Petroleum Science 22 (2025) 5068–5083

Contents lists available at ScienceDirect

Petroleum Science

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

Original Paper

Mechanism-guided data-driven model for optimized completion design

Shi-Meng Hu [a,b,c], Mao Sheng [a,b,c,*], Bing-Bing Liu [a,b,c], Jie Li [c,d], Shou-Ceng Tian [a,b,c], Xiao-Dong He [c,d], Gen-Sheng Li [a,b,c]

[a] College of Artificial Intelligence, China University of Petroleum (Beijing), Beijing, 102249, China
[b] State Key Laboratory of Deep Geothermal Resources, China University of Petroleum (Beijing), Beijing, 102249, China
[c] Research Center for Intelligent Drilling & Completion Technology and Equipment, China University of Petroleum (Beijing), Beijing, 102249, China
[d] The Oil Production Technology Research Institute of Xinjiang Oilfield Company, PetroChina Xinjiang Oilfield Company, Karamay, 834000, Xinjiang, China

ARTICLE INFO

Article history:
Received 21 February 2025
Received in revised form
16 September 2025
Accepted 16 September 2025
Available online 22 September 2025

Edited by Jia-Jia Fei

Keywords:
Unconventional reservoirs
Intelligent fracturing
Completion design optimization
Mechanical specific energy
Dynamic programming
Multi-objective optimization
Stage and cluster optimization

ABSTRACT

Effective completion design in hydraulic fracturing (HF) is crucial for optimizing production in unconventional reservoirs. Traditional geometric designs often fail to account for geological and engineering heterogeneity, leading to suboptimal stimulation. This study introduces a mechanism-guided data-driven model for optimized completion design that covers the entire process from sweet spot evaluation to stage and cluster optimization. For geological sweet spot evaluation, a mechanism-guided weighted *K*-medoids clustering model was developed by assigning weights to petrophysical parameters based on their correlation with production profiles. Engineering sweet spots were characterized using bottomhole mechanical specific energy (MSE_b) and minimum horizontal in-situ stress (S_{hmin}). The completion design optimization employed dynamic programming and a hybrid multi-objective optimization approach (NSGA-II), integrating geological and engineering sweet spots with operational constraints. The study showed a positive correlation between high-quality geological sweet spots and production (average correlation coefficient of 0.34), and a negative correlation between fluid allocation and engineering sweet spots (correlation coefficient of −0.46). Field application in the Jimsar Sag, Xinjiang, demonstrated that the proposed model significantly outperforms traditional geometric designs. Test wells showed an average 186% increase in cumulative production per 100 m over three months compared to conventional wells. The key findings of this work provide a novel technical pathway for optimized completion design of unconventional reservoirs with significant engineering applicability.

1. Introduction

Unconventional oil and gas resources led by shale play an increasingly important role in global energy supply. Multistage hydraulic fracturing (MSHF) is a key technology for developing unconventional resources, and its completion design significantly impacts the effectiveness of reservoir stimulation. A prerequisite for effective completion design is the accurate evaluation of sweet spots—zones with favorable geological and engineering characteristics. Identifying these sweet spots and optimizing completion design are essential for targeting high-quality reservoirs and achieving uniform fracture initiation and propagation.

The concept of sweet spots is widely utilized in oil and gas exploration and development. Early studies primarily focused on geological sweet spots, which describe blocks with relatively good reservoir quality (Prise et al., 1993; Hart et al., 2000; Zhang et al., 2000; Zhao et al., 2004). With the advancement of unconventional reservoir development, the definition has expanded to include engineering sweet spots, which emphasize geomechanical properties conducive to HF (Pitcher et al., 2012; Chorn et al., 2013; Bhattacharya and Carr, 2019; Liao et al., 2020; Sun et al., 2023). Geological sweet spots refer to zones with high hydrocarbon content and favorable petrophysical properties, while engineering

* Corresponding author.
E-mail address: shengmao@cup.edu.cn (M. Sheng).
Peer review under the responsibility of China University of Petroleum (Beijing).

https://doi.org/10.1016/j.petsci.2025.09.031

sweet spots denote areas conducive to efficient HF (Chen et al., 2018; Liao et al., 2020; Sun et al., 2023). Geological sweet spots are typically evaluated using parameters such as clay content, total organic carbon, porosity, and permeability (Ko et al., 2016; Buller et al., 2010), whereas engineering sweet spots rely on geomechanical indicators like Young's modulus, Poisson's ratio, and in-situ stresses (Liao et al., 2014; Chen et al., 2017). However, the strong nonlinear relationships and the complexity of these parameters make it challenging to discern their relative importance, highlighting the necessity of identifying key parameters as a primary task.

Until now, studies that comprehensively considers both geological and engineering sweet spots and their application in completion design is still limited compared to those focusing solely on geological or engineering sweet spots. In these studies, a common approach involves using methods such as single-factor evaluation, multi-factor correlation analysis method or radar area models to determine classification thresholds for the optimal parameters, thereby constructing coupled classification arrays for geological and engineering sweet spots to form a comprehensive 'double sweet spot' evaluation table (Zhou et al., 2023; Zhang et al., 2024; Shen et al., 2024). Based on this evaluation table, with perforation clusters increased in fracturing stage where 'double sweet spot' was favorable and sparsified in less favorable fracturing stage. Nonetheless, it is still hard to resolve the issue of achieving balanced fracture initiation. On one hand, these classification thresholds often vary depending on the selected methodology and reservoir characteristics, resulting in challenges such as non-uniqueness and limited regional applicability. On the other hand, while both geological and engineering sweet spots were considered and the perforation cluster density is increased in high-quality stages, a geometric completion design was applied for perforation cluster positioning within each stage, without accounting for the variability in fracture initiation pressure between clusters.

In recent years, with the application of artificial intelligence (AI) in oil and gas exploration and development (Li et al., 2024, 2025; Sheng et al., 2024; Li et al., 2022), clustering algorithms have emerged as a promising alternative for sweet spot evaluation and have been utilized in completion design optimization. Multidimensional parameters are input into an unsupervised clustering model, and the clustering results are analyzed to identify reservoir sweet spots, allowing perforation clusters to be placed in high-quality sweet spot well intervals (Zhang et al., 2021; Hu et al., 2023a, b). Despite its potential, this purely data-driven method struggles to distinguish and integrate geological and engineering sweet spots, requiring substantial effort to incorporate mechanistic insights for result interpretation, which limits its ability to comprehensively consider 'double sweet spot' in optimized completion design. Hu et al. proposed a clustering model aimed at achieving balanced fracturing by evaluating initiation pressure using bottomhole mechanical specific energy (MSE_b), but geological sweet spots were not considered (Hu et al., 2023a, b). Therefore, a mechanism-guided AI approach is still needed to evaluate both geological and engineering sweet spots and integrate 'double sweet spot' for optimizing completion design.

To address these limitations, this study proposes a mechanism-guided, data-driven model that integrates geological and engineering sweet spot evaluation with intelligent completion design. The model combines mechanism-guided clustering with geomechanical strength indicators, and incorporates dynamic programming and multi-objective optimization to automate fracturing stage design and perforation cluster placement. Unlike conventional methods that treat geological and engineering factors separately or rely on empirical geometric designs, this approach enables a unified and physically consistent optimization workflow. Built entirely on standard drilling and logging data with low computational cost, the proposed model provides a practical and effective solution for unconventional reservoir development.

The paper is organized as follows. First, we introduce the data and preprocessing methods employed in this study. Next, a mechanism-guided weighted clustering model is developed to accurately identify and evaluate geological sweet spots by emphasizing influential reservoir petrophysical parameters. Then, geomechanical strength is adopted as the engineering sweet spot parameter, and a dynamic programming algorithm is introduced to optimize the design of fracturing stages. Subsequently, a hybrid multi-objective optimization algorithm is employed to integrate the 'double sweet spot' and engineering constraints, thereby optimizing perforation cluster design. Finally, the proposed model is comprehensively evaluated—with model performance comparison, monitoring data validation, and production monitoring of field pilot well tests—to demonstrate its effectiveness in sweet spot evaluation and production enhancement via optimized completion design.

2. Methodology

To further illustrate how logging and drilling data can be combined with AI algorithm for geological and engineering sweet spot evaluation and optimized completion design. This entire procedure consists of three parts: (1) Data preprocessing; (2) Model and algorithm construction; (3) Validation and application.

(1) **Data preprocessing:** ① Exploratory data analysis was conducted on raw drilling and logging datasets, and the resolution was unified, missing values and outliers were processed. ② Reservoir petrophysical and geomechanical parameters were calculated, followed by smoothing and noise reduction techniques to enhance data reliability.

(2) **Model and algorithm construction**: ① Correlation analysis between production fluid profiles and reservoir petrophysical parameters was performed to construct a mechanism-guided weighted geological sweet spot clustering model, enabling intelligent hierarchical evaluation of geological sweet spots. ② Reservoir geomechanical strength was characterized using bottom-hole mechanical specific energy and minimum horizontal principal stress, establishing an engineering sweet spot evaluation indicator. ③ A dynamic programming-based algorithm was developed to optimize fracturing stage design, targeting balanced fracture initiation. ④ A hybrid multi-objective optimization algorithm integrating geological sweet spots, engineering sweet spots, and physical constraints was proposed to optimize perforation cluster design.

(3) **Validation and application:**① The efficiency and practical advantages of the proposed model were validated through a performance comparison with traditional models and algorithms. ② The effectiveness of geological and engineering sweet spot evaluation models was validated against monitoring datasets, demonstrating robust consistency between predictions and field observations. ③ Comparative analysis of production performance between optimized designed test wells and conventionally designed wells confirmed the method's capability to enhance production in unconventional reservoir stimulation.

The whole workflow is shown in Fig. 1.

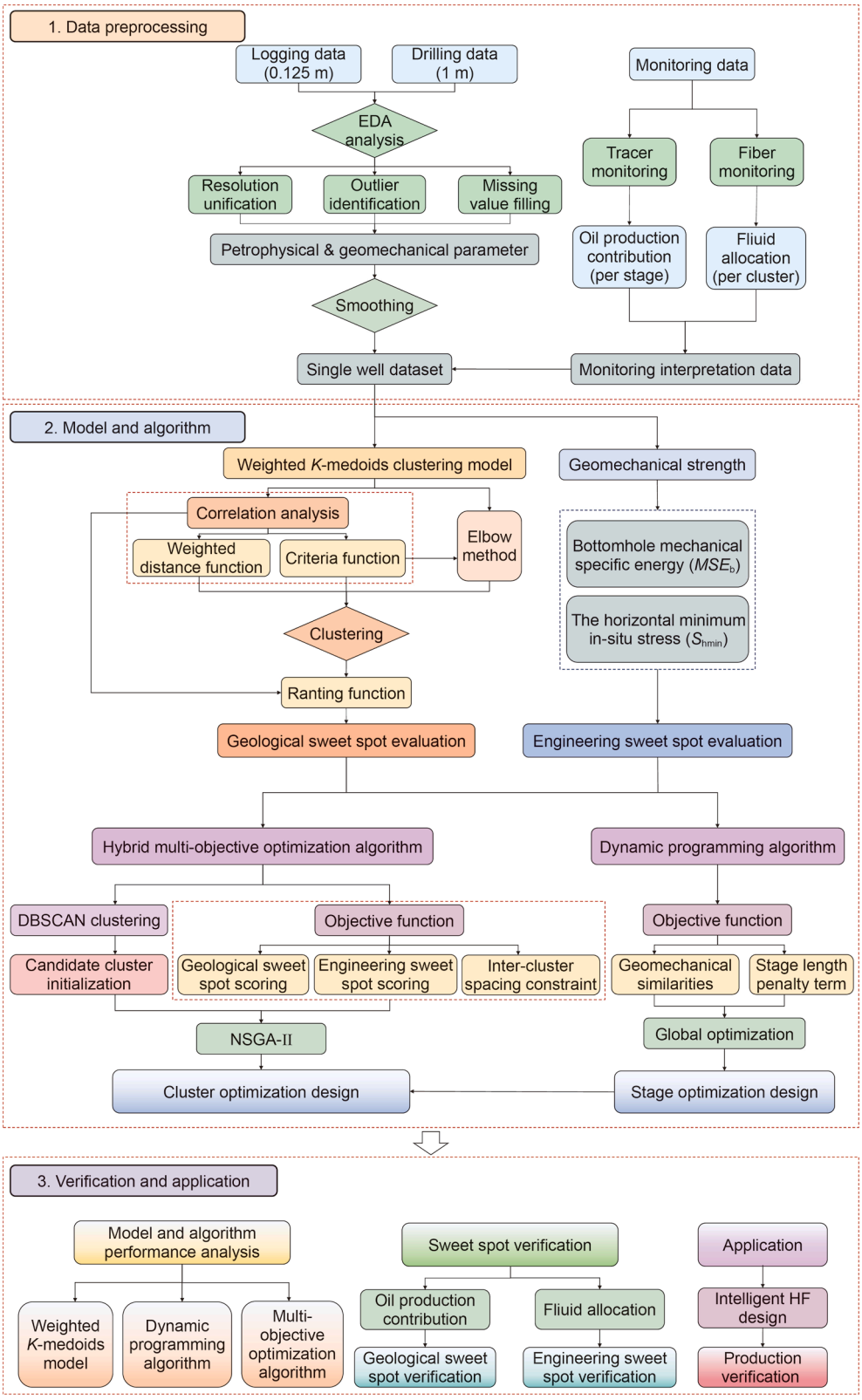

Fig. 1. Workflow for a mechanism-guided data-driven model for optimized completion design.

2.1. Dataset selection and preprocessing

2.1.1. Data selection

The effectiveness of MSHF in horizontal wells is governed by a complex interplay of reservoir physical properties, geomechanical characteristics, and completion design. During drilling, completion, and production, valuable data can be collected including logging data, drilling data, and post-fracturing monitoring data, with the monitoring data comprising stage production profiles through tracer diagnostics and fluid allocation through fiber

monitoring. We have selected a suite of physically meaningful parameters to evaluate sweet spot, which satisfies the dual criteria: On the one hand, these parameters include reservoir petrophysical and geomechanical parameters, which can effectively characterize the reservoir flow capacity, storage performance and mechanical characteristics. On the other hand, all selected parameters are computationally derivable from oil field logging and drilling data without requiring specialized logging programs, ensuring practical applicability for large-scale field implementations.

Geological sweet spot evaluation primarily relies on petrophysical parameters quantifying reservoir flow capacity and storage performance, including porosity, oil saturation, permeability, and clay volume. Engineering sweet spot assessment focuses on geomechanical indicators comprising in-situ stress profiles and bottomhole mechanical specific energy. The original datasets and corresponding parameters are systematically presented in Table 1.

2.1.2. Data preprocessing and smoothing

Data quality significantly impacts the performance of AI algorithms in modeling. Raw datasets often exhibit discontinuities and anomalies caused by environmental factors, instrument performance, or human errors. Therefore, raw data must be preprocessed before it can be utilized in subsequent machine learning processes. In this work, we primarily employed exploratory data analysis (EDA) through statistical methods and visualization to understand the structure, distribution, and quality of the dataset (Behrens, 2012; Zhu et al., 2024). The preprocessing procedure is as follows:

1) Exploratory data analysis (EDA): An initial EDA was conducted to gain a fundamental understanding of the dataset, including its structure, statistical distribution, and quality issues.
2) Resolution unification: Data acquired with varying measurement accuracies were first unified in resolution. For instance, drilling data recorded at 1 m intervals were interpolated to match the 0.125 m resolution of the well logging data.
3) Data cleaning: Outliers and missing values were then addressed. Missing values were imputec using the *k*-nearest neighbors (KNN) method (Halder et al., 2024). Outliers were identified using the interquartile range (IQR) method (Reimann et al., 2017): the IQR is computed as the difference between the third quartile ($Q3$) and the first quartile ($Q1$), and any data points outside the range [$Q1-1.5 \times$ IQR, $Q3+1.5 \times$ IQR] were classified as outliers. These outliers were replaced by the value of their nearest neighbor to minimize bias.
4) Parameter calculation: From the preprocessed raw data, the dataset of petrophysical and geomechanical parameters is constructed via the above process. The data are stored in order of ascending well depth at intervals of 0.125 m.

Table 1
Data and parameters.

Category		Attribute
Raw data	Logging data	Gamma ray logs (GR) Density logs (DEN) Sonic logs (DT) Neutron logs (CNL) Resistivity logs (RT)
	Drilling data	Weight on bit (WOB) Torque (T) Rotational speed (N) Rate of penetration (ROP) Area of bit (A_b)
	Monitoring data	Tracer monitoring data Fiber monitoring data
Parameter	Petrophysical parameter	Porosity (ϕ) Permeability (K) Shale content (V_{sh}) Oil saturation (S_o)
	Geomechanical parameter	In-situ stress (S_{hmin}) Bottomhole mechanical specific energy (MSE_b)
	Monitoring interpretation parameters	Oil production profiles Fluid allocation

As shown in Fig. 2, the petrophysical and geomechanical parameters directly calculated from the raw data exhibits considerable fluctuations and numerous discontinuous points. Hence, it is difficult to distinguish the macroscopic differences of various parameters along the horizontal wellbore. To reduce the noise and enhance the reliability of subsequent clustering evaluations, a data smoothing procedure was implemented. Given that all of the parameter curves represent spatially ordered data along the wellbore depth, the depth axis was treated analogously to a time series. The wavelet transform (WT) method was selected owing to its excellent time-frequency localization and multi-resolution analysis capabilities, which improve data precision while preserving the underlying trend. This method has been widely used in geophysical signal processing, including well logging data denoising (Li et al., 2024).

The smoothing process retains the overall trends of all parameters and the data distribution before and after smoothing are consistent with each other (Fig. 2), Moreover, the number of isolated points is significantly reduced (Fig. 3). Such data processing provides a desirable dataset and serves as a robust foundation for building clustering model development.

2.2. Model and algorithm

2.2.1. Geological sweet spot evaluation: A mechanism-guided weighted clustering model

The primary objective of geological sweet spot identification is to delineate reservoir intervals that simultaneously exhibit excellent reservoir flow capacity and storage performance, such as high porosity, high oil saturation, high permeability, and low clay content. It is necessary to classify well intervals with similar reservoir properties into the same category. Commonly, clustering algorithms are used. However, unsupervised clustering techniques based on Euclidean distance typically assign equal weights to reservoir petrophysical parameters, disregarding their differential contributions to production performance. To address this limitation, we propose a mechanism-guided weighted geological sweet spot clustering model.

Weighted *K*-medoids clustering algorithm. The objective of clustering in this study is to classify geological sweet spots along horizontal interval using log-derived petrophysical parameters under mechanism-guided principles. Clustering algorithms are commonly grouped into partition-based (e.g., *K*-means, *K*-medoids), density-based (e.g., DBSCAN, HDBSCAN), and model-based methods (e.g., Gaussian Mixture Models, GMM). The input dataset in this context consists of depth-ordered, continuous petrophysical curves with relatively uniform data density and no distinct spatial gaps. Moreover, mechanism-guided feature weighting is required to prioritize parameters with stronger correlations to production performance (e.g., porosity or saturation). These considerations make distance-based methods more appropriate than density- or model-based alternatives, which are designed for irregularly distributed data and offer limited flexibility in incorporating custom distance metrics. Among partition-based approaches, *K*-medoids was selected over *K*-means due to its greater robustness to outliers. While *K*-means uses the mean of all

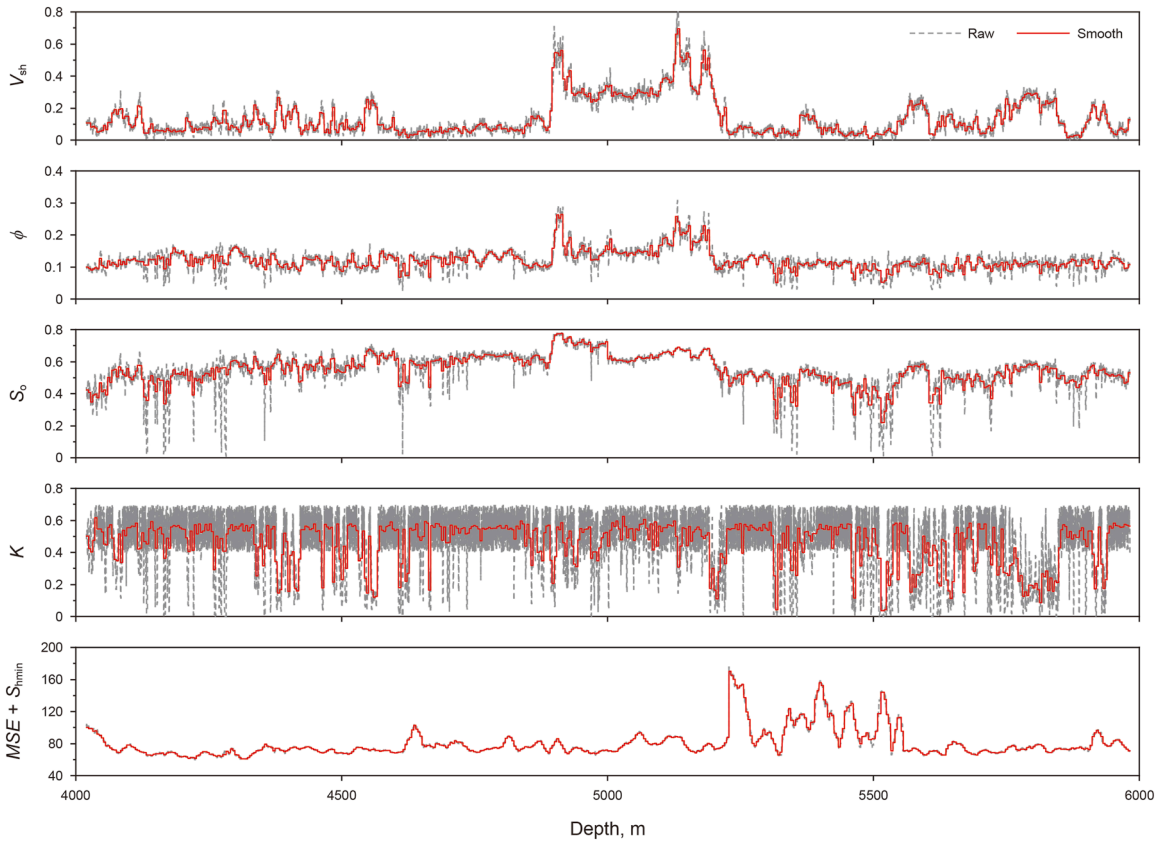

Fig. 2. Wavelet transform to smooth reservoir property curves.

points as the cluster center—making it sensitive to extreme values—*K*-medoids selects actual data points (medoids) as cluster centers, enhancing interpretability and stability (Awad and Hamad, 2022). We performed correlation analysis based on the production fluid profile to construct a weighted distance function and criterion function that emphasize key reservoir petrophysical parameters. This dual focus on both the cluster centers and the weighting function enhances robustness against outliers and better captures the underlying geological mechanisms. The weighted *K*-medoids clustering algorithm divides the *n*-sample dataset $\boldsymbol{X}$, which comprises four-dimensional reservoir petrophysical parameters along the horizontal well depth, into *K* clusters, $\{ck\}_{k=1}^{n}$. The weighted euclidean distance function is used to evaluate the distance between each sample point and its corresponding cluster center within each cluster. The objective of clustering is to find the optimal clustering solution $C = \left\{c_j\right\}_{j=1}^{n}$ by minimizing the criterion function. The feature matrix $\boldsymbol{X}$ is defined as follows:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & x_{n3} & x_{n4} \end{bmatrix} = \begin{bmatrix} \phi_1 & K_1 & S_{o1} & V_{sh1} \\ \phi_2 & K_2 & S_{o2} & V_{sh2} \\ \vdots & \vdots & \vdots & \vdots \\ \phi_n & K_n & S_{on} & V_{shn} \end{bmatrix} \tag{1}$$

where each row vector x_i corresponds to petrophysical parameters (ϕ, S_o, K, V_{sh}) of the *i*-th sample.

Mechanism-guided weight allocation. Establishing a reasonable distance function is key to guiding the weighted *K*-medoids clustering model. Correlation analysis with oil production profiles is conducted to determine the weights of the reservoir petrophysical parameters. Since the relationship between reservoir petrophysical properties and production profile data is often nonlinear, the Spearman method was chosen for the correlation analysis (Jiang et al., 2024). Unlike the Pearson correlation coefficient, which assumes linear relationships and normality of data distribution and is sensitive to outliers, Spearman's method is a non-parametric statistical approach that evaluates monotonic relationships (linear or nonlinear) without requiring normality assumptions and is robust to outliers. Its calculation formula is given as follows:

$$\rho = 1 - \frac{6\sum d_i^2}{n\left(n^2 - 1\right)} \tag{2}$$

where d_i is the rank difference between the *i*-th paired observations of reservoir properties and production data, and *n* is the sample size. Since production performance is influenced by a wide range of geological and engineering factors, some reservoir parameters may inherently show only limited correlation with production. To filter out parameters with negligible or ambiguous relationships, a threshold of $|\rho| \geq 0.2$ was adopted. This corresponds to the commonly accepted threshold for weak correlations in Spearman coefficient interpretation, where 0.8–1.0 indicates very strong correlation, 0.6–0.8 strong, 0.4–0.6 moderate, 0.2–0.4 weak, and 0.0–0.2 very weak or no correlation (Hu et al., 2023a, b). Therefore, only those parameters exhibiting at least a weak monotonic relationship ($|\rho| \geq 0.2$) were retained. This selection strategy ensures the physical relevance and statistical robustness of the weighted clustering model, while minimizing the influence of noise and overfitting risk.

Since the derived weights are assumed to be locally applicable within the same lithological unit of the evaluation block, production profile data from a single representative well is sufficient to guide the weighting process. The final weights are assigned based on the absolute values of the Spearman correlation coefficients and normalized to ensure that their sum equals 1:

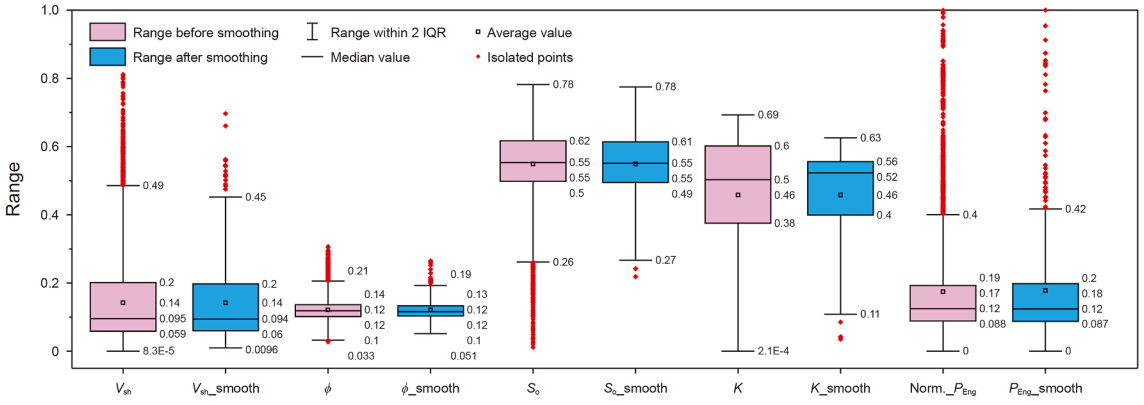

Fig. 3. Box plot of parameter distribution before and after smoothing.

$$w_j = \frac{\rho_j}{\sum_{j=1}^{p} |\rho_j|} \tag{3}$$

where $|\rho_j|$ is the absolute correlation coefficient of the j-th parameter and p is the total number of parameters. It is important to note that the derived weights are assumed to be locally applicable within the same lithological unit of the evaluation block. Therefore, production profile data from a single representative well is sufficient to guide the weighting process for clustering across the block

Weighted euclidean distance function. To enhance the influence of key petrophysical parameters in the clustering process, the weighted euclidean distance is adopted as the similarity measure among all samples (Fig. 4). The weighted Euclidean distance function is given by Eq. (4).

$$D\left(x_i, c_j\right) = \sqrt{\sum_{k=1}^{p} |w_k| \left(x_{ik} - c_{jk}\right)^2} \tag{4}$$

where x_i is the parameter vector of the i-th sample, c_j is the centroid of the j-th cluster, and w_k is the normalized weight of the k-th parameter.

Criteria function. The K-medoids model uses the criterion function to evaluate the clustering effect. The weighted sum of squared errors (WSSE) of all K clusters is defined by Eq. (5).

$$\mathrm{WSSE} = \sum_{j=1}^{K} \sum_{x_i \in C_j} \sum_{k=1}^{p} |w_k| (x_{ik} - c_{ik})^2 \tag{5}$$

where K is the predetermined number of clusters, and C_j represents all samples belonging to the j-th cluster. The optimal number of clusters K is determined using the elbow method. When K is below an appropriate value, increasing K results in a significant reduction in WSSE. However, when K exceeds that value, the decrease in WSSE becomes negligible, and this value is regarded as the optimal K. Fig. 5 presents an illustrative example of the WSSE curve, where a clear elbow is observed at $K = 3$.

Rating function. The horizontal well interval is classified into K different clusters, but their relative quality cannot be directly assessed. To quantitatively assess reservoir quality and link clustering results to geological mechanisms, a weighted geological sweet spot index (WGSSI) is proposed. This index integrates weights derived from production data correlations to reflect the relative contributions of petrophysical parameters to reservoir

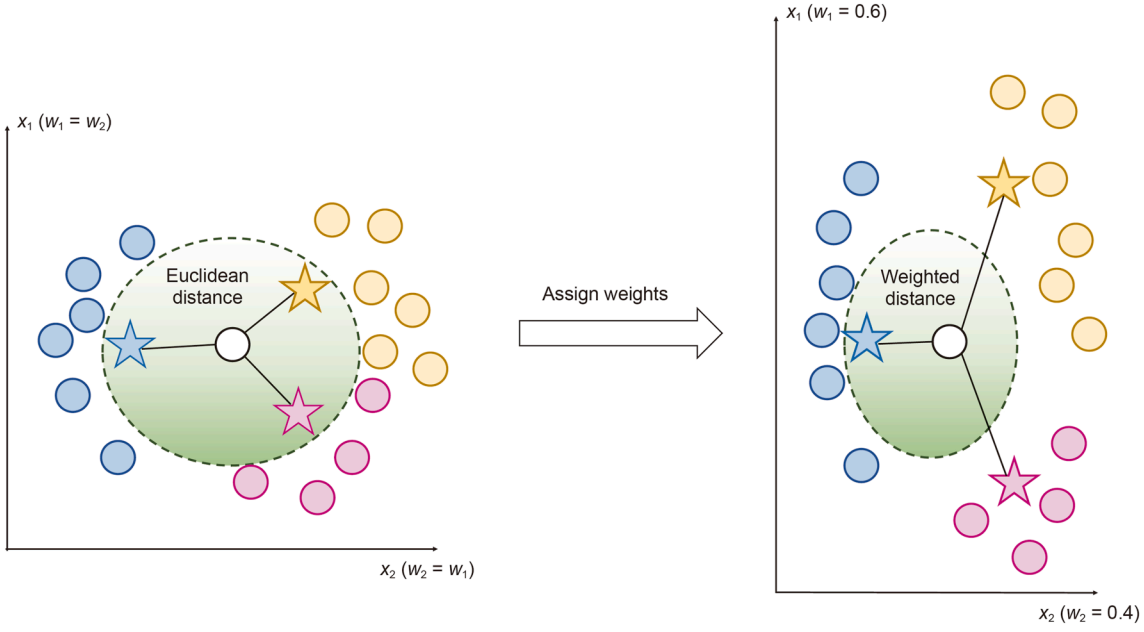

Fig. 4. Euclidean distance vs weighted Euclidean distance.

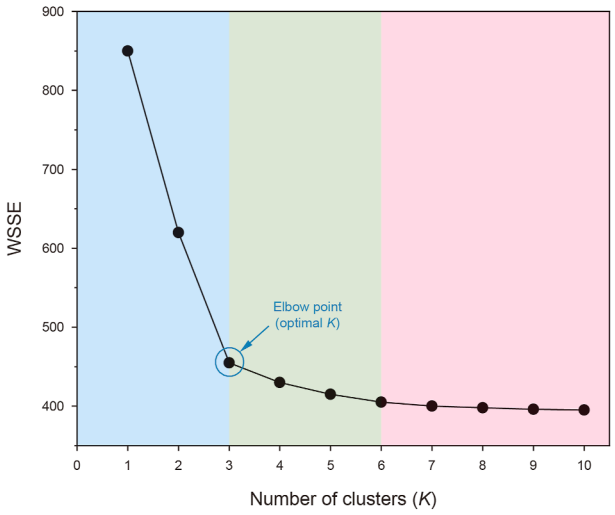

Fig. 5. WSSE curve for candidate K values. The elbow point suggests $K = 3$ as the optimal cluster number.

performance. To minimize the influence of outliers, the analysis focuses on the interquartile range (25th to 75th percentiles) of the normalized values for each parameter within each cluster. The WGSSI is formulated as:

$$\mathrm{WGSSI}_j = \sum_{k\in\{\phi,S_o,K,V_{sh}\}} w_k \cdot \left(\frac{p_{jk}^{25\%} + p_{jk}^{75\%}}{2}\right) \tag{6}$$

where WGSSI_j represents WGSSI for cluster j, $p_{jk}^{25\%}$ and $p_{jk}^{75\%}$ represents 25th and 75th percentiles of parameter k in cluster j. A larger WGSSI value indicates a better geological sweet spot, suggesting a more favorable reservoir quality for production.

The weighted K-medoids clustering algorithm steps are as follows:

1) **Initialization:** K cluster centers were randomly selected from the dataset.
2) **Weighted Euclidean distance assignment:** Calculate the weighted Euclidean distance from each sample point to the K cluster centers and divide it into the cluster with the smallest weighted Euclidean.
3) **Cluster center update:** For each cluster, a new center was selected by minimizing the criteria function WSSE defined as the total weighted within-cluster variance.
4) **Iteration:** Steps 2–3 were repeated until convergence (no change in cluster centers) or a maximum iteration threshold was reached.
5) **Result:** Final clustering results were labeled using a rating function based on WGSSI, categorizing geological sweet spots into K classes ranging from Class I (excellent) to Class K (poor).

To provide an intuitive summary of the proposed method, Fig. 6 shows the overall workflow of the mechanism-guided weighted K-medoids clustering model. The upper part highlights the modular steps from input data preparation to clustering and rating, while the lower panel shows an example of clustering results along the horizontal well interval.

2.2.2. Geological sweet spot evaluation: A mechanism-guided weighted clustering model

Hydraulic fracture initiation requires overcoming the strength of the rock and in-situ stress barriers. Each perforation cluster within a fracturing stage along the horizontal well has a corresponding fracture initiation pressure, and when the wellbore pressure exceeds this threshold, the corresponding fracture will be initiated. The fracture initiation condition can be expressed as:

$$P_w > T_{str,eff,c} + \sigma_{init,c} + \Delta\sigma_{e,c} + \Delta\sigma_{i,c} \tag{7}$$

where P_w represents the pressure within the wellbore (MPa), $T_{str,eff,c}$ represents the effective tensile strength (MPa), $\Delta\sigma_{e,c}$ the stress shadow from the prior stage (MPa), $\Delta\sigma_{i,c}$ is the stress shadow from other fractures within the same stage (MPa) and $\Delta\sigma_{init,c}$ is the initial magnitude of the minimum principal stress at the cluster (MPa).

Engineering sweet spot evaluation parameter. To quantitatively assess geomechanical strength, we introduce a composite parameter P_{Eng}, which integrates bottomhole mechanical specific energy (MSE_b) to characterize reservoir rock strength while accounting for the influence of in situ-stress on fracture initiation and propagation. MSE_b serves as a crucial indicator of rock mechanical strength under in-situ stress, natural fractures, and pore pressure conditions. As demonstrated by Hu et al. (2023a, b), MSE_b exhibits a strong correlation with both fracture initiation and propagation. The P_{Eng} parameter is formulated as:

$$P_{Eng} = MSE_b + S_{hmin} \tag{8}$$

where P_{Eng} represents the geomechanical strength (MPa), MSE_b represents downhole mechanical specific energy (MPa), S_{hmin} represents the horizontal minimum in-situ stress (MPa).

During the fracturing initiation stage, when fractures have not yet extensively propagated and interference between fractures can be neglected, the fracturing initiation pressure condition can be further simplified as:

$$P'_W \geq P_{Eng} = MSE_b + S_{hmin} \tag{9}$$

where P'_W represents the initiation pressure (MPa).

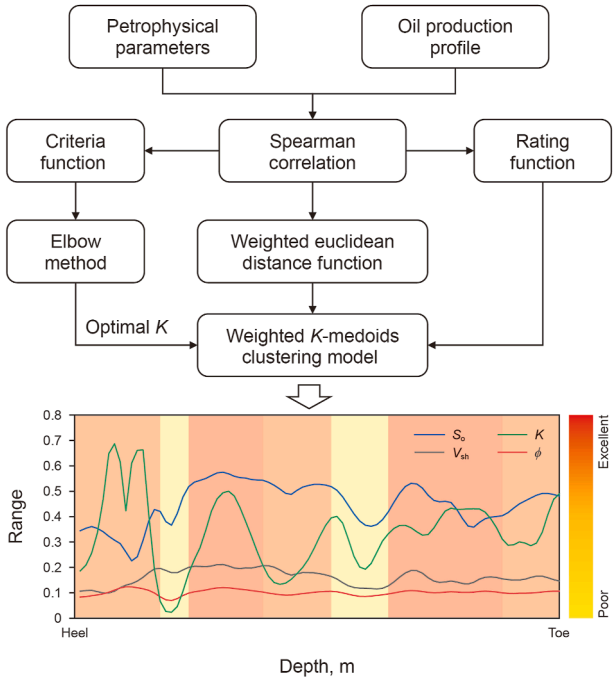

Fig. 6. Workflow of the mechanism-guided clustering method for geological sweet spot evaluation.

The geomechanical strength parameter P_{Eng}, by incorporating both the rock mechanical strength (represented by MSE_{b}) and the horizontal minimum in-situ stress (S_{hmin}), serves as a critical control on fracture initiation. A higher P_{Eng} indicates greater resistance to fracture activation, requiring higher wellbore pressure to overcome both rock and stress barriers. As illustrated in Fig. 7, despite uniform perforation spacing in geometric design, the actual initiation behavior is strongly modulated by the heterogeneity of geomechanical strength along the horizontal well. Specifically, clusters 2 and 3 are located in zones with elevated P_{Eng}, resulting in significantly higher fracture initiation thresholds. In contrast, clusters 1 and 4, positioned in mechanically weaker regions, exhibit earlier and more favorable fracture initiation. This imbalance highlights the inadequacy of purely geometric cluster placement strategies in heterogeneous reservoirs, and underscores the need for geomechanical-guided stage and cluster optimization to ensure simultaneous initiation and effective stimulation.

To achieve balanced fracture initiation within each fracturing stage, this study proposes a dynamic programming (DP)-optimized fracturing stage design method that clusters reservoir intervals with similar geomechanical strength into identical fracturing stages. The methodology integrates well depth data and geomechanical strength indices while maintaining spatial continuity of reservoir intervals. Stage length constraints are systematically imposed in accordance with engineering requirements. This approach establishes both theoretical foundations and technical support for geological and engineering dual sweet spot identification and optimized HF cluster design.

DP algorithm. The DP algorithm is an effective approach for solving sequential optimization problems, particularly suited for seeking a globally optimal segmentation in a nonconvex solution space. Given a dataset $\boldsymbol{X} = (x_1, x_2, \ldots, x_n)$, where each sample x irepresents the geomechanical strength at depth i, the goal is to divide the wellbore into a set of fracturing stages $\boldsymbol{S} = (S_1, S_2, \ldots, S_m)$ such that the overall objective function is minimized, ensuring that each stage consists of points with similar geomechanical properties and satisfies predefined stage length constraints. The objective function $J(j,i)$ for a candidate stage from index j to i is composed of two terms:

(1) Geomechanical similarity term: This term measures the variability of geomechanical strength among sample points within the same stage. It is quantified using the sum of squared differences:

$$J_{\mathrm{geo}}^{(j,i)} = \sum_{t=j}^{i} \left(x_t - \bar{x}_{j,i}\right)^2 \tag{10}$$

where $\bar{x}_{j,i}$ is the mean geomechanical strength over the interval $[j, i]$.

(2) Stage length penalty term: To ensure that the depth range $L_{j,i}$ of each stage falls within a target interval $[a,b]$, a penalty is imposed when $L_{j,i}$ deviates from the target stage length L_{target} (e.g., the midpoint of $[a,b]$):

$$J_{\mathrm{len}}^{(j,i)} = \left(L_{j,i} - L_{\mathrm{target}}\right)^2 \tag{11}$$

where $L_{j,i} = \mathrm{Depth}(i) - \mathrm{Depth}(j)$.

By integrating these two terms, the total objective function J is formulated as:

$$J(j,i) = J_{\mathrm{geo}}^{(j,i)} + \lambda J_{\mathrm{len}}^{(j,i)} \tag{12}$$

where λ is a weighting factor that balances the two terms. In this study, λ is set to 1, considering that geomechanical constraints and engineering constraints are equally important.

The DP algorithm steps are as follows:

1) State definition: Define $D(i)$ as the minimum cumulative objective function value required to optimally segment the first i samples $(x_1, x_2, \ldots, x_i)$. To enable backtracking, an auxiliary array seg[i] is maintained, where seg[i] records the optimal segmentation point (i.e., the index of the previous stage boundary) that leads to the minimum cost at position i. The initial state is defined as $(0) = 0$, and all other $D(i)$ values are initialized as infinity.
2) State transition: For each sample index $i \in [1, n]$, all valid starting indices $K(i) = \left\{j | L_{\min} \le L_{j+1,i} \le L_{\max}\right\}$ denotes the set of feasible previous stage endpoints satisfying the stage length constraint. The cost of a candidate segment from $j+1$ to i is given by the objective function $J(j+1, i)$, and the transition relation is:

$$D(i) = \min_{j \in K(i)} \{D(j) + J(j+1, i)\} \tag{13}$$

The optimal segmentation point is simultaneously recorded as:

$$seg[i] = \arg\min_{j \in K(i)} [D(j) + J(j+1, i)] \tag{14}$$

where the objective function $J(j+1,i)$ is defined in Eq. (12), combining the geomechanical similarity term and the stage length penalty term.

3) Global optimal solution and backtracking. Once all D(i) values are computed, the optimal segmentation path is extracted via

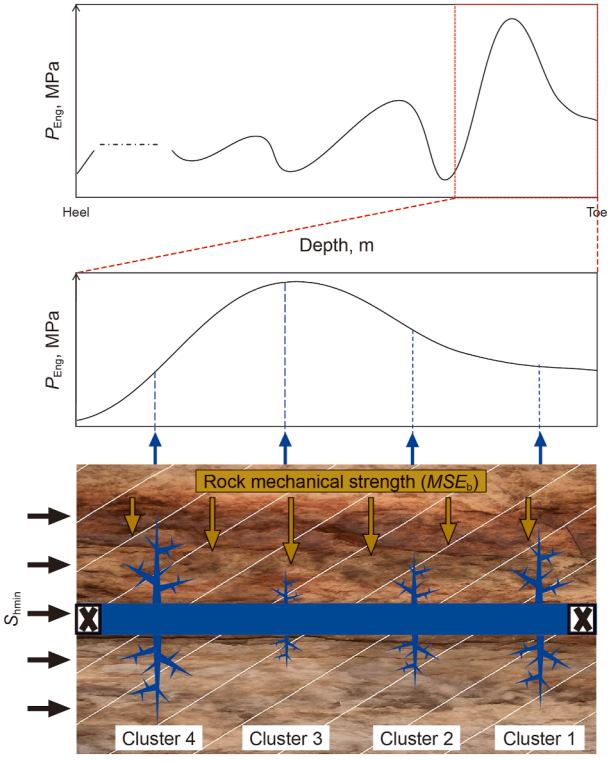

Fig. 7. Schematic illustration of engineering sweet spot distribution and fracture initiation behavior along a horizontal well.

backtracking from $i = n$ using the stored indices in $seg[i]$. This recursive trace yields the set of optimal stage boundaries that minimize the total cost while enforcing geomechanical consistency and engineering length constraints.

Fig. 8 shows the overall workflow for determining the optimal fracturing stage division. The DP algorithm efficiently explores all possible stage divisions while avoiding redundant computations through recursive updates. Given n total samples and K stages, the computational complexity is approximately $O(n^2)$, making it feasible for practical fracturing stage design applications.

2.2.3. Perforation cluster design: A hybrid multi-objective optimization method

Simultaneous development of high-quality reservoirs and balanced fracture initiation and propagation is a prerequisite for designing perforation clusters. The design of cluster locations requires the consideration of geological sweet spots, engineering sweet spots, and fracturing construction requirements. Consequently, this is not a simple linear problem but a typical nonlinear optimization challenge, where the key lies in constructing appropriate objective functions and constraints and employing optimization algorithms to obtain a set of optimal solutions. Thus, we propose a hybrid multi-objective optimal method based on the dual geological–engineering sweet spot for the intelligent optimization design of perforation clusters. The method comprises two main steps:

(1) **Candidate cluster initialization:** Candidate cluster centers are automatically identified and generated within the candidate region using the DBSCAN clustering algorithm.
(2) **Multi-objective optimization screening:** The NSGA-II multi-objective optimization algorithm is used to further select the optimal clusters from the initial candidate clusters under the premise of satisfying physical constraints.

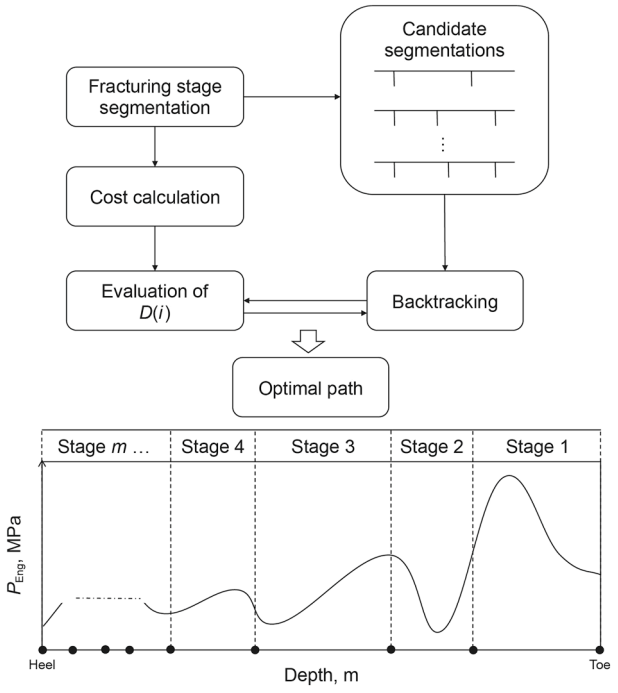

Fig. 8. Workflow of the dynamic programming algorithm for fracturing stage segmentation.

This method integrates mechanism-guided clustering techniques with multi-objective optimization, ensuring that the optimal clusters exhibit similar geomechanical strength while striving to maximize the quality of geological sweet spots, thereby enhancing the stimulation effectiveness of the fracturing stages.

Data processing and candidate regions establishment. The input dataset comprises four key parameters: well depth, geological sweet spot labels (classes), engineering sweet spot labels (geomechanical strength), and stage labels. The process includes: first, constructing geological sweet spot weights based on the reverse order of the geological sweet spot category labels (Eq. (15)), which will be used in the subsequent formulation of the geological sweet spot scoring objective function.

$$w_k = K - \text{class}_k + 1 \tag{15}$$

where K is the total number of geological sweet spot classes, and class_k is the integer label assigned to *class k* (e.g., 1 for Class I, 2 for Class II, etc.). This weighting ensures that Class I receives the maximum weight, while lower-quality classes receive proportionally smaller weights.

Then, stage labels are used to identify the fracturing stages and to impose constraints on the clusters within each stage, such that all clusters within a stage must reside within the predefined stage depth range:

$$c_i^{\text{depth}} \in \left[Stage_{n,\text{start}}, Stage_{n,\text{end}}\right] \tag{16}$$

where c_i^{depth} denotes the depth coordinate of cluster i within the n-th stage.

Finally, candidate regions are generated using the engineering sweet spot labels (Eq. (17)). Unlike traditional fixed-threshold methods, the candidate region is dynamically defined: samples with engineering sweet spot values within $\mu \pm 10\%$ of the mean μ for the n-th stage are identified as high-quality engineering sweet spot zones and are the only samples used for subsequent clustering (Fig. 9).

$$P_{\text{Eng}} \in [0.95\mu_n, 1.05\mu_n] \tag{17}$$

Candidate cluster initialization based on the DBSCAN algorithm. To optimize the selection of initial cluster centers within candidate regions, the DBSCAN clustering algorithm is employed. As a density-based clustering method, DBSCAN does not require predefining the number of clusters and autonomously identifies high-density regions, thereby preventing candidate clusters from spanning well intervals with significant variations in geomechanical strength. In this approach, two-dimensional features—well depth and engineering sweet spot labels (geomechanical strength) —are standardized to ensure that the clustering results are not affected by differences in scale. The key hyperparameters of the DBSCAN algorithm include the neighborhood radius (eps) and the minimum number of samples (min_samples). The parameter eps defined the reachable density range for data points.

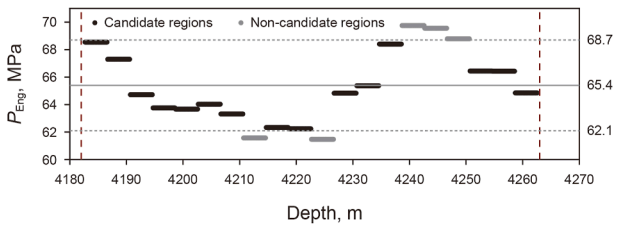

Fig. 9. Cluster candidate regions.

The optimal value of eps = 0.3 is determined through the *k*-distance graph method (Ester et al., 1996) and the elbow method. The parameter min_samples, defined as the minimum number of samples required to form a cluster to ensure that only sufficiently dense regions are recognized as valid clusters, is usually set to twice the dimensionality of the input dataset, and here it is set to 4. The candidate clusters identified by DBSCAN are subsequently used as the input for the NSGA-II multi-objective optimization process, thereby enhancing the convergence properties of the multi-objective optimization (Fig. 10).

Optimal cluster selection based on the NSGA-II multi-objective optimization algorithm. To further select the optimal cluster set from the initial candidate clusters, this study employed the elitist non-dominated sorting genetic algorithm II (NSGA-II) (Deb et al., 2002). Unlike traditional single-objective methods, NSGA-II simultaneously optimizes multiple competing objectives and utilizes Pareto front searching to generate a set of non-dominated solutions. Additionally, adaptive penalty functions are incorporated to enforce spatial constraints. The specific optimization objectives and constraints are as follows:

Objective 1: Minimization of geomechanical strength variance. This objective minimizes the variance in geomechanical strength among clusters to ensure that the clusters exhibit similar mechanical properties:

$$f_1 = \sum_{i \in S_j} \left(x_i - \mu_j \right)^2 \tag{18}$$

where S_j is the set of samples in cluster j, x_i represents the geomechanical strength of sample i, and μ_j is the mean geomechanical strength value of cluster j.

Objective 2: Maximization of geological sweet spot score. This objective maximizes the geological sweet spot scores to ensure that the selected cluster centers correspond to regions with favorable reservoir conditions. The geological sweet spot score is calculated as the sum of geological sweet spot weights of the samples within each cluster as follows:

$$f_2 = \sum_{i \in S_j} w_i \tag{19}$$

where w_i represents the geological sweet spot weight of sample i.

Constraint condition: Spatial separation of cluster. To prevent the selected clusters from being too close or too far apart, adjacent clusters must maintain a spacing of 10–30 m to balance fracture interference and stimulation efficiency. A fixed penalty is applied when the depth difference is less than 10 m or greater than 30 m, guiding the optimization process toward solutions that satisfy physical constraints:

$$10\ \text{m} \le |c_i - c_{i+1}| \le 30\ \text{m} \tag{20}$$

where c_i represents the depth coordinate of cluster i.

The NSGA-II algorithm was implemented under the DEAP model, following these steps:

1) Population encoding and initialization: Candidate solutions were encoded as binary vectors, with each gene corresponding to a candidate cluster (1 for selection, 0 for exclusion). The initial population size was set to 100, initialized with a 50% biased random distribution.
2) Fitness evaluation: For each individual, the geological sweet spot scores and geomechanical strength variance were calculated by assigning data to clusters based on the nearest depth. Spatial constraints were integrated into the objective function values.
3) Non-dominated sorting and crowding distance: All individuals were sorted using non-dominated sorting to generate Pareto fronts. Within each Pareto front, crowding distance was calculated to maintain population diversity.
4) Selection, crossover, and mutation: Binary tournament selection with crowding comparison was used. Simulated binary crossover (SBX) and random index mutation were applied to generate offspring. The crossover probability was set to 0.7, and the mutation probability was 0.3.
5) Population update and iteration: Parent and offspring populations were combined, and the next generation was selected using non-dominated sorting and crowding distance. This process was iterated for 100 generations or until convergence.
6) Extraction and post-processing of optimal solutions: The Pareto optimal set was extracted from the final population, from which one optimal individual was selected. Minor random perturbations were applied to the selected clusters to achieve the final optimized cluster set (Fig. 11).

3. Results

3.1. Field background and data

To verify the effectiveness and applicability of the proposed model in sweet spot evaluation and completion design, a representative horizontal shale oil well—Well X1 from the Chang 7 Member of the Yanchang Formation in the Ordos Basin—was selected as the study subject. The total vertical depth of the well is 2278 m, and the horizontal interval extends from 4010 to 5982 m. This well is equipped with tracer-based production profile monitoring and distributed acoustic sensing (DAS), enabling quantitative validation of fracturing effectiveness. In addition, field application was conducted in the Jimsar Sag of the Junggar Basin, Xinjiang, where four horizontal wells were designed using the proposed optimization framework. Comparative analysis with offset wells designed using traditional geometric methods was performed to evaluate production performance. This dual validation strategy ensures both process-level accuracy and field-scale applicability of the proposed intelligent fracturing design.

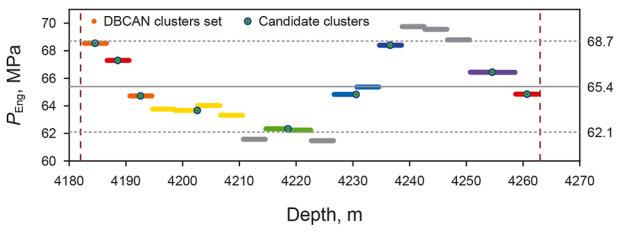

Fig. 10. Candidate clusters based on DBSCAN algorithm.

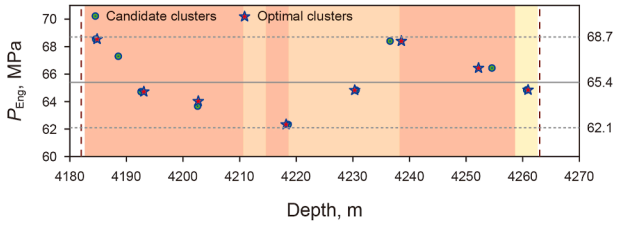

Fig. 11. Optimal clusters based on the NSGA-II multi-objective optimization algorithm.

3.2. Model and algorithm effect verification

3.2.1. Performance analysis of the clustering model

By comparing the geologic sweet spot clustering results of the conventional *K*-medoids model with those of the proposed mechanism-guided weighted *K*-medoids model, we verified the advantages of the latter in evaluating geologic sweet spots The conventional model adopts the Euclidean distance as a metric, assigning equal importance to each parameter, whereas the weighted model employs correlation analysis to quantify the relationships between reservoir petrophysical properties and production profile data. Based on these quantified relationships, a mechanism-guided weighted Euclidean distance is constructed, thereby yielding sweet spot classifications more consistent with geological principles. Fig. 12 presents the SHAP correlation coefficients of reservoir petrophysical parameters along with their corresponding weight allocations in the weighted *K*-medoids clustering. For Well X1, the optimal *K* value obtained by the elbow method is 4 (Fig. 13).

Fig. 14 shows the geological sweet spot profiles of Well X1 derived from two clustering algorithms, and divides the geological sweet spots into four classes (I, II, III, and IV) from "excellent" to "poor". Fig. 14(a) shows the result using the conventional *K*-medoids clustering algorithm, which treats all petrophysical parameters equally when calculating similarity. In contrast, Fig. 14(b) presents the result obtained from the mechanism-guided weighted *K*-medoids algorithm, which incorporates parameter weights derived from production correlation analysis. The two methods yield substantially different classification results, particularly regarding the distribution and continuity of the intervals for each class. To further investigate the underlying causes of these differences, we analyzed the distribution of petrophysical parameters across different sweet spot classes derived from each clustering method.

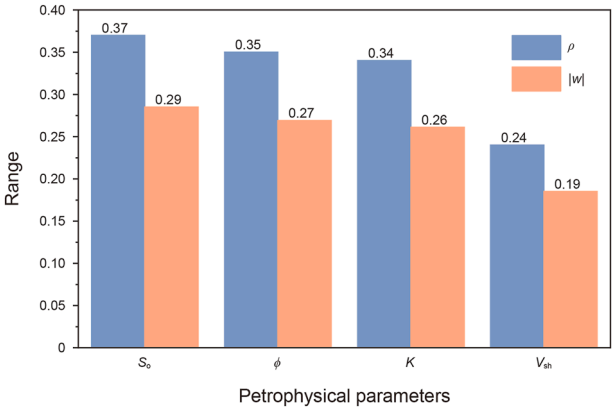

Fig. 12. SHAP correlation coefficients and weight of reservoir petrophysical parameters of Well X1.

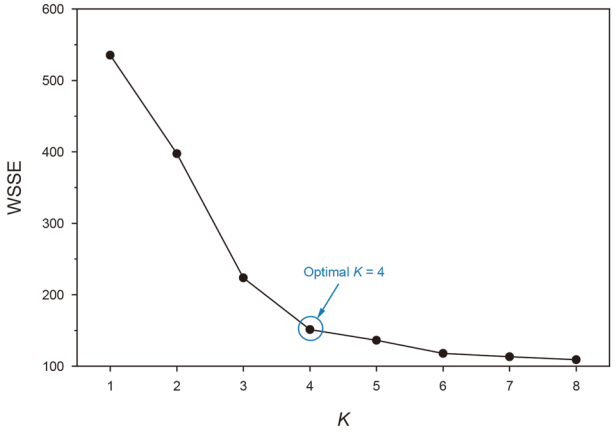

Fig. 13. Elbow method to optimize *K* of Well X1.

Fig. 15(a) shows the geologic sweet spots classes from the conventional *K*-medoids model, where the delineation of Class IV appears inconsistent. Although Class IV exhibits higher V_{sh} values—consistent with the characteristics of low-quality geologic sweet spots—the distributions of ϕ, S_o, and *K* exceeding those of Classes I–III, which contradicts the established geological mechanisms. This discrepancy primarily arises because the conventional clustering model relies solely on the Euclidean distance to measure sample similarity, treating all petrophysical parameters with equal weight. Consequently, parameters with a broader range of numerical values can dominate the clustering results, thereby masking the distributional features of other key parameters. In contrast, Fig. 15(b) shows geologic sweet spots classes from the weighted *K*-medoids model. By enhancing the influence of critical parameters during the clustering process, the weighted model distinctly delineates the petrophysical characteristics of geologic sweet spots, yielding a more rational classification across Classes I–IV. Moreover, the clustering results from the weighted model are highly consistent with the outcomes of the correlation analysis and align well with the underlying geological mechanisms, thereby providing a more robust basis for the classification and evaluation of geologic sweet spots.

3.2.2. Performance analysis of the DP algorithm

To validate the applicability of the proposed DP algorithm in fracturing stage design optimization, we compared it against a geomechanics-based unsupervised clustering approach (Hu et al., 2023a, b). Both methods aim to group well intervals with similar geomechanical strength into the same fracturing stage, but they differ in methodology: the geomechanics-based unsupervised clustering model primarily relies on sample similarity measures to cluster well intervals exhibiting comparable geomechanical strength, whereas the DP optimization algorithm, in formulating its objective function, not only takes into account geomechanical strength but also integrates constraints on stage length and spatial continuity. This enables the DP method to automatically produce a rational fracturing stage design without manual intervention.

Fig. 16(a) shows the fracturing stage division derived from the unsupervised clustering method. Although this approach effectively groups well intervals with similar geomechanical strength, it still requires manual adjustments to accommodate stage length and spatial continuity, thus failing to generate fully automated fracturing stages design that meet operational requirements. By contrast, Fig. 16(b) shows the fracturing stage division obtained via the DP optimization algorithm. Different colors denote distinct fracturing stages composed of well intervals that not only exhibit comparable geomechanical strength but also satisfy stage length constraints while maintaining spatial continuity. Thus, the DP algorithm ensures geomechanical consistency while automatically accounting for stage length and segment continuity, thereby demonstrating its feasibility and superiority in practical fracturing stage design.

3.2.3. Performance analysis of the hybrid multi-objective optimization algorithm

To verify the applicability of the proposed multi-objective optimization algorithm in hydraulic HF cluster design, we compared it with the threshold-based HF cluster design method

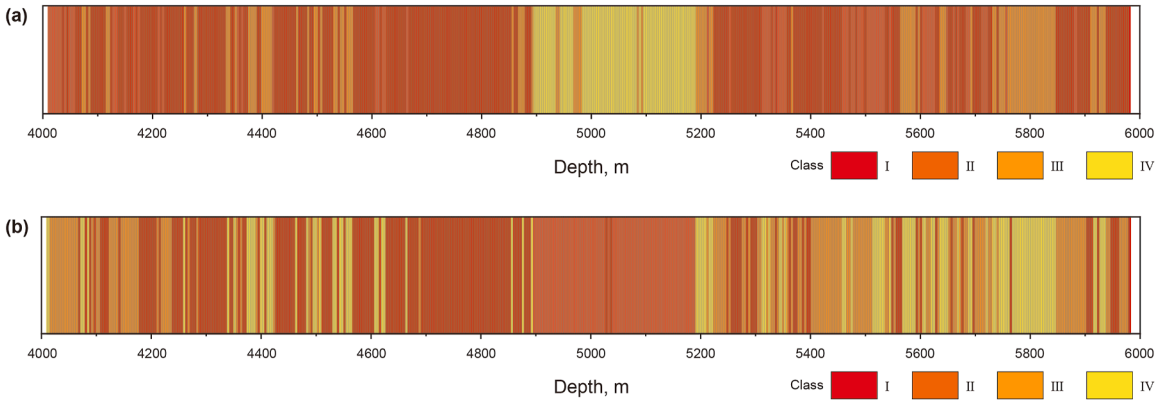

Fig. 14. Geological sweet spot profile of Well X1. (**a**) *K*-medoids clustering model vs. (**b**) weighted *K*-medoids clustering model.

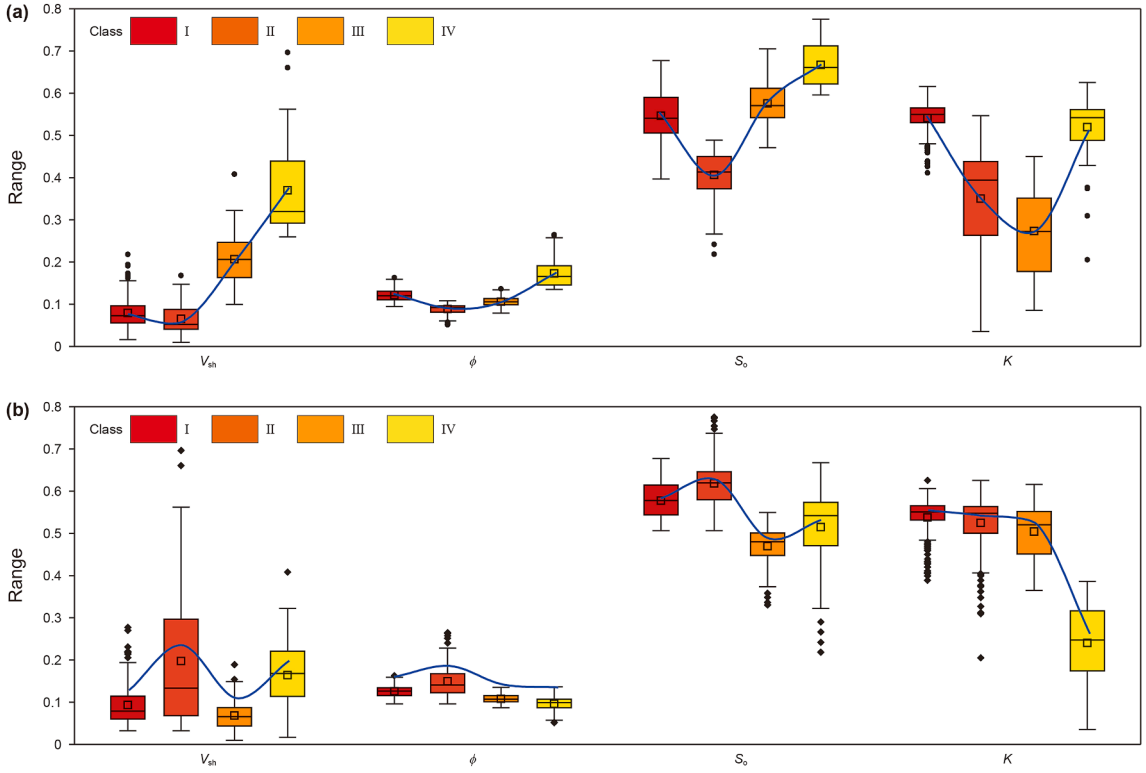

Fig. 15. Box plot of parameter distribution of the geologic sweet spots classes from the (**a**) *K*-medoids clustering model vs. (**b**) weighted *K*-medoids clustering model.

(Hu et al., 2023a, b). Although both approaches aim to arrange perforation clusters in intervals exhibiting similar geomechanical strength, their specific implementations differ. In the threshold method, feasible zones are defined by setting an upper and lower bound of ±5% around the mean geomechanical strength for each fracturing stage, and perforation clusters are then designed with consideration for cluster spacing. By contrast, the multi-objective optimization algorithm builds upon the threshold approach by incorporating geologic sweet spots, engineering sweet spots, and HF operational requirements into its objective function and penalty terms. This ensures not only that candidate clusters are arranged under similar geomechanical conditions but also that they are preferentially located within high-quality geologic sweet spots, thereby enhancing the effectiveness of HF.

Fig. 17(a) shows the HF cluster selection results derived from the threshold method. Although this method can group well intervals with similar geomechanical strength into the same cluster set, it fails to adequately account for geologic sweet spots and still requires manual arrangement of perforation clusters according to the cluster spacing range, making it impossible to automatically generate perforation clusters that satisfy all constraints. In contrast, Fig. 17(b) presents the HF cluster arrangement obtained using the multi-objective optimization algorithm. The selected cluster positions not only fulfill the requirement for similar geomechanical strength but also prioritize placement in high-quality geologic sweet spot intervals (such as clusters 4 and 6), while automatically accommodating cluster spacing and spatial continuity. Consequently, the multi-objective optimization algorithm can simultaneously address geomechanical similarity, geologic sweet spot quality, and cluster spacing constraints in practical HF cluster design, demonstrating superior feasibility and effectiveness.

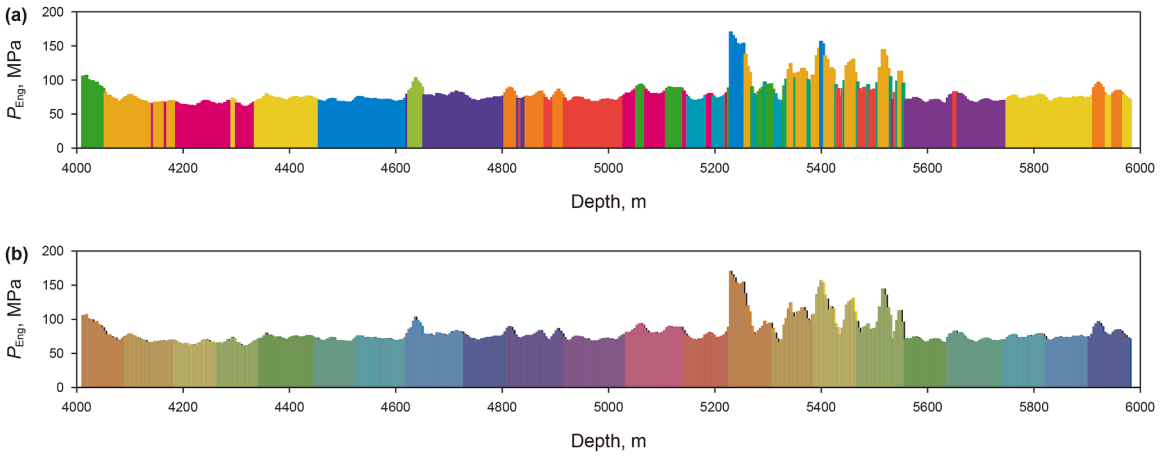

Fig. 16. Fracturing stage design of Well X1 based on (**a**) clustering algorithm vs. (**b**) DP algorithm.

3.3. Verification based on monitoring data

This study utilized tracer monitoring and fiber-optic monitoring data from Well X1 to obtain valid oil contribution rates for 16 stages and fluid allocation profiles for 14 stages. Based on these data, we analyzed the proportion of geological sweet spot classes within each fracturing stage and investigated the correlation between oil production contribution rates and geological sweet spot classes. Additionally, the relationship between engineering sweet spots (i.e., geomechanical strength) and cluster fluid allocation was explored using fiber-optic monitoring results.

3.3.1. Verification of geological sweet spot evaluation

The oil production contribution rates and the proportion of different geological sweet spot classes within each of the 16 stages in Well X1 were statistically analyzed. As shown in Fig. 18(a), it was observed that stages with higher oil production contribution rates exhibited a higher proportion of Class I and Class II high-quality sweet spots, while the proportion of Class III and Class IV sweet spots was lower. Conversely, stages with lower oil production contribution rates showed a higher proportion of Class III and Class IV sweet spots and a lower proportion of Class I and Class II sweet spots. To quantify the correlation between geological sweet spot classes and oil contribution, Pearson's correlation coefficient was employed. The Pearson correlation coefficient ranges from −1 to 1, where a stronger relationship between variables corresponds to a larger absolute value of the coefficient (closer to 1), and a weaker relationship corresponds to a value closer to 0. The sign of the coefficient indicates positive or negative correlations. The results revealed that oil contribution was positively correlated with Class I and Class II sweet spots, with correlation coefficients of 0.17 and 0.51 (average correlation coefficient of 0.34), respectively, while it was negatively correlated with Class III and Class IV sweet spots, with coefficients of −0.6 and −0.72 (average correlation coefficient of −0.655), respectively (Fig. 18(b)). These findings validate the effectiveness of the proposed mechanism-guided weighted geological sweet spot clustering model. In completion design, prioritizing stages with high-quality geological sweet spots (Class I and II) is beneficial for enhancing production in unconventional reservoirs.

3.3.2. Verification of engineering sweet spot evaluation

The variation in inter-cluster fluid allocation directly reflects fracture initiation and propagation. By analyzing the fluid allocation profiles and geomechanical strength (P_{Eng}) within each of the 14 stages in Well X1, both parameters were normalized. The results demonstrated a clear trend (Fig. 19): generally, clusters with higher P_{Eng} exhibited relatively lower fluid allocation, while those with lower strength showed higher fluid allocation. Pearson's correlation analysis was applied to assess the relationship between P_{Eng} and fluid allocation of each stage, revealing a moderate negative correlation in Well X1, with a mean correlation coefficient of −0.46. This result confirms the validity of using geomechanical strength as an engineering sweet spot indicator. In completion design, it is essential to select stages with similar geomechanical strength for perforation cluster placement to promote balanced fracture initiation and propagation.

3.4. Field application and production verification

The optimized completion design model presented in this study was applied to four shale oil horizontal wells in the Jimsar Sag, Xinjiang, with two geometric designed wells as a comparison. All wells are located in the Jimsar Sag, with depths ranging from

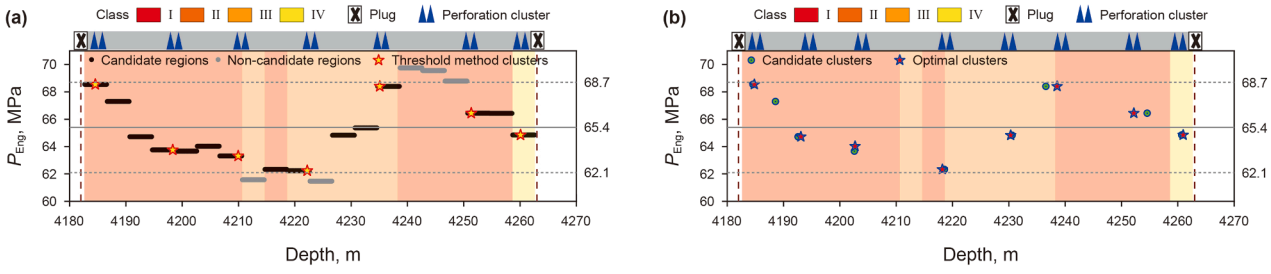

Fig. 17. Optimal clusters based on the (**a**) threshold-based method vs. (**b**) NSGA-II multi-objective optimization algorithm.

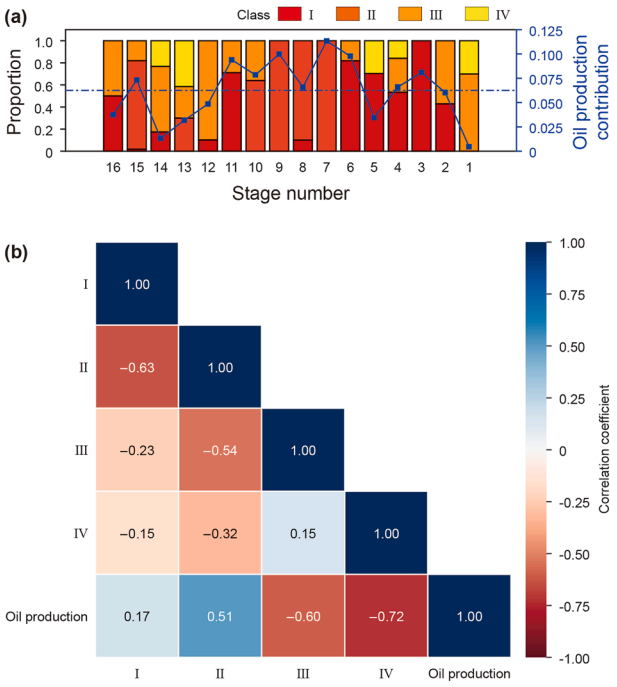

Fig. 18. (**a**) Proportion of geological sweet spot classes and oil production contribution in the fracturing stages. (**b**) Heat map of the correlation between oil production and geological sweet spot classes.

4150 to 4300 m and horizontal section lengths between 1584 and 1634 m. The two geometric designed wells (S1, S2) adopted a geometric/single completion design, while the four test wells (O1, O2, O3, O4) employed the optimized completion design. All wells were drilled in the same formation and utilized identical

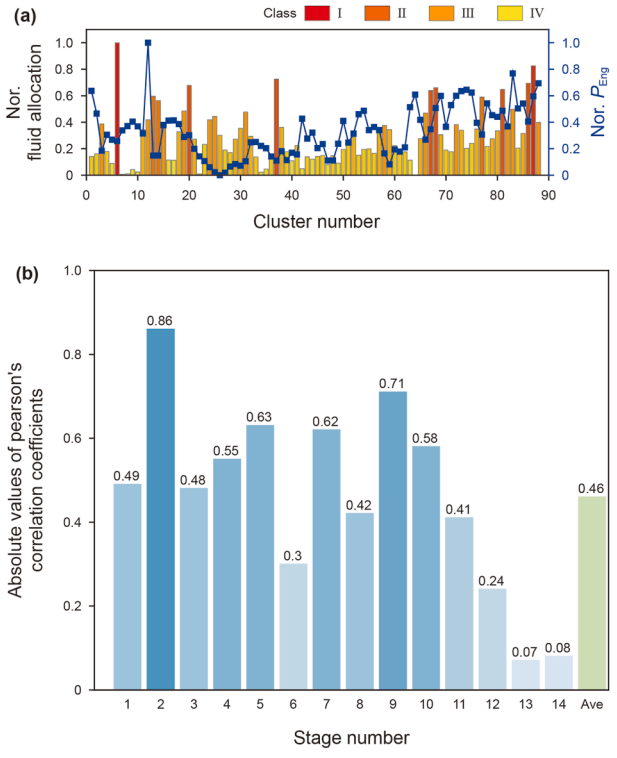

Fig. 19. (**a**) Fluid allocation profiles and P_{Eng} within perforation clusters. (**b**) Absolute correlation coefficient between fluid allocation and P_{Eng} in each fracturing stages.

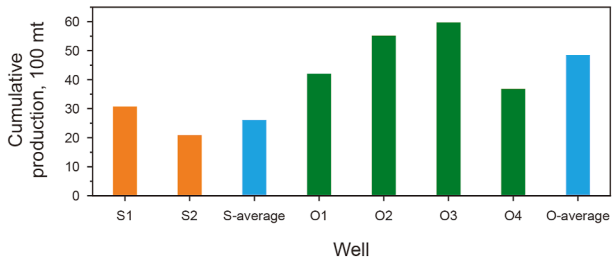

Fig. 20. Diagram of average cumulative production in the first three months comparison.

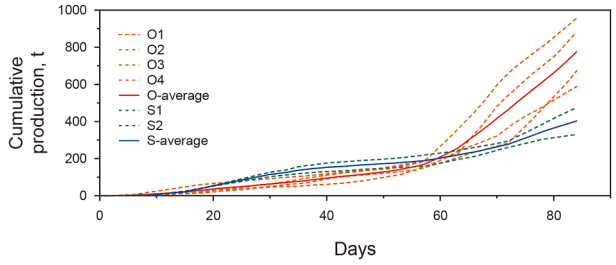

Fig. 21. Diagram of cumulative production growth.

fracturing equipment and proppant loading, providing an ideal opportunity to compare the geometric and optimized completion design methods.

Due to variations in horizontal section lengths among the wells, the average cumulative production per 100 m was compared. After three months, the average cumulative production per 100 m for the four test wells was 48.3 t (O-average, blue bar), while the geometric designed wells averaged 25.9 t, representing a 186% average increase for the test wells (Fig. 20), signifying optimized completion design is effective in increasing production.

We also compared the cumulative production trends within the first three months (Fig. 21), including individual test wells (orange dashed lines), geometric designed wells (green dashed lines), and the average cumulative production for the two methods (red and blue lines, respectively). Both methods exhibit a gradual increase in cumulative production during the first 60 days, followed by a steeper upward trend. However, after day 60, the production increase for the test wells becomes more pronounced, as indicated by the higher slope of the red curve compared to the blue curve. By the end of the observation period, the average cumulative production of the optimized completion design wells is significantly higher than that of the geometric designed wells, demonstrating the superior performance and production efficiency of the optimized completion design strategy.

4. Conclusions

In this paper, an integrated data-mechanism fusion model for intelligent hydraulic fracturing (HF) design was proposed, which covers the entire process from sweet spot evaluation to stage and cluster optimization. The model employs a mechanism-guided weighted *K*-medoids clustering model for geological sweet spot evaluation, a DP algorithm for fracturing stage design, and a hybrid multi-objective optimization method for perforation cluster design. Then, we used the model performance comparison, monitoring data validation, and production monitoring of field intelligent completion design test wells to demonstrate that the proposed model significantly outperforms traditional methods in terms of sweet spot evaluation accuracy, stage and cluster design

rationality, and production enhancement. The major conclusions can be drawn as follows:

(1) The geological sweet spot classification results obtained from the mechanism-guided weighted *K*-medoids clustering model were well validated by the production profile data. High-quality geological sweet spots (Classes I and II) showed a positive correlation with oil production, while low-quality geological sweet spots (Classes III and IV) exhibited a negative correlation. Compared to traditional clustering models, the results of this model are more in line with geological principles. Therefore, the mechanism-guided data-driven clustering model can achieve better results than pure data-driven methods.
(2) The geomechanical strength, characterized by downhole mechanical specific energy and minimum horizontal principal stress, had a high correlation with the fluid allocation results obtained from fiber-optic monitoring. In most stages, clusters with higher fluid intake had lower geomechanical strength values. This can serve as an effective evaluation parameter for engineering sweet spots. Moreover, this parameter can be directly calculated from field drilling and logging data, making it easily accessible.
(3) Aiming to develop high-quality reservoirs and achieve balanced fracture initiation, the DP algorithm and the multi-objective optimization algorithm (NSGA-II) intelligently design stages and clusters by considering 'double sweet spot' and physical constraints. The entire process is automated and intelligent, allowing engineers to adjust physical constraints according to the construction requirements of the block to realize intelligent completion design.
(4) Comprehensive evaluation through model performance comparison, monitoring data validation, and field pilot well tests demonstrated the superiority of the proposed model. The optimized completion design significantly outperformed traditional methods in terms of sweet spot evaluation accuracy, stage and cluster design rationality, and production enhancement. In the field application of four shale oil test wells in the Jimsar Sag, Xinjiang, the average cumulative production per 100 m in the first three months increased by 186% compared to geometric designed wells. This verified the practical feasibility and application value of the model in the development of unconventional reservoirs.
(5) There are still some limitations in this model. The model relies on drilling and logging data, and the quality, resolution, and consistency of data collection directly affect the model evaluation results. Some blocks may not be fully applicable due to limitations in data acquisition conditions. During the weighted clustering, DP, and multi-objective optimization processes, the parameters involved in the model, such as weight allocation, objective function construction, and constraint conditions, have a certain degree of subjectivity and are sensitive to the results. They may need to be optimized for different reservoir conditions. Although this study has attempted to combine geological and engineering sweet spots, there are still shortcomings in the multi-scale and multi-physics coupling mechanism. In the future, the coupling relationship between mechanistic data and production responses needs to be further strengthened.

CRediT authorship contribution statement

Shi-Meng Hu: Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Mao Sheng:** Writing – review & editing, Project administration, Methodology, Funding acquisition, Conceptualization. **Bing-Bing Liu:** Software, Investigation, Formal analysis, Data curation. **Jie Li:** Resources, Methodology, Investigation. **Shou-Ceng Tian:** Resources, Funding acquisition, Conceptualization. **Xiao-Dong He:** Resources, Methodology, Investigation. **Gen-Sheng Li:** Supervision, Resources.

Acknowledgments

The authors would like to thank the financial project supported by the National Natural Science Foundation of China (Grant/Award Nos.: 52310001009 and 52122401).

References

Awad, F.H., Hamad, M.M., 2022. Improved *k*-means clustering algorithm for big data based on distributed smartphoneneural engine processor. Electronics 11 (6), 883. https://doi.org/10.3390/electronics11060883.
Behrens, J.T., DiCerbo, K.E., Yel, N., Levy, R., 2012. Exploratory data analysis. In: Weiner, I.B. (Ed.), Handbook of Psychology, second ed., vol. 2. Research Methods in Psychology. Wiley, Hoboken. https://doi.org/10.1002/9781118133880.hop202002.
Bhattacharya, S., Carr, T.R., 2019. Integrated data-driven 3D shale lithofacies modeling of the Bakken Formation in the Williston basin. J. Pet. Sci. Eng. 178, 703–715. https://doi.org/10.1016/j.petrol.2019.02.036.
Buller, D., Hughes, S.N., Market, J., et al., 2010. Petrophysical Evaluation for Enhancing Hydraulic Stimulation in Horizontal Shale Gas Wells. SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/132990-MS.
Chen, S., Zhao, W., Ouyang, Y., et al., 2017. Comprehensive prediction of shale gas sweet spots based on geophysical properties: a case study of the Lower Silurian Longmaxi Fm in Changning block, Sichuan Basin. Nat. Gas. Ind. 37 (5), 20–30. https://doi.org/10.3787/j.issn.1000-0976.2017.05.003.
Chen, Z., Yang, C., Jiang, C., et al., 2018. Production characteristics and sweet-spots mapping of Bakken Formation tight oil. Petrol. Explor. Dev. 45 (4), 626–635. https://doi.org/10.1016/S1876-3804(18)30070-3.
Chorn, L., Yarus, J., del Rosario-Davis, S., Pitcher, J., 2013. Identification of Shale Sweet Spots Using Key Property Estimates from Log Analysis and Geostatistics. Unconventional Resources Technology Conference. https://doi.org/10.1190/urtec2013-154.
Deb, K., Pratap, A., Agarwal, S., et al., 2002. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation 6 (2), 182–197. https://doi.org/10.1109/4235.996017.
Ester, M., Kriegel, H.-P., Sander, J., et al., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. https://doi.org/10.5555/3001460.3001507.
Halder, R.K., Uddin, M.N., Khraisat, A., et al., 2024. Enhancing *K*-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. J. Big Data 11 (1), 113. https://doi.org/10.1186/s40537-024-00973-y.
Hart, B.S., Pearson, R.A., Herrin, J.M., et al., 2000. Horizon Attributes and fracture-swarm Sweet Spots in low-permeability Gas Reservoirs. SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/63207-MS.
Hu, S., Sheng, M., Qin, S., et al., 2023a. Unsupervised cluster model of formation fracability based on drill-log data. Pet. Sci. Bull. 5, 767–774. https://doi.org/10.3969/j.issn.2096-1693.2021.01.001.
Hu, S., Sheng, M., Shi, S., et al., 2023b. Optimization of fracturing stages/clusters in horizontal well based on unsupervised clustering of bottomhole mechanical specific energy on the bit. Nat. Gas. Ind. 43 (7), 55–62. https://doi.org/10.3787/j.issn.1000-0976.2023.07.006.
Jiang, J., Zhang, X., Yuan, Z., 2024. Feature selection for classification with Spearman's rank correlation coefficient-based self-information in divergence-based fuzzy rough sets. Expert Syst. Appl. 249, 123633. https://doi.org/10.1016/j.eswa.2024.123633.
Ko, K.N., Jeong, T., Park, Y.J., et al., 2016. Shale gas field sweet spot determination process. J. Geol. Soc. Korea 52 (6), 799–814. https://doi.org/10.14770/jgsk.2016.52.6.799.
Li, G., Song, X., Tian, S., et al., 2022. Intelligent drilling and completion: a review. Engineering 18 (11), 33–48. https://doi.org/10.1016/j.eng.2022.07.014.
Li, G., Tian, S., Sheng, M., et al., 2025. Research progress and prospect of intelligent hydraulic fracturing technologies. Drill. Prod. Technol. 48 (1), 1–9. https://doi.org/10.3969/J.ISSN.1006-768X.
Li, L., Wang, Z., Yin, S., et al., 2024. Selection and application of wavelet transform in high-frequency sequence stratigraphy analysis of coarse-grained sediment in rift basin. Pet. Sci. 21 (5), 3016–3028. https://doi.org/10.1016/j.petsci.2024.06.020.
Liao, D., 2020. Evaluation methods and engineering application of "double sweet spots" in shale gas reservoirs. Pet. Drill. Tech. 48 (4), 94–99. https://doi.org/10.11911/syztjs.2020063.

Liao, D., Xiao, L., Zhang, Y., 2014. Evaluation model for shale brittleness index model based on mineral content and fracture toughness. Pet. Drill. Tech. 42 (4), 37–41. https://doi.org/10.3969/j.issn.1001-0890.2014.04.007.

Prise, G.J., Stewart, D.R., Bird, T.M., et al., 1993. Successful Completion Operations on Ravenspurn North Development. SPE Offshore Europe Conference and Exhibition. https://doi.org/10.2118/26744-MS.

Pitcher, J., Kwong, S., Yarus, J., Mullen, M., 2012. Exploring shale basins using existing wells. In: SPE/EAGE European Unconventional Resources Conference. https://doi.org/10.2118/152579-MS.

Reimann, C., Filzmoser, P., Hron, K., et al., 2017. A new method for correlation analysis of compositional (environmental) data-a worked example. Sci. Total Environ. 607–608, 965–971. https://doi.org/10.1016/j.scitotenv.2017.06.063.

Shen, Y., Liu, B., Jiang, H., et al., 2024. Geological and engineering integration fracturing design in Weixinan Sag. J. Mar. Sci. Eng. 12 (10), 1821. https://doi.org/10.3390/jmse12101821.

Sheng, M., Tian, S., Zhu, D., et al., 2024. Design concept and key technologies of intelligent decision-making system for hydraulic fracturing pumping. Nat. Gas. Ind. 44 (9), 108–113. https://doi.org/10.3787/j.issn.1000-0976.2024.09.010.

Sun, L., Zhao, W., Liu, H., et al., 2023. Concept and application of “sweet spot” for shale oil. Acta Pet. Sin. 44 (1), 1–13. https://doi.org/10.7623/syxb202301001.

Zhang, J., Jin, Z., Pang, X., 2000. Formation conditions and internal features of deep basin gas accumulations. Exp. Pet. Geol. 22 (3), 210–214. https://doi.org/10.11781/sysydz200003210.

Zhang, B., Zhou, Z., Mi, H., et al., 2024. A method for evaluating geology-engineering sweet spots of tight sandstone gas reservoirs considering the influence of thin coal seams: a case study on Shenfu Block of the Ordos Basin. Nat. Gas. Ind. 44 (10), 170–181. https://doi.org/10.3787/j.issn.1000-0976.2024.10.01.

Zhang, D., Yu, Y., Li, S., et al., 2021. Staging optimization of multi-stage perforation fracturing based on unsupervised machine learning. J. China Univ. Pet. (Ed. Nat. Sci.) 45 (4), 59–66. https://doi.org/10.3969/j.issn.1673-5005.2021.04.007.

Zhao, W., Zou, C., Wang, Z., et al., 2004. The intention and significance of “Sag-wide Oil-Bearing Theory” in the hydrocarbon-rich depression with terrestrial origin. Petrol. Explor. Dev. 31 (2), 5–13. https://doi.org/10.3321/j.issn:1000-0747.2004.02.002.

Zhou, X., Wei, J., Fahimeh, H., et al., 2023. “Double sweet spot” identification method via cluster analysis of petrophysical properties from well logg data: a case study of a tight sandstone reservoir. Geoenergy Sci. Eng. 227, 211821. https://doi.org/10.1016/j.geoen.2023.211821.

Zhu, J., Niu, B., Cai, P., et al., 2024. Towards automated cross-domain exploratory data analysis through large language models. https://doi.org/10.48550/arXiv.2412.07214.