



Study on gradational optimization of oil reservoir streamline field based on an artificial intelligence algorithm

Qi Guo¹ · Shumei He¹ · Lixin Meng¹

Received: 22 August 2018 / Accepted: 2 November 2018 / Published online: 17 November 2018
© The Author(s) 2018

Abstract

Reservoir streamline numerical simulation, as an effective means to describe the trajectory of reservoir fluid and evaluate the flow size of different regions, has been widely used in various oil fields. To reflect the quantitative description of the streamline field in different development states, taking the different spatial positions and streamline attributes as the evaluation index, the reservoir flow field was clustered and evaluated combined with an artificial intelligence algorithm. The clustering results of different streamline clusters were evaluated using an evaluation coefficient. Finally, a quantitative evaluation method of reservoir streamline field simulation was formed. The results show that the peak density algorithm based on the evaluation coefficient can run efficiently in the reservoir streamline field and can quickly determine the optimal clustering number. The evaluation method was applied to the Nm3-4-1 layer of the Gang Dong oilfield. The current streamline field was divided into 14 grades. The average oil–water mobility ratio of the first kind of streamline cluster was much higher than that of the 14th kind of streamline cluster. After adjusting the streamline field, the optimal number of clusters was calculated by the evaluation coefficient again, and the number of streamline clusters was changed into seven groups, which indicated that the whole reservoir driving energy was more balanced, and the utilization degree of the reservoir was clearly improved.

Keywords Artificial intelligence algorithm · Evaluation coefficient · Streamline field · Streamline numerical simulation · Streamline cluster · Clustering analysis

Introduction

Streamline numerical simulation is widely used as an important means to characterize fluid flow paths and describe the intensity of the flow field during the development of an oil field. By establishing the pressure equation on the grid and quadrature of the streamline, we can get the corresponding pressure equipotential surface and establish a natural migration network. The fluid moves along the streamline, so as to track the migration of oil, gas, and water in the reservoir. Compared with the traditional finite difference numerical

simulation method, modern reservoir streamline simulation in many aspects has achieved great breakthroughs, such as the Pollock streamline tracing method (Jackson et al. 1988), data transfer method between the network and grid (Batycky et al. 1997), and the history matching method based on streamline simulation (Wang 2002). Thiele and Batycky (2003) established water injection optimization model by streamline numerical simulation technology. Cheng et al. (2007) established the historical fitting method of streamline field numerical simulation through three-dimensional three-phase reservoir streamline simulation. Zhao et al. (2016) predicted the connectivity between oil and water wells based on the interwell connectivity inversion model. Wang et al. (2017) characterized the water drive flow field through the water drive characteristic curve and the interwell connectivity model. However, the current streamline field simulation can only qualitatively judge the distribution of the current flow field by the streamline density in each area of the reservoir; it cannot quantitatively evaluate the streamline field. Therefore, we need to explore a method for quantitative evaluation of reservoir streamline field.

✉ Qi Guo
qqqqguoqi@163.com
Shumei He
ggggguoqi@163.com
Lixin Meng
18954656977@163.com

¹ Petro China Da gang Oilfield Exploration and Development Research Institute, Tianjin 300280, People's Republic of China

Clustering analysis is an important statistical method in research on classification; one clustering method classifies the data through unsupervised ways and tests the similarity and difference between data. In the end, classification results of the large similarity in each cluster and large differences between clusters are formed (Agarwal et al. 1999). With the development of data mining technology, clustering methods have been used to analyze data sets with multiple attributes and complex distribution structures (Grabmeier and Rudolph 2002). In recent years, quantum clustering, spectral clustering, granularity clustering, probabilistic graph clustering, and synchronous clustering have also become popular. Aminzadeh et al. (2013) used tomographic inversion, fuzzy clustering, and shear wave splitting to obtain reliable characteristics about fractured areas; it can be used to optimize drilling targets or stimulation jobs to reduce costs and maximize production. Tan et al. (2016) introduced the application of data mining to petroleum exploration and development to obtain high-performance predictive models and optimal classifications of geology, reservoirs, reservoir beds, and fluid properties. Aliyarov and Ramazanov (2016) used fuzzy clustering to predict reservoir rock properties. However, under the influence of the difficulty in extracting the streamline data of the reservoir, there is no relevant literature to introduce the clustering evaluation method of streamline field based on reservoir numerical simulation at present. How to extract and cluster streamline data and determine the optimal number of clusters is the focus of this study.

In this paper, the data of the streamline field were extracted, and the streamline field was evaluated by the peak density clustering algorithm. An evaluation coefficient was used to evaluate the clustering effect of different streamline cluster numbers, and ultimately we achieved the purpose of streamline grading optimization. The results show that, for the current streamline field of the Nm3-4-1 layer in the Gang dong oil field, when the evaluation coefficient we specified is the minimum, the clustering result is the best. The grading results clearly show the distribution of displacing efficiency at the present time of reservoir streamline field. The new method provides a way to regulate the streamline field and enhance oil recovery.

The establishment of the peak density algorithm model

Rodriguez and Laio (2014) published “Machine learning. Clustering by fast search and find of density peaks” in the magazine “Science”. Their peak density algorithm provided a new way for clustering evaluation. They considered that the cluster centers should have the characteristics of high density and far away from the data point that is larger than it

is. The difference between density-based and distance-based clustering algorithms is that the density-based clustering algorithm determines clusters of arbitrary shapes, while the distance-based clustering algorithm determines a spherical cluster. This plays an important role in the evaluation of data sets with noise points and density differences.

The calculation flow of the peak density algorithm is shown in Fig. 1. The basic principle is to measure sample density based on the number of similar samples. Select the maximum density sample of the local area as the clustering center, and ensure that the distance between the sample and other samples with a larger density is large enough.

The specific steps for the peak density algorithm are:

1. Standardize data sets to eliminate the effects of different data units.
2. Determine the local density and separation distance of each sample separately.
3. The product of local density and separation distance is used as the evaluation value, and the sample evaluation values are arranged in descending order. The samples with the top rank are selected as the clustering centers.
4. Assign the remaining points to the higher density clusters that are adjacent to it.

In the calculation process, the formula for calculating the local density of item i is:

$$\rho_i = \sum_j (d_{ij} - d_c) \quad i = 1, 2, 3 \dots n, \quad (1)$$

where ρ_i is the local density of sample i , d_{ij} is the distance between sample i and sample j , d_c is the cut-off distance, and n is the total sample size.

To ensure a large separation distance between sample i and larger density samples, we need to evaluate the separation distance of the sample. The expression is:

$$\delta_{q_i} = \min_{q_j, j < i} d_{q_i, q_j}, \quad (2)$$

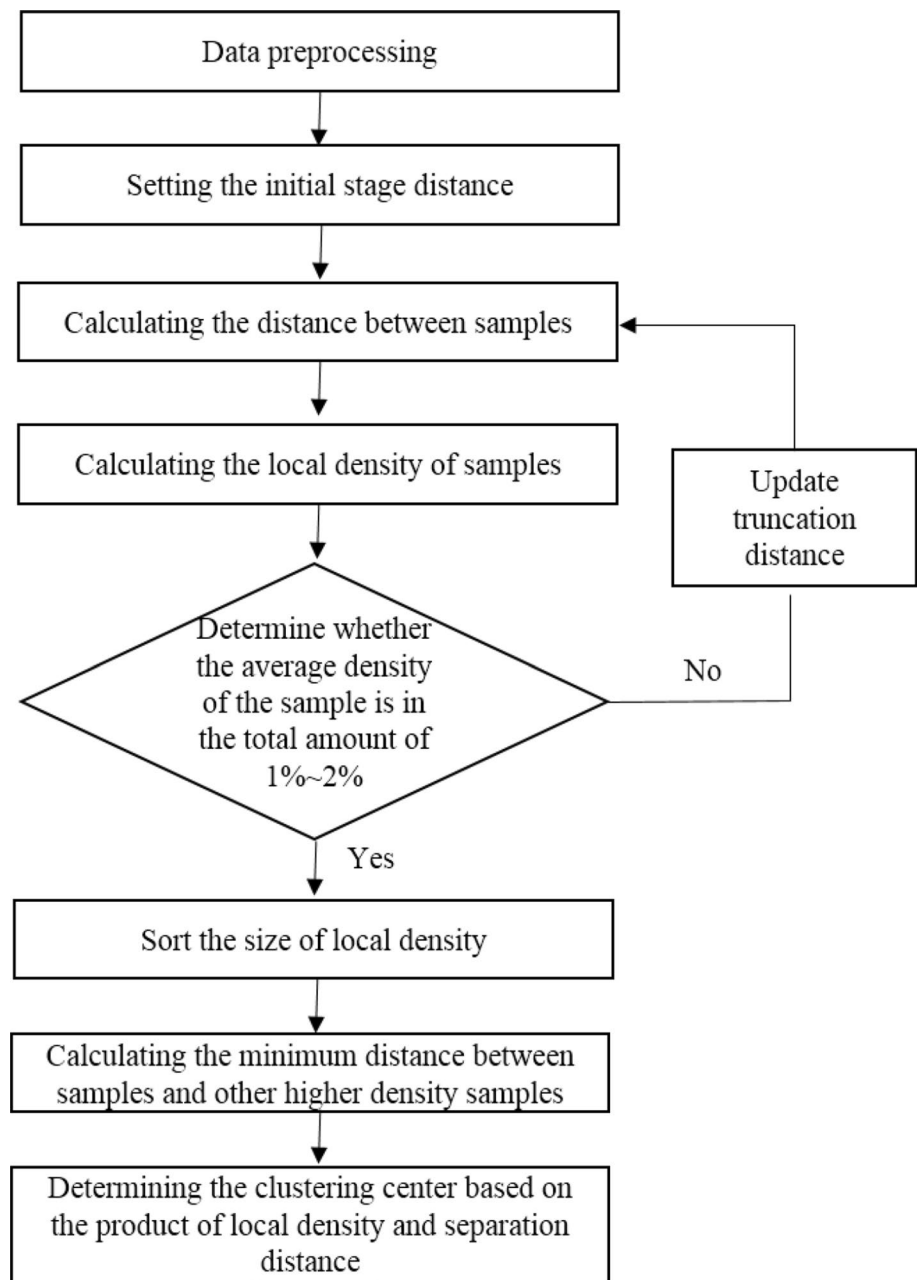
where δ_{q_i} is the separation distance of samples and d_{q_i, q_j} is the distance between sample q_i and sample q_j .

For a cluster center, it should have two attributes with larger local density in its own interior and larger separation distance from other cluster centers. Therefore, the cluster center evaluation formula for the samples i can be determined as follows:

$$\gamma_i = \rho_i \cdot \delta_{q_i}, \quad (3)$$

where γ_i is the cluster center evaluation value of samples i .

Assume that the total data are divided into n group of cluster center, we can descend the γ calculated by all the data, and the former n sample point is the cluster center

Fig. 1 Calculation flow chart of peak density algorithm

point; then the cluster can be determined according to the control range which the user set up at the density center point.

The peak density algorithm has strong ability to distinguish data sets. When the reservoir reaches the high water cut stage, the distribution of streamline field is complex, the streamline attributes are different in density, and so we choose the peak density algorithm to evaluate the streamline field on the basis of density distribution difference.

Determination of evaluation index for clustering effect

The peak density algorithm is used to classify the reservoir streamline field, which can quickly determine the collection of the same kind of streamline. However, the quality of the clustering effect can directly affect the evaluation result. Too many or too few streamline clusters are meaningless; it is necessary to evaluate the effectiveness of the clustering results. Many scholars put forward different algorithms to evaluate

the clustering effect from the compactness and separation degree of the cluster. Compactness is a measure of whether a set of samples is compact enough through the average distance, variance, and other parameters of the cluster center, while the degree of separation is to measure whether the distance between one cluster and another cluster is far enough.

Calinski determined the effectiveness of the clustering effect by evaluating the square sum coefficient between different clusters in 1974. When Calinski–Harabasz index reaches the maximum, clustering is the best (Caliński and Harabasz 1974). Its expression is:

$$\text{Calinski – Harabasz index} : \frac{\sum_i n_i d^2(c_i, c)/(NC - 1)}{\sum_i \sum_{x \in C_i} d^2(x, c_i)/(n - NC)} \quad (4)$$

In the above expressions, n_i is the number of objects in C_i ; $d(x, y)$: distance between x and y ; c_i is the center of C_i ; NC is the clusters number; n is the number of objects in sample set; C_i is the i -th cluster; i is the code for different cluster clusters.

Dunn (1974) used the minimum distance between different clusters as the evaluation of the internal separation degree, and the maximum diameter between different clusters was used to evaluate the compactness of the cluster. When Dunn index reaches the maximum, clustering is the best. The specific expression is as follows:

$$\text{Dunn index} : \min_i \left\{ \min_j \left(\frac{\min_{x \in C_i, y \in C_j} d(x, y)}{\max_k \{ \max_{x, y \in C_k} d(x, y) \}} \right) \right\}, \quad (5)$$

where j is the code for different cluster clusters.

In 1985, Hubert proposed to evaluate the clustering effect by calculating the inconsistencies between different cluster data. The Modified Hubert coefficients corresponding to different clusters are numerically plotted; when the inflection point appears in the curve, the corresponding number of clusters is the best (Hubert and Arabie 1985). The expression is:

Modified Hubert coefficient :

$$\frac{2}{n(n-1)} \sum_{x \in D} \sum_{y \in D} d(x, y) \quad d_x \in C_i, d_y \in C_j(c_i, c_j), \quad (6)$$

where D is the data sets; n is the number of objects in D ; c_i, c_j is the center of C_i, C_j , respectively.

Rousseeuw and Sihouettes (1987) defined the contour coefficient, which is based on the distance difference between and within the cluster to evaluate the clustering effect. When the contour coefficient value is the largest, the corresponding number of clusters is the best.

$$\text{Contour coefficient} : \frac{1}{NC} \sum_i \left\{ \frac{1}{n_i} \sum_{x \in C_i} \frac{b(x) - a(x)}{\max[b(x), a(x)]} \right\}, \quad (7)$$

where $a(x) = \frac{1}{n_i - 1} \sum_{y \in C_i, y \neq x} d(x, y)$ and $b(x) = \min_{j, j \neq i} \left[\frac{1}{n_j} \sum_{y \in C_j} d(x, y) \right]$.

In 1996, Mark Marcucci proposed the method of root-mean-square standard deviation to evaluate the clustering effect. It mainly measured the uniformity of the cluster. When the curve formed by the numerical connection of different clustering numbers occurs at inflection point, the corresponding clustering number is the best (Marcucci 1996). The expression is:

Root-mean-square standard deviation :

$$\left\{ \sum_i \sum_{x \in C_i} x - c_i^2 / \left[P \sum_i (n_i - 1) \right] \right\}^{\frac{1}{2}} \quad (8)$$

In the above expressions, P is the attributes number of D (in this streamline clustering evaluation process, the value of P is 4, including centers X and Y , and oil/water phase velocities).

The SD index method is based on the concept of the average scattering and the total separation of the clusters. The concept of average scattering is based on variance clustering of an object to calculate the compactness, while the total separation is calculated according to the distance between different cluster centers. The value of SD index is the summation of these two terms. When the index reaches minimum, the corresponding cluster number is the best (Halkidi et al. 2000), and the expression is:

$$\text{SD index} : \text{Dis}(NC_{\max}) \text{Scat}(NC) + \text{Dis}(NC), \quad (9)$$

where $\text{Scat}(NC) = \frac{1}{NC} \sum_i \sigma(C_i)/\sigma(D)$ and $\text{Dis}(NC) = \frac{\max_{i,j} d(c_i, c_j)}{\min_{i,j} d(c_i, c_j)} \sum_i \left(\sum_j d(c_i, c_j) \right)^{-1}$.

In the above expressions, $\sigma(C_i)$ is the variance vector of C_i ; $\sigma(D)$ is the variance vector of D ; NC_{\max} is the maximum clustering number.

The R-squared index is the ratio of the sum of squares between clusters to the total sum of squares of the whole data set. It is used to measure the degree of difference between clusters. When the index is the largest, the number of clusters corresponding to it is the best. Its expression is:

$$\text{R-squared index} : \sum_{x \in D} x - c^2 - \sum_i \sum_{x \in C_i} x - c_i^2 / \sum_{x \in D} x - c^2 \quad (10)$$

The basic idea of the S_Dbw coefficient is that there is at least one point in all cluster centers whose density is larger than the density of their midpoint. Then, the SD algorithm is used to determine the clustering effect (Halkidi et al. 2000). When the S_Dbw coefficient is the smallest, the number of

clusters corresponding to it is the best. The expression of the S_Dbw coefficient is: Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

$$S_Dbw \text{ coefficient} : \text{Scat}(\text{NC}) + \text{Dens_bw}(\text{NC}), \quad (11)$$

$$\text{where Dens_bw}(\text{NC}) = \frac{1}{\text{NC}(\text{NC}-1)} \sum_i \left[\sum_{\substack{j \neq i}} \frac{\sum_{x \in C_i \cup C_j} d(x, u_{i,j})}{\max\{\sum_{x \in C_i} d(x, c_i), \sum_{x \in C_j} d(x, c_j)\}} \right]$$

Maulik defined I index in 2002 to determine the best number of clusters. It represents the degree of separation by the maximum distance between cluster centers; when I index reaches the minimum, the number of clusters corresponding to it is the best (Maulik and Bandyopadhyay 2002). The expression is as follows:

$$I \text{ index} : \left(\frac{1}{\text{NC}} \cdot \frac{\sum_{x \in D} d(x, c)}{\sum_i \sum_{x \in C_i} d(x, c_i)} \cdot \max_{i,j} d(c_i, c_j) \right)^P \quad (12)$$

In the high water cut stage, the distribution of streamline field is complex, and the data are greatly influenced by

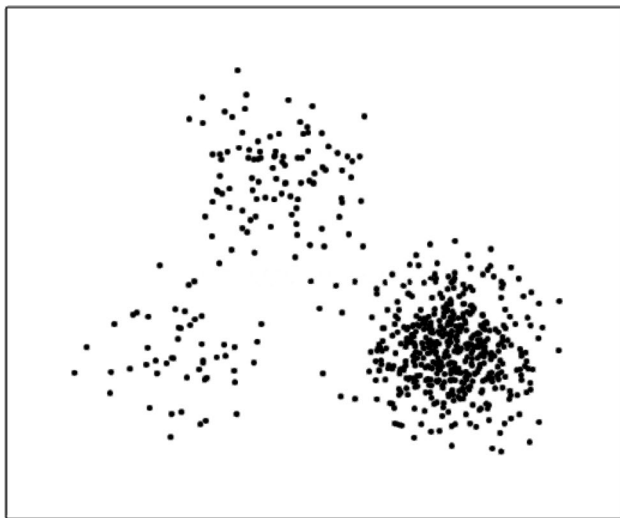


Fig. 2 2-D sample data set

streamline density. Different clustering validation algorithms may lead to differences in prediction accuracy.

The peak density algorithm was applied to two-dimensional data sets as shown in Figs. 2, 3, 4, 5, 6, 7 and 8; clusters were formed respectively. The effectiveness of different clustering groups was evaluated by the above clustering validation method. As we can see, judging by the naked eye, the data set can be roughly divided into three clusters, and the density of each cluster is different, which will also result in different verifying results from different verification methods. Through the above verification method, the clustering effects of different cluster numbers were evaluated as shown in Table 1. The data in bold font in each line were the best number of clusters for each verification method. We can see that the various verification methods are different in determining the optimal number of clusters. The conclusions obtained by several validation methods are inaccurate. The verification methods that can get the correct conclusions are Calinski–Harabasz index, Dunn index, contour coefficient, SD index and S_Dbw coefficient. The S_Dbw coefficient considers the density factor of the data set as the evaluation criterion of the clustering separation degree, and the evaluation is accurate. Therefore, the S_Dbw coefficient was used as the criterion to evaluate the number of clusters in the streamline field.

The procedure of applying artificial intelligence algorithm to analyze the streamline field

The current commercial numerical simulation software can only determine the streamline field intensity of the reservoir by the density of the streamlines in the various regions, and the corresponding oil–water saturation attribute and flow velocity attribute on the streamline cannot be classified; these can all lead to the failure to evaluate the streamline field quantitatively. Making full use of the spatial position of each streamline coordinate point and its corresponding oil and water attribute parameters, we

Table 1 Comparison of clustering effect verification methods under different numbers of clusters

Number of clusters	2	3	4	5	6	7	8
Calinski–Harabasz index	1152	1177	1103	916	797	722	646
Dunn index	0.0485	0.0751	0.0047	0.0048	0.0048	0.0026	0.0025
Modified Hubert coefficient	2922	3704	3939	4268	4269	4270	4272
Contour coefficient	0.577	0.635	0.455	0.366	0.307	0.273	0.24
Root-mean-square standard deviation	28.012	20.45	14.577	3.147	3.029	2.907	2.786
SD index	0.693	0.365	0.661	0.68	0.936	1.172	1.084
R-squared index	0.616	0.787	0.884	0.977	0.978	0.979	0.979
S_Dbw coefficient	0.593	0.27	0.394	0.361	0.307	0.293	0.286
I index	118.1	102.5	91.9	77.3	58.9	55.1	44

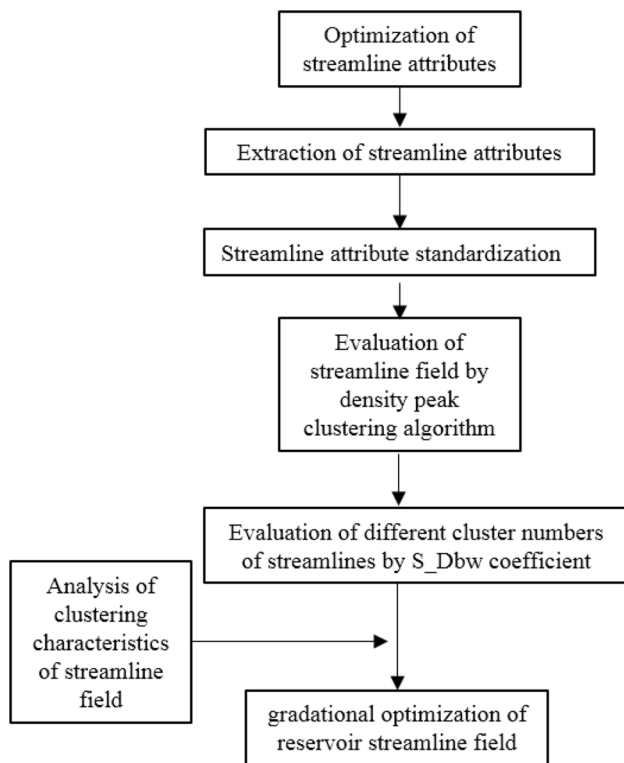
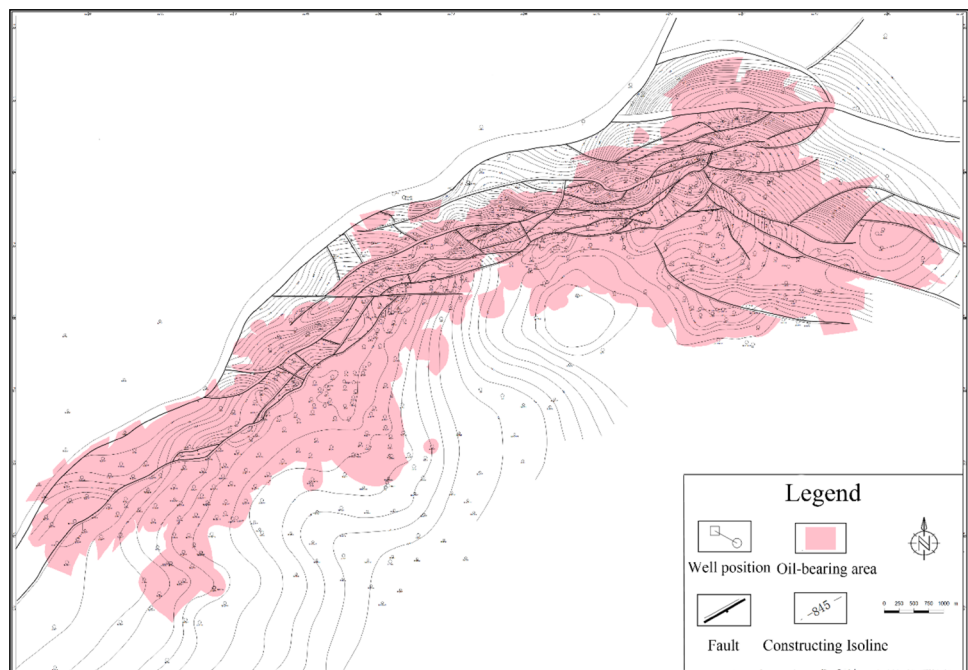


Fig. 3 Application of the peak density clustering algorithm to evaluate the streamline field

used the peak density algorithm to classify the rank of the streamline cluster and determined the optimal number of clusters by the S_Dbw coefficient. The basic process is

Fig. 4 Location map of study area in Gang dong oil field, China



shown in Fig. 3. Based on the peak density algorithm and the S_Dbw coefficient evaluation method, the reservoir streamline field evaluation software was developed. The Petrel_RE-based streamline field evaluation plug-in was formed. The software can effectively access the software of the commercial reservoir numerical simulation and provide technical support for the analysis of the streamline field of the reservoir by the peak density clustering algorithm. Here we described the following key steps in the whole evaluation process.

Data optimization and extraction of reservoir streamline field

The extraction of streamline data is mainly to obtain the spatial coordinates of the streamline data and their corresponding attribute data, thus applying these data to the cluster analysis. Because the commercial numerical simulation software cannot directly derive the relevant parameters of the streamline, this study has developed the relevant software through the ocean platform of the Schlumberger Company. The X and Y spatial coordinates and corresponding water phase flow velocity and oil phase flow velocity data were extracted from the target block streamlines. The spatial coordinate data were to reflect streamline clusters with similar positions in space, and the water phase and oil phase flow velocities were to ensure that the streamline clusters obtained by clustering not only had similar positions, but also had similar flow capacity.

Fig. 5 Streamline distribution of the Nm3-4-1 layer in the Gang Dong oil field

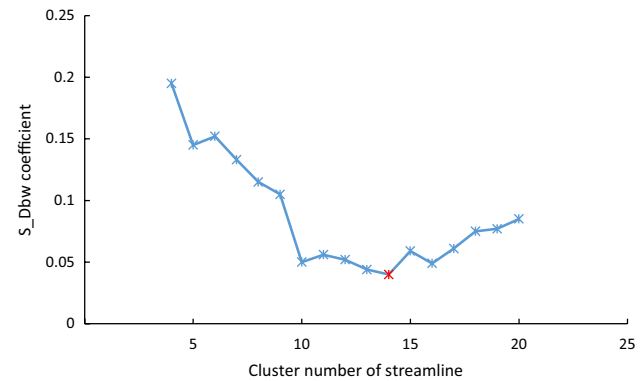
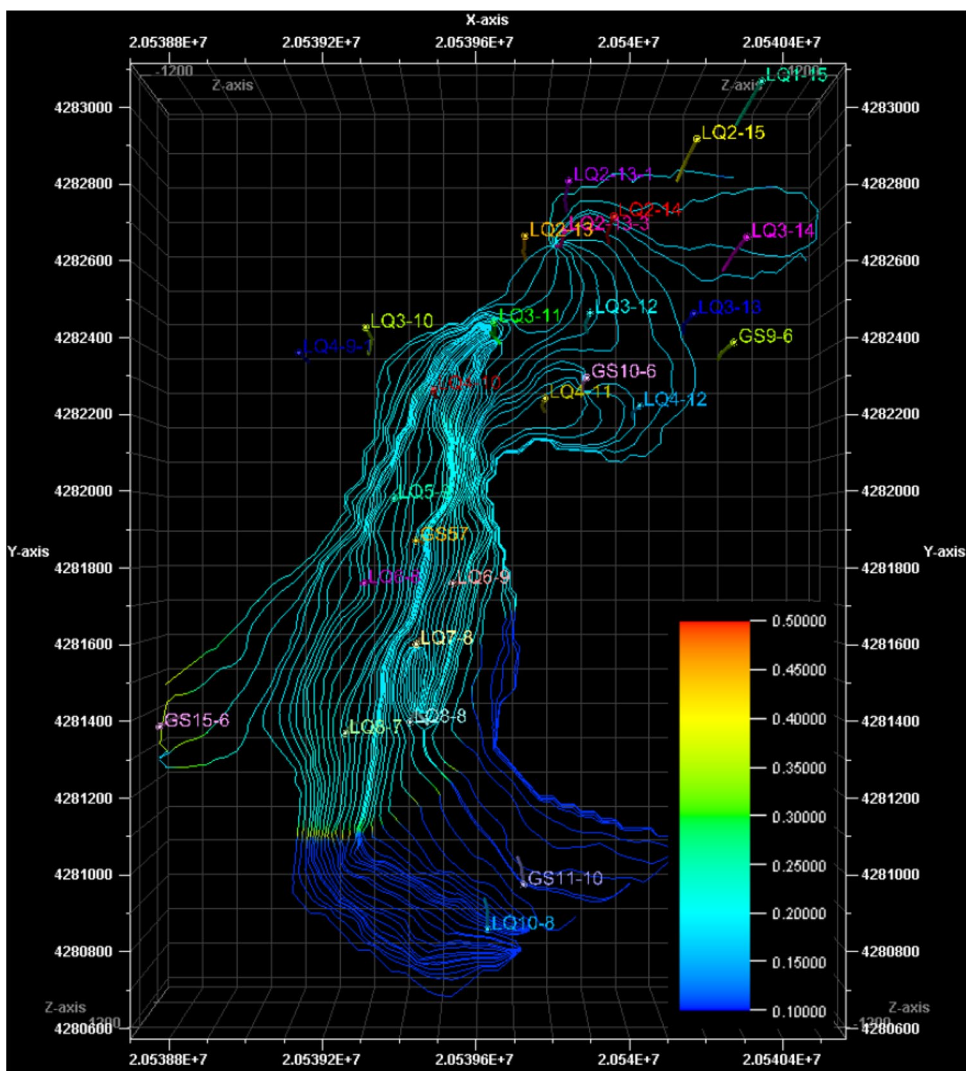


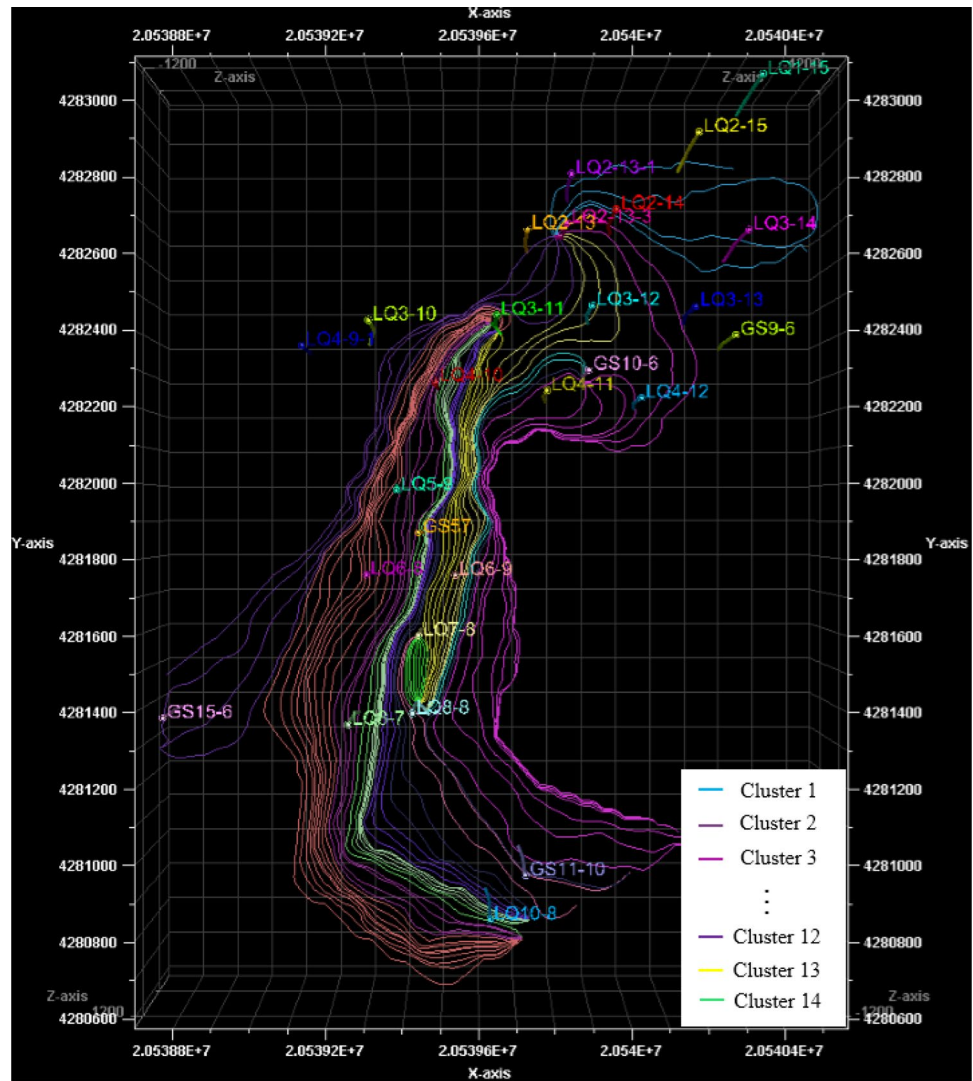
Fig. 6 Re between cluster number of streamline cluster and S_Dbw coefficient

Standardization of the streamline data

Because all kinds of streamline attributes have different data ranges, if they are not standardized, the weight of some attributes will be reduced or increased, which will affect the clustering results. After standardization of streamline coordinates and attributes, the distribution of various parameters can be reflected reasonably. In this study, the location of the X and Y coordinates at the middle point, the oil phase flow velocity and the water phase flow velocity of one single streamline were obtained, and the standard deviation was used to standardize the data. The method of obtaining each parameter is shown as follows:

X coordinates of the middle point position of the streamline:

Fig. 7 Gradational evaluation of Nm3-4-1 layer streamline field in the Gang Dong oil field



$$X_a = \sum_1^{n_a} x_i/n_a \tag{13}$$

Y coordinates of the middle point position of the streamline:

$$Y_a = \sum_1^{n_a} y_i/n_a \tag{14}$$

Average water phase flow velocity on a streamline:

$$V_o = \sum_1^{n_a} v_{oi}/n_a \tag{15}$$

Average oil phase flow velocity on a streamline:

$$V_w = \sum_1^{n_a} v_{wi}/n_a \tag{16}$$

In all the above expressions, X_a is the X-axis coordinates of the middle point position of a single streamline. x_i is the x coordinates of a single coordinate point on a streamline. n_a is the total number of data points on a single streamline. Y_a is the Y-axis coordinates of the middle point position of a single streamline. y_i is the y coordinates of a single coordinate point on a streamline. V_o is the average oil phase velocity of a single streamline. v_{oi} is the oil phase velocity of a single coordinate point on a streamline. V_w is the average water phase velocity of a single streamline. v_{wi} is the water phase velocity of a single coordinate point on a streamline.

Cluster evaluation of reservoir streamline field

We can evaluate the spatial coordinates and the oil–water flow ratio of the streamline field through the peak density algorithm mentioned in “The establishment of the

peak density algorithm model”. The optimal streamline clustering number is determined by the S_{Dbw} coefficient mentioned in “Determination of evaluation index for clustering effect”. On the basis of characteristic analysis of clustering results, the reservoir streamline gradational evaluation map is formed. The core content of this study is how to extract the effective information of the streamline field and form the best method of clustering and evaluation.

Example application

The Gang Dong oil field, as shown in Fig. 4, is located in the middle part of the Huang Hua depression. Its main oil-bearing formations are Minghua formation and Guantao formation. The average permeability is $800 \times 10^{-3} \mu\text{m}^2$, which corresponds to a typical medium high-permeability reservoir. To date, the comprehensive water cut has reached 95.68%, and the recovery degree has reached 43.18%. At present, there are serious problems such as invalid reservoir water circulation and difficulty in subsequent development. The distribution of streamline field is complex, and it is hard to improve oil recovery. In this paper, the applicability of the gradational optimization of the streamline field was discussed through the example of the Nm3-4-1 layer of the Gang Dong oil field.

Analysis of the distribution of the streamline field in the Nm3-4-1 layer

According to the streamline field corresponding to the oil saturation property of Fig. 5, it can be seen that the overall oil saturation of the Nm3-4-1 layer is close to the residual oil saturation. Only the edge of the reservoir has some areas with high oil saturation. For such low-potential reservoirs with high water cut, an important reason for the difficulty of tapping is that the distribution of the streamline field is complex and unclear. Therefore, it is necessary to apply the method of peak density clustering to optimize the streamline field of the reservoir.

Extraction and evaluation of the reservoir streamline field

Extracting the spatial coordinates, and the oil and water flow ratio of each position point in the Nm3-4-1 layer of the Gang Dong oil field as the parameters of the cluster analysis, after standardization, the parameters were

evaluated by the peak density clustering method, and 4–20 groups of streamline cluster number were formed. The S_{Dbw} coefficient was used to evaluate the clustering effect of different streamline cluster number, as shown in Fig. 6. With the increase of the streamline cluster number, the S_{Dbw} coefficient first decreased, indicating that the effect of clustering was gradually changed. When the number of cluster was 14, the S_{Dbw} coefficient was the smallest, and its value was 0.04, then the S_{Dbw} coefficient began to increase. So the best number of streamline cluster was 14 and we chose to divide the current streamline field of the Nm3-4-1 layer into 14 sets of streamline cluster.

Analysis of clustering characteristics of the streamline field

The Nm3-4-1 layer streamline field was divided into 14 types using the peak density clustering algorithm, as shown in Fig. 7. Different classifications represent the rank of streamline clusters, of which the first type of streamline clusters indicates that the oil–water flow ratio of the group is the largest and the development potential is the maximum. In contrast, the 14th class of streamline clusters represents the minimum oil–water flow ratio and the least development potential. Sorting the different streamline cluster regions, the smallest oil–water flow ratio area is distributed between Lq7-8 and Lq8-8 wells. In contrast, the largest oil–water flow ratio area is located near the northern Lq2-15 well area. The grade of streamline in the middle of the oil reservoir is obviously higher than the streamline at the edge of the reservoir.

Comparison of the different types of streamline attributes in the 14 clusters is shown in Table 2. In the first types of streamline clusters, the average oil saturation, the flow rate of oil phase fluid, and the oil–water flow ratio are much higher than that of the 14th type of streamline cluster. The above parameters decrease with the increase of streamline classification. The average water saturation and the flow rate of water phase fluid in different streamline clusters increase with the increase of streamline classification. We can accurately determine the optimal number of streamline clusters using the S_{Dbw} coefficient. Meanwhile, the streamline field evaluation by the peak density clustering algorithm can well reflect the current development status of the reservoir.

When a medium–high-permeability reservoir reaches the high water cut stage, the potential of residual oil in the layer is small, but this does not mean that there is no potential in the target layer. Through the evaluation of the streamline field, we can get the fluid flow path at the current time. At the same time, based on the streamline attribute clustering

Table 2 Streamline attribute statistics with different levels

Streamline cluster levels	Average oil saturation	Fluid flow rate in oil phase	Average water saturation	Fluid flow rate in water phase	Oil–water flow ratio
1	0.39	244.44	0.61	54.03	3.32
2	0.37	792.62	0.63	248.24	2.92
3	0.33	226.33	0.67	92.39	1.91
4	0.32	80.65	0.68	40.44	1.77
5	0.32	86.48	0.68	56.74	1.43
6	0.31	395.95	0.69	461.32	0.82
7	0.28	132.76	0.72	175.41	0.75
8	0.28	167.29	0.72	214.77	0.40
9	0.27	72.62	0.73	297.78	0.20
10	0.26	10.00	0.74	34.53	0.14
11	0.24	19.72	0.76	82.19	0.07
12	0.26	75.72	0.74	482.82	0.05
13	0.23	30.05	0.77	247.24	0.07
14	0.21	12.10	0.79	428.74	0.01

method, we can classify the streamlines on different flow trajectories, which can reflect the current flow state of the reservoir and determine the potential of different regions. All these provide the basis for the adjustment of the streamline field in the high water cut period.

Optimization of the reservoir streamline field

According to the current streamline field distribution of the Nm3-4-1 layer, the streamline field was adjusted to weaken the flow intensity between Lq7-8 and Lq8-8 wells, and enhance the flow intensity in the Lq2-15 well area. The other areas were adjusted according to the level of different streamline clusters. We varied the production and injection rates in each well and the equilibrium displacement was achieved according to the classification of the streamline field. The streamline field distribution of the adjustment scheme was predicted after 10 years, and the cluster evaluation and effectiveness verification were carried out again. The new streamline field was formed as shown in Fig. 8. After adjustment, the optimal clustering number for the streamline field was seven groups. Compared with 14 groups of streamline clusters before adjustment, the reduction of the number of new streamline clusters after adjustment meant that the internal differences of the whole streamline field were weakened after the new cluster evaluation and effect verification, and the displacement energy in each region was more balanced. By comparing the development indicators before and after adjustment in Table 3, the variation coefficient of oil–water mobility ratio was reduced by 0.33, the water

cut was reduced by 3.8%, the cumulative oil production was increased by 95,000 tons, the remaining oil saturation changed from 0.42 before adjustment to 0.35 after adjustment, and the development effect was improved significantly.

Conclusions

We used the peak density clustering algorithm to classify the streamline field and determined the best streamline cluster number by the S_Dbw coefficient. These methods take full account of the internal attributes of streamline data, reduce the human factors of qualitative evaluation of reservoir streamline field and determine the direction for remaining oil potential in the high water cut stage. The new streamline field grading method classifies the displacement energy according to the streamline spatial coordinates and the oil and water flow ratio, and clearly indicates the distribution position of the weak drive streamline and the strong drive streamline, making the streamline numerical simulation not only relying on the streamline density to determine the flow intensity qualitatively, but can reach the quantitative evaluation. The Nm3-4-1 layer of the Gang Dong oilfield is in the late stage of high water cut development. The distribution of the streamline field is complex. By adjusting the streamline field in time, the flow strength between Lq7-8 and Lq8-8 wells is decreased. The streamline field distribution is more uniform, and the development effect is improved.

Fig. 8 Gang Dong oilfield Nm3-4-1 layer streamline field gradational evaluation after adjustment

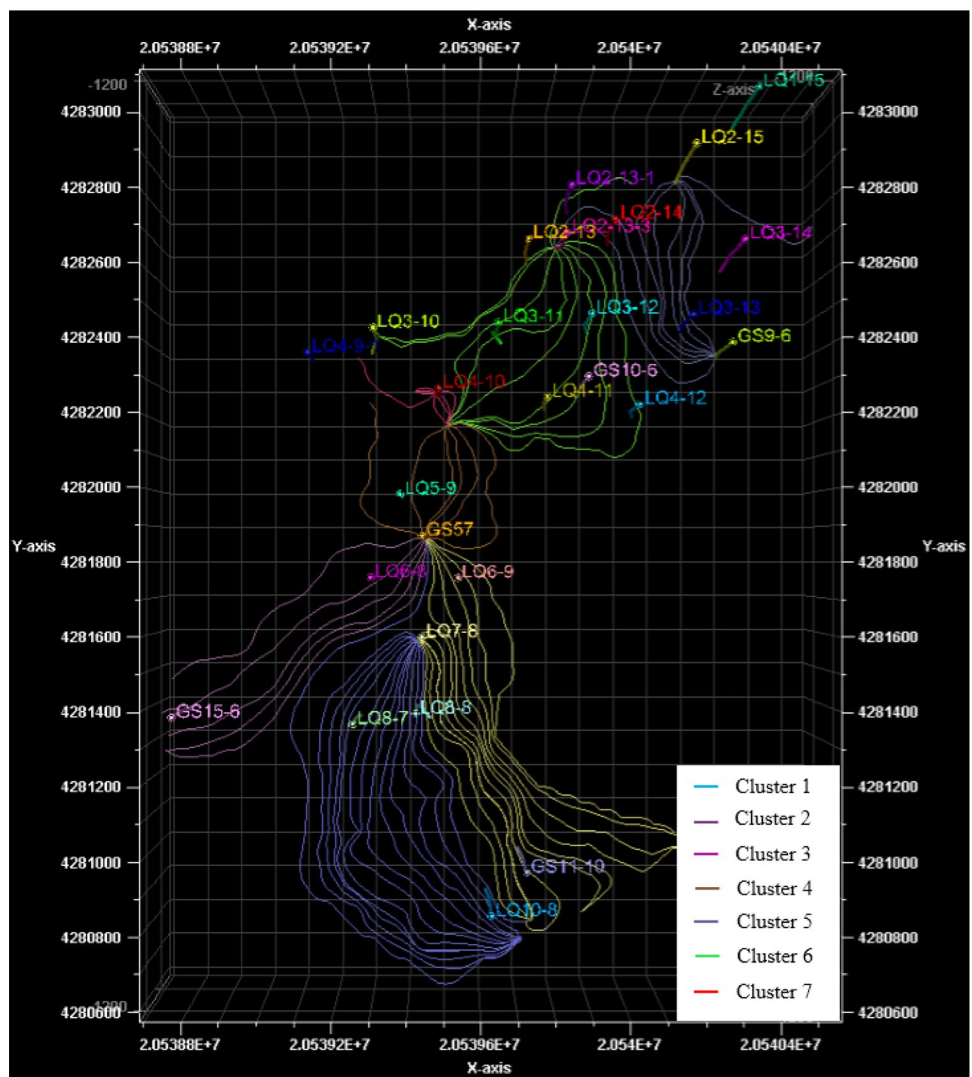


Table 3 Comparison of parameters before and after adjustment of the Nm3-4-1 layer streamline field

Water cut, %		Accumulation of the oil production, Million tons	
Before adjustment	After adjustment	Before adjustment	After adjustment
95.3	91.5	49.1	58.6
Residual oil saturation, decimal		Variation coefficient of the oil and water flow ratio, decimal	
Before adjustment	After adjustment	Before adjustment	After adjustment
0.42	0.35	0.96	0.63

Acknowledgements We gratefully acknowledge the Fundamental Research Funds for the National Science and Technology Major Project (2017ZX05009001); the National Natural Science Foundation of China (51674279). The funders had no conflict of interest or any role in the study design, data collection, and analysis, decision to publish, or preparation of the manuscript. The authors also would like to acknowledge the technical support of PETREL and ECLIPSE in this paper.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Agarwal PK, Guibas LJ, Edelsbrunner H (1999) Data clustering: a review. *ACM Comput Surv* 31(3):264–323
- Aliyarov RY, Ramazanov RA (2016) Prediction of multivariable properties of reservoir rocks by using fuzzy clustering. *Proc Comput Sci* 102:434–440
- Aminzadeh F, Tafti TA, Maity D (2013) An integrated methodology for sub-surface fracture characterization using microseismic data: a case study at the NW Geysers. *Comput Geosci* 54(54):39–49
- Batycky RP, Blunt MJ, Thiele MR (1997) A 3D field-scale streamline-based reservoir simulator. *Spe Reserv Eng* 12(4):246–254
- Caliński T, Harabasz J (1974) A dendrite method for cluster analysis. *Commun Stat* 3(1):1–27
- Cheng H, Oyerinde AS, Datta-Gupta A et al (2007) Compressible streamlines and three-phase history matching. In: SPE/DOE symposium on improved oil recovery, 22–26 April, Tulsa, Oklahoma, USA, SPE 99465, pp 121–127. <https://doi.org/10.2118/99465-MS>
- Dun JC (1974) Well-separated clusters and optimal fuzzy partitions. *J Cybern* 4(1):95–104
- Grabmeier J, Rudolph A (2002) Techniques of cluster algorithms in data mining. *Data Min Knowl Discov* 6(4):303–360
- Halkidi M, Vazirgiannis M, Batistakis Y (2000) Quality scheme assessment in the clustering process [M]//principles of data mining and knowledge discovery. Springer, Berlin, pp 265–276
- Hubert L, Arabie P (1985) Comparing partitions. *J Classif* 2(1):193–218
- Jackson AS, Pollock ML, Graves JE et al (1988) Reliability and validity of bioelectrical impedance in determining body composition. *J Appl Physiol* 64(2):529–534
- Marcucci M (1996) Applied multivariate techniques. *Technometrics* 39(1):100–101
- Maulik U, Bandyopadhyay S (2002) Performance evaluation of some clustering algorithms and validity indices. *IEEE Trans Pattern Anal Mach Intell* 24(12):1650–1654
- Rodriguez A, Laio A (2014) Clustering by fast search and find of density peaks. *Science* 344(6191):1492
- Rousseeuw PJ, Silhouettes (1987) A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20(20):53–65
- Tan F, Luo G, Wang D et al (2016) Evaluation of complex petroleum reservoirs based on data mining methods. *Comput Geosci* 2016:1–15
- Thiele MR, Batycky R (2003) Water injection optimization using a streamline-based workflow. In: SPE annual technical conference and exhibition, 5–8 October, Denver, Colorado, SPE 84080. <https://doi.org/10.2118/84080-MS>
- Wang WC (2002) Simulation-facilitated model for assessing cost correlations. *Comput Aided Civ Infrastruct Eng* 17(5):368–380
- Wang J, Shi C, Ji S et al (2017) New water drive characteristic curves at ultra-high water cut stage. *Pet Explor Dev* 44(6):955–960
- Zhao H, Kang Z, Sun H et al (2016) An interwell connectivity inversion model for waterflooded multilayer reservoirs. *Pet Explor Dev* 43(1):99–106

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.