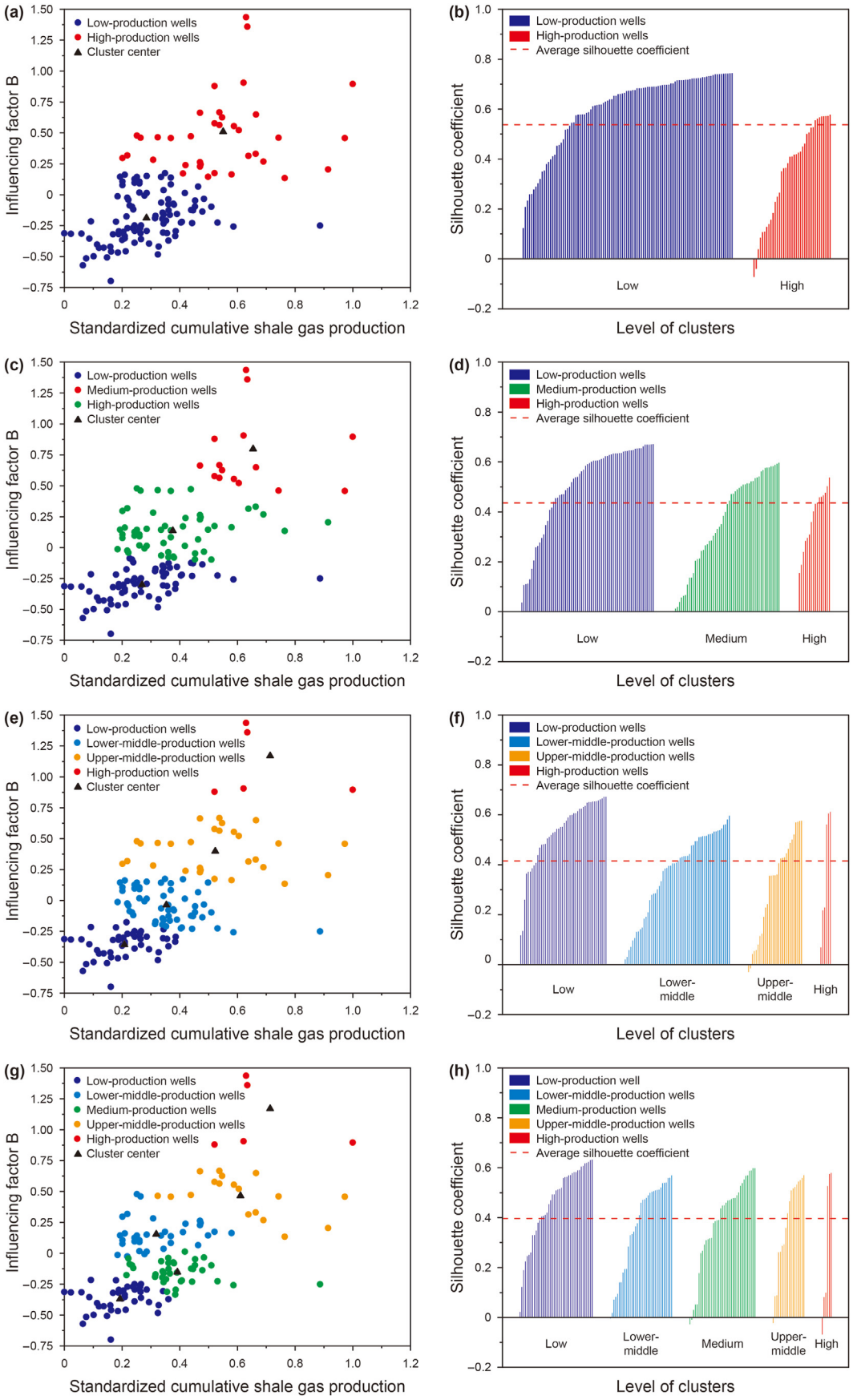
Fig. 13. Data visualization: (**a**) all factors for cumulative gas production, (**b**) the factors correlated with cumulative gas production based on Spearman correlation coefficient analysis.

eliminated from the related factors.

Then, the Mahalanobis distance method is further applied to check the abnormal values of various factors. It is found that the abnormal values continue to appear in some wells for multiple times so that these wells can be deleted. Spearman correlation analysis is performed again after deleting outliers (Fig. 12). The correlation results of only few factors are changed after deleting outliers compared with Fig. 11, indicating that the proposed data processing method used in this work is reliable.

Table 1
Silhouette coefficients corresponding to clusters.

Number of clusters	Silhouette coefficient
2	0.5373
3	0.4363
4	0.4148
5	0.3964


(a)
Low-production wells
High-production wells
Cluster center
Influencing factor B
Standardized cumulative shale gas production
(b)
Low-production wells
High-production wells
Average silhouette coefficient
Silhouette coefficient
Low
High
Level of clusters
(c)
Low-production wells
Medium-production wells
High-production wells
Cluster center
Influencing factor B
Standardized cumulative shale gas production
(d)
Low-production wells
Medium-production wells
High-production wells
Average silhouette coefficient
Silhouette coefficient
Low
Medium
High
Level of clusters
(e)
Low-production wells
Lower-middle-production wells
Upper-middle-production wells
High-production wells
Cluster center
Influencing factor B
Standardized cumulative shale gas production
(f)
Low-production wells
Lower-middle-production wells
Upper-middle-production wells
High-production wells
Average silhouette coefficient
Silhouette coefficient
Low
Lower-middle
Upper-middle
High
Level of clusters
(g)
Low-production wells
Lower-middle-production wells
Medium-production wells
Upper-middle-production wells
High-production wells
Cluster center
Influencing factor B
Standardized cumulative shale gas production
(h)
Low-production well
Lower-middle-production wells
Medium-production wells
Upper-middle-production wells
High-production wells
Average silhouette coefficient
Silhouette coefficient
Low
Lower-middle
Medium
Upper-middle
High
Level of clusters

4.2. Data visualization

Firstly, the PCA is carried out for all factors (F1) and the factors related to the target (F2) respectively. Two influencing factors (A and B) characterizing the two kinds of factors are determined. Then, the cumulative gas production of shale gas wells and the two influencing factors are used for data visualization (Fig. 13). The data dispersion is minor by using the influencing factor (B) (Fig. 13a and b), indicating the clustering performance is also better (Fig. 13b) and the modelling accuracy becomes higher.

4.3. Shale gas production level evaluation

The Kmeans algorithm is used to perform clustering for the F2. The cumulative gas production is used as the classification evaluation standard, and the silhouette coefficient is selected as the clustering index. Firstly, the number of clusters is divided into 2 to 5 clusters, and the silhouette coefficients from 2 to 5 clusters are shown in Table 1.

Fig. 14 shows the silhouette coefficient analysis for Kmeans clustering on sample data from 2 to 5 clusters. Left figures are the visualized clustered data for different clusters, and right figures are the silhouette coefficient histograms for different clusters. The two clusters show the highest average silhouette coefficients after clustering and the number of samples in the cluster corresponding to the low-production wells. Therefore, two clusters are preferred and achieve high accuracy for shale gas wells with low production. Although the average silhouette coefficients gradually decrease with the increase of cluster number, the silhouette coefficients of more than half of the data are still higher than the average value for three and four clusters. Thus three or four clusters can be used if low-production wells need to be investigated. For five clusters, the average silhouette coefficients is further decreased, and the cluster center difference between lower-middle and medium is minor so that the recommended final cluster number is two to four clusters.

4.4. Shale gas production prediction

The KNN algorithm is used to predict the production level based on the production level evaluation results.

Cross-validation. Cross validation is used to verify the stability and generalization of the developed model to effectively avoid the error caused by selecting different testing sets. The cross-validation accuracy for two clusters, three clusters, four clusters are 99.09%, 97.23%, and 96.27% respectively. The accuracy of random testing set (accounting for 20% data) for two clusters, three clusters, four clusters reach to 100%, 96.42%, and 96.43%. It is obvious that the validation accuracy for different clusters is very high since the shale gas production level prediction model is established on the corresponding production level evaluation model. New data can be input to forecast shale gas production level after the production levels are determined.

Shale gas production regression prediction. The data set after basic data processing is regarded as the basic data set and directly used for modelling and parameters adjustment (without considering correlation and using PCA) to ensure the accuracy (Baseline). Other complicated data processing steps are divided into three categories according to whether the correlation and PCA are considered. The first one is considering correlation and not using PCA ($Score_1$). The second one is using both correlation analysis and PCA ($Score_2$). The third one is using the PCA while the Spearman correlation analysis is not performed ($Score_3$). Three data processing methods are used for modeling and parameter adjustment respectively. The accuracy of the corresponding models is compared with Baseline, and the optimal model is selected as the regression model.

Two indicators (MSE and R^2) are regarded as the evaluation indexes during parameter adjustment (Oke et al., 2020). MSE can reflect whether the model is fitted with the correct value (Eq. (10)), and R^2 shows whether the model is fitted with enough information to capture the data quality (Eq. 11).

$$\mathrm{MSE} = \frac{1}{z}\sum_{i=1}^{z}(y_i - \widehat{y}_i)^2 \tag{9}$$

where MSE is the mean squared error; z is the number of samples in the test set; y_i is the actual value of the sample label in the test set; and $\widehat{y}_i$ is the predicted value of the sample label in the test set.

$$R^2 = 1 - \frac{\sum_{i=0}^{z}(y_i - \widehat{y}_i)^2}{\sum_{i=0}^{z}(y_i - \overline{y})^2} \tag{10}$$

where R^2 is the fitting-quality coefficient; and $\overline{y}$ is the average sample labels in the test set.

The optimal model is established using the factors related to cumulative gas production, and data set is processed using PCA to reduce the error caused by data noise and collinearity. The optimal models based on MSE and R^2 are the same (Table 2**)**. Data processing methods may be different for different data sets so that three data processing methods are required to obtain the optimal one.

The optimal model is used for cross validation. The R^2 and MSE equal 83.07% and 0.0078. 41 wells are randomly selected to predict cumulative gas production using the optimal regression model. The prediction results are shown in Fig. 15. The R^2 is 92.8% which indicates the prediction results are accurate and can be used for cumulative gas production prediction of shale gas wells.

5. Conclusions

This work develops a novel shale gas production evaluation framework, including dominated factor system of shale gas production, data visualization and data processing methods as well as production prediction model to evaluate shale gas well production more accurately. Field application is performed based on the shale gas reservoirs in China using the proposed evaluation framework.

(1) The shale gas production evaluation framework is developed to evaluate shale gas production qualitatively and quantitatively, including dominated factor system, data processing workflow, data visualization, production evaluation and prediction model.
(2) A comprehensive dominated factor system for shale gas production is developed by incorporating the full-cycle features (i.e., geological factors, drilling factors, fracturing factors, and production factors) and introducing the new indicators related to frac hits and well shut in, improve the accuracy of data-driven models.

Fig. 14. Silhouette coefficient analysis for Kmeans clustering on sample data: (**a**) clustered data visualization for two clusters, (**b**) silhouette coefficients for two clusters, (**c**) clustered data visualization with three clusters, (**d**) silhouette coefficients for three clusters, (**e**) clustered data visualization with four clusters, (**f**) silhouette coefficients for four clusters, (**g**) clustered data visualization with five clusters, (**h**) silhouette coefficients for five clusters.

Table 2
MSE and R^2 after parameter adjustment.

Cases	Evaluation index	Cross-validation		Optimal parameters				
		MSE	R^2	Number of trees	Maximum depth	Minimum samples split	Minimum samples leaf	Maximum features
Baseline	MSE	0.0125	0.5111	49	14	2	1	20
Baseline	R^2	0.0124	0.5510	111	13	2	2	20
$Score_1$	MSE	0.0150	0.4563	71	13	11	4	8
$Score_1$	R^2	0.0151	0.4556	70	7	11	4	8
$Score_2$	MSE	0.0078	0.8307	91	17	2	1	4
$Score_2$	R^2	0.0079	0.8307	91	12	2	1	4
$Score_3$	MSE	0.0121	0.4348	101	18	2	1	6
$Score_3$	R^2	0.0126	0.4583	91	9	2	1	6

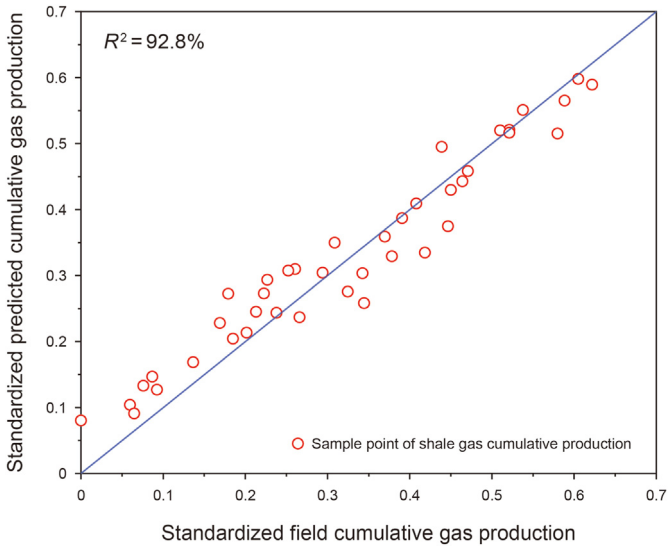


Fig. 15. The prediction results of cumulative gas production using the optimal regression model.

(3) A data processing workflow is proposed including basic data processing steps (missing value interpolation, data standardization, correlation analysis, and outlier analysis) and complicated data processing methods (Spearman correlation analysis and PCA) to improve data quality and modeling accuracy.
(4) Field application indicates that evaluation accuracy of shale gas production level exceeds 95%. Forty-one wells are randomly to predict cumulative gas production using the optimal regression model and the R^2 achieves 92.8%, which indicates the prediction results are accurate and can be used for shale gas production prediction.

The shale gas production evaluation framework can quickly evaluate and predict the shale gas production, effectively improve the development efficiency of shale gas reservoirs and reduce the development cost.

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.

Acknowledgements

This work was funded by National Natural Science Foundation of China (52004238) and China Postdoctoral Science Foundation (2019M663561).

References

Agarwal, R.G., Gardner, D.C., Kleinsteiber, S.W., Fussell, D.D., 1999. Analyzing well production data using combined-type-curve and decline-curve analysis concepts. SPE Reservoir Eval. Eng. 2 (5), 478–486. https://doi.org/10.2118/57916-PA.
Ali, T., Sheng, J., Soliman, M., 2014. New production-decline models for fractured tight and shale reservoirs. In: SPE Western North American and Rocky Mountain Joint Meeting. https://doi.org/10.2118/169537-MS.
Al-Rbeawi, S., 2020. The impact of spatial and temporal variability of anomalous diffusion flow mechanisms on reservoir performance in structurally complex porous media. J. Nat. Gas Sci. Eng. 78, 103331. https://doi.org/10.1016/j.jngse.2020.103331.
Ambrose, R.J., Hartman, R.C., Diaz-Campos, M., Akkutlu, I.Y., Sondergeldet, C.H., 2010. New pore-scale considerations for shale gas in place calculations. In: SPE Unconventional Gas Conference. https://doi.org/10.2118/131772-MS.
Arps, J.J., 1945. Analysis of decline curves. Trans. AIME 160 (1), 228–247. https://doi.org/10.2118/945228-G.
Azom, P.N., Javadpour, F., 2012. Dual-continuum modeling of shale and tight gas reservoirs. In: SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/159584-MS.
Baihly, J.D., Malpani, R., Altman, R., Lindsay, G., Clayton, R., 2015. Shale gas production decline trend comparison over time and basins—revisited. In: SPE/AAPG/SEG Unconventional Resources Technology Conference. https://doi.org/10.15530/URTEC-2015-2172464.
Bhattacharya, S., Carr, T.R., Pal, M., 2016. Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: case studies from the Bakken and Mahantango-Marcellus Shale, USA. J. Nat. Gas Sci. Eng. 33, 1119–1133. https://doi.org/10.1016/j.jngse.2016.04.055.
Bhattacharya, S., Ghahfarokhi, P.K., Carr, T.R., Pantaleone, S., 2019. Application of predictive data analytics to model daily hydrocarbon production using petrophysical, geomechanical, fiber-optic, completions, and surface data: a case study from the Marcellus Shale, north America. J. Pet. Sci. Eng. 176, 702–715. https://doi.org/10.1016/j.petrol.2019.01.013.
Bowker, K.A., 2007. Barnett shale gas production, Fort Worth Basin: issues and discussion. AAPG (Am. Assoc. Pet. Geol.) Bull. 91 (4), 523–533. https://doi.org/10.1306/06190606018.
Cao, C., Jia, P., Cheng, L.S., Jin, Q.S., Qi, S.C., 2022. A review on application of data-driven models in hydrocarbon production forecast. J. Pet. Sci. Eng. 110296. https://doi.org/10.1016/j.petrol.2022.110296.
Carter, R.D., 1985. Type curves for finite radial and linear gas-flow systems: constant-terminal-pressure case. SPE J. 25 (5), 719–728. https://doi.org/10.2118/12917-PA.
Chen, Y.T., Jiang, S., Zhang, D.X., Liu, C.Y., 2017. An adsorbed gas estimation model for shale gas reservoirs via statistical learning. Appl. Energy 197, 327–341. https://doi.org/10.1016/j.apenergy.2017.04.029.
Cipolla, C., Wallace, J., 2014. Stimulated reservoir volume: a misapplied concept? SPE hydraulic fracturing technology conference. https://doi.org/10.2118/168596-MS.
Clark, A.J., Lake, L.W., Patzek, T.W., 2011. Production forecasting with logistic growth models. In: SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/144790-MS.
Clarkson, C.R., Haghshenas, B., Ghanizadeh, A., Qanbari, F., Williams-Kovacs, J.D., Riazi, N., Debuhr, C., Deglint, H.J., 2016. Nanopores to megafractures: current challenges and methods for shale gas reservoir and hydraulic fracture characterization. J. Nat. Gas Sci. Eng. 31, 612–657. https://doi.org/10.1016/j.jngse.2016.01.041.
Danielsson, P.E., 1980. Euclidean distance mapping. Comput. Graph. Image Process. 14 (3), 227–278. https://doi.org/10.1016/0146-664X(80)90054-4.
Duong, A.N., 2010. An unconventional rate decline approach for tight and fracture-dominated gas wells. In: Canadian Unconventional Resources and International Petroleum Conference. https://doi.org/10.2118/137748-MS.
EIA, 2015. World shale resource assessments. https://www.eia.gov/analysis/studies/

worldshalegas/.
Eisner, L., Staněk, F., 2018. Microseismic data interpretation—what do we need to measure first? First Break 36 (2), 55–58. https://doi.org/10.3997/1365-2397.n0069.
Fan, C.H., Li, H., Qin, Q.R., He, S., Zhong, C., 2020. Geological conditions and exploration potential of shale gas reservoir in Wufeng and Longmaxi formation of southeastern Sichuan Basin, China. J. Pet. Sci. Eng. 191, 107138. https://doi.org/10.1016/j.petrol.2020.107138.
Fetkovich, M.J., 1973. Decline curve analysis using type curves. Fall Meeting of the Society of Petroleum Engineers of AIME. https://doi.org/10.2118/4629-MS.
Fulford, D.S., Blasingame, T.A., Evaluation of time-rate performance of shale wells using the transient hyperbolic relation. SPE Unconventional Resources Conference. https://doi.org/10.2118/167242-MS.
Ganzer, L.J., 2002. Simulating fractured reservoirs using adaptive dual continuum. In: SPE/DOE Improved Oil Recovery Symposium. https://doi.org/10.2118/75233-MS.
Grieser, W.V., Bray, J.M., 2007. Identification of production potential in unconventional reservoirs. Prod. Operation Symp. https://doi.org/10.2118/106623-MS.
Guindon, L., 2015. Determining interwell connectivity and reservoir complexity through fracturing pressure hits and production-interference analysis. J. Can. Pet. Technol. 54, 88–91. https://doi.org/10.2118/0315-088-JCPT.
Guo, J.C., Lu, Q.L., He, Y.W., 2022. Key issues and explorations in shale gas fracturing. Nat. Gas. Ind. 42 (8), 148–161. https://doi.org/10.3787/j.issn.1000-0976.2022.08.012 (in Chinese).
Guo, X.Y., Jin, Y., Zi, J.Y., Lin, B., 2021. Numerical investigation of the gas production efficiency and induced geomechanical responses in marine methane hydrate-bearing sediments exploited by depressurization through hydraulic fractures. Energy Fuel. 35 (22), 18441–18458. https://doi.org/10.1021/acs.energyfuels.1c02569.
Han, D., Kwon, S., 2021. Application of machine learning method of data-driven deep learning model to predict well production rate in the shale gas reservoirs. Energies 14 (12), 3629. https://doi.org/10.3390/en14123629.
He, Y.W., He, Z.Y., Tang, Y., Qin, J.Z., Song, J.J., Wang, Y., 2021. Shale gas well production evaluation and prediction based on machine learning. Oil Drill. Prod. Technol. 43 (4), 518–524. https://doi.org/10.13639/j.odpt.2021.04.016.
He, Y.W., Guo, J.C., Tang, Y., Xu, J.L., Li, Y.C., Wang, Y., Lu, Q.L., Patil, S., Rui, Z.H., Sepehrnoori, K., 2020. Interwell fracturing interference evaluation of multi-well pads in shale gas reservoirs: a case study in WY basin. In: SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/201694-MS.
He, Y.W., Xu, Y.J., Tang, Y., Qiao, Y., Yu, W., Sepehrnoori, K., 2022. Multi-phase rate transient behaviors of the multi-fractured horizontal well with complex fracture networks. J. Energy Resour. Technol. 144 (4), 043006. https://doi.org/10.1115/1.4053247.
Hu, B.W., Wang, J.G., Ma, Z.G., 2020. A fractal discrete fracture network based model for gas production from fractured shale reservoirs. Energies 13 (7), 1857. https://doi.org/10.3390/en13071857.
Hu, S.Y., Zhu, Q., Guo, J.J., Tang, B., 2017. Production rate analysis of multiple-fractured horizontal wells in shale gas reservoirs by a trilinear flow model. Environ. Earth Sci. 76 (11), 1–12. https://doi.org/10.1007/s12665-017-6728-0.
Huang, J.W., Jin, T.Y., Chai, Z., Barrufet, M., Killough, J., 2019. Compositional simulation of fractured shale reservoir with distribution of nanopores using coupled multi-porosity and EDFM method. J. Pet. Sci. Eng. 179, 1078–1089. https://doi.org/10.1016/j.petrol.2019.05.012.
Huang, S., Li, W.G., Wu, J.F., Zhang, H.J., Luo, Y.P., 2020. A study of the mechanism of nonuniform production rate in shale gas based on nonradioactive gas tracer technology. Energy Sci. Eng. 8, 2648–2658. https://doi.org/10.1002/ese3.691.
Ilk, D., Rushing, J.A., Perego, A.D., Blasingame, T.A., 2008. Exponential vs. hyperbolic decline in tight gas sands: understanding the origin and implications for reserve estimates using Arps' decline curves. In: SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/116731-MS.
Jacobs, T., 2017. Oil and gas producers find frac hits in shale wells a major challenge. J. Petrol. Technol. 69 (4), 29–34. https://doi.org/10.2118/0417-0029-JPT.
Jia, C.Q., Huang, Z.Q., Sepehrnoori, K., Yao, J., 2021b. Modification of two-scale continuum model and numerical studies for carbonate matrix acidizing. J. Pet. Sci. Eng. 197, 107972. https://doi.org/10.1016/j.petrol.2020.107972.
Jia, C.Q., Sepehrnoori, K., Huang, Z.Q., Yao, J., 2021a. Modeling and analysis of carbonate matrix acidizing using a new two-scale continuum model. SPE J. 26, 2570–2599. https://doi.org/10.2118/205012-PA.
Jia, P., Cheng, L.S., Clarkson, C.R., Huang, S.J., Wu, Y.H., Williams-Kovacs, J.D., 2018. A novel method for interpreting water data during flowback and early-time production of multi-fractured horizontal wells in shale reservoirs. Int. J. Coal Geol. 200, 186–198. https://doi.org/10.1016/j.coal.2018.11.002.
Johnson, N.L., Currie, S.M., Ilk, D., Blasingame, T.A., 2009. A simple methodology for direct estimation of gas-in-place and reserves using rate-time data. In: SPE Rocky Mountain Petroleum Technology Conference. https://doi.org/10.2118/123298-MS.
Josh, K., Lee, J., 2013. Comparison of various deterministic forecasting techniques in shale gas reservoirs. In: SPE Hydraulic Fracturing Technology Conference. https://doi.org/10.2118/163870-MS.
Langmuir, I., 1918. The adsorption of gases on plane surfaces of glass, mica and platinum. J. Am. Chem. Soc. 40 (9), 1361–1403. https://pubs.acs.org/doi/pdf/10.1021/ja02242a004.
Lawal, H., Jackson, G., Abolo, N., Flores, C., 2013. A novel approach to modeling and forecasting frac hits in shale gas wells. In: EAGE Annual Conference & Exhibition Incorporating SPE Europec. https://doi.org/10.2118/164898-MS.
Li, C.X., Kong, L.Y., Ostadhassan, M., Gentzis, T., 2019. Nanoscale pore structure characterization of tight oil formation: a case study of the bakken formation. Energy Fuel. 33 (7), 6008–6019. https://doi.org/10.1021/acs.energyfuels.9b00514.
Li, J.W., Yu, W., Wu, K., 2018. Analyzing the impact of fracture complexity on well performance and wettability alteration in Eagle Ford shale. In: Unconventional Resources Technology Conference. https://doi.org/10.15530/urtec-2018-2899349.
Liu, W.C., Zhang, Q.T., Zhu, W.Y., 2019. Numerical simulation of multi-stage fractured horizontal well in low-permeable oil reservoir with threshold pressure gradient with moving boundary. J. Pet. Sci. Eng. 178, 1112–1127. https://doi.org/10.1016/j.petrol.2019.04.033.
Liu, Y.Y., Ma, X.H., Zhang, X.W., Guo, W., Kang, L.X., Yu, R.Z., Sun, Y.P., 2021. A deep-learning-based prediction method of the estimated ultimate recovery (EUR) of shale gas wells. Petrol. Sci. 18 (5), 1450–1464. https://doi.org/10.1016/j.petsci.2021.08.007.
Makhotin, I., Orlov, D., Koroteev, D., Burnaev, E., Karapetyan, A., Antonenko, D., 2022. Machine learning for recovery factor estimation of an oil reservoir: a tool for derisking at a hydrocarbon asset evaluation. Petroleum 8 (2), 278–290. https://doi.org/10.1016/j.petlm.2021.11.005.
Makinde, I., Lee, W.J., 2017. Forecasting production of liquid rich shale (LRS) reservoirs using simple models. J. Pet. Sci. Eng. 157, 461–481. https://doi.org/10.1016/j.petrol.2017.07.049.
Mao, S.W., Siddhamshetty, P., Zhang, Z., Yu, W., Chun, T., Kwon, J.S., Wu, K., 2020. Impact of proppant pumping schedule on well production for slickwater fracturing. SPE J. 26 (1), 342–358. https://doi.org/10.2118/204235-PA.
Maraggi, L.R., Lavia, M.A., Savioli, G.B., 2016. Production decline analysis in the vaca muerta formation. The application of modern time-rate relations using public data. In: SPE Argentina Exploration and Production of Unconventional Resources Symposium. https://doi.org/10.2118/180971-MS.
Mattar, L., Moghadam, S., 2009. Modified power law exponential decline for tight gas. In: Canadian International Petroleum Conference. https://doi.org/10.2118/2009-198.
McClure, M.W., Babazadeh, M., Shiozawa, S., Huang, J., 2016. Fully coupled hydromechanical simulation of hydraulic fracturing in 3D discrete-fracture networks. SPE J. 21 (4), 1302–1320. https://doi.org/10.2118/173354-PA.
McKenna, S.A., Fusco, F., Eck, B.J., 2014. Water demand pattern classification from smart meter data. Procedia Eng. 70, 1121–1230. https://doi.org/10.1016/j.proeng.2014.02.124.
Mehana, M., Guiltinan, E., Vesselinov, V., Middleton, R., Hyman, J.D., Kang, Q.J., Viswanathan, H., 2021. Machine-learning predictions of the shale wells' performance. J. Nat. Gas Sci. Eng. 88, 103819. https://doi.org/10.1016/j.jngse.2021.103819.
Miao, Y.N., Zhao, C.J., Wu, K.L., Li, X.F., 2019. Analysis of production prediction in shale reservoirs: influence of water film in inorganic matter. J. Nat. Gas Sci. Eng. 63, 1–9. https://doi.org/10.1016/j.jngse.2019.01.002.
Mishra, S., 2012. A new approach to reserves estimation in shale gas reservoirs using multiple decline curve analysis models. In: SPE Eastern Regional Meeting. https://doi.org/10.2118/161092-MS.
Moinfar, A., Narr, W., Hui, M.H., Mallison, B., Lee, S.H., 2011. Comparison of discrete-fracture and dual-permeability models for multiphase flow in naturally fractured reservoirs. SPE Reserboir Simul. Symp. https://doi.org/10.2118/142295-MS.
Moos, D., Lacazette, A., Vassilellis, G.D., Cade, R., Franquet, J.A., Bourtembourg, E., Daniel, G., 2011. Predicting shale reservoir response to stimulation: the Mallory 145 multi-well project. In: SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/145849-MS.
Ning, Y., Zhang, K.Y., He, S., Chen, T.L., Wang, H.Y., Qin, G., 2019. Numerical modeling of gas transport in shales to estimate rock and fluid properties based on multiscale digital rocks. Energy Proc. 158, 6093–6098. https://doi.org/10.1016/j.egypro.2019.01.505.
Ojo, K.P., Osisanya, S.O., 2006. Material balance revisited. In: Nigeria Annual International Conference and Exhibition. https://doi.org/10.2118/105982-MS.
Oke, E.O., Nwosu-Obieogu, K., Ude, J.C., 2020. Experimental study and exergy efficiency prediction of three-leaved yam (Dioscorea dumetorum) starch drying. Int. J. Exergy 33 (4), 427–443. https://doi.org/10.1504/IJEX.2020.111690.
Ouarda, S., Souad, B., 2022. Euclidean distance versus Manhattan distance for skin detection using the SFA database. Int. J. Biometrics. 14 (1), 46–60. https://doi.org/10.1504/IJBM.2022.119553.
Palacio, J.C., Blasingame, T.A., 1993. Unavailable-decline-curve analysis with type curves-analysis of gas well production data. Low Permrability Reservoir. Sym. https://doi.org/10.2118/25909-MS.
Pang, H., Pang, X.Q., Dong, L., Zhao, X., 2018. Factors impacting on oil retention in lacustrine shale: permian lucaogou formation in jimusaer depression, junggar basin. J. Pet. Sci. Eng. 163, 79–90. https://doi.org/10.1016/j.petrol.2017.12.080.
Patil, V.V., McPherson, B.J., Priewisch, A., Moore, J., Moodie, N., 2017. Factors affecting self-sealing of geological faults due to CO_2-leakage. Greenh Gases 7 (2), 273–294. https://doi.org/10.1002/ghg.1673.
Qin, J.Z., Zhong, Q.H., Tang, Y., Yu, W., Sepehrnoorib, K., 2022. Well interference evaluation considering complex fracture networks through pressure and rate transient analysis in unconventional reservoirs. Petrol. Sci. https://doi.org/10.1016/j.petsci.2022.09.029.
Ross, D.J., Bustin, R.M., 2007. Impact of mass balance calculations on adsorption capacities in microporous shale gas reservoirs. Fuel 86 (17), 2696–2706. https://doi.org/10.1016/j.fuel.2007.02.036.

Sardinha, C., Petr, C., Lehmann, J., Pyecroft, J., 2014. Determining interwell connectivity and reservoir complexity through frac pressure hits and production interference analysis. In: SPE/CSUR Unconventional Resources Conference–Canada. https://doi.org/10.2118/171628-MS.

Shi, F., 2021. XFEM-based numerical modeling of well performance considering proppant transport, embedment, crushing and rock creep in shale gas reservoirs. J. Pet. Sci. Eng. 201, 108523. https://doi.org/10.1016/j.petrol.2021.108523.

Song, L.S., Martin, K., Carr, T.R., Ghahfarokhi, P.K., 2019. Porosity and storage capacity of Middle Devonian shale: a function of thermal maturity, total organic carbon, and clay content. Fuel 241, 1036–1044. https://doi.org/10.1016/j.fuel.2018.12.106.

Syed, F.I., AlShamsi, A., Dahaghi, A.K., Neghabhan, S., 2020. Application of ML & AI to model petrophysical and geo-mechanical properties of shale reservoirs–A systematic literature review. Petroleum. 8 (2), 158–166 doi:10.1016/j.petlm.2020.12.001.

Tian, W., Liu, H.Q., 2020. Insight into the adsorption of methane on gas shales and the induced shale swelling. ACS Omega 5 (49), 31508–31517. https://doi.org/10.1021/acsomega.0c02980.

Tontiwachwuthikul, P., Chan, C.W., Zeng, F., Liang, Z.W., Sema, T., Chao, M., 2020. Recent progress and new developments of applications of artificial intelligence (AI), knowledge-based systems (KBS), and machine learning (ML) in the petroleum industry. Petroleum 6 (4), 319–320. https://doi.org/10.1016/j.petlm.2020.08.001.

Umar, I.A., Negash, B.M., Quainoo, A.K., Ayoub, M.A., 2021. An outlook into recent advances on estimation of effective stimulated reservoir volume. J. Nat. Gas Sci. Eng. 88 (3), 103822. https://doi.org/10.1016/j.jngse.2021.103822.

Valko, P.P., 2009. Assigning value to stimulation in the Barnett Shale: a simultaneous analysis of 7000 plus production hystories and well completion records. In: SPE Hydraulic Fracturing Technology Conference. https://doi.org/10.2118/119369-MS.

Vikara, D., Remson, D., Khanna, V., 2020. Machine learning-informed ensemble framework for evaluating shale gas production potential: case study in the Marcellus Shale. J. Nat. Gas Sci. Eng. 84, 103679. https://doi.org/10.1016/j.jngse.2020.103679.

Wang, D.D., Shao, L.Y., Li, Z.X., Li, M.P., Lv, D.W., Liu, H.Y., 2016. Hydrocarbon generation characteristics, reserving performance and preservation conditions of continental coal measure shale gas: a case study of Mid-Jurassic shale gas in the Yan'an Formation, Ordos Basin. J. Pet. Sci. Eng. 145, 609–628. https://doi.org/10.1016/j.petrol.2016.06.031.

Wang, F.S., Chen, J.G., Liu, F.R., 2021a. Keyframe generation method via improved clustering and silhouette coefficient for video summarization. J. Web Eng. 20 (1), 147–170. https://doi.org/10.13052/jwe1540-9589.2018.

Wang, H., Chen, L., Qu, Z.G., Yin, Y., Kang, Q.J., Yu, B., Tao, W.Q., 2020a. Modeling of multi-scale transport phenomena in shale gas production—a critical review. Appl. Energy 262, 114575. https://doi.org/10.1016/j.apenergy.2020.114575.

Wang, J.G., Gu, D.H., Guo, W., Zhang, H.J., Yang, D.Y., 2019. Determination of total organic carbon content in shale formations with regression analysis. J. Energy Resour. Technol. 141 (1), 012907. https://doi.org/10.1115/1.4040755.

Wang, K., Li, H.T., Wang, J.C., Jiang, B.B., Bu, C.Z., Zhang, Q., Luo, W., 2017a. Predicting production and estimated ultimate recoveries for shale gas wells: a new methodology approach. Appl. Energy 206, 1416–1431. https://doi.org/10.1016/j.apenergy.2017.09.119.

Wang, K., Liu, H., Luo, J., Wu, K.L., Chen, Z.X., 2017b. A comprehensive model coupling embedded discrete fractures, multiple interacting continua, and geomechanics in shale gas reservoirs with multiscale fractures. Energy Fuel. 31 (8), 7758–7776. https://doi.org/10.1021/acs.energyfuels.7b00394.

Wang, K., Jiang, B.B., Li, H.T., Liu, Q., Bu, C.Z., Wang, Z.Q., Tan, Y.S., 2020b. Rapid and accurate evaluation of reserves in different types of shale-gas wells: production-decline analysis. Int. J. Coal Geol. 218, 103359. https://doi.org/10.1016/j.coal.2019.103359.

Wang, L., Yang, S.L., Meng, Z., Chen, Y.Z., Qian, K., Han, W., Wang, D.F., 2018. Time-dependent shape factors for fractured reservoir simulation: effect of stress sensitivity in matrix system. J. Pet. Sci. Eng. 163, 556–569. https://doi.org/10.1016/j.petrol.2018.01.020.

Wang, S., Qin, C.X., Feng, Q.H., Javadpour, F., Rui, Z.H., 2021b. A framework for predicting the production performance of unconventional resources using deep learning. Appl. Energy 295, 117016. https://doi.org/10.1016/j.apenergy.2021.117016.

Wang, Y., Liu, L.F., Cheng, H.F., 2021c. Gas adsorption characterization of pore structure of organic-rich shale: insights into contribution of organic matter to shale pore network. Nat. Resour. Res. 30 (3), 2377–2395. https://doi.org/10.1007/s11053-021-09817-5.

Wei, S.M., Jin, Y., Xia, Y., Lin, B.T., 2020. The flowback and production analysis in subsaturated fractured shale reservoirs. J. Pet. Sci. Eng. 186, 106694. https://doi.org/10.1016/j.petrol.2019.106694.

Wu, K.L., Chen, Z.X., Li, X.F., Xu, J.Z., Li, J., Wang, K., Wang, H., Wang, S.H., Dong, X.H., 2017. Flow behavior of gas confined in nanoporous shale at high pressure: real gas effect. Fuel 205, 173–183. https://doi.org/10.1016/j.fuel.2017.05.055.

Wu, Y.H., Cheng, L.S., Ma, L.Q., Huang, S.J., Fang, S.D., Killough, J., Jia, P., Wang, S.R., 2021. A transient two-phase flow model for production prediction of tight gas wells with fracturing fluid-induced formation damage. J. Pet. Sci. Eng. 199, 108351. https://doi.org/10.1016/j.petrol.2021.108351.

Xu, S.Q., Feng, Q.H., Wang, S., Javadpour, F., Li, Y.Y., 2018. Optimization of multistage fractured horizontal well in tight oil based on embedded discrete fracture model. Comput. Chem. Eng. 117, 291–308. https://doi.org/10.1016/j.compchemeng.2018.06.015.

Xue, L., Gu, S.H., Jiang, X.E., Liu, Y.T., Yang, C., 2021a. Ensemble-based optimization of hydraulically fractured horizontal well placement in shale gas reservoir through Hough transform parameterization. Petrol. Sci. 18, 1–13. https://doi.org/10.1007/s12182-021-00560-3.

Xue, L., Liu, Y.T., Xiong, Y.F., Liu, Y.L., Cui, X.H., Lei, G., 2021b. A data-driven shale gas production forecasting method based on the multi-objective random forest regression. J. Pet. Sci. Eng. 196, 107801. https://doi.org/10.1016/j.petrol.2020.107801.

Yan, B.C., Chen, B.L., Harp, D.R., Jia, W., Pawar, R.J., 2022a. A robust deep learning workflow to predict multiphase flow behavior during geological CO_2 sequestration injection and post-injection periods. J. Hydrol. 607, 127542. https://doi.org/10.48550/arXiv.2107.07274.

Yan, B.C., Harp, D.R., Chen, B.L., Pawar, R., 2022b. A physics-constrained deep learning model for simulating multiphase flow in 3D heterogeneous porous media. Fuel 313, 122693. https://doi.org/10.1016/j.fuel.2021.122693.

Yin, S., Lv, D.W., Jin, L., Ding, W.L., 2018. Experimental analysis and application of the effect of stress on continental shale reservoir brittleness. J. Geophys. Eng. 15 (2), 478–494. https://doi.org/10.1088/1742-2140/aaa5d2.

Yu, S.Y., Miocevic, D.J., 2013. An improved method to obtain reliable production and EUR prediction for wells with short production history in tight/shale reservoirs. In: SPE/AAPG/SEG Unconventional Resources Technology Conference. https://doi.org/10.1190/urtec2013-003.

Yu, W., Xu, Y.F., Weijermars, R., Wu, K., Sepehrnoori, K., 2018. A numerical model for simulating pressure response of well interference and well performance in tight oil reservoirs with complex-fracture geometries using the fast embedded-discrete-fracture-model method. SPE Reservoir Eval. Eng. 21, 489–502. https://doi.org/10.2118/184825-PA.

Yuan, Y.Z., Qi, Z.L., Chen, Z.X., Yan, W.D., Zhao, Z.H., 2020. Production decline analysis of shale gas based on a probability density distribution function. J. Geophys. Eng. 17 (2), 365–376. https://doi.org/10.1093/jge/gxz122.

Zeng, J., Liu, J.S., Li, W., Leong, Y.K., Elsworth, D., Guo, J.C., 2021. Shale gas reservoir modeling and production evaluation considering complex gas transport mechanisms and dispersed distribution of kerogen. Petrol. Sci. 18 (1), 195–218. https://doi.org/10.1007/s12182-020-00495-1.

Zhang, H., Rietz, D., Cagle, A., Cocco, M., Lee, J., 2016. Extended exponential decline curve analysis. J. Nat. Gas Sci. Eng. 36, 402–413. https://doi.org/10.1016/j.jngse.2016.10.010.

Zhang, Q., He, F., He, Y.W., 2022. Well interference evaluation and prediction of shale gas wells based on machine learning. Reservoir Eval. Eng. 12 (3), 487–495. https://doi.org/10.13809/j.cnki.cn32-1825/te.2022.03.011.

Zhang, Y., Di, Y., Yu, W., Sepehrnoori, K., 2017. A comprehensive model for investigation of CO_2-EOR with nanopore confinement in the Bakken tight oil reservoir. In: SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/187211-MS.

Zhao, J.Z., Ren, L., Jiang, T.X., Hu, D.F., Wu, L.Z., Wu, J.F., Y, et al., 2021. Ten years of gas shale fracturing in China: review and prospect. NGI 41 (8), 121–142. https://doi.org/10.1016/j.ngib.2022.03.002.

Zuo, L.H., Yu, W., Wu, K., 2016. A fractional decline curve analysis model for shale gas reservoirs. Int. J. Coal Geol. 163, 140–148. https://doi.org/10.1016/j.coal.2016.07.006.