

Article

Optimization of Staged Fracturing and Cluster Design for Horizontal Wells in Low-Permeability Reservoirs Based on Mechanical Specific Energy

Yuan Pan ¹ , Min Liu ¹, Xuwei Liu ¹, Feixiang Qin ¹, Xiaoting Gou ¹, Yingxi Zhang ¹, Xiangyun Sui ¹, Peipei Liu ¹ and Mingxiu Zhang ^{2,*}

¹ Petroleum Engineering Research Institute, PetroChina Dagang Oilfield Company, Tianjin 300280, China; panyuan_yangtzeu@163.com (Y.P.)

² College of Petroleum Engineering, Northeast Petroleum University, Daqing 163318, China

* Correspondence: zmx_0924@163.com

Abstract

The fracturing design of horizontal wells in low-permeability reservoirs mostly relies on static data such as logging data, and insufficient utilization of while-drilling information from drilling and mud logging leads to an unreliable basis for staged fracturing and cluster design, making it difficult to achieve efficient reservoir stimulation. To address this issue, this paper takes mechanical specific energy (MSE) as the core engineering parameter. Based on the drilling and mud logging data from the horizontal section of an example well, a mechanical specific energy model is established after data preprocessing. The K-medoids unsupervised clustering algorithm is adopted to realize the fine division of fracturing stages, and the optimal number of stages is determined using the elbow method. A fracturing sweet spot index (FSI) is constructed using the random forest algorithm. This index integrates four key parameters: oil saturation, porosity, permeability, and mechanical specific energy to optimize perforation cluster placement. The proposed method is validated through finite element numerical simulation and field data comparison. The research results show that the optimal division of the example well is determined as 11 stages with six clusters per stage, and the matching degree between stage-cluster division and reservoir heterogeneity is significantly improved. Compared with the original scheme, the stimulated reservoir volume (SRV) of the optimized construction scheme is increased by 19.6%, and the treatment pressure exhibits an obvious overall reduction trend, as demonstrated by the construction pressure curves. The research indicates that this method can make full use of drilling while-drilling data to realize integrated geological and engineering evaluations of fracturing and optimization of staged fracturing and cluster designs, and can provide effective technical support for the fracturing optimization design of horizontal wells in low-permeability reservoirs lacking complete logging data.



Academic Editor: Hossein Hamidi

Received: 3 June 2026

Revised: 26 June 2026

Accepted: 29 June 2026

Published: 20 July 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: low-permeability reservoirs; horizontal wells; staged fracturing and cluster design; mechanical specific energy; intelligent optimization; numerical simulation

1. Introduction

Staged multi-cluster fracturing stimulation of horizontal wells is the core technology for efficient development of low-permeability reservoirs [1–4]. Long horizontal sections in low-permeability reservoirs exhibit significant heterogeneity, and the collaborative optimization of fracturing stage-cluster layout and construction parameters is the key to

achieving efficient production of low-permeability reservoirs [5–7]. Limited by on-site operation and data acquisition conditions, traditional fracturing design mostly relies on static geological data and numerical simulation. Most horizontal wells in low-permeability reservoirs are designed only based on logging data, making it difficult to fully integrate real-time drilling parameters [8–10]. Therefore, constructing the mapping relationship between reservoir stimulation effects and engineering sweet spot parameters such as mechanical specific energy, and realizing accurate staged fracturing and cluster design based on while-drilling data to improve stimulation effects and construction efficiency, has shown good application prospects in reservoir stimulation of low-permeability oil and gas reservoirs.

In terms of theoretical research on mechanical specific energy, Teale [11] first proposed the “mechanical specific energy” theory and constructed a mechanical specific energy model, which was the first systematic theoretical framework formed around “mechanical specific energy” in the drilling field. Although this formula is convenient for calculation, many data cannot be obtained from field data, and complex factors such as bit type, lithology changes, and hydraulic parameters are not considered. Meng Yingfeng et al. [12] established a new mechanical specific energy model considering the influence of hydraulic energy based on Teale’s original specific energy model. Cui Meng et al. [13] introduced two key parameters, the bit rock-breaking efficiency coefficient and the sliding friction coefficient, and established a functional relationship between torque and weight on bit, solving the problem of difficult torque acquisition during drilling.

In terms of the combination of mechanical specific energy and fracturing, Robert D et al. [14] proposed the relationship between fracture response and rock properties obtained from drilling. By clarifying the relationship between rock characteristics, completion design, and fracture height growth, they re-evaluated future drilling targets and modified treatment designs to maintain maximum fracturing energy in the target area. Furthermore, it was proved that optimizing stage boundaries and perforation cluster positions is beneficial to completion operations. Z. Peter Wang et al. [15] proposed a method combining MSE with logging parameters to improve fracture design. This study showed that drilling data can be effectively used in fracturing design to achieve significant results in production. In addition, this study helped locate the optimal positions for fracturing engineering, guide the determination of perforation positions, and improve the fracturing design of horizontal wells. Shimeng H et al. [16] established an unsupervised clustering model of bottomhole mechanical specific energy of the bit in horizontal well sections, considering the influences of drill string friction, compound drilling, and jet-assisted rock breaking on bottomhole mechanical specific energy, and obtained the bottomhole mechanical specific energy distribution and clustering categories with decimeter-level spatial resolution. Zhuang X et al. [17] adopted a variety of machine learning techniques to predict rock fracture pressure in hydraulic fracturing based on the concept of mechanical specific energy. Zeng Fanhui et al. [18] established an integrated geological and engineering sweet spot evaluation method for shale gas horizontal well reservoirs based on mechanical specific energy, and added geological and engineering sweet spots into the evaluation system. Subsequently, an unsupervised clustering algorithm was used to finely classify the reservoirs, and a data-driven intelligent optimization method for fracturing parameters of shale gas horizontal wells was formed based on the classification results, providing data support for the selection of on-site fracturing parameters. After years of development, the mechanical specific energy theory has achieved remarkable results in identifying drilling conditions, evaluating formation drillability, and judging bit status, and is an important theory that reflects formation mechanical properties and guides drilling parameter optimization. Despite these advances, several critical gaps remain. At present, this theory is mainly used in the drilling link, and its application in the fracturing design link is still in the development

stage. The internal relationship between mechanical specific energy and reservoir physical properties has not been truly established, resulting in a large amount of original drilling data failing to play its due role and having little effect on the optimization of fracturing construction parameters.

To address the above gaps, this study proposes a novel data-driven optimization method that integrates MSE-based clustering and random forest-based sweet spot evaluation for horizontal well fracturing design. Compared with existing workflows, this method offers several distinctive features: it utilizes while-drilling data rather than relying solely on logging data; it automates stage-cluster design through unsupervised clustering and machine learning; and it integrates geological and engineering evaluations into a unified framework. This paper takes the low-permeability reservoirs in the Bohai Bay Basin as the research object, and guides the optimization design of fracturing stage-cluster positions and construction parameters combined with numerical simulation methods based on the mechanical specific energy model data of the example well in this area. Firstly, the mechanical specific energy model of the example well is established. Based on this model, Pearson analysis is used to screen the main controlling factors of the reservoir; K-medoids unsupervised clustering is adopted to realize fine division of fracturing stages, the fracturing sweet spot index is constructed combined with Euclidean proximity and random forest to complete perforation cluster optimization, and fracturing parameter optimization is realized through finite element numerical simulation. Finally, the effectiveness of the design method is verified through comparisons with field fracturing data.

2. Materials and Methods

2.1. Optimization of Mechanical Specific Energy Model

Mechanical specific energy (MSE) is the total mechanical work required to break a unit volume of rock, which can reflect rock strength, drillability, and fracability characteristics in real time.

Interference from non-rock-breaking related factors such as friction between the drill string and the wellbore, wear of bit cutting teeth, and stick-slip vibration makes it difficult for collected drilling and mud logging data to accurately reflect the mechanical specific energy state at the actual bottomhole position where the bit is located. To accurately obtain the real energy consumption of the bit breaking rock and reflect the reservoir rock strength characteristics, it is necessary to correct the ground-collected parameters such as weight on bit, torque, and rotational speed to bottomhole conditions according to the traditional mechanical specific energy model. The specific practices are as follows [16]:

1. Weight on bit correction: Comprehensively considering the frictional resistance generated by the contact between the drill string and the wellbore and the rock-breaking effect of jet impact force, the effective bottomhole weight on bit W_e is obtained:

$$W_e = W_b - \mu_s \gamma_b - \eta \frac{\rho_d Q^2}{A_0} \quad (1)$$

where W_e is the effective bottomhole weight on bit, kN; μ_s is the friction coefficient between the drill string and the wellbore, usually ranging from 0.25 to 0.4, taken as 0.35; γ_b is the well deviation angle, rad; η is the bit hydraulic power coefficient, related to nozzle flow velocity, nozzle size, and position; ρ_d is the drilling fluid density, g/cm³; Q is the drilling fluid flow rate, L/s; and A_0 is the bit nozzle exit area, cm². These parameter values are selected based on the established literature and industry standards for similar reservoir conditions.

2. Torque correction: The bit torque T_b can be calculated from the effective bit weight on bit and the bit friction coefficient as follows:

$$T_b = \int_0^{d_b} \int_0^{2\pi} \rho^2 \frac{4\mu W_e}{\pi d_b^2} d\rho d\theta = \frac{\mu W_e d_b}{3} \quad (2)$$

where T_b is the bit torque, kN·m; d_b is the bit diameter, m; μ is the bit sliding friction coefficient, and the sliding friction coefficient of PDC bits is usually taken as 0.8.

3. Rotational speed correction: The bit rotational speed (N_b) is where the screw drill motor converts the hydraulic energy of circulating drilling fluid into kinetic energy to provide the driving force for the bit, and its output theoretical rotational speed is only related to the flow rate through the drill tool and the displacement per revolution of the screw as follows:

$$N_b = N + N_L = N + \frac{60Q}{q} \quad (3)$$

where N_b is the bit rotational speed, r/min; N_L is the theoretical self-rotational speed output by the screw drill motor, r/min; and q is the displacement per revolution of the screw, which is only related to the geometric size and linear type of the drill tool, L/s, taken as 16.5 L/s by checking the table (screw manual).

4. Hydraulic energy: Considering the hydraulic parameters, the hydraulic energy (E_b) caused by hydraulic power during the bit rock-breaking process is as follows:

$$E_b = \frac{\eta P_b}{A_b V_{pc}} = \frac{\eta \Delta p_b Q}{A_b V_{pc}} \quad (4)$$

The improved mechanical specific energy model MSE_b is obtained:

$$MSE_b = \frac{W_e}{A_b} + \frac{120\pi N_b T_b}{A_b V_{pc}} + \frac{\eta \Delta p_b Q}{A_b V_{pc}} \quad (5)$$

where MSE_b is the mechanical specific energy, MPa; W_e is the effective weight on bit, kN; A_b is the bit cross-sectional area, mm²; N is the rotational speed, r/min; T_b is the bit torque, kN·m; V_{pc} is the rate of penetration, m/h; η is the bit hydraulic power coefficient; Q is the drilling fluid flow rate, L/s; P_b is the hydraulic power, kW; and Δp_b is the bit pressure drop, MPa.

2.2. Data Preprocessing

The original drilling and mud logging data are time-domain data, containing both drilling and non-drilling information. The same well depth often corresponds to different parameters at multiple time points, and some well depths have problems such as missing parameters and insufficient accuracy. It is necessary to convert them into continuous depth-domain data, process multi-group parameters at the same depth, non-drilling data, missing values, and outliers to obtain pure drilling and mud logging data for calculating the variation profile of mechanical specific energy along the well depth in the horizontal section. The data cleaning and smoothing process is shown in Figure 1.

Taking the data of the example well as a case, the well has a total vertical depth (TVD) of 5242 m, a horizontal interval from 4275.0 m to 5239.0 m (measured depth), and a maximum hole deviation of 86.42° at 4920 m (measured depth). A set of drilling and mud logging data was established with a 1 m interval along the increasing measured depth.

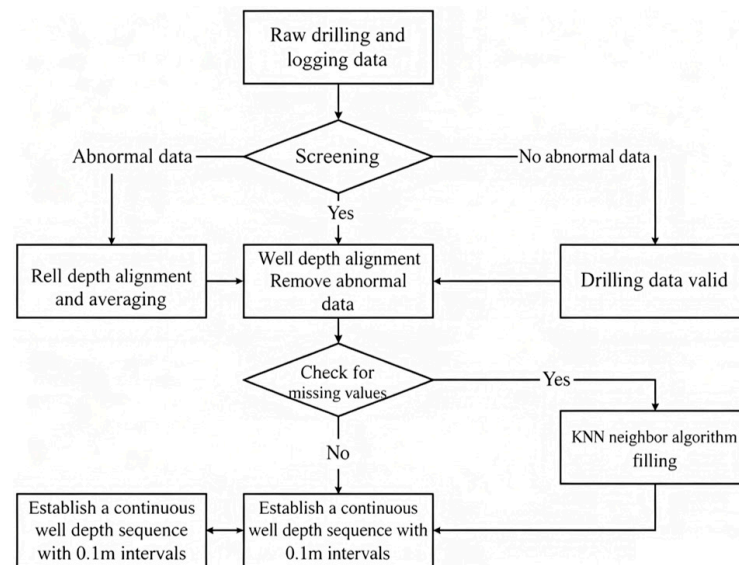


Figure 1. Flowchart of data cleaning and smoothing.

(1) Extract data points that meet the actual drilling conditions from the original drilling and mud logging records. Filter out records where the three parameters of weight on bit, rotational speed, and standpipe pressure are all greater than zero, and exclude data from non-drilling states (such as tripping, circulating drilling fluid, and other operation stages) to ensure that the selected data can truly reflect the rock-breaking state during drilling.

The drilling condition judgment formula is as follows:

$$WOB > 0 \cap N > 0 \cap SPP > 0 \quad (6)$$

where SPP is the standpipe pressure, KMa.

(2) There may be multiple sets of drilling and mud logging parameter records at the same well depth position. For different parameter values at the same depth, the arithmetic mean is directly taken. In this way, the results of multiple measurements are integrated to obtain a more representative parameter value for the depth point.

(3) The well depth alignment average formula is as follows:

$$\bar{P}_d = \frac{1}{n_d} \sum_{i=1}^{n_d} p_{d,i} \quad (7)$$

where $\bar{P}_d$ is the average parameter value at depth d ; n_d is the number of measurements at depth d ; and $P_{d,i}$ is the i -th measured parameter value at depth d .

When the same reservoir, the same bottom hole assembly, and similar rock-breaking conditions are met, the corresponding drilling and mud logging parameters should have similarity. Therefore, the K-nearest neighbors (KNN) algorithm [19] is used to fill in missing values. KNN is adopted because it preserves local data structures and is robust to outliers, making it suitable for drilling data where variations often reflect lithology changes rather than random noise. KNN uses the local similarity between data to estimate missing values. Compared with conventional methods such as mean filling, mode filling, or regression filling, it gives more reliable and robust results and can better retain the original structure and distribution characteristics of the data.

After determining the k-nearest neighbors, inverse distance weighting is used to calculate the filling value. The missing value estimation formula is as follows:

$$\hat{P}_{miss} = \frac{\sum_{i=1}^k w_i \cdot P_i}{\sum_{i=1}^k w_i} \quad (8)$$

The weight calculation formula is as follows:

$$w_i = \frac{1}{d_i + \varepsilon} \quad (9)$$

where $\hat{P}_{miss}$ is the estimated missing parameter value; P_i is the parameter value of the i-th nearest neighbor; d_i is the Euclidean distance from the missing point to the i-th nearest neighbor; ε is a very small constant (usually taken as 10^{-6}) to prevent division by zero error; and k is the number of nearest neighbors (usually taken as 3–5, adjusted according to data density).

After data processing, a drilling and mud logging parameter dataset increasing with well depth is established. This dataset has an interval of 1 m and provides a continuous and complete sequence of drilling and mud logging parameters, which provides high-quality data support for applications such as formation identification and reservoir staged fracturing and cluster design. Partial data before and after drilling and mud logging data cleaning and missing value filling are shown in Tables 1 and 2. The data cleaning and imputation process significantly improved data quality. After processing, the number of valid depth points increased, and the data distribution became more continuous and stable. The standard deviations of key drilling parameters were notably reduced, as summarized in the statistical comparison shown in Figure 2.

Table 1. Drilling and mud logging data before cleaning and missing value filling.

DEPT (m)	WOB (kN)	TORQUE (kN·m)	ROP (m/h)	FLOW RATE (L/s)	SPP (kN)
4555	100.78	34.07	15.77	8.02	21.42
4556	105.84	34.39	16.68	3.65	22.36
4557	121.27	29.24	22.24	0	22.19
4558	0	31.53	14.83	0	22.02
4559	94.66	33.1	15.35	0.42	22.24
4560	0	33.3	17.35	0	21.95
4561	94.12	27.62	20.4	1.52	22.05

Table 2. Drilling and mud logging data after cleaning and missing value filling.

DEPT (m)	WOB (kN)	TORQUE (kN·m)	ROP (m/h)	FLOW RATE (L/s)	SPP (kN)
4555	100.78	34.07	15.77	8.02	21.42
4556	105.84	34.39	16.68	3.65	22.36
4557	121.27	29.24	22.24	3.34	22.19
4558	99.87	31.53	14.83	1.46	22.02
4560	95.65	33.3	17.35	1.26	21.95
4561	94.12	27.62	20.4	1.52	22.05
4561	94.12	27.62	20.4	1.52	22.05

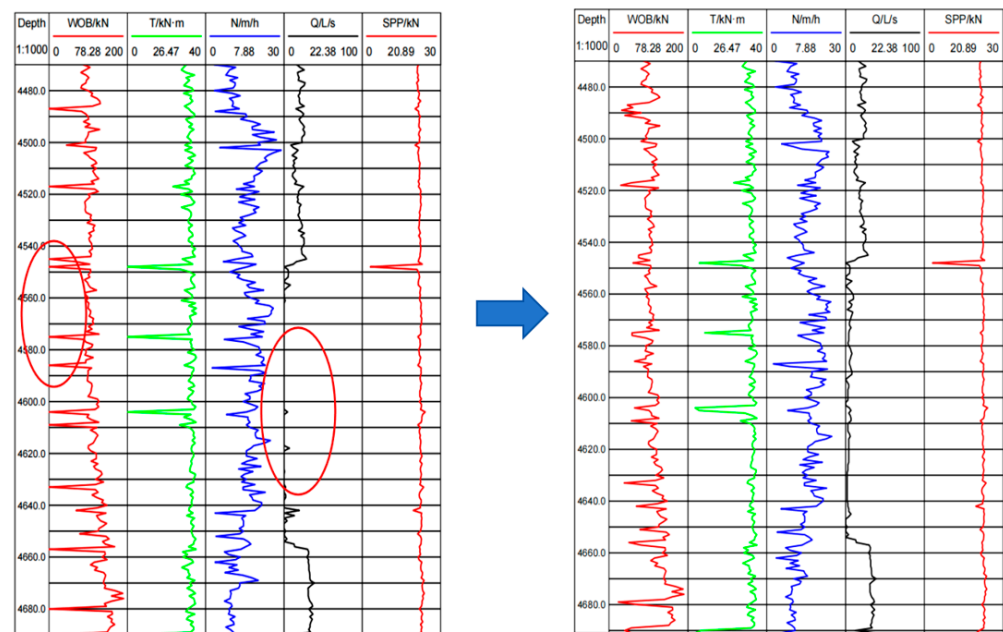


Figure 2. Comparison of drilling and mud logging data before and after cleaning and missing value filling.

After data screening, depth alignment, KNN-based imputation, and smoothing, the mechanical specific energy profile is obtained.

According to the corrected mechanical specific energy model and after data preprocessing, the mechanical specific energy data of the horizontal section of the example well are calculated. The original mechanical specific energy and processed mechanical specific energy data of the horizontal well section are shown in Figure 3.

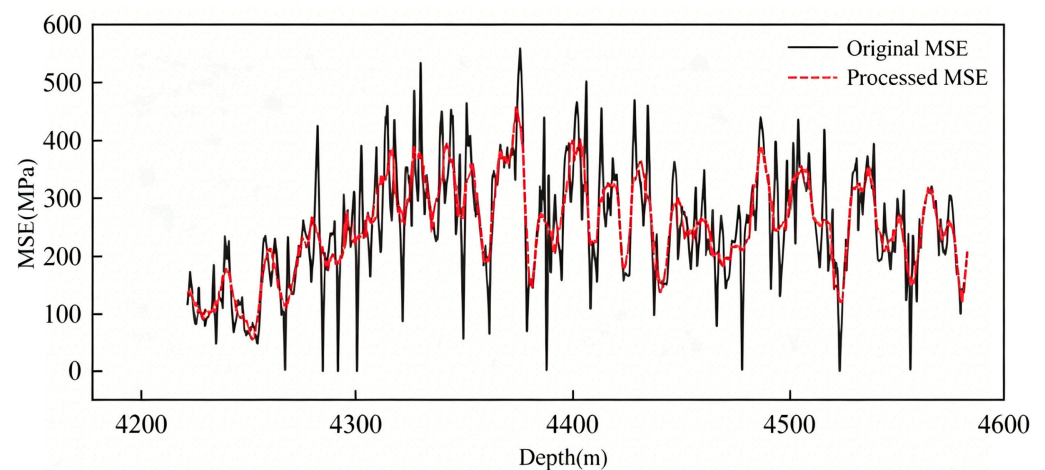


Figure 3. Schematic diagram of mechanical specific energy variations with well depth for the example well after the sequential preprocessing workflow.

2.3. Stage and Cluster Design Based on Mechanical Specific Energy

Figure 4's schematic diagram of staged multi-cluster fracturing of horizontal wells shows the division of the horizontal wellbore into multiple stages, each equipped with several perforation clusters for sequential multi-cluster fracturing treatment from the toe to the heel.

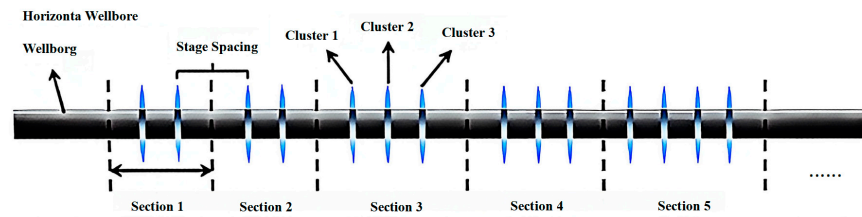


Figure 4. Schematic diagram of staged multi-cluster fracturing of horizontal wells.

2.4. Stage Design Using K-Medoids Clustering Algorithm

The K-medoids algorithm, also known as the PAM algorithm (Partitioning Around Medoids), is a partitioning unsupervised clustering algorithm based on medoids. It is a classic improved scheme for the K-means algorithm's poor noise resistance and the lack of physical meaning of centroids [20], and avoids the defect that the mean centroid of K-means is easily disturbed by outliers, with a more sensitive response to local fluctuations. The basic flow of the K-medoids clustering algorithm is shown in Figure 5.

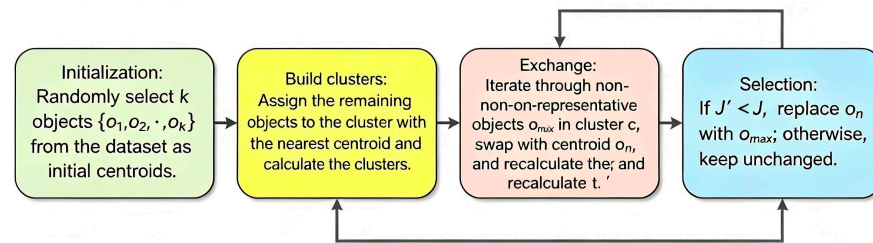


Figure 5. Basic flow of K-Medoids clustering algorithm.

The purpose of the K-medoids algorithm is to divide the dataset into K clusters $C = \{c_1, c_2, c_k\}$, where c_i represents the i -th cluster. The most central object in the cluster is taken as the representative object, i.e., the medoid; other objects are non-representative objects, and the sum of dissimilarities between all non-representative objects and the medoid of their cluster is minimized.

The expression for measuring the clustering effect with a function is [21]:

$$J = \sum_{j=1}^k \sum_{x_i \in c_j} D(s_i, o_j) \quad (10)$$

where s_i is the data object, o_j is the clustering center of cluster c_j , and $D(s_i, o_j)$ is the spatial Euclidean distance between s_i and o_j .

In the design of horizontal well staged fracturing, whether the number of stages and stage boundaries are reasonable is directly related to the fracturing effect. To determine the optimal number of clusters, this paper uses the elbow method for error analysis. As the K value gradually increases, the sum of the distances from samples to cluster centers first decreases rapidly and then tends to be flat. The K value corresponding to the inflection point of the curve is the appropriate number of clusters. The K-medoids clustering algorithm is applied to the mechanical specific energy data, and the Sum of Absolute Distances (SAD) is used as the clustering index with the following calculation formula:

$$SAD = \sum_{i=1}^K \sum_{x \in c_i} ||x - m_i||^2 \quad (11)$$

where m_i is the sample point of the i -th cluster.

Calculate the SAD values when K ranges from 2 to 15 and draw the elbow curve, as shown in Figure 6. The results show that the curve has an obvious inflection point at

$K = 11$, where the SAD value drops sharply from $K = 2$ to $K = 11$ and then flattens out with only minor decreases afterward, indicating that the marginal improvement of increasing the number of stages becomes negligible beyond this point. Therefore, $K = 11$ is determined to be the optimal number of stages.

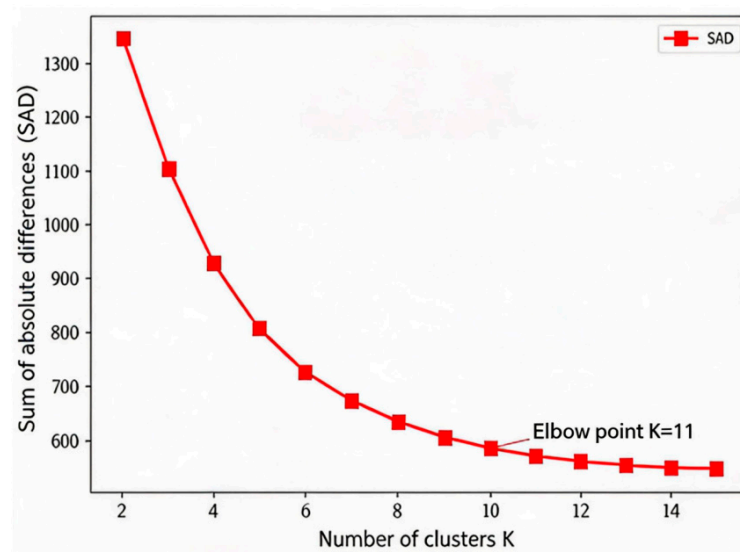


Figure 6. SAD curve of K-medoids algorithm.

The difference in elbow points between the two algorithms reflects their different responses to data distribution characteristics. Considering engineering practice comprehensively, the K-medoids clustering result is $K = 11$ for stage division; that is, the fracturing stages of the example well are divided into 11 stages.

Mechanical specific energy itself can reflect the lithology and physical property differences in the reservoir, so only this one parameter is used for stage division. Although reservoir heterogeneity is multidimensional, MSE serves as a practical indicator for stage division because it is continuously available from drilling data and is strongly correlated with rock strength and fracability. The horizontal section of the example well has a depth range of 4575–5239 m and a horizontal section length of 664 m.

The horizontal section stage division data are obtained as shown in Figure 7. It can be seen from Figure 7 that the stage lengths are generally distributed between 26.5 m and 86.0 m, which is reasonable and meets the conventional requirements for stage length in horizontal well staged fracturing technology in this block. The slightly shorter first stage is because the clustering algorithm automatically identifies the local mutation of the MSE curve, ensuring the uniformity of lithology and mechanical properties within the stage. The average mechanical specific energy values of each stage are shown in Figure 8.

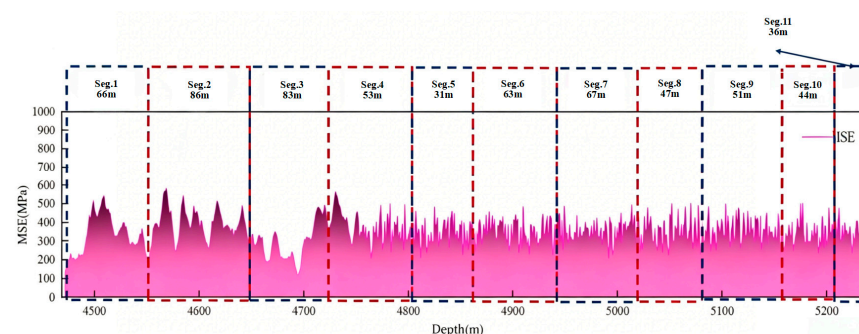


Figure 7. Schematic diagram of stage division results by K-medoids algorithm.

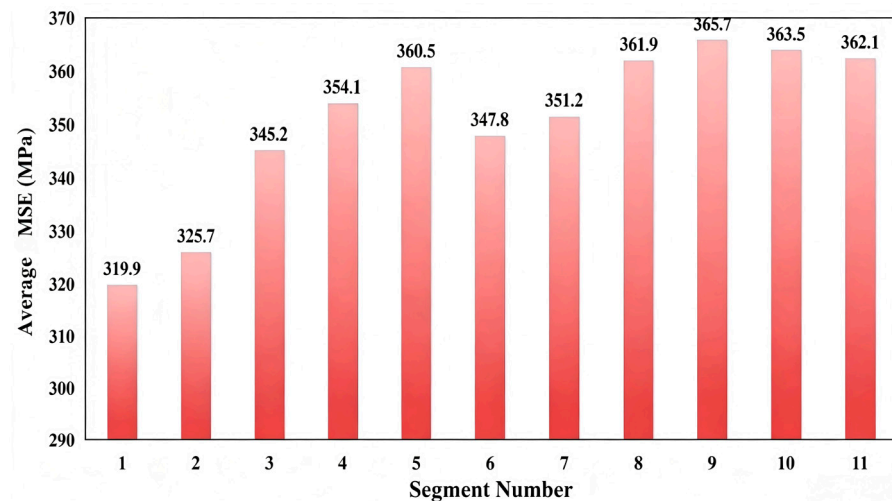


Figure 8. Average mechanical specific energy values of each stage.

2.5. Cluster Design Using Random Forest Scoring Method

After completing the stage division, the cluster positions within each stage are optimized based on the specified four parameters. To provide a clear understanding of the evaluation process, the following steps describe how the random forest model integrates the four parameters and generates the fracturing sweet spot index. The core idea is as follows: taking each potential perforation point (usually one sampling point per meter) as a sample and using the random forest algorithm to synthesize the performance of the four parameters, calculate a “fracturing sweet spot index”, and determine the final cluster positions according to the local peak points of the index.

(1) Label preparation: Within each divided fracturing stage, depth point data are collected along the well trajectory. Each depth point constitutes a sample, and each sample contains four characteristic parameters: oil saturation, porosity, permeability, and mechanical specific energy.

(2) Construction of random forest evaluation model: The advantage of random forest is that it can handle nonlinear relationships between features [22,23]. The random forest algorithm is used to fuse these four features to generate a comprehensive “fracturing sweet spot index” and give the importance ranking of each feature.

The four parameters of oil saturation, porosity, permeability, and mechanical specific energy at each depth point are combined into a four-dimensional vector as the feature matrix X .

According to the interpretation of different horizons in logging interpretation data, the four parameters are linearly weighted as the initial label, and the random forest model is trained with the FSI label. The random forest model was trained on a dataset consisting of depth points from the horizontal section of the example well at 1 m intervals. After data cleaning and quality control, 892 valid samples were retained for model development. These samples were divided into training and testing subsets, with the training portion used for model fitting and the testing portion used for performance evaluation.

(3) Calculation of fracturing sweet spot index and feature analysis: Based on the constructed evaluation model, the obtained FSI values are normalized data, which can accurately reflect the comprehensive reservoir quality along the well trajectory. The model output is designed to be interpretable, with feature importance scores reflecting the contribution of each parameter to the overall reservoir quality assessment. After the training of the random forest model is completed, the importance ranking of the four features will be output. If the order of oil saturation and permeability is significantly higher than the other two, it indicates that the reservoir productivity in this block is mainly restricted by

oil-bearing property and seepage capacity, and these areas should be preferentially perforated during cluster design. If the importance of mechanical specific energy is prominent, it means that the in situ stress heterogeneity is strong, and high mechanical specific energy areas should be avoided during cluster design to prevent failure to initiate fractures after fracturing construction.

(4) Determination of final perforation cluster positions: After obtaining the continuous FSI data, the cluster positions are optimized, and candidate intervals are delineated. Within each fracturing stage, an FSI threshold is set to screen out areas with a high sweet spot index as “candidate areas”. In the candidate areas, the local maximum points of the FSI curve are found. These points represent the positions with the best comprehensive geological and engineering characteristics in the stage and can usually be designed as the center positions of perforation clusters. If the local peak points are too dense, it is necessary to perform equal or unequal spacing thinning among these peak points according to the engineering optimal cluster spacing limit. The flow of the random forest scoring method is shown in Figure 9.

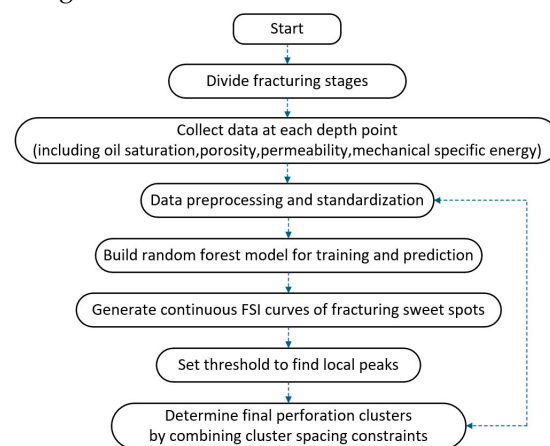


Figure 9. Flowchart of random forest scoring method.

The above procedure simplifies the fracturing design workflow in three aspects. First, it directly utilizes while-drilling data, avoiding dependence on complete logging data. Second, it automatically determines stage-cluster layouts through clustering and machine learning, minimizing manual intervention. Third, it provides quantitative guidance for cluster placement through the FSI, reducing the need for empirical trial and error.

Final principle: On the premise of meeting the minimum cluster spacing, the local peak points with the highest FSI values are preferentially selected as the final perforation cluster positions.

Taking the 10th stage in the stage design results of the K-medoids algorithm as an example, the four parameters of oil saturation, porosity, permeability, and mechanical specific energy in the stage are based on the results output by the random forest model, and a linear weighting formula of the fracturing sweet spot index (FSI) is constructed to quantitatively characterize the reservoir quality. The formula is as follows:

$$FSI = 0.32por_{norm} + 0.29so_{norm} + 0.27\log(perm)_{norm} + 0.12MSE_{norm} \quad (12)$$

where ϕ is porosity, %; so is oil saturation, %; $perm$ is permeability, mD; and MSE is the mechanical specific energy in the stage, MPa.

The weights were derived from the normalized feature importance scores output by the trained random forest model, representing the relative contribution of each parameter to the fracturing sweet spot index. FSI indicates that porosity and oil-bearing properties are the key factors that determine the comprehensive quality of the reservoir in this fracturing

stage; seepage capacity plays an auxiliary role, and in situ stress heterogeneity has a weak influence on sweet spot distribution. Therefore, in the optimization of perforation clusters, intervals with high porosity and oil saturation should be preferentially selected to ensure the advantages of reservoir storage capacity and oil-bearing conditions; permeability can be used as a supplementary screening index, and mechanical specific energy should be properly avoided in local extremely high value areas. The cluster design results and schematic diagram are shown in Table 3 and Figure 10.

Table 3. Cluster design results by random forest evaluation method.

Cluster No.	Perforation Cluster Center Position (m)	FSI Value
1	5098	0.130361931
2	5108	0.996541214
3	5124	0.907627576
4	5135	0.982058515
5	5146	0.937237936
6	5159	0.922867813

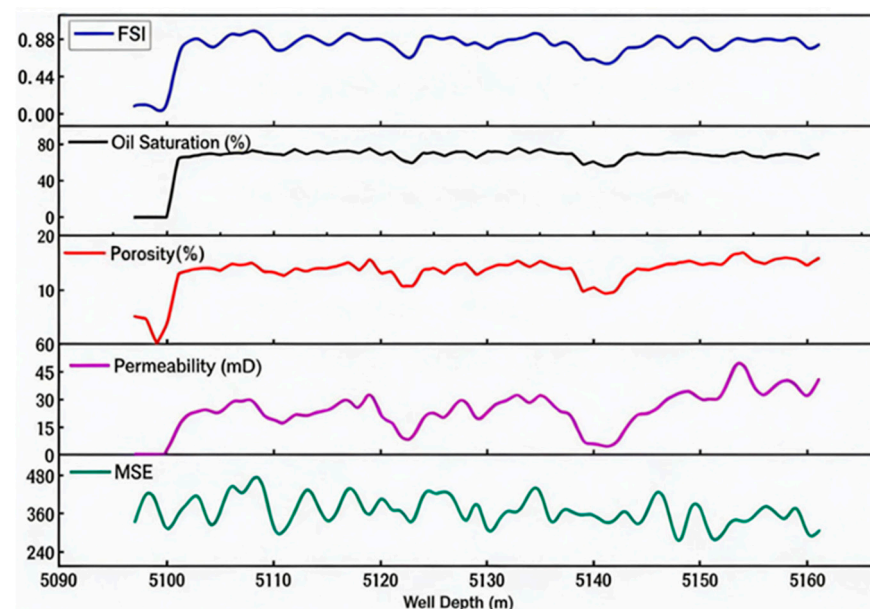


Figure 10. Schematic diagram of cluster design results by random forest evaluation method.

3. Results and Discussion

On the basis of the stage-cluster division results, a fracturing model of the example well was constructed to further verify the rationality of the division results. The numerical simulations were performed using the Petrel platform (Schlumberger) with the Kinetix module for hydraulic fracture propagation modeling. The implementation effect of the scheme was then evaluated by comparison with field fracturing construction data, providing a reliable basis for on-site fracturing design.

In the numerical model, fluid flow within the fracture is governed by the Reynolds lubrication equation, and rock deformation follows linear elastic fracture mechanics. The model domain was discretized using quadrilateral elements with local mesh refinement near the fracture tip. Roller supports were applied as boundary conditions, with in situ stress prescribed as initial conditions.

Based on the optimization results of fracturing construction parameters, the fracturing target layer of the example well is at 4487.8–5192 m. To improve the degree of reservoir

stimulation, plug-and-perf staged fracturing technology is adopted, and proppant fracturing is carried out by casing injection. The fracturing fluid adopts a slickwater system. The original fracturing construction scheme divides the fracturing target layer into nine stages and 54 clusters. The optimized perforation parameters are shown in Table 4.

Table 4. Optimized perforation parameters.

Perforation Stage No.	Well Section (m)	Thickness (m)	Number of Clusters	Perforations per Cluster	Total Perforations per Stage	Plug Position (m)
11	4469.6–4535.5	65.9	5	4	20	4545
10	4568.9–4654.9	89.6	7	3	27	4658
9	4656.3–4698.9	42.6	3	8	24	4701
8	4701.1–4773.8	72.7	7	3	21	4773
7	4778.9–4830.3	51.4	6	4	24	4837
6	4831.6–4894.4	62.8	5	5	25	4896
5	4898.9–4946.3	47.4	5	5	25	4951
4	4956–5003.8	47.8	6	4	24	5009
3	5013.5–5095.4	81.9	7	3	21	5096
2	5097.3–5161.5	64.2	6	4	24	5163
1	5163.5–5190	26.5	3	10	30	/

During on-site construction, construction parameters such as proppant volume, sand ratio, fluid volume, and displacement should be adjusted in time according to pressure changes. If the on-site construction pressure is high and proppant addition is difficult, the proppant addition mode and pumping schedule should be adjusted according to the on-site situation.

Fracturing result analysis is carried out by comparing with on-site construction data. Table 5 shows the optimization results of fracturing parameters for each stage of the example well. The example well has 11 fracturing stages, with a small number of perforation clusters in the first stage and a fracturing stage length of 26.5 m, and the remaining stage lengths are between 42 m and 90 m. The total fracturing fluid volume is 20,711.0 m³, the total proppant volume is 1792.8 m³, and the average fracture length of each stage is between 190 m and 280 m. The reservoir is fully stimulated, and the simulation results are in good agreement with the design results. The fracture width of the horizontal section is shown in Figure 11.

Table 5. Fracturing optimization results.

Stage No.	Stage Length (m)	Number of Clusters	Average Fracture Height (m)	Average Fracture Length (m)	Total Fracture Area (10 ² m ²)	Stimulated Reservoir Volume (SRV) (10 ⁴ m ³)
1	26.5	3	22.36	191.34	320.21	26.92
2	64.2	6	21.77	273.46	1080.20	41.22
3	81.9	7	20.96	223.68	1136.21	35.35
4	47.8	6	19.03	256.34	475.23	28.13
5	47.4	5	21.56	198.43	417.36	32.36
6	62.8	5	23.42	213.20	998.64	34.51
7	51.4	6	18.96	213.56	756.32	25.94
8	72.7	7	22.38	186.57	990.20	33.92
9	42.6	3	19.61	257.36	401.27	25.79
10	89.6	7	20.39	248.37	983.23	32.47
11	65.9	5	22.30	224.13	1002.10	34.22

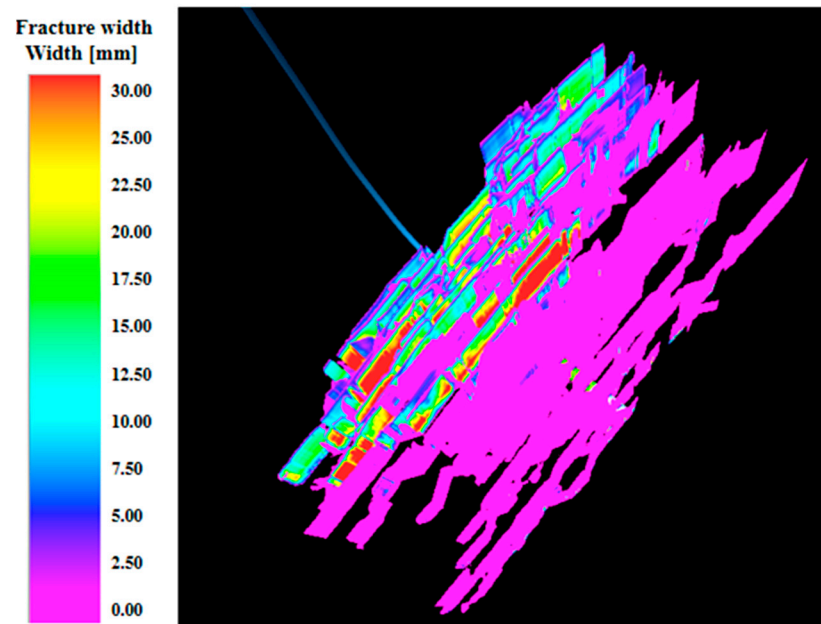


Figure 11. Fracture width diagram after optimization.

Meanwhile, during on-site construction, the second stage of the example well had continuous pressure increase during fracturing because the fracturing target layer was mudstone with a high shale content and strong plasticity, and only four clusters with 32 perforations were perforated, resulting in a small number of perforation clusters. In this optimization, the perforation clusters are changed to six clusters with 60 perforations, the viscosity of slickwater is increased, the perforation friction is reduced, and the construction success rate is improved. Comparing Figures 12 and 13, it can be seen that the construction pressure is significantly reduced, and the effect of the optimized perforation scheme is obvious.

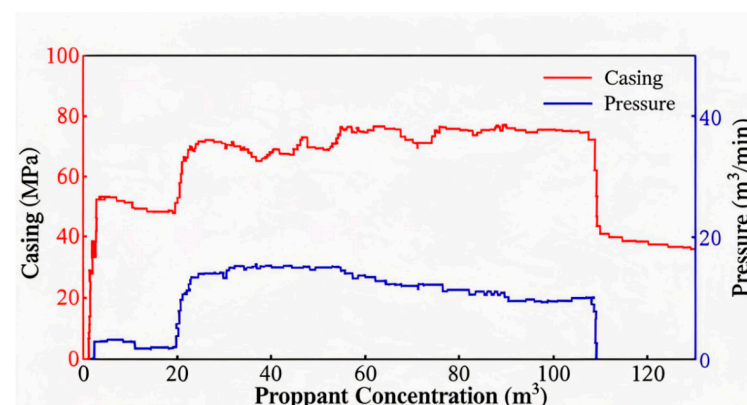


Figure 12. Field fracturing construction curve of the second stage.

The fracturing optimization scheme is compared with the field construction results, and the optimization scheme is higher than the field construction in all key indicators. The comparison results before and after optimization are shown in Figure 14.

The average fracture height of the optimization scheme is 21.16 m, and the field construction result is 19.01 m, which is 2.15 m higher than the optimization scheme, with a relative increase of 11.3%; the average fracture length of the optimization scheme is 226.06 m, and the field construction result is 208.16 m, which is 17.90 m higher than the optimization scheme, with a relative increase of 8.6%; the total fracture area of

the optimization scheme reaches $778.27 \times 10^2 \text{ m}^2$, and the field construction result is $684.88 \times 10^2 \text{ m}^2$, which is $93.39 \times 10^2 \text{ m}^2$ higher than the optimization scheme, with a relative increase of 13.6%. In terms of stimulated reservoir volume, the optimized scheme is $31.89 \times 10^4 \text{ m}^3$, and the field construction result is $26.66 \times 10^4 \text{ m}^3$, which is $5.23 \times 10^4 \text{ m}^3$ higher than the optimization scheme, with a relative increase of 19.6%. From fracture height, fracture length, total fracture area to stimulated reservoir volume, all values of the optimization scheme are greater than the field construction results. Compared with the on-site construction scheme, this optimization design shows obvious advantages in the vertical and horizontal expansion size of fractures, fracture coverage range, and overall reservoir stimulated volume.

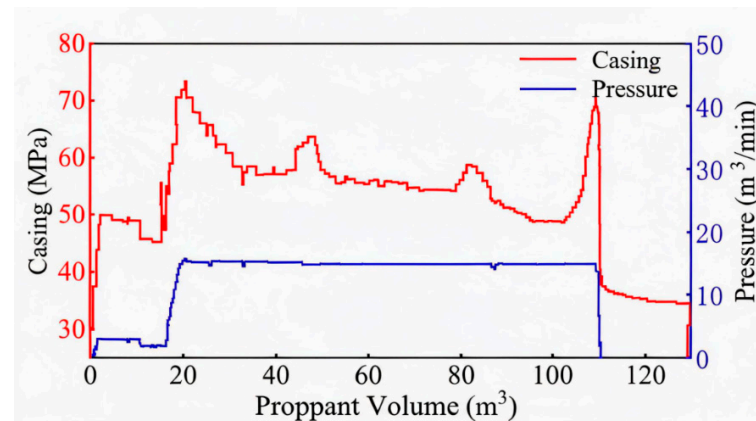


Figure 13. Simulated fracturing construction curve of the second stage after optimization.

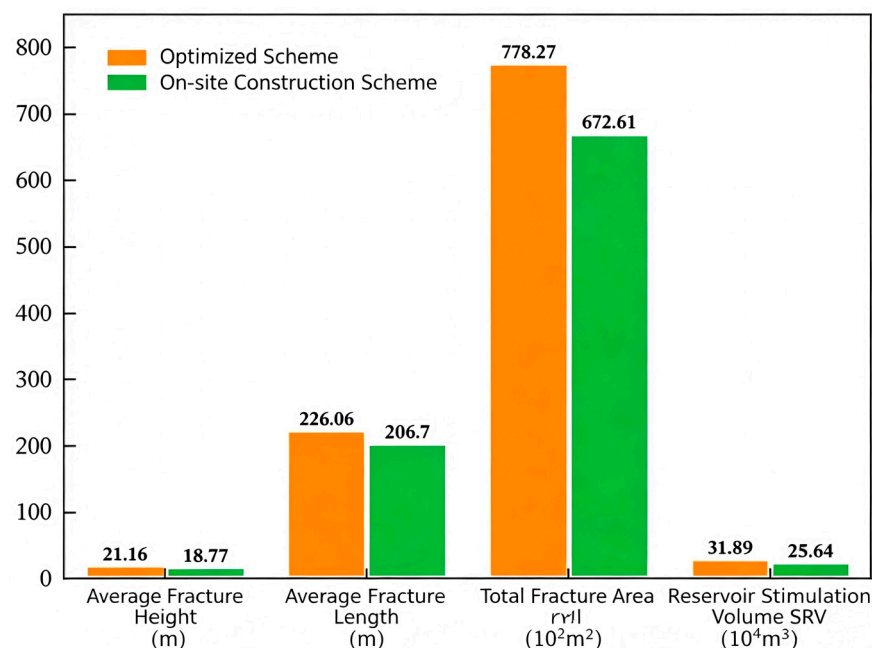


Figure 14. Comparison diagram of results before and after optimization.

4. Conclusions

This paper takes bottomhole mechanical specific energy as the core parameter, combines clustering intelligent algorithms to form an optimization design method for staged fracturing and cluster design of horizontal wells in low-permeability reservoirs, and obtains the following conclusions after verification by numerical simulation and field data.

(1) Through the correction of drill string friction, hydraulic energy, bit torque, and rotational speed, an effective bottomhole mechanical specific energy model is established, which eliminates the interference of non-rock-breaking factors on drilling and mud logging data and can more truly reflect the energy consumption of bit breaking rock and reservoir rock strength characteristics.

(2) The original drilling and mud logging data are cleaned and smoothed, the optimal number of clusters is determined combined with the elbow method, and unsupervised clustering stage division of mechanical specific energy is realized by using the K-medoids algorithm. The stage length distribution is reasonable, ensuring the uniformity of lithology and mechanical properties within the stage.

(3) Based on the random forest algorithm, four parameters of oil saturation, porosity, permeability, and mechanical specific energy are fused to construct a fracturing sweet spot index (FSI) evaluation model, which realizes automatic optimization of perforation cluster positions within stages, can differentially design cluster spacing, and effectively guides the refined formulation of fracture placement schemes.

(4) Numerical simulation and field construction comparison verification show that the optimized perforation parameters significantly improve the fracturing stimulation effect: the average fracture length, total fracture area, and stimulated reservoir volume are increased by 8.6%, 13.6%, and 19.6%, respectively, compared with the field scheme, and the construction pressure is reduced at the same time, verifying the rationality and engineering practicability of the stage-cluster division method based on mechanical specific energy.

(5) It should be acknowledged that the validation in this study is based on a single example well, which may limit the generalizability of the proposed method. Future work will focus on applying this method to multiple wells in different reservoirs to further validate its applicability and robustness.

Author Contributions: Conceptualization, M.L. and M.Z.; Methodology, Y.P. and M.Z.; Software, M.Z.; Validation, X.G. and Y.Z.; Formal analysis, X.L. and Y.Z.; Investigation, P.L.; Resources, Y.P. and F.Q.; Data curation, X.S.; Writing—original draft, M.Z.; Writing—review & editing, M.Z.; Supervision, Y.P. and X.L.; Project administration, Y.P.; Funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by “Personalized fracturing technology and test of multi type shale oil”, grant number 2023ZZ28YJ04.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Authors Yuan Pan, Min Liu, Xuewei Liu, Feixiang Qin, Xiaoting Gou, Yingxi Zhang, Xiangyun Sui and Peipei Liu were employed by the Petroleum Engineering Research Institute, PetroChina Dagang Oilfield Company. 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

1. Xu, D.J.; Yin, W.L.; Liu, W.P.; Guo, Z.; Wu, K.; He, F.; Niu, S.; Xu, Z. Research progress of hydraulic fracture monitoring technology. *Sci. Technol. Eng.* **2026**, *26*, 901–913.
2. Yuan, G.J.; Fu, L.; Wang, Y.; Guo, K.; Chen, G. Current status and development suggestions of drilling and completion technologies for economic and effective development of unconventional oil and gas in China. *Pet. Drill. Tech.* **2022**, *50*, 1–12.
3. Wang, G.F.; Li, F.X.; Wang, H.B.; Zhou, T.; Zhang, Y.; Wang, R.; Li, N.; Chen, Y.; Xiong, X. Practice and understanding of integrated geological-engineering fracturing in unconventional gas reservoirs in Sichuan Basin. *Oil Gas Geol.* **2022**, *43*, 1221–1237.
4. Jiang, W.Y.; Yu, H.Y.; Chen, C.W.; Song, S.; Gao, L.; Wang, X.; Liu, G.; Feng, J. Exploration and practice of large-scale benefit production of continental shale oil. *Mod. Ind. Econ. Informationization* **2023**, *13*, 249–252.

5. Zou, C.N.; Yu, R.Z.; Dong, D.Z.; Zhang, X.; Chen, Y.; Zheng, M.; Liu, H.; Gao, J. Geological characteristics, development technologies and development prospects of unconventional natural gas in China. *Pet. Explor. Dev.* **2026**, *53*, 659–673.
6. Pu, X.G.; Han, W.Z.; Xu, J.; Shi, Z.; Dong, J.; Wei, J.; Zhang, W.; Wang, N.; Ji, L. Formation and enrichment conditions and exploration and development practice of shale oil in faulted lake basins in eastern China: A case study of shale oil in Huanghua Depression, Bohai Bay Basin. *Reserv. Eval. Dev.* **2026**, *16*, 238–255.
7. Jiang, T.X.; Zhou, J. Discussion on the evolution path of stimulation theory and technology for shale oil and gas reservoirs. *Oil Gas Geol.* **2025**, *46*, 2041–2058.
8. Jiang, J. Research on Large-Scale Fracturing Development Strategy and Economic Effect of Thin Interbedded Tight Oil Reservoirs. Ph.D. Thesis, Northeast Petroleum University, Daqing, China, 2025.
9. Shen, B.J.; Li, D.; He, J.H.; Xu, B.; Wu, Y.; Wang, R.; Jiang, R.; Li, R.; Huo, Z.; Liu, K. Intelligent prediction method of in-situ stress for single well in shale reservoirs driven by multi-source data and its application. *J. Geomech.* **2026**, *32*, 227–244.
10. Liu, C.L.; Liu, X.J.; Zhang, H.J.; Fan, L.; Yang, X.; Zang, Q.; Dai, B.; Meng, Y.; Huo, H.; Wang, F. Practice of integrated geological-engineering technology in shale oil in Ansai area, Ordos Basin. *Oil Gas Geol.* **2022**, *43*, 1238–1248.
11. Teale, R. The concept of specific energy in rock drilling. *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.* **1965**, *2*, 57–73. [[CrossRef](#)]
12. Meng, Y.F.; Yang, M.; Li, G.; Li, Y.; Tang, S.; Zhang, J.; Lin, S. A new method for while-drilling evaluation and optimization of drilling efficiency based on mechanical specific energy theory. *J. China Univ. Pet. (Ed. Nat. Sci.)* **2012**, *36*, 110–114.
13. Cui, M.; Li, J.J.; Ji, G.D.; Chen, Y. Optimization method of compound drilling parameters based on mechanical specific energy theory. *Pet. Drill. Tech.* **2014**, *42*, 66–70.
14. Downie, R.; Daves, D. Improving Hydraulic Fracturing Performance and Interpreting Fracture Geometry Based on Drilling Measurements. In Proceedings of the SPE Hydraulic Fracturing Technology Conference and Exhibition, SPE, Woodlands, TX, USA, 5–7 February 2019.
15. Wang, Z.P.; Li, J.K.; Ma, W.Y.; Li, Y.; Qiu, S.; Wu, J.; Zhang, W. Optimization of Horizontal Well Fracturing Design Using Mechanical Specific Energy (MSE). In *Proceedings of the International Field Exploration and Development Conference 2021, IFEDC 2021*; Springer: Singapore, 2022.
16. Hu, S.M.; Mao, S.; Shi, S.Z.; Sheng, M.; Xu, B.; Li, G. Optimization of fracturing stages/clusters in horizontal well based on unsupervised clustering of bottomhole mechanical specific energy on the bit. *Nat. Gas Ind. B* **2023**, *10*, 583–590. [[CrossRef](#)]
17. Zhuang, X.; Liu, Y.; Hu, Y.; Guo, H.; Nguyen, B.H. Prediction of rock fracture pressure in hydraulic fracturing with interpretable machine learning and mechanical specific energy theory. *Rock Mech. Bull.* **2025**, *4*, 68–72. [[CrossRef](#)]
18. Zeng, F.H.; Hu, D.G.; Guo, J.C.; Zhang, N.; Zheng, B.; Bai, X.; Chen, Z. Intelligent optimization of staged multi-cluster differential fracturing parameters for shale gas horizontal wells based on drilling and mud logging data. *Nat. Gas Ind.* **2025**, *45*, 84–94.
19. Li, D.; Chi, J.J.; Xiang, B.; Wang, M. Petroleum data missing filling algorithm based on SMOTE and KNN. *Math. Pract. Theory* **2019**, *49*, 187–195.
20. Wang, L.F. Research on K-Medoids Clustering Algorithm and Its Application in Text Clustering. Master's Thesis, Chongqing University of Technology, Chongqing, China, 2017.
21. Zhang, L.; Song, C.Y.; Zou, Y.J.; Hong, C.; Wang, Z. Bearing performance degradation assessment based on Renyi entropy and K-medoids clustering. *J. Vib. Shock.* **2020**, *39*, 24–31+46.
22. Abdulsattar, Y.S.; Hossam, R.A. Improved Random Forest Algorithm Performance For Big Data. *J. Phys. Conf. Ser.* **2021**, *1897*, 012035. [[CrossRef](#)]
23. Liliya, D.; Ivkina, M.S. An approach to determining the search range boundaries of optimal parameters values for the random forest algorithm. *J. Phys. Conf. Ser.* **2021**, *1902*, 012047. [[CrossRef](#)]

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.