WEP Warwick Evans Publishing

International Journal of Natural Resources and Environmental Studies

ISSN: 3006-2012 (Print), ISSN: 3006-0834 (Online) | Volume 7, Number 1, Year 2025
DOI: https://doi.org/10.62051/ijnres.v7n1.02
Journal homepage: https://wepub.org/index.php/IJNRES/index

# Classification of Shale Oil Horizontal Well Production: A Case Study of the Chang-7 Interval in the Ordos Basin

Jiaming Liu, Yuechen Li, Yutong Li and Ruifei Wang

College of Petroleum Engineering, Xi'an Shiyou University, Xi 'an, Shaanxi 710065, China

## ABSTRACT

This study analyzes 70 shale-oil horizontal wells in the Chang-7 interval of the Qingcheng Oilfield, Ordos Basin, using the K-means clustering algorithm to evaluate production capacity. The objective is to provide a foundation for differentiated development and refined management. First-year cumulative oil production was selected as the primary classification metric. The silhouette coefficient and elbow methods were used to determine the optimal cluster number. Results indicate that a three-class division is most appropriate: Class I wells (> 4,300 t, 21.7%), Class II wells (2,900-4,300 t, 40.8%), and Class III wells (< 2,900 t, 37.5%). Correlation analysis revealed distinct controlling factors for each well type. Class I wells are jointly influenced by engineering and geological conditions, with key parameters including drilled lateral length, proppant volume, and log-interpreted Class I intervals. Class II wells are primarily affected by engineering parameters and production practices, notably injected fluid volume, cluster number, pump rate, and shut-in duration. Class III wells are mainly governed by geological factors, with permeability, porosity, and average total organic carbon content being the most critical. The findings suggest that fracture-design and production strategies should be tailored to the specific controlling factors of each well type to enhance shale-oil development efficiency and achieve targeted exploitation. This approach offers a practical technical pathway and theoretical reference for classification management and optimization of shale-oil horizontal wells in the region.

## KEYWORDS

Shale Oil; Horizontal Well; Classification and Evaluation; K-means Algorithm.

## 1. INTRODUCTION

The Ordos Basin is one of China's key hydrocarbon-rich regions, characterized by abundant shale oil resources with considerable development potential. In recent years, the widespread application of horizontal drilling and volumetric fracturing has markedly enhanced shale oil recovery efficiency [1, 2]. However, complex geological conditions and varying engineering practices have led to pronounced differences in horizontal well production, creating major challenges for development management and resource evaluation [3-6]. Therefore, developing scientific classification methods for shale-oil horizontal wells has become a critical requirement for achieving differentiated development in current shale-oil exploitation.

Current research on shale-well production differences has mainly focused on individual geological or engineering factors, while systematic classification and evaluation methods remain underdeveloped [7, 8]. This limitation hinders the precise alignment of development strategies with the characteristics of different well types. Although clustering-based methods have been applied in petroleum engineering, their use in shale-well production classification remains limited, especially for systematic evaluations integrating multi-source data.

Against this background, this study investigates Chang-7 shale-oil horizontal wells in the Ordos Basin, using first-year cumulative oil production as the primary classification metric. K-means clustering is applied to 70 horizontal wells, and optimization techniques are employed to determine the optimal number of clusters. Based on this, the production characteristics of different well types are analyzed to provide theoretical support and technical guidance for classification management and efficient shale-oil development.

# 2. SHALE-OIL HORIZONTAL WELL PRODUCTION CLASSIFICATION METHOD

## 2.1. K-means algorithm

The K-means algorithm is a widely used unsupervised learning method. Its core principle is to partition a dataset into K clusters through iterative optimization. In each iteration, samples are assigned to the nearest cluster centroid, and centroid positions are updated repeatedly until convergence is achieved.

Let the dataset be $X = \{x^{(1)}, \dots, x^{(i)}, \dots, x^{(m)}\}$, containing m samples where $x^{(i)} \in R^n$. The K-means procedure is described as follows:

1. Randomly initialize k cluster centroids ($\mu = \{\mu_1, \dots, \mu_j, \dots, \mu_k\}$), where $\mu_j \in R^n$. Each centroid corresponds to a cluster $C_j$, yielding k clusters in total.
2. Compute the distance from each data sample to the k centroids: $d_j^{(i)} = \left\|x^{(i)} - \mu_j\right\|^2$, where $d_j^{(i)}$ denotes the squared distance between sample $x^{(i)}$ and centroid $\mu_j$. Denote $D^{(i)} = \{d_1^{(i)}, \dots, d_j^{(i)}, \dots, d_k^{(i)}\}$.
3. Assign each data sample to the cluster whose centroid is nearest: $c^i := arg\min_j D^{(i)}$. Here $C_j$ indicates the nearest cluster index for sample i, with $c^i \in \{1,2, \dots, j, \dots, k\}$.
4. After all samples have been assigned, update each cluster centroid:$\mu_j = \frac{1}{|C_j|}\sum_{x^{(i)} \in C_j} x^{(i)}$.
5. Repeat steps (2)-(4) until the centroids no longer change or a preset maximum number of iterations is reached.

## 2.2. Evaluation methods and metrics

In the K-means algorithm, the number of clusters (k) is a critical parameter that directly influences clustering performance. This study employs both the silhouette coefficient and elbow methods to evaluate clustering quality, thereby identifying the optimal cluster number and defining categories of shale-oil horizontal well productivity.

The silhouette coefficient evaluates clustering performance by considering both within-cluster cohesion and between-cluster separation. For each sample, the silhouette value is computed as in Equation (1):

$$s^{(i)} = \frac{b^{(i)} - a^{(i)}}{\max\{a^{(i)}, b^{(i)}\}} \tag{1}$$

Here, $a^{(i)}$ is the average distance between sample $x^{(i)}$ and other samples within the same cluster, representing cohesion. $b^{(i)}$ is the average distance between $x^{(i)}$ and samples in the nearest neighboring cluster, representing separation. The silhouette coefficient ranges from −1 to 1. Values near 1 indicate that a sample is well matched to its cluster and clearly separated from others, reflecting strong clustering. Values near −1 suggest poor clustering and the need to reassess assignments. Values

close to 0 indicate boundary samples with ambiguous clustering quality. The average silhouette coefficient across all samples is used to assess overall clustering quality. In K-means clustering, the optimal k is typically the value that maximizes the mean silhouette coefficient.

The elbow method evaluates clustering compactness using the within-cluster sum of squared errors (SSE), defined as in Equation (2):

$$\mathrm{SSE} = \sum_{j=1}^{k} \sum_{x \in c_j} \left\| x - \mu_j \right\|^2 \tag{2}$$

When k is smaller than the optimal cluster number, increasing k markedly improves compactness, leading to sharp decreases in SSE. Once k reaches the optimal value, further increases yield only marginal gains, and the SSE curve flattens. The optimal cluster number is identified at the "elbow point," where the SSE curve shows a distinct inflection.

# 3. CLASSIFICATION ANALYSIS OF SHALE-OIL HORIZONTAL WELLS IN AREA Q

## 3.1. Classification criteria for shale-oil horizontal wells in area Q

This study analyzes 70 shale-oil horizontal wells from the Qingcheng Oilfield. Silhouette coefficients and within-cluster sum of squared errors (SSE) were calculated for cluster numbers ranging from k = 2 to 10. As shown in Fig.1, the silhouette coefficient varies with the number of clusters, reaching a maximum of 0.599 at k = 3. At k = 2, the coefficient is 0.589, which is close to the value for k = 3. Therefore, both k = 2 and k = 3 yield comparable silhouette values, suggesting reasonably good clustering performance. Fig.2 presents the variation of SSE with respect to cluster centroids across different values of k. When k < 3, SSE decreases rapidly as k increases, whereas for k > 3, the rate of decrease becomes gradual. According to the elbow method, the SSE curve exhibits a distinct elbow at k = 3, indicating that three clusters represent the optimal solution. By integrating the silhouette coefficient and elbow method results, this study identifies three clusters as the optimal classification for shale-oil horizontal wells.

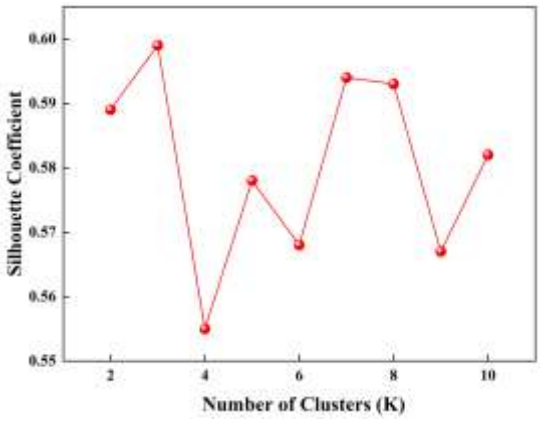


**Fig. 1** Silhouette coefficients for different numbers of clusters

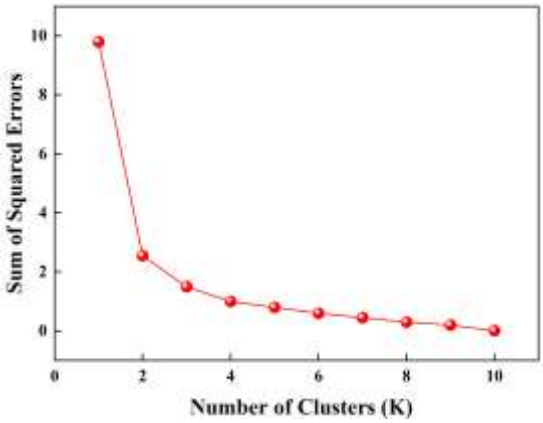


**Fig. 2** Sum of squared errors for different numbers of clusters

Based on first-year cumulative oil production as the classification criterion, the shale-oil horizontal wells in area Q were divided into three classes according to the optimal number of clusters. The classification criteria are summarized in Table 1. Fig. 3 presents the normalized classification results, where Class I, II, and III wells account for 21.7%, 40.8%, and 37.5% of the total, respectively.

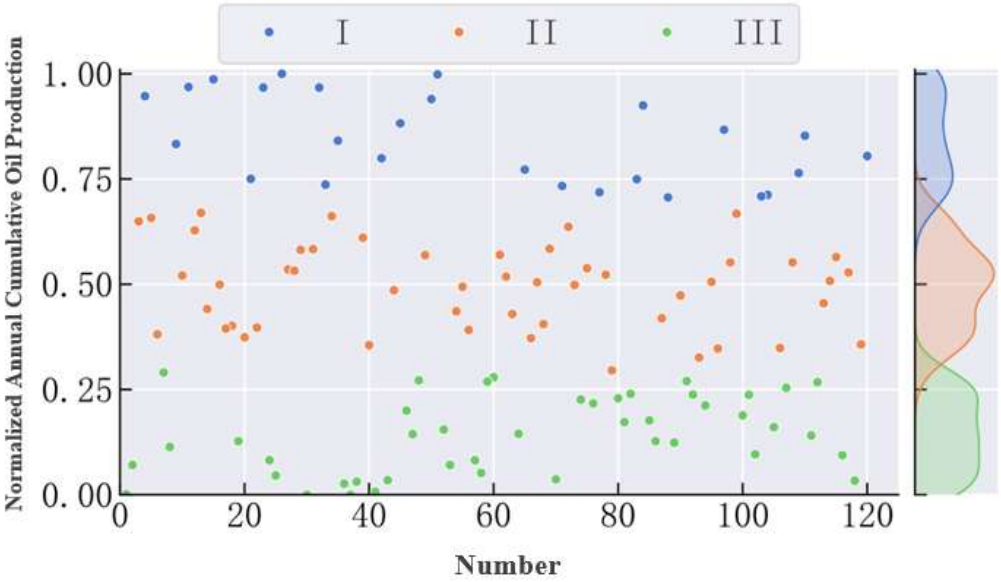


**Fig. 3** Classification map of shale-oil horizontal wells in Qingcheng Oilfield

**Table 1** Classification criteria for shale-oil horizontal wells

| Category | First-year cumulative oil production (t) |
|---|---|
| Class I | >4300 |
| Class II | 2900-4300 |
| Class III | <2900 |

### 3.2. Analysis of factors influencing production capacity of different well types

The correlation coefficient is a statistical measure that quantifies the relationship between two random variables, typically used to describe both the direction and strength of their association. The correlation coefficient r, first proposed by Pearson, measures the degree of linear association between variables. It ranges from −1 to 1 under the assumptions of approximate normality and nonzero standard deviation of the variables. Different definitions of correlation exist depending on the context; the Pearson correlation coefficient is most commonly used. Common interpretation thresholds for |r| are: ≥ 0.8, very strong correlation; 0.6–0.8, strong correlation; 0.4–0.6, moderate (weak) correlation; < 0.4, very weak or no correlation. The Pearson correlation coefficient is defined as:

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \cdot \sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{3}$$

where r is the correlation coefficient; $\bar{x}$ and $\bar{y}$ are the means of variables x and y, respectively; and $x_i$ and $y_i$ are the i-th observations of x and y.

Using first-year cumulative oil production as the evaluation metric, correlation analyses were conducted for Class I, II, and III wells with respect to geological, engineering, and production-operational factors. Class I wells. The top five factors by overall correlation are: actual drilled lateral length > proppant volume > log-interpreted Class I interval > injected fluid volume > number of stages. Engineering factors (ranked): actual drilled lateral length > proppant volume > injected fluid volume > number of stages > number of clusters > sand ratio > pump rate. Geological factors (ranked): log-interpreted Class I interval > porosity > oil saturation > permeability > average total organic carbon (TOC) > drilling encounter rate. Production dynamics & operational factors (ranked): shut-in days > backflow rate when water cut = 90% > backflow rate when water cut = 60% > backflow rate

when water cut = 30%. Class II wells. Engineering parameters and production regimen factors show the highest correlations. The top five factors overall are: injected fluid volume > number of clusters > pump rate > shut-in days > proppant volume. Engineering factors (ranked): injected fluid volume > number of clusters > pump rate > proppant volume > actual drilled lateral length > sand ratio > number of stages. Geological factors (ranked): oil saturation > permeability > porosity > average TOC > log-interpreted Class I interval > drilling encounter rate. Production dynamics & operational factors (ranked): shut-in days > backflow rate at 60% water cut > backflow rate at 30% water cut > backflow rate at 90% water cut. Class III wells. Geological factors dominate. The top five factors overall are: permeability > average TOC > porosity > oil saturation > actual drilled lateral length. Engineering factors (ranked): actual drilled lateral length > number of clusters > pump rate > number of stages > injected fluid volume > proppant volume > sand ratio. Geological factors (ranked): permeability > porosity > average TOC. Production dynamics & operational factors (ranked): backflow rate at 60% water cut > backflow rate at 30% water cut > drilling encounter rate > shut-in days > backflow rate at 90% water cut.

## 4. SUMMARY

1. Based on 70 shale-oil horizontal wells in the Qingcheng Oilfield, the K-means clustering algorithm was applied to evaluate silhouette coefficients and clustering errors for cluster counts k = 2–10. Using first-year cumulative oil production as the classification criterion and selecting the optimal number of clusters, the wells were categorized into three classes: Class I — >4,300 t; Class II — 2,900–4,300 t; and Class III — <2,900 t.

2. The factors influencing production capacity vary significantly across well types. Class I wells are jointly controlled by engineering and geological conditions, with key drivers including actual drilled lateral length, proppant volume, and log-interpreted Class I intervals. This highlights the importance of both reservoir characterization and engineering optimization. Class II wells are primarily influenced by engineering and production-operational parameters, with injected fluid volume, number of fracturing clusters, pump rate, and shut-in days as dominant factors, underscoring the need for refined fracturing design and production management. Class III wells are governed mainly by geological attributes, where permeability, porosity, and average total organic carbon (TOC) exert the strongest control, indicating that development should emphasize reservoir quality assessment and selective deployment. In light of these differences, differentiated development strategies should be implemented to enhance overall recovery efficiency, providing both theoretical guidance and technical support for precision shale-oil development.

## REFERENCES


[1] Li, E., Wang, Z., Guo, Y., Zhang, J., Bi, S., & Sun, B. (2025). Research on the synergistic inhibition of wax formation in shale oil system using efficient wax inhibitors: Experiments and mechanisms. Chemical Engineering Science, 121902.

[2] Chen, Y. J., Wang, H. Y., & Sharma, M. (2025). The benefits and challenges of well monitoring of Gulong shale oil. Earth Energy Science.

[3] Cao, J., Cai, M., Li, J., Guo, Z., Jiang, X., & Dong, G. (2025). Optimization of Volumetric Fracturing Stages and Clusters in Continental Shale Oil Reservoirs Based on Geology-Engineering Integration. Energies, 18(12), 3066.

[4] Liu, J., Wang, R., Song, P., Li, Y., & Zheng, S. (2025). Quantitative Characterization and Flow Simulation of Micropore Structure in Clastic Gas Reservoirs Based on Micron CT Scanning. ACS omega, 10(20), 20686-20700.

[5] Zhang, Z., Hu, J., & Zhang, Y. (2025). A semi-analytical model for fractured horizontal wells production considering imbibition during shut-in periods. Petroleum Science and Technology, 43(15), 1891-1909.

[6] Chaikine, I. A., & Gates, I. D. (2021). A machine learning model for predicting multi-stage horizontal well production. Journal of Petroleum Science and Engineering, 198, 108133.

[7] Liang, Z., Li, X., Zhou, H., Meng, L., Sun, A., Wu, Q., & Wen, H. (2025). Continental Shale Oil Reservoir Lithofacies Identification and Classification with Logging Data—A Case Study from the Bohai Bay Basin, China. Minerals, 15(5), 484.

[8] Chen, S., Wang, X., Li, X., Sui, J., Yang, Y., Yang, Q., ... & Dai, C. (2024). Geophysical prediction technology for sweet spots of continental shale oil: A case study of the Lianggaoshan Formation, Sichuan Basin, China. Fuel, 365, 131146.