

Article

A Comprehensive Evaluation of Shale Oil Reservoir Quality

Fuchun Tian ^{1,2}, Yongqiang Fu ¹, Xuwei Liu ^{1,*}, Dongping Li ¹, Yunpeng Jia ¹, Lifei Shao ¹, Liyong Yang ¹, Yudong Zhao ¹, Tao Zhao ¹, Qiwu Yin ¹ and Xiaoting Gou ¹

¹ Petroleum Engineering Research Institute, PetroChina Dagang Oilfield Company, Tianjin 300280, China; tianfuchun@petrochina.com.cn (F.T.); fuyq@petrochina.com.cn (Y.F.); lidping@petrochina.com.cn (D.L.); jiaypeng@petrochina.com.cn (Y.J.); shaolfei@petrochina.com.cn (L.S.); yanglyong@petrochina.com.cn (L.Y.); zhaoydong@petrochina.com.cn (Y.Z.); zhaotao001@petrochina.com.cn (T.Z.); yinqwu@petrochina.com.cn (Q.Y.); gouxtng@petrochina.com.cn (X.G.)

² College of Petroleum Engineering, China University of Petroleum, Beijing 102249, China

* Correspondence: 202121000998@stu.swpu.edu.cn

Abstract: To enhance the accuracy of the comprehensive evaluation of reservoir quality in shale oil fractured horizontal wells, the Pearson correlation analysis method was employed to study the correlations between geological parameters and their relationship with production. Through principal component analysis, the original factors were linearly combined into principal components with clear and specific physical meanings, aiming to eliminate correlations among factors. Furthermore, Gaussian membership functions were applied to delineate fuzzy levels, and the entropy weight method was used to determine the weights of principal components, establishing a fuzzy comprehensive evaluation model for reservoir quality. Without using principal component analysis, the correlation coefficient between production and evaluation results for the 40 wells in the Cangdong shale oil field was only 0.7609. However, after applying principal component analysis, the correlation coefficient increased to 0.9132. Field application demonstrated that the average prediction accuracy for the cumulative oil production per kilometer of fractured length over 12 months for the 10 applied wells was 91.8%. The proposed comprehensive evaluation method for reservoir quality can guide the assessment of reservoir quality in shale oil horizontal wells.

Keywords: shale oil; principal component analysis; reservoir quality; fuzzy comprehensive evaluation



Citation: Tian, F.; Fu, Y.; Liu, X.; Li, D.; Jia, Y.; Shao, L.; Yang, L.; Zhao, Y.; Zhao, T.; Yin, Q.; et al. A

Comprehensive Evaluation of Shale Oil Reservoir Quality. *Processes* **2024**, *12*, 472. <https://doi.org/10.3390/pr12030472>

Academic Editor: Qingbang Meng

Received: 13 January 2024

Revised: 23 February 2024

Accepted: 23 February 2024

Published: 26 February 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Following the unconventional energy revolution, shale oil development has become a crucial support for increased petroleum production [1]. To overcome the characteristics of low porosity, low permeability, and tightness in reservoirs and achieve the economic development of shale oil, extensive hydraulic fracturing modifications are necessary [2]. Reservoir quality determines the success of hydraulic fracturing measures for production enhancement [3,4]. Many factors influence reservoir quality, which can be categorized into source rock lithology and fracability [5]. Source rock lithology reflects the oil content and organic maturity of the reservoir, while fracability determines the distribution and complexity of reservoir fractures after hydraulic fracturing. Some factors are interrelated, and their contributions to reservoir quality (production) are unequal [6]. Only a few factors cannot fully reflect the true condition of the reservoir. However, when considering multiple factors, it is challenging to establish an effective comprehensive evaluation model for accurately assessing reservoir quality [7].

To overcome the aforementioned challenges, various artificial intelligence methods have been developed for comprehensive reservoir quality assessment. These methods include Artificial Neural Networks (ANNs) [8], Support Vector Machines (SVMs) [9], Genetic Algorithms (GAs) [10], and Fuzzy Logic Systems (FLSs) [11]. ANNs have the capacity to capture intricate non-linear relationships within data. Nevertheless, an extensive corpus

of annotated data is requisite for training, a task that may prove challenging in certain domains [12]. Concurrently, the model lacks interpretability, rendering comprehension of the decision-making process arduous [13]. SVMs exhibit efficacy within high-dimensional spaces and are applicable to intricate decision boundaries. However, akin to ANNs, the interpretability of the model poses challenges, especially within high-dimensional spaces. GAs are well-suited for optimization quandaries and the exploration of solution spaces. However, there exists a propensity for entrapment in local optima under specific circumstances. FLSs possess the capability to model uncertainty and imprecision within data, offering interpretability through linguistic rules [14,15]. The formulation of pertinent fuzzy rules necessitates domain-specific expertise, potentially impeding the handling of complex non-linear relationships inherent in the data [16].

Zoveidavianpoor et al. [17] used Gaussian distribution membership functions to input seven geological factors such as permeability and skin coefficient, determined the weights according to expert experience, and established a fuzzy score for preselected hydraulic fracturing wells. Davarpanah et al. [18] analyzed and compared five indicators affecting the effect of hydraulic fracturing through the Fuzzy Analytic Hierarchy Process (FAHP) and Fuzzy Technique for Order Preference by Similarity to Ideal Solution (FTOPSIS) techniques. Gou et al. [15] divided geological and construction parameters into six categories, screened and retained the parameters with the greatest correlation with production as the model input through grey clustering analysis, and established a reservoir quality evaluation model of fractured wells by using the normal distribution membership function. Xie et al. [19] established a shale hydraulic fracturing brittleness evaluation model using the FAHP, taking into account factors such as shale brittle mineral content, porosity, and confining pressure. However, the above-mentioned studies directly analyze factors without considering the reduction in model accuracy caused by multicollinearity resulting from the strong correlation among production-influencing factors. Additionally, a limited number of input parameters fail to adequately represent the geological features of the reservoir, significantly impacting the accuracy of the model [20].

In summary, the geological parameters determining reservoir quality are intricate, and they possess fuzzy characteristics that are challenging to accurately describe. In current fuzzy comprehensive evaluation models, the weights of various factors are provided by empirical coefficients or experts. Moreover, factors related to reservoir fracturing capability are seldom considered, diminishing the accuracy and scientific validity of the model.

In this research, based on data samples from the Cangdong shale oil field, geological parameters influencing reservoir quality were selected using geological information and well logging data. Through Pearson correlation analysis, the correlations between geological parameters and production were determined. Additionally, by employing the principal component analysis method, factors influencing production were linearly combined into multiple comprehensive indicators, thereby enhancing the model's accuracy. Utilizing feature data after dimension reduction, a principal component membership matrix was established. Combining Gaussian membership functions for delineating fuzzy levels and incorporating the entropy weight method to determine the weights of each principal component, a comprehensive evaluation model for the reservoir quality of shale oil fractured horizontal wells was developed, thereby facilitating the high-quality development of shale oil.

2. Analysis of Main Control Factors of Shale Oil Reservoir Quality

2.1. Extraction of Geological Feature Parameters

To accurately assess the reservoir quality of shale oil horizontal wells, it is imperative to identify the key influencing factors. The evaluation indicators for shale oil reservoir quality can be categorized into two aspects: source lithology and fracability. Source lithology reflects the physical–chemical characteristics and hydrocarbon content of shale oil reservoirs, encompassing natural gamma radiation (GR), total organic carbon content (TOC), organic matter thermal maturity (R_o), the pyrolysis parameter (S_1), and the oil saturation index (OSI). Fracability, on the other hand, signifies the capacity of shale oil

reservoirs to form complex fractures after hydraulic fracturing. Characterizing indicators include the mineral brittleness index (MBI), Young's modulus (E), Poisson's ratio (ν), minimum horizontal principal stress (σ_h), and the coefficient of variation of horizontal principal stress (σ_{dif}). High-quality shale oil reservoirs necessitate both a high abundance of hydrocarbons and excellent fracability. Consequently, a comprehensive evaluation considering both organic source rock properties and fracability characteristics is imperative for assessing shale oil reservoir quality.

The maturity and abundance of organic matter are focal points in the analysis of organic source rock properties. Liquid hydrocarbons at low maturity stages exhibit higher molecular weights and increased viscosity and density values, making their migration through micro-nano pores in shale challenging and resulting in lower shale oil production. As maturity increases, oil undergoes secondary cracking, yielding hydrocarbons with lower polarity, smaller molecular weights, and improved fluidity. Natural gamma radiation (GR) effectively identifies hydrocarbon source rocks, and within the context of organic source rock properties, both GR and thermal maturity (R_o) reflect hydrocarbon maturity, information obtainable from geological data [21]. R_o indicates the hydrocarbon generation potential of shale oil and is instrumental in evaluating its fluidity. Only within the oil and gas generation window can a significant quantity of oil and gas be formed. S_1 represents the residual hydrocarbon content, and its value, relative to total organic carbon (TOC), characterizes the heterogeneity of hydrocarbon accumulation in shale formations. TOC can be extracted from geological data, while S_1 and the oil saturation index (OSI) can be obtained from well logging data. Together, they allow the elucidation of the abundance of hydrocarbons in the shale formation.

Shale reservoirs, characterized by low permeability and tightness, necessitate the construction of a complex network of fractures to establish efficient fluid pathways toward the well for economically viable shale oil development. The fracability indicators of the reservoir reflect the degree and capability of fracture network formation through hydraulic fracturing [22]. Poisson's ratio and Young's modulus are key parameters for assessing shale stress and fractures [23]. However, the prediction and computation of the brittleness index often rely on rock physics parameters obtained from well logging data, without considering the influence of brittle mineral content [24]. Research indicates that as the horizontal principal stress decreases and the coefficient of variation of horizontal principal stress increases, the primary fractures formed after hydraulic fracturing become larger and more extended. These fractures are more prone to communicating with natural fractures in the deeper reservoir, leading to the creation of a complex fracture network. A higher fracture network index indicates a more intricate network. However, a larger coefficient of variation of horizontal principal stress makes the formation of a complex fracture network near the wellbore more challenging. Among the fracability indicators, the mineral brittleness index can be estimated from mineral content, while the remaining indicators can be calculated from well logging data [25].

Production is the most direct indicator of reservoir quality. However, short-term production may not adequately reflect reservoir quality due to variations in geological parameters and production systems among wells. Simultaneously, to mitigate the impact of horizontal segment length on production, the cumulative oil production over 12 months per kilometer (Q_{12mon}) was employed as the reservoir quality assessment indicator. Forty wells from the Cangdong sag shale oil field were selected as the subjects for this study (Table 1).

Table 1. Parameters of the Cangdong sag shale oil field.

Well No.	GR (API)	TOC (%)	Ro	S ₁ (mg/g)	OSI	MBI	E (MPa)	ν	σ_h (MPa)	σ_{dif}	Q _{12mon} (t/km)
W-1	104.87	2.67	1.12	2.30	0.86	0.80	37,671.1	0.228	83.85	0.28	4449.0
W-2	98.44	2.23	1.55	3.05	1.37	0.78	37,379.8	0.228	83.04	0.28	2920.6
W-3	107.60	1.94	1.28	2.74	1.41	0.78	37,671.1	0.228	83.85	0.28	3167.9
W-4	91.07	4.48	1.03	1.97	0.44	0.76	27,198.6	0.231	56.84	0.24	3466.0
W-5	99.36	3.35	1.19	2.38	0.71	0.80	25,681.7	0.229	91.56	0.14	3977.1
W-6	96.49	5.05	1.38	9.90	1.96	0.80	25,681.7	0.229	91.56	0.14	6744.6
W-7	85.40	4.90	1.54	12.10	2.47	0.80	27,895.0	0.220	95.90	0.13	9984.7
W-8	114.31	3.62	1.87	10.50	2.90	0.80	27,895.0	0.220	95.90	0.13	7106.0
W-9	91.76	4.04	1.67	4.95	1.23	0.81	35,254.0	0.228	83.25	0.27	8604.5
W-10	95.09	3.00	1.55	3.57	1.19	0.81	33,530.4	0.229	83.55	0.26	7499.8
W-11	92.10	3.83	1.73	4.51	1.18	0.80	36,607.8	0.228	79.76	0.28	7321.0
W-12	108.99	4.20	1.72	3.24	0.77	0.76	32,778.4	0.229	81.96	0.26	2928.3
W-13	102.99	3.70	1.12	4.67	1.26	0.76	36,914.7	0.227	88.41	0.27	2435.9
W-14	85.28	3.96	1.18	4.38	1.11	0.76	37,138.5	0.228	85.19	0.28	2271.4
W-15	91.12	3.51	1.15	2.97	0.85	0.76	34,369.7	0.228	84.48	0.26	2047.5
W-16	98.14	5.00	1.52	6.13	1.23	0.80	33,530.4	0.229	83.55	0.26	8415.2
W-17	97.30	2.94	1.33	6.76	0.97	0.77	33,332.3	0.225	76.08	0.21	4078.9
W-18	94.87	3.43	1.64	5.58	1.21	0.77	34,092.1	0.226	72.42	0.24	4539.6
W-19	97.38	3.48	1.81	7.91	0.58	0.78	30,028.6	0.229	68.53	0.22	5777.9
W-20	106.72	3.35	1.73	7.86	1.01	0.78	29,855.9	0.229	70.53	0.21	5568.2
W-21	101.61	3.36	1.54	7.17	1.08	0.78	33,080.5	0.224	76.73	0.23	5448.1
W-22	99.29	3.51	1.24	6.99	1.38	0.77	34,335.2	0.225	74.19	0.20	3826.4
W-23	105.55	3.31	1.64	6.20	1.31	0.77	33,139.1	0.224	71.22	0.21	4882.8
W-24	99.64	3.72	1.71	8.44	1.54	0.78	33,364.9	0.223	75.24	0.23	6355.9
W-25	97.01	3.19	1.39	6.65	0.94	0.78	33,379.5	0.227	70.63	0.20	5214.2
W-26	100.66	3.45	1.72	6.56	1.52	0.77	34,079.4	0.225	74.28	0.22	4514.4
W-27	100.05	3.18	1.35	6.73	1.18	0.77	32,492.1	0.226	73.43	0.21	4613.0
W-28	100.15	3.48	1.77	6.69	1.48	0.77	32,724.0	0.226	73.93	0.21	4622.5
W-29	93.73	3.42	1.24	6.13	1.04	0.77	33,020.2	0.227	70.75	0.22	4366.2
W-30	98.90	3.12	1.54	6.95	1.32	0.77	32,904.5	0.225	75.65	0.21	4063.9
W-31	95.67	3.60	1.68	8.08	0.87	0.78	32,822.4	0.226	71.35	0.21	6756.1
W-32	101.48	3.34	1.39	6.12	2.22	0.77	33,620.3	0.224	79.11	0.18	3201.2
W-33	98.76	3.73	1.72	5.90	1.75	0.78	33,324.2	0.223	76.58	0.20	5103.3
W-34	99.79	3.25	1.25	7.66	1.50	0.78	31,993.0	0.227	72.64	0.20	4262.1
W-35	102.04	3.91	1.29	8.24	1.29	0.78	32,668.4	0.227	69.38	0.20	6167.0
W-36	97.48	3.70	1.62	7.25	1.49	0.78	33,849.5	0.224	72.20	0.22	6827.0
W-37	100.43	3.83	1.48	7.70	1.14	0.79	34,857.2	0.223	74.23	0.22	7082.4
W-38	103.57	3.62	1.55	8.39	0.78	0.78	31,146.0	0.227	74.39	0.23	5514.4
W-39	94.41	3.53	1.62	6.49	1.58	0.78	33,559.9	0.224	72.81	0.22	5183.4
W-40	96.95	3.46	1.84	6.64	0.91	0.78	34,783.9	0.226	70.91	0.24	6809.9

2.2. Analysis of Main Controlling Factors

From Table 1, it is evident that the Cangdong Sag shale oil field exhibits a multitude of basic parameters with varying ranges. Some parameters demonstrate substantial variations; for instance, the natural gamma radiation spans from 85.28 to 114.31 API, while S₁ ranges from 1.97 to 12.1. However, certain parameters exhibit minimal variations, such as Poisson's ratio, which ranges from 0.22 to 0.23, and the mineral brittleness index, which varies within the range of 0.76 to 0.81. To accurately identify the correlations among geological parameters and their relationship with production, a Pearson correlation coefficient method was employed for single-factor correlation analysis.

The Pearson correlation coefficient is computed as the covariance between two factors divided by the product of their respective standard deviations. It serves to quantify the

strength and direction of the linear relationship between the selected variables. The formula for calculating the Pearson correlation coefficient is [26]:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X\sigma_Y} = \frac{m \sum_{i=1}^m x_i y_i - \sum_{i=1}^m x_i \sum_{i=1}^m y_i}{\sqrt{m \sum_{i=1}^m x_i^2 - (\sum_{i=1}^m x_i)^2} \sqrt{m \sum_{i=1}^m y_i^2 - (\sum_{i=1}^m y_i)^2}} \quad (1)$$

where $\rho_{X,Y}$ is the Pearson correlation coefficient between factor X and factor Y, dimensionless; $Cov(X, Y)$ is the covariance between factor X and factor Y, dimensionless; σ_X and σ_Y are the standard deviations of factor X and factor Y, dimensionless; and m is the number of wells, dimensionless.

Figure 1 depicts a heatmap of the correlation analysis among factors. The analysis revealed a strong correlation between mineral brittleness, S_1 , TOC, Ro, and production, while the correlation between minimum horizontal principal stress and natural gamma radiation was relatively weak. Simultaneously, significant correlations exist among certain production-influencing factors. For instance, the correlation coefficients between the coefficient of variation of horizontal principal stress and Young’s modulus, as well as between the coefficient of variation of horizontal principal stress and S_1 , are 0.801 and -0.663 , respectively, indicating multicollinearity. This can lead to distortion in the comprehensive evaluation model, affecting its accuracy. Therefore, this study introduced principal component analysis to address this issue.

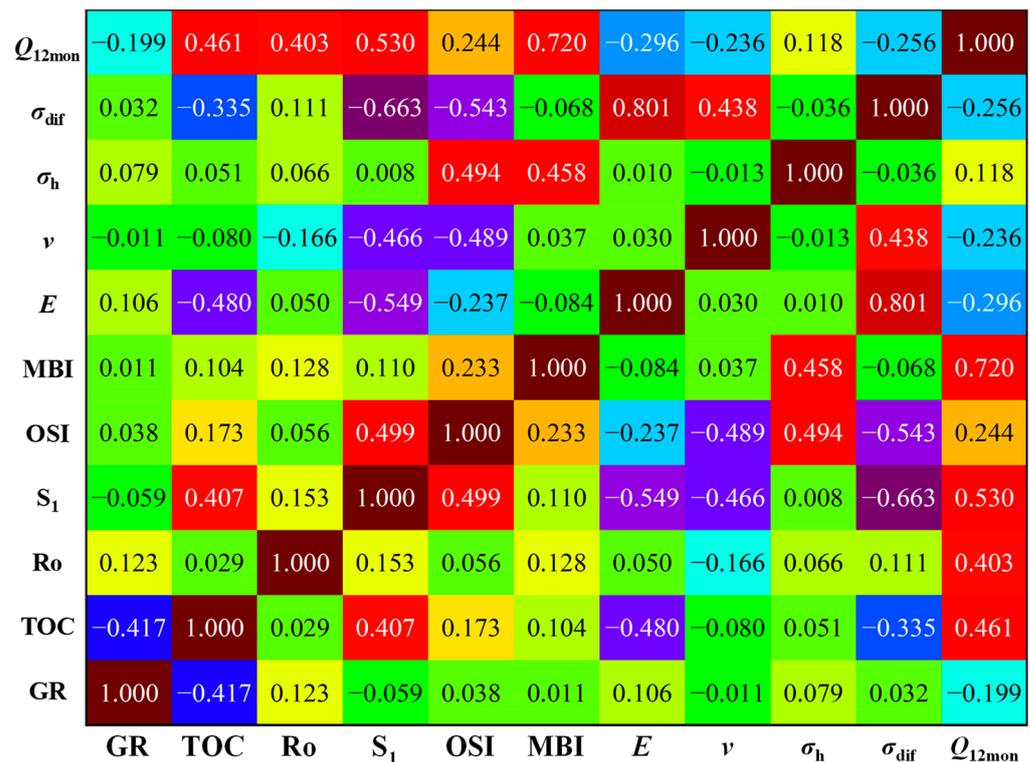


Figure 1. Pearson correlation heatmap. The values at the intersection of row and column variables represent the correlation coefficient between the two factors, with a deeper shade indicating a stronger correlation between the two.

2.3. Principal Component Analysis

Principal component analysis transforms the original factors into mutually orthogonal and independent principal components through linear combinations. It selects features based on the contribution of variance from the principal components, reducing the dimen-

sionality of the dataset and eliminating correlations among factors without losing essential information. After principal component analysis, a dataset with n factors generates n principal components. Principal Component 1 consistently represents the direction of maximum variance in the dataset, Principal Component 2 represents the second-largest variance direction, and so forth [27].

Due to the disparate scales of feature data, making direct comparisons and applying weighted processing is challenging. Prior to dimensionality reduction, it is necessary to standardize the data. Subsequently, the covariance matrix of the standardized data should be computed to delve into the interrelationships among variables. Then, one should proceed with the calculation of eigenvalues and the corresponding eigenvectors of the covariance matrix. Eigenvalues elucidate the variance explained by each principal component, while eigenvectors dictate the orientation of these components. Finally, prioritization based on eigenvalues should be undertaken, opting for a subset of principal components, capturing a pre-defined percentage of the total variance for further scrutiny.

The Z-Score normalization method is employed for this purpose:

$$x^* = \frac{x - \omega}{\delta} \quad (2)$$

where x^* is the normalized value for from 0 to 1; x is the value of the factor affecting the fracturing effect; ω is the mean of the sample data; and δ is the standard deviation.

The covariance matrix R should be calculated from the normalized dataset:

$$R = (r_{ij})_{n \times n} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}_{n \times n} \quad (3)$$

The eigenvalue of the matrix R , $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$, and the corresponding eigenvector v_1, v_2, \cdots, v_n , where $v_j = (v_{1j}, v_{2j}, \cdots, v_{nj})^T$, v_{nj} is the n -th component of the j -th feature, should then be calculated, where the principal component after dimension reduction is as follows [28]:

$$\begin{cases} z_1 = v_{11}x_1 + v_{21}x_2 + \cdots + v_{n1}x_n \\ z_2 = v_{12}x_1 + v_{22}x_2 + \cdots + v_{n2}x_n \\ \vdots \\ z_n = v_{1n}x_1 + v_{2n}x_2 + \cdots + v_{nn}x_n \end{cases} \quad (4)$$

where z_n is the n -th principal component and x_n is the n -th factor. The variance ratio of each principal component and their cumulative variance contribution rate are calculated as follows [29]:

$$\tau_j = \frac{\lambda_j}{\sum_{k=1}^n \lambda_k} \quad (j = 1, 2, \cdots, n) \quad (5)$$

$$\alpha_p = \frac{\sum_{k=1}^p \lambda_k}{\sum_{k=1}^n \lambda_k} \quad (p \leq n) \quad (6)$$

where τ_j is the variance contribution rate, dimensionless; and α_p is the cumulative contribution rate, $\alpha_p (p \leq n)$, dimensionless.

The results of principal component analysis are depicted in Figure 2. The general criteria for principal component selection are that the eigenvalue should be greater than 1, and the cumulative contribution rate of the selected principal components should exceed

85%. However, due to the relatively modest size of the sample and the dataset in this study, all 10 principal components are retained [28].

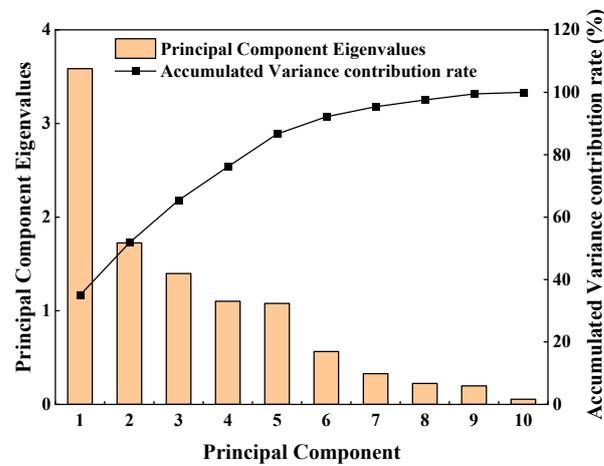


Figure 2. Initial factor feature matrix. Eigenvalues signify the extent to which the introduction of a principal component can elucidate, on average, the information contained in the original scalars. The variance contribution rate represents the proportion of variance that a principal component can explain relative to the total variance.

The principal component feature vectors are illustrated in Figure 3. Based on the coefficients of each factor corresponding to each principal component, expressions for the principal components can be constructed. For example, Principal Component 1 can be expressed as: $PC1 = -0.012GR - 0.155TOC - 0.126Ro + 0.141S_1 + 0.114OSI + 0.007MBI - 0.602E - 0.222v - 0.016\sigma_h + 0.708\sigma_{dif}$. Mechanical parameters such as Young’s modulus and the coefficient of variation of horizontal principal stress contribute the most to Principal Component 1, indicating that production is a comprehensive reflection influenced by multiple factors. Although the correlation between reservoir mechanical parameters and production is relatively weak, it still contains a considerable amount of significant feature information.

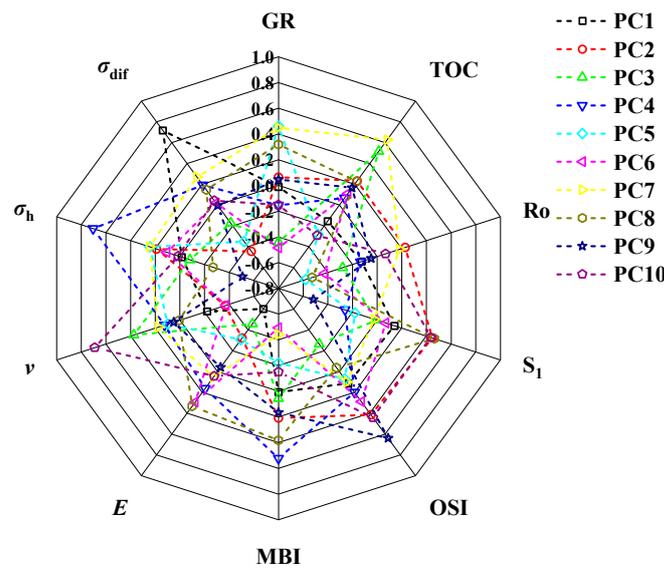


Figure 3. Principal component extraction and analysis. The numerical values corresponding to factors represent the coefficients of each factor in the process of composing principal components. A larger coefficient indicates a higher importance of the corresponding factor in the respective principal component.

3. Fuzzy Comprehensive Evaluation of Reservoir Quality

Due to the disparate contributions of individual principal components to production, precise characterization of the heterogeneity in productivity for horizontally fractured wells proves challenging with quantitative assessment methods. Fuzzy comprehensive evaluation employs membership degrees to signify the extent to which each parameter belongs to different reservoir quality grades, endowing each parameter with an inherently ambivalent nature, suitable for modeling nonlinear functions of arbitrary complexity. The fuzzy comprehensive evaluation model comprises three components: the factor set, the evaluation set, and the weight set.

3.1. Principal Component Membership

The degree of subordination is jointly determined by parameter values and membership functions. Traditional fuzzy planning problems commonly employ linear membership functions such as triangles and trapezoids. However, for the complex and nonlinear issue of reservoir quality assessment involving multiple parameters, these approaches prove inadequate. Gaussian distribution, characterized by continuity and adjustability, offers a more suitable alternative. It aptly reflects the characteristics of data variations and has demonstrated commendable application outcomes. Gaussian membership functions were assigned to each variable to capture the degree of membership of an element to a set. The functions were defined based on the mean and standard deviation of each variable. Therefore, this study employs Gaussian distribution as the membership function [30].

$$\mu(x, \lambda, \delta) = e^{-\frac{(x-\lambda)^2}{2\delta^2}} \quad (7)$$

where μ represents the degree of membership; λ is the principal component value at the peak of the Gaussian distribution; and δ is the standard deviation.

The evaluation set is utilized to discern the excellence or inferiority of determining factors. For the sake of providing an intuitive scoring system, the following definitions were established: Grade I = “Excellent” = 100, Grade II = “Good” = 75, Grade III = “Ordinary” = 50, Grade IV = “Poor” = 25. In accordance with the evaluation set and its partitioning of reservoir quality, these four evaluation grades correspond to numerical ranges of principal component values (0 to 0.25, 0.25 to 0.50, 0.50 to 0.75, and 0.75 to 1). Based on Equation (7), Gaussian-shaped membership functions for each grade can be constructed, as illustrated in Figure 4.

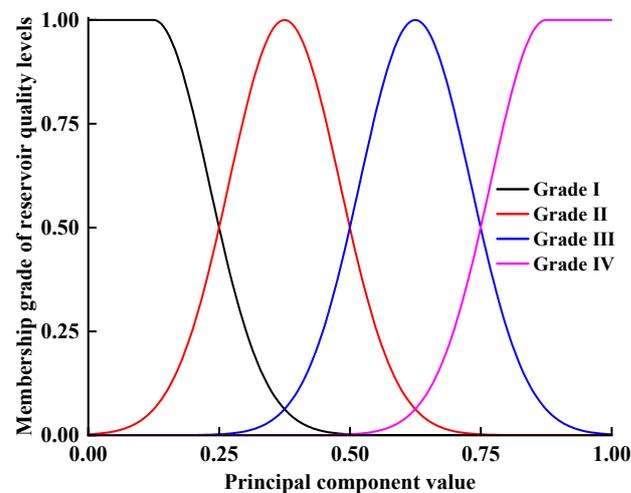


Figure 4. Gaussian membership function. The values corresponding to the intersections with the curves representing the four levels, namely “Excellent”, “Good”, “Ordinary”, and “Poor”, signify the membership degree of the principal component for the respective level.

The membership degree matrix of each principal component for each grade of each well can be constructed:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ \vdots & \vdots & \vdots & \vdots \\ a_{k,1} & a_{k,2} & a_{k,3} & a_{k,4} \end{bmatrix}_{k \times 4} \quad (8)$$

where $a_{k,1}$ refers to the membership degree of the k -th principal component to its level I.

3.2. Entropy Weight Method

The entropy weight method objectively assigns weights to the overall impact of the system based on the relative change of indicators. It accurately reflects the inherent relationships of information within the system, eliminating the irrationality of subjective experiential judgment. Therefore, in this study, this method was employed to objectively and quantitatively determine the weights of the principal components based on actual data, avoiding errors introduced by the traditional fuzzy logic approach, which relies on expert experience to determine weights [31].

Based on k principal components of m data points, a standardized evaluation matrix was constructed after normalization according to Formula (3):

$$X = \{x_j(i)\}_{m \times k} = \begin{bmatrix} x_1(1) & \cdots & x_j(1) & \cdots & x_k(1) \\ \vdots & & \vdots & & \vdots \\ x_1(i) & \cdots & x_j(i) & \cdots & x_k(i) \\ \vdots & & \vdots & & \vdots \\ x_1(m) & \cdots & x_j(m) & \cdots & x_k(m) \end{bmatrix}_{m \times k} \quad i = 1, 2, \dots, m; j = 1, 2, \dots, k \quad (9)$$

The proportion of each data point in a single principal component to the sum of all data points of the principal component—that is, entropy p_{ij} —is as follows:

$$p_{ij} = \frac{x_j(i)}{\sum_{i=1}^m x_j(i)} \quad (10)$$

The entropy of each principal component e_j is as follows [32]:

$$e_j = -\frac{1}{\ln m} \sum_{i=1}^m p_{ij} \cdot \ln p_{ij} \quad (11)$$

where p_{ij} is the ratio of each data point in a single principal component to the sum of all data points of the principal component; e_j is the entropy of the main component, $0 \leq e_j \leq 1$; and m is the number of data points.

The entropy weight method assumes that the greater the difference of the data samples of each evaluation index, the greater the weight given to the index should be. However, the higher the difference of the samples, the smaller the entropy of the index. Therefore, the entropy weight of each principal component is as follows [32]:

$$c_j = \frac{1 - e_j}{\sum_{j=1}^k (1 - e_j)} \quad j = 1, 2, \dots, k \quad (12)$$

where c_j is the weight of the j -th principal component, dimensionless; and e_j is the entropy of the j -th evaluation index, dimensionless.

Therefore, the principal component weight matrix can be established:

$$C = [c_1 \ \cdots \ c_j \ \cdots \ c_k]_{1 \times k} \quad (13)$$

Figure 5 displays the weights of the principal components. Principal Component 1 has the highest weight proportion, reaching 26.82%, followed by Principal Component 4. For Principal Component 4, the minimum horizontal principal stress and brittleness index contribute the most, further illustrating that production is influenced by a combination of geological and mechanical parameters.

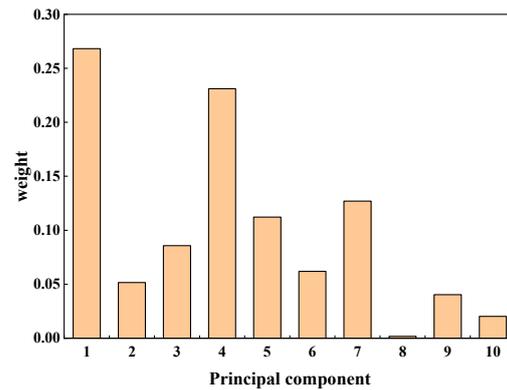


Figure 5. Principal component weight. Entropy values were calculated, reflecting the amount of uncertainty or disorder associated with each variable. Variables with higher entropy were given lower weights, emphasizing the more informative features.

3.3. Fuzzy Comprehensive Evaluation Model

Single-factor fuzzy evaluation only reflects the impact of one factor on the evaluation results, without capturing the comprehensive influence of all factors. Fuzzy comprehensive evaluation, based on selected parameters and production data, establishes rank divisions, membership functions, and factor weight rules. By inputting the data of a new shale oil well into these rules, the membership degrees for each grade are calculated. Considering the contribution of all factors, and employing the matrix multiplication and summation method, the weight matrix C is combined with the principal component membership matrix A , resulting in the fuzzy comprehensive evaluation matrix B for reservoir quality assessment [33]:

$$B = [c_1 \ \cdots \ c_j \ \cdots \ c_k]_{1 \times k} \circ \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ \vdots & \vdots & \vdots & \vdots \\ a_{k,1} & a_{k,2} & a_{k,3} & a_{k,4} \end{bmatrix}_{k \times 4} = [b_1 \ b_2 \ b_3 \ b_4]_{1 \times 4} \quad (14)$$

where b_1 is the membership of the well to grade I. Further, according to the corresponding score of each grade, the fuzzy comprehensive evaluation results are transformed into fuzzy comprehensive scores [34]:

$$r = \frac{\sum_{i=1}^4 (b_i \times u_i)}{\sum_{i=1}^4 u_i} \quad (15)$$

The scores of b_i corresponding comment sets I, II, III and IV are 100, 75, 50 and 25 respectively.

Table 2 presents the membership degrees and fuzzy comprehensive evaluation results for the Cangdong shale oil wells. Higher membership degrees for higher grades indicate better reservoir quality for the shale oil well. Consequently, the higher the fuzzy compre-

hensive evaluation score, the better the reservoir quality. The correspondence between the production and evaluation score reflects the accuracy of the model.

Table 2. Evaluation results of Cangdong sag.

No. Well	Membership Degree				Score
	I	II	III	IV	
W-1	0.339	0.093	0.250	0.318	61.30
W-2	0.047	0.295	0.174	0.483	47.69
W-3	0.032	0.205	0.296	0.468	45.04
W-4	0.189	0.183	0.358	0.270	57.28
W-5	0.249	0.103	0.245	0.403	54.93
W-6	0.150	0.438	0.354	0.058	67.01
W-7	0.411	0.346	0.117	0.126	76.05
W-8	0.322	0.157	0.317	0.205	64.87
W-9	0.376	0.244	0.259	0.121	71.90
W-10	0.283	0.233	0.279	0.205	64.86
W-11	0.286	0.256	0.250	0.208	65.51
W-12	0.091	0.166	0.271	0.471	46.93
W-13	0.019	0.141	0.400	0.440	43.46
W-14	0.061	0.273	0.234	0.432	49.09
W-15	0.000	0.200	0.281	0.520	42.00
W-16	0.340	0.296	0.295	0.069	72.69
W-17	0.062	0.158	0.645	0.135	53.66
W-18	0.138	0.241	0.414	0.208	57.71
W-19	0.340	0.197	0.301	0.162	67.91
W-20	0.154	0.302	0.470	0.073	63.43
W-21	0.059	0.409	0.479	0.054	61.81
W-22	0.053	0.246	0.569	0.132	55.48
W-23	0.079	0.371	0.491	0.059	61.75
W-24	0.121	0.430	0.426	0.023	66.23
W-25	0.062	0.399	0.389	0.150	59.31
W-26	0.035	0.378	0.504	0.084	59.08
W-27	0.049	0.299	0.565	0.087	57.76
W-28	0.024	0.440	0.413	0.123	59.14
W-29	0.060	0.295	0.457	0.188	55.67
W-30	0.040	0.258	0.614	0.088	56.26
W-31	0.196	0.375	0.387	0.042	68.14
W-32	0.013	0.128	0.667	0.192	49.02
W-33	0.076	0.377	0.389	0.158	59.26
W-34	0.092	0.198	0.614	0.096	57.16
W-35	0.141	0.310	0.504	0.045	63.66
W-36	0.167	0.343	0.478	0.013	66.57
W-37	0.249	0.265	0.438	0.048	67.89
W-38	0.119	0.475	0.343	0.064	66.20
W-39	0.160	0.196	0.592	0.053	61.55
W-40	0.258	0.247	0.305	0.191	64.30

Figure 6 depicts the relationship between $Q_{12\text{mon}}$ and the comprehensive evaluation score without using principal component analysis [34]. When compared to the analysis model using PCA (Figure 7), it is evident that the direct analysis of factors on reservoir quality has a correlation coefficient of only 0.7609. In contrast, the model developed in this study achieves a higher correlation coefficient of 0.9132. The good consistency between reservoir quality evaluation results and production indicates that the model built in this study can effectively reflect the enhanced production potential of the reservoir.

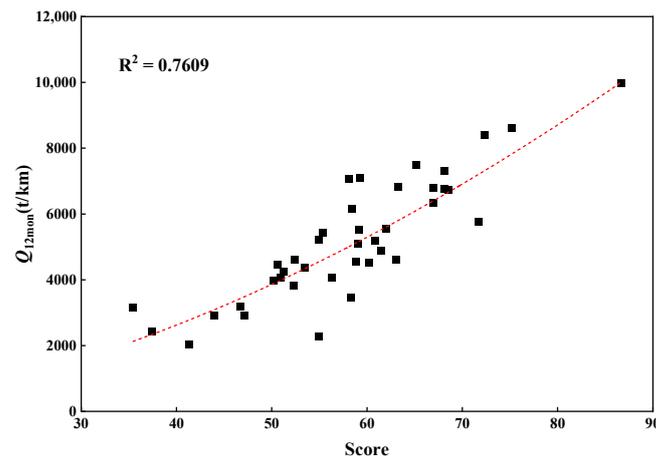


Figure 6. $Q_{12\text{mon}}$ and Score correlation without PCA. At this juncture, without transforming the evaluation indicators into principal components, the relatively low correlation coefficient implies a significant adverse impact of inter-factor correlations on the evaluation effectiveness.

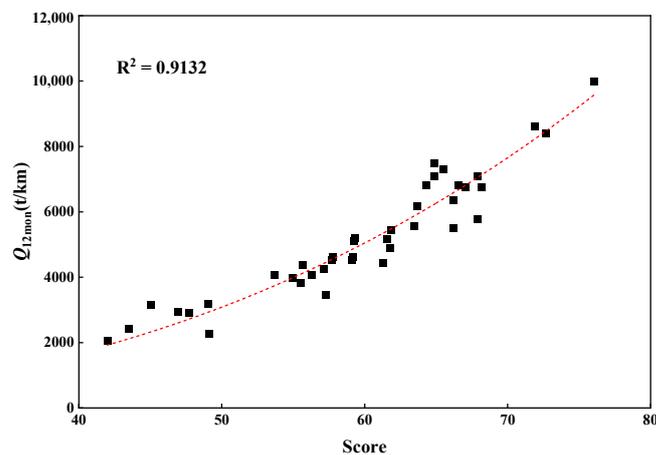


Figure 7. $Q_{12\text{mon}}$ and Score correlation with PCA. The correlation coefficient between production and the comprehensive score, obtained by transforming multiple reservoir quality assessment indicators using this model, is 0.9132. This value indicates a robust evaluation performance of the model.

4. Field Application

The model developed in this study was applied to assess reservoir quality and predict production for 10 wells in the Cangdong shale oil field, utilizing the basic parameters of the wells as shown in Table 3. By substituting the basic parameters of the wells into the evaluation model established using the 40 sample wells in the previous analysis, comprehensive scores were obtained. Further, based on the fitted curve between production and scores shown in Figure 7, the production for the applied wells was predicted, as indicated in Table 3. Among the 10 application wells, the T-10 well exhibits the highest production, reaching 7744.5 t/km. However, the T-3 well, despite having seemingly similar parameters to other wells, demonstrates a production discrepancy close to threefold when compared to the T-10 well. This further underscores that production is the outcome of a multifactorial influence.

Figure 8 illustrates the comparative analysis between the actual cumulative oil production per kilometer of fractured length over 12 months for the 10 applied wells and the corresponding predicted results. It is evident that the relative prediction error ranges from 0.39% to 12.96%, with an average relative error of 8.2%. This indicates that the method proposed in this study can elucidate the potential correlations between source lithology, fracability, and production.

It effectively addresses the intricate issue of reservoir comprehensive evaluation, providing a quantitative assessment of shale oil reservoir quality.

Table 3. Parameters of application wells in Cangdong sag.

Well No.	GR (API)	TOC (%)	Ro	S ₁ (mg/g)	OSI	MBI	E (MPa)	ν	σ_h (MPa)	σ_{dif}	Score	Q _{12mon} (t/km)	
												Predictive	Actual
T-1	110.33	2.59	1.10	2.25	0.87	0.79	37,671.1	0.228	83.85	0.28	54.73	3901.9	3759.9
T-2	96.22	3.95	0.87	5.04	2.25	0.77	34,676.1	0.224	77.16	0.18	50.19	3196.1	3183.7
T-3	103.67	3.02	2.00	4.78	1.58	0.76	36,482.2	0.228	86.43	0.27	46.53	2720.5	2982.2
T-4	102.20	3.13	1.10	4.57	1.46	0.81	39,669.4	0.228	89.63	0.29	59.78	4873.0	4371.5
T-5	101.63	3.41	0.91	5.82	1.62	0.78	32,064.6	0.227	72.46	0.19	51.67	3410.9	3687.1
T-6	103.16	3.64	0.90	9.28	0.77	0.77	30,964.3	0.230	70.29	0.22	55.84	4098.0	4415.3
T-7	94.77	3.81	0.87	7.79	0.99	0.78	31,955.3	0.223	73.09	0.23	59.62	4839.8	5214.5
T-8	99.18	3.52	0.92	5.42	1.17	0.78	32,863.4	0.224	63.05	0.17	59.68	4852.2	5536.2
T-9	96.21	3.71	1.10	11.09	2.99	0.80	25,681.7	0.229	91.56	0.14	65.59	6292.7	7229.6
T-10	101.61	3.63	2.00	3.85	1.06	0.81	36,607.8	0.228	79.76	0.28	67.81	6937.9	7744.5

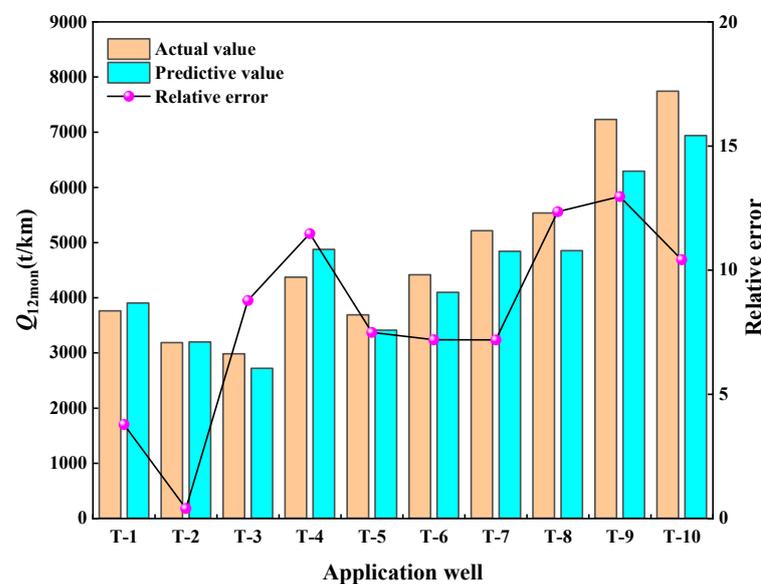


Figure 8. Comparative between actual production and predicted production. By inputting the evaluation factors of the 10 wells into the model to predict production, the relative errors range from 0.39% to 12.96% when compared with the actual production. This suggests a promising prospect for the practical application of the model in the field.

The proposed comprehensive evaluation model combines the strengths of various methods to create a versatile and interpretable framework. Its advantages lie in its ability to handle both quantitative and qualitative aspects, offering a holistic evaluation approach. However, potential limitations include the need for a comprehensive understanding of its components and potential parameter tuning. The choice between models ultimately depends on the specific characteristics and requirements of the application domain.

5. Conclusions

- (1) To accurately assess the reservoir quality of shale oil fractured horizontal wells, a fuzzy comprehensive evaluation model based on principal component analysis and the entropy weight method is proposed. The contribution of parameters to production was determined through Pearson analysis. The linear combination of factors and the elimination of correlations between factors were achieved by incorporating principal component analysis. The entropy weight method was introduced to determine the weights of principal components, establishing the fuzzy comprehensive evaluation model for reservoir quality.
- (2) Analysis of 40 wells in the Cangdong shale oil field revealed that the correlation coefficient between single-factor assessment of production and the evaluation result

for reservoir quality is only 0.7609. However, after principal component analysis, the correlation coefficient increased to 0.9132, significantly enhancing the model's accuracy.

(3) Applying the model developed in this study to assess reservoir quality and predict production for 10 wells in the Cangdong shale oil field, the results indicate that the average relative error between the actual cumulative oil production per kilometer of fractured length over 12 months for the applied wells and the corresponding predicted results is 8.2%. This suggests that the method proposed in this study can elucidate the potential correlations between source lithology, fracability, and production.

Author Contributions: Writing—original draft, F.T., Y.F., X.L., D.L., Y.J., L.S., L.Y., Y.Z., T.Z., Q.Y. and X.G.; Writing—review and editing, F.T., Y.F., X.L. and D.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding authors.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Xu, G.J.; Ren, Z.Y.; Wang, Z.P.; Cui, L.; Su, J.Z.; Meng, X.L.; Chen, P.L.; Li, P.; Wang, N.N.; Hao, X.; et al. Petroleum-like fuels with substantially enriched branched iso-paraffins and benzenes via boehmite-assisted pyrolysis of oil shale. *Fuel* **2024**, *358*, 130324. [[CrossRef](#)]
- Zhao, G.X.; Yao, Y.D.; Zhang, T.; Wang, L.; Adenutsi, C.D.; Nassar, N.N. An Integrated Approach for History Matching of Complex Fracture Distributions for Shale Oil Reservoirs Based on Improved Adaptive Particle Filter. *SPE J.* **2023**, *28*, 594–613. [[CrossRef](#)]
- Lin, H.Y.; Mao, L.J.; Mai, Y. Influence of multistage fracturing in shale gas wells on the casing deformation of horizontal wells. *Pet. Sci. Technol.* **2024**, *42*, 56–80. [[CrossRef](#)]
- Hyunjun, I.; Hyongdoo, J.; Erkan, T.; Micah, N. Long- and Short-Term Strategies for Estimation of Hydraulic Fracturing Cost Using Fuzzy Logic. *Minerals* **2022**, *12*, 715–735.
- Iyare, U.C. Fracability evaluation of the upper Cretaceous Naparima Hill Formation, Trinidad. *J. Pet. Sci. Eng.* **2022**, *208*, 109599. [[CrossRef](#)]
- Zheng, D.Z.; Miska, S.; Ozbayoglu, E.; Zhang, J.G. Combined Experimental and Well Log Study of Anisotropic Strength of Shale. In Proceedings of the SPE Annual Technical Conference and Exhibition, San Antonio, TX, USA, 16–18 October 2023.
- Huang, Y.Y.; Zhang, Y.; Xi, J.H.; Huang, L.L.; Wang, S.; Zhang, Y.L.; Lai, J.; Jiang, C.Z. Logging evaluation of pore structure and reservoir quality in shale oil reservoir: The Fengcheng Formation in Mahu Sag, Junggar Basin, China. *Mar. Pet. Geol.* **2023**, *156*, 106454. [[CrossRef](#)]
- Liu, Y.Z.; Zeng, J.H.; Qiao, J.C.; Yang, G.Q.; Liu, S.N.; Cao, W.F. An advanced prediction model of shale oil production profile based on source-reservoir assemblages and artificial neural networks. *Appl. Energy* **2023**, *333*, 120604. [[CrossRef](#)]
- Wen, S.Q.; Wei, B.; He, Y.J.; Xin, J.; Varfolomeev, M.A. Forecasting oil production in unconventional reservoirs using long short term memory network coupled support vector regression method: A case study. *Petroleum* **2023**, *9*, 647–657. [[CrossRef](#)]
- Qin, X.Z.; Hu, X.H.; Liu, H.; Shi, W.Y.; Cui, J.S. A Combined Gated Recurrent Unit and Multi-Layer Perception Neural Network Model for Predicting Shale Gas Production. *Processes* **2023**, *11*, 806. [[CrossRef](#)]
- Liu, W.C.; Yang, Y.J.; Qiao, C.C.; Liu, C.; Lian, B.; Yuan, Q.W. Progress of Seepage Law and Development Technologies for Shale Condensate Gas Reservoirs. *Energies* **2023**, *16*, 2446. [[CrossRef](#)]
- Mu, J.F.; Qiao, H.J.; Guo, Q. A Fracture Toughness-Based Evaluation Method for Deep Shale Oil Reservoir Compressibility. *Chem. Technol. Fuels Oils* **2022**, *58*, 880–886. [[CrossRef](#)]
- Feng, C.; Feng, Z.Y.; Mao, R.; Li, G.L.; Zhong, Y.T.; Ling, K.G. Prediction of vitrinite reflectance of shale oil reservoirs using nuclear magnetic resonance and conventional log data. *Fuel* **2023**, *339*, 127422. [[CrossRef](#)]
- You, X.T.; Liu, J.Y.; Jia, C.S.; Li, J.; Liao, X.Y.; Zheng, A.W. Production data analysis of shale gas using fractal model and fuzzy theory: Evaluating fracturing heterogeneity. *Appl. Energy* **2019**, *250*, 1246–1259. [[CrossRef](#)]
- Gou, B.; Wang, C.; Yu, T.; Wang, K.J. Fuzzy logic and grey clustering analysis hybrid intelligence model applied to candidate-well selection for hydraulic fracturing in hydrocarbon reservoir. *Arab. J. Geosci.* **2020**, *13*, 1–13. [[CrossRef](#)]
- Okwu, M.O.; Nwachukwu, A.N. A review of fuzzy logic applications in petroleum exploration, production and distribution operations (Review). *J. Pet. Explor. Prod. Technol.* **2019**, *9*, 1555–1568. [[CrossRef](#)]
- Zoveidavianpoor, M.; Gharibi, A. Applications of type-2 fuzzy logic system: Handling the uncertainty associated with candidate-well selection for hydraulic fracturing. *Neural Comput. Appl.* **2016**, *27*, 1831–1851. [[CrossRef](#)]

18. Davarpanah, A.; Shirmohammadi, R.; Mirshekari, B.; Aslani, A. Analysis of hydraulic fracturing techniques: Hybrid fuzzy approaches. *Arab. J. Geosci.* **2019**, *12*, 1–8. [[CrossRef](#)]
19. Xie, J.Y.; Zhang, J.J.; Fang, Y.P.; Cao, J.X.; Deng, J.X. Quantitative Evaluation of Shale Brittleness Based on Brittle-Sensitive Index and Energy Evolution-Based Fuzzy Analytic Hierarchy Process. *Rock Mech. Rock Eng.* **2023**, *56*, 3003–3021. [[CrossRef](#)]
20. Verma, A.K.; Singh, T.N. A neuro-fuzzy approach for prediction of longitudinal wave velocity. *Neural Comput. Appl.* **2013**, *22*, 1685–1693. [[CrossRef](#)]
21. Sylwan, C.; Vernengo, L.; Sharma, R.K.; Chopra, S.; Trincherro, E. Reducing uncertainty in characterization of Vaca Muerta Formation Shale with poststack seismic data. *Lead. Edge* **2015**, *34*, 1462–1467.
22. Foster, D.; Yenugu, M.; Sayers, C.M. Introduction to this special section: Resource plays II: Geophysics. *Lead. Edge* **2015**, *34*, 1440–1441.
23. Rickman, R.; Mullen, M.; Petre, E.; Grieser, B.; Kundert, D. A Practical Use of Shale Petrophysics for Stimulation Design Optimization: All Shale Plays Are Not Clones of the Barnett Shale. In Proceedings of the SPE Annual Technical Conference & Exhibition, Denver, CO, USA, 21–24 September 2008.
24. Zhang, W.; Huang, F.; Li, Z.; He, X.; He, Y. Shale-gas reservoir-prediction study in Daanzhai, Eastern Sichuan Basin. *Lead. Edge* **2014**, *33*, 526–534.
25. Dashtian, H.; Jafari, G.R.; Sahimi, M.; Masihi, M. Scaling, multifractality, and long-range correlations in well log data of large-scale porous media. *Phys. A Stat. Mech. Its Appl.* **2011**, *390*, 2096–2111. [[CrossRef](#)]
26. Fu, Y.K.; Dehghanpour, H.; Ezulike, D.O.; Jones, R.S. Estimating Effective Fracture Pore Volume From Flowback Data and Evaluating Its Relationship to Design Parameters of Multistage-Fracture Completion. *SPE Prod. Oper.* **2017**, *32*, 423–439. [[CrossRef](#)]
27. Rios, E.H.; Azeredo, R.B.D.V.; Moss, A.K.; Pritchard, T.N.; Domingues, A.B.G. Estimating the Permeability of Rocks by Principal Component Regressions of NMR and MICP Data. *Petrophysics* **2022**, *63*, 442–453.
28. Wang, L.; Yao, Y.; Wang, K.; Adenutsi, C.D.; Zhao, G. Combined Application of Unsupervised and Deep Learning in Absolute Open Flow Potential Prediction: A Case Study of the Weiyuan Shale Gas Reservoir. In Proceedings of the SPE/AAPG/SEG Asia Pacific Unconventional Resources Technology Conference, Virtual, 16–18 November 2021.
29. Li, Y.; Li, B.Z.; Liu, L.; Xiong, L.H.; Luo, H.; Peng, H.; Wang, D.G. Case study of a super-giant field rejuvenation. In Proceedings of the 78th EAGE Conference and Exhibition 2016: Efficient Use of Technology—Unlocking Potential, Vienna, Austria, 30 May–2 June 2016.
30. Liu, Z.Q.; Wang, X. The distributivity of extended uninorms over extended overlap functions on the membership functions of type-2 fuzzy sets. *Fuzzy Sets Syst.* **2022**, *448*, 94–106. [[CrossRef](#)]
31. Gao, C.L.; Li, S.C.; Wang, J.; Li, L.P.; Lin, P. The Risk Assessment of Tunnels Based on Grey Correlation and Entropy Weight Method. *Geotech. Geol. Eng.* **2017**, *36*, 1621–1631. [[CrossRef](#)]
32. Ju, Y.; Wu, G.J.; Wang, Y.L.; Liu, P.; Yang, Y.M. 3D Numerical Model for Hydraulic Fracture Propagation in Tight Ductile Reservoirs, Considering Multiple Influencing Factors via the Entropy Weight Method. *SPE J.* **2021**, *26*, 2685–2702. [[CrossRef](#)]
33. Zeng, F.H.; Cheng, X.Z.; Guo, J.C.; Tao, L.; Chen, Z.X. Hybridising human judgment, ahp, grey theory, and fuzzy expert systems for candidate well selection in fractured reservoirs. *Energies* **2017**, *10*, 447. [[CrossRef](#)]
34. Zeng, F.H.; Guo, J.C.; Long, C. A hybrid model of fuzzy logic and grey relation analysis to evaluate tight gas formation quality comprehensively. *J. Grey Syst.* **2015**, *27*, 87–98.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.