Optimization of Gas Hydrate Well Placement Based on K-Means Cluster Analysis and Genetic Algorithm

Hao Fu †, Pengwan Yi †,*, Zheng Liang †, Lijia Xing

School of Information Science and Technology, Xizang University, Lhasa, China, 850000

*Corresponding author: yipengwan2023@stu.utibet.edu.cn

† These authors contributed equally.

Abstract. Natural gas hydrate, also known as "combustible ice", is an efficient and clean backup energy source. The purpose of this study is to conduct a comprehensive quantitative assessment of natural gas hydrate resources in a sea area, and to propose a site selection plan for additional wells. Firstly, this paper obtains the geological data of 14 exploration wells in this sea area, establishes a K-means clustering model, and derives the clustering of well data into three different regions, and the distribution range of natural gas hydrate is obtained by calculating the location of the center of each cluster. Secondly, combined with GIS tools, spatial interpolation analysis was carried out to generate the spatial distribution map of each parameter (effective thickness, porosity, hydrate saturation) in the study area. Finally, this study proposed to add five more wells in the region, and the optimal spatial coordinates of the five well locations were obtained by constructing a multi-objective optimization model using a genetic algorithm. This study creatively proposes a dual-stage optimization model integrating K-means clustering and genetic algorithms, which not only addresses the gap in comprehensive decision-making frameworks but also enhances exploration accuracy in complex marine environments and promotes the commercialization development process.

Keywords: Exploration wells, well placement optimization, K-means clustering, genetic algorithm.

1. Introduction

Natural gas hydrate (commonly known as "combustible ice") is regarded as a strategic resource for alleviating the energy crisis and realizing the low-carbon transition due to its large reserves and clean combustion properties [1]. In recent years, with the advancement of offshore exploration technology, the attention to its commercial exploitation has increased significantly worldwide. However, traditional well planning methods mostly rely on empirical judgment of single geological parameters (e.g., stratigraphic depth or rock type), which is prone to lead to problems such as low exploration accuracy and incomplete resource coverage in complex marine environments [2].

Gao Bin et al. [3] proposed a resource distribution prediction model based on Kriging interpolation, which can generate spatial distribution maps, but did not combine with cluster analysis to identify resource-rich areas, resulting in a lack of systematic layout of new wells; Li En et al. [4] used ISCA-BP neural network to construct the hydrate prediction model, but did not solve the problem of multi-objective optimization, which makes it difficult to balance the efficiency of resource development and economic costs. In addition, existing studies mostly focus on single technical means (e.g., statistical methods or spatial interpolation [5]), and have not yet formed a comprehensive decision-making framework that integrates multidimensional parameters and intelligent optimization algorithms [2, 6-7].

Based on this, this study proposes a two-step method of "cluster analysis + intelligent optimization" to optimize the layout of gas hydrate exploration wells. In this paper, K-means clustering algorithm, GIS spatial interpolation [8-10] and multi-objective genetic algorithm [2, 6] are integrated for the first time, integrating multi-dimensional parameters such as rock type, porosity, hydrate saturation and so on, and constructing a full-process decision-making framework of "resource zoning - dynamic optimization". The framework of "resource zoning - dynamic optimization" is constructed to systematically solve the problem of well location planning in complex marine environment.

2. Optimal well placement based on K-means clustering algorithm or genetic algorithm

2.1. Data sources

Geologic data, including but not limited to rock type, mineral content, and stratigraphic depth, are available for each of the exploration wells in this study. Specific data are available on the website: https://www.nmmcm.org.cn/notice_detail/332.

In order to conduct a more refined survey of the region, five additional wells are proposed in the region with the objective of determining the optimal well placement to maximize reserve estimates or achieve optimal resource development efficiency.

2.2. Optimal selection of well location based on K-means clustering algorithm

2.2.1 Modeling Principles

K-means clustering method was utilized to perform cluster analysis of the existing 14 exploration wells to determine the distribution characteristics of the resource concentration areas and well locations, so as to better arrange the optimal well locations after the proposed addition of 5 wells.

First, this research team already has geological data for each exploration well, including but not limited to rock type, mineral content, and stratigraphic depth.

Next, the Elbow Method is used to determine the K-value, i.e., the number of clusters. This method evaluates the relationship between the number of clusters and the total sum of squared errors (SSE), and selects an "elbow point", i.e., a point where the drop in SSE suddenly slows down, as the K-value.

Next, the preprocessed data is clustered using the K-means clustering algorithm. The execution flow of the algorithm is as follows:

Inputs: set of samples to be clustered $D = \{x_1, x_2, \cdots, x_m\}$, number of clusters$K$, maximum number of iterations n.

Output: clustered clusters $C = \{c_1, c_2, \cdots, c_k\}$

The steps are as follows:

Stepl: Randomly select$K$ clustering centers from the dataset D and let the corresponding vector be $(\mu_1, \mu_2, \cdots, \mu_k)$.

Step2: Calculate the distance of each sample point to the$K$ clustering center $\|x_i - \mu_k\|_2^2$, and group each sample point into the cluster closest to it.

Step3: Calculate the center vectors of K clusters obtained in step 2.

$$\mu_k = \frac{1}{M_k} \Sigma_{x_i \in C_k} x_i \tag{1}$$

Use it as a new clustering center.

Step4: Repeat Step2 and Step3 until the maximum number of iterations is satisfied n or the center vectors of all clusters no longer change, output the clusters that are well clustered $C = \{c_1, c_2, \cdots, c_k\}$.

2.2.2 Modeling and Solving

Cluster analysis is used to determine the location of the centers of each cluster and identify areas of resource concentration. The preprocessed data is brought into the above model and the following operations are performed through python:

For a dataset of n points, k from 1 to n is computed iteratively, and the sum of the squares of the distances from each point to the center of the cluster to which it belongs is computed each time clustering is completed; the sum of the squares becomes progressively smaller until it reaches 0 when k==n, because each point is the center of the cluster to which it belongs itself. In this sum of squares change process, there will be an inflection point also known as the "elbow" point, the rate of decline suddenly slowed down that is considered to be the optimal value of k. K-means clustering elbow rule results in the graph shown in Figure 1:

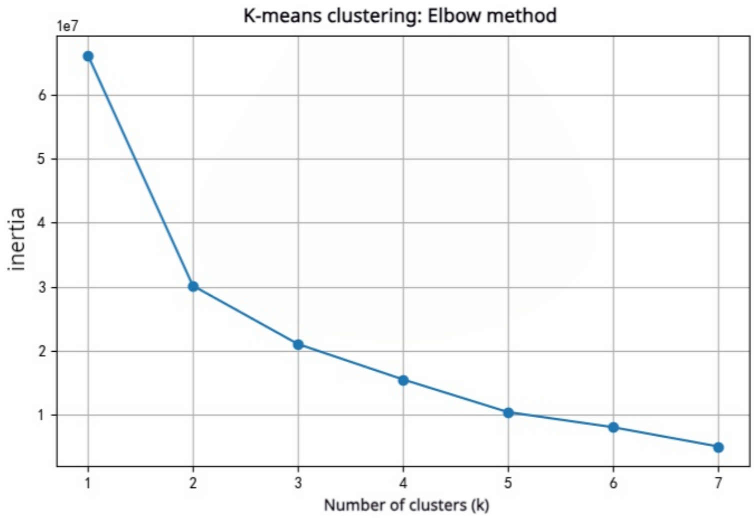

Figure 1. K-means clustering elbow rule result plot

From Figure 1, it can be observed that when k increases from 1 to 2, the SSE decreases very significantly. When k continues to increase, the decrease of SSE slows down gradually, especially after k=3, the decrease is much smaller. According to the elbow rule, the team chose the point where the decline of SSE slows down significantly as the optimal k value. In this figure, k=3 appears to be an appropriate choice because the decrease in SSE is no longer significant after this point. k=3 well clustering results are plotted in Figure 2:

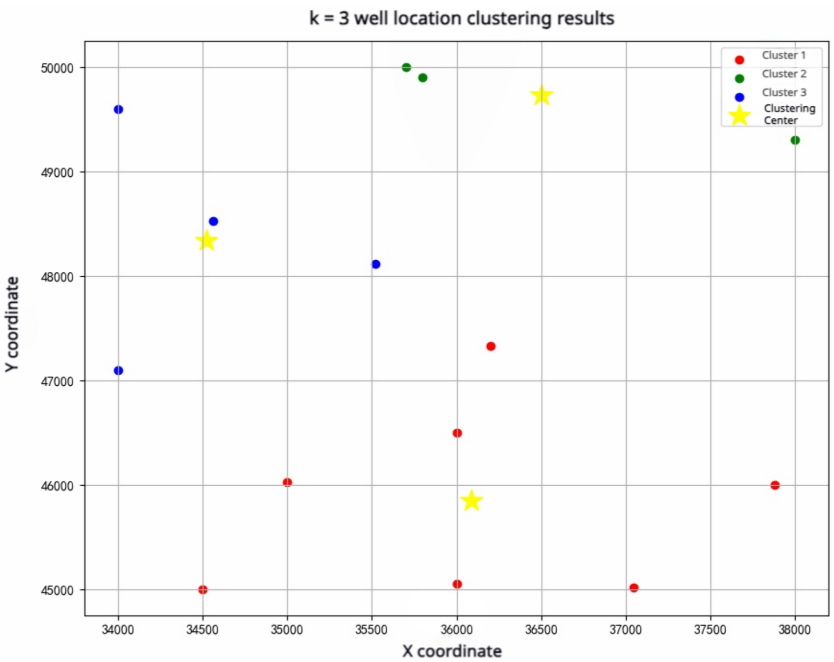

Figure 2. k=3 clustering results for well locations

Figure 2 shows how the data points are categorized into three clusters at k = 3. Each cluster is represented by a different color (red, green, and blue), and the center of mass of each cluster is marked by a yellow asterisk; the distribution of the clusters shows a relatively clear spatial separation, and the data points in each cluster are relatively concentrated in the two-dimensional space, and the clusters are farther apart from each other, which indicates that the clustering is more effective.

2.2.3 Analysis of results

The K-means clustering method was used to analyze the clustering of the 14 existing exploration wells and to determine the distribution characteristics of the wells and the areas of resource concentration. The clustering results show that the existing well locations can be divided into three main clustering areas, and the distribution of resources within each clustering area is relatively concentrated.

Based on the results of K-means clustering analysis, combined with GIS tools, this research team performed spatial interpolation analysis to generate spatial distribution maps of each parameter (effective thickness, porosity, hydrate saturation) in the study area. These spatial distribution maps helped this research team to identify potential high resource areas not covered by existing exploration wells.

Based on the spatial distribution map and the distribution of existing well locations, this research team proposes an optimal siting plan for five new wells. The specific coordinates of the new well locations are listed in Table 1.

These well locations were designed to be distributed in a circle around the center of the existing area to cover the entire area and fully explore the resource potential in different directions.

Table 1. Optimal siting coordinates for five additional wells

X	Y
38729.29	47391.43
36656.34	50244.60
33302.23	49154.78
33302.23	45628.07
36656.34	44538.26

2.3. Optimal selection of well location based on genetic algorithm modeling

2.3.1 Modeling Principles

Based on Genetic Algorithm (GA), multi-objective optimization is used to determine the location of new wells in conjunction with existing survey data in the region. This model will consider address data to predict potential resource distribution and find the most favorable drilling location.

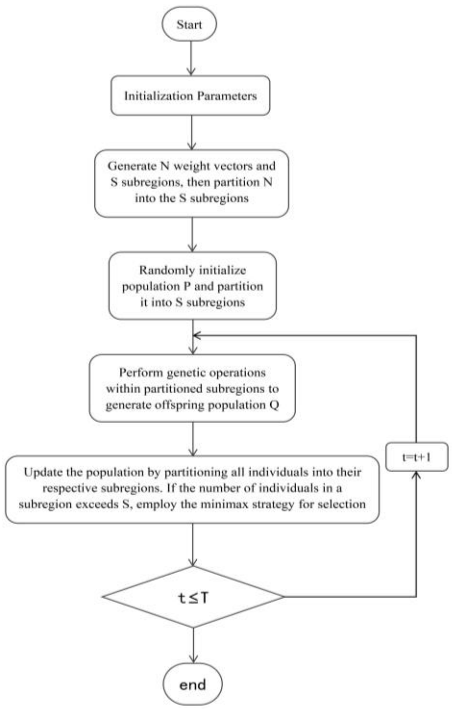

Figure 3. Flowchart of subregion-based genetic algorithm

This problem is solved using genetic algorithm for the described modeling problem. In order to guarantee the convergence and distribution of the multi-objective genetic algorithm while maintaining the diversity and distribution of the solution population, this research team applies a sub-region based evolutionary algorithm to cope with this multi-objective problem. The core idea of this algorithm is to subdivide the target region space into multiple subregions. Before updating the population, all contemporary individuals and their offspring are assigned to subpopulations in each subregion. If the number of individuals in a subregion does not exceed a preset upper limit U, all individuals are retained; if it exceeds U, the U best individuals are screened by a very large very small strategy to populate the subpopulations in that subregion. During the execution of the algorithm, the parent individuals of the hybridization operation must be selected from the same subregion. The specific execution process is shown in Figure 3.

The entire solution steps are organized as follows based on the above flowchart:

Step1: Initialize parameters. Initialize the population size N, the number of subregions S, the maximum number of iterations T, the probability of hybridization P_c and the probability of mutation P_m, generate S a vector of subregion centers $W^1, W^2, \ldots, W^S$, generate N a vector of weights $V = (v_1{}^i, v_2{}^j), (1,2,\ldots,N)$.

$$\mathrm{v} = \left(\mathrm{v}_1^{\mathrm{i}}, \mathrm{v}_2^{\mathrm{j}}\right), (1,2, \cdots \mathrm{N}) \tag{2}$$

Divide the N weight vectors into the S subregions.

Step2: Initialize the population. Adopt the design scheme of this paper to randomly generate $2N$ individuals, and use the above regional division method to divide the population into sub-populations ofS regions. If the number of individuals in a region does not exceed S, then all the individuals in this region are assigned to the sub-population of this region, if the number of individuals in a region exceeds S, then use the very large and very small selection operator to select S better individuals into the sub-population.

Step3: Perform the hybridization mutation operation on individuals from each region separately to ensure that both parents involved in the hybridization originate from the same region.

Step4: Update the population. The offspring produced after hybridization and mutation, as well as the individuals in all subregions of the current generation, are rationally divided into each corresponding region. If the number of individuals in a region does not exceed S, assign all the individuals in this region to the sub-population of that region, and if the number of individuals in a region exceeds S, select S better individuals into the sub-population by using the very large very small selection operator.

Step5: If $t \le T, t = t + 1$, return to step Step3, otherwise output non-inferior solution.

$$\mathrm{t} \le \mathrm{T}, \mathrm{t} = \mathrm{t} + 1 \tag{3}$$

2.3.2 Modeling and Solving

In order to solve the well location planning problem after the addition of five wells, a new drilling program is planned by using the optimal well locations determined by the genetic algorithm model in the following steps:

1) Initialize the population: At the beginning, a set of initial solutions are randomly generated that represent the possible locations of the drilling wells. These initial solutions constitute the first generation of the population.

2) Evaluating fitness: For each individual (drilling location), a fitness function is used to evaluate its degree of merit in the problem space. The fitness function can take into account a variety of factors such as distance from the target geologic formation, drilling costs, and expected output.

3) Selection operation: according to the evaluation results of the fitness function, select the better individual as the parent of the next generation. Commonly used selection methods include roulette selection and ranking selection.

4) Crossover operation: By crossover operation, the chromosomes of selected individual parents are crossed over to produce a new individual. The location and manner of crossover can be designed according to the characteristics of the specific problem.

5) Mutation operation: Introduce a certain degree of random variation in the individuals after crossover to increase the diversity of the population. The variation operation helps to jump out of the local optimal solution and makes the algorithm more global search ability.

6) Generate the next generation: Generate the next generation of the population through selection, crossover and mutation operations. Repeat this process until the stopping conditions are met (e.g., the maximum number of iterations is reached, the fitness reaches a threshold, etc.).

7) Convergence and result output: as the iteration proceeds, the individuals in the population gradually converge, and one or more better drilling location solutions are finally obtained. As needed, the optimal solution or a collection of optimal solutions can be output.

2.3.3 Analysis of results

In order to arrange the well location based on the addition of five wells, the distribution range data obtained from Problem 1 should first be spatially analyzed to identify the resource-rich and under-explored areas. Building on this foundation, the genetic algorithm model was implemented, and the obtained results are presented in Table 2.

Table 2. Coordinates and fitness values for the five additional well locations

x	y	fitness
37300	49000	0.983635997
37600	47900	0.839609769
37670	49850	0.88722716
36480	48610	0.952538086
38000	47130	0.929110263

The optimal well locations determined by the model are used to plan new drilling programs to ensure efficient resource development and maximize economic benefits. Through a combination of genetic algorithms and geological data analysis, this research team was able to scientifically guide the placement of new wells to optimize resource exploration and development. This approach not only improves the accuracy of exploration, but also provides strong decision support for future development activities.

3. Conclusion

In this paper, based on the address data of 14 exploration wells in this sea area, the distribution range of natural gas hydrate was obtained by establishing a K-means clustering model and calculating the location of each cluster center. The spatial distribution map of each parameter in the study area is derived by spatial interpolation analysis using GIS tools, and the optimal spatial coordinates of five well locations are obtained by constructing a multi-objective optimization model combined with genetic algorithm. There are two aspects that need to be improved in the research of this paper: the first one is the insufficient consideration of dynamic geological constraints, and the effective thickness, porosity and hydrate saturation are selected as the parameter basis for calculation in this paper, and the existing model does not fully integrate the dynamic evolution parameters of the reservoir, such as formation pressure gradient and porosity time-varying characteristics, which may lead to the well location optimization scheme facing the risk of stability in the long-term exploitation. The second is the algorithm convergence and accuracy limitation, K-means clustering is sensitive to the initial center, which may lead to deviation in the identification of resource-rich zones, and the genetic algorithm still has the risk of premature convergence under the high-dimensional solution space. This study will introduce a dynamic coupling model in the follow-up, which will incorporate a variety of dynamic parameters that need to be considered in the actual development into the fitness function of the genetic algorithm, in addition to the hybrid optimization of the algorithms used in this paper, in order to achieve the purpose of improving the accuracy of the assessment.

References

[1] QIN Ping, WANG Qingtao. Layout in advance to fill the gap of scientific and technological research and development to support the comprehensive utilization of combustible ice [J]. China Metrology, 2023, (11): 46-48.

[2] LIU Fan, CUI Jinfeng, WU Ming, et al. An automatic paper grouping method based on cognitive diagnosis and large model optimization genetic algorithm [J]. Experimental Technology and Management, 2025, 42 (04): 233-238.

[3] GAO Bin, ZHAO Scape, WANG Huajie, et al. Evaluation and prediction of natural gas hydrate distribution based on Kriging interpolation [J]. Technology and Market, 2024, 31 (12): 15-26.

[4] LI En, LIU Yun, WU Forest, et al. Natural gas hydrate prediction model based on ISCA-BP algorithm [J]. Chemical Engineering, 2022, 50 (08): 62-67.

[5] Li, H., Shao, Z.. A review of spatial interpolation analysis algorithms [J]. Computer System Applications, 2019, 28 (07): 1-8.

[6] ZHAO Leyi, ZHANG Meng, SUN Chengtian, et al. Fast optimization of reservoir well location based on hybrid genetic algorithm [J]. Petroleum Geology and Engineering, 2025, 39 (02): 89-94.

[7] YANG Junbang, ZHAO Chao. A research review of K-Means clustering algorithm [J]. Computer Engineering and Applications, 2019, 55 (23): 7-14+63.

[8] Zheng Yuefeng, Yu Siyuan, Wang Xinwei. Airport roadway planning and construction management based on BIM and GIS [J]. Architecture, 2025, (04): 120-122.

[9] LONG Wenjia, ZHANG Xiaofeng, ZHANG Chain. A business process clustering method based on k-means and elbow rule [J]. Journal of Jianghan University (Natural Science Edition), 2020, 48 (01): 81-90.

[10] He Xiansen, He Fan, Xu Li, et al. Determination of optimal number of clusters for K-Means algorithm [J]. Journal of University of Electronic Science and Technology, 2022, 51 (06): 904-912.