Study on the Site Selection of Gas Company Maintenance Duty Point in Random Uncertain Environment

Xianfeng Ding, Qian Yin, Kang Cen

Abstract—**With the development of economy, the utilization rate of natural gas in urban areas increases year by year, and so does the incidence of gas events. Therefore, it is very important to set the location of duty points reasonably. The selection of duty points is related to the distribution of emergency needs. However, most of the existing emergency facility site selection methods divide the entire area into grid forms and use GIS software for site selection, ignoring the rationality of regional division and the randomness of natural gas events in the distribution of the entire demand space. This paper uses Birch clustering algorithm for spatial clustering based on the consequence characteristics of natural gas events and relevant historical data. Considering the randomness of urban transportation and gas events, the opportunity constraint model of gas company duty point under uncertain environment is constructed, and the demand weight is determined combined with the pipeline hidden danger risk assessment level. The SA algorithm with global optimization ability and PSO with random search ability were used to solve the model under different risk coefficients, and the optimal position of the guard point was obtained. By comparing the experimental data with the mean value model and the P-median model, the model established in this paper is more suitable for the actual environment and has advantages over other models.**

Index Terms—**spatial regional division; Birch clustering algorithm; demand randomness; uncertain environment; demand weight; SAPSO**

I. Introduction

According to the 2022 BP Statistical Review of World Energy [1], global natural gas demand increased by 5.3%, returning to the level before the 2019 new crown epidemic, and China has surpassed Japan to become the world's largest importer of liquefied natural gas, accounting for nearly 60% of global liquefied natural gas demand growth, and is an important driving force for the growth of the global natural gas industry. Among them, urban natural gas is developing rapidly, and gas pipelines are all over the streets of the city, bringing great convenience to the life of urban residents. However, with the increase of pipeline service time, urban construction and traffic development, gas accidents due to corrosion, third-party damage and other factors of gas pipelines often occur. According to statistics [2], there were 3943 urban gas explosion accidents in China from 2016 to 2020, with 4424 people were injured and 458 died, which means that China has 33 accidents per hour. By the end of 2021, the total mileage of natural gas pipelines managed by the three major platform companies of Southwest Oil and Gas Field Company has reached 45,923.11 kilometers, and some pipelines have entered a stage of high incidence of accidents. In addition, geological disasters, corrosion, illegal occupation pressure, illegal construction, man-made destruction and other natural environment changes and human factors have led to accidents occur from time to time of natural gas pipeline leakage and rupture. The gas pipelines of the three platform companies failed 1276 times in 2021, with a failure rate of 28.38 times per thousand kilometers per year. In recent years, gas incidents have occurred frequently, and governments in various provinces and cities across the country have issued and improved the gas emergency plans for gas emergencies, and gas companies in each province and city have relatively complete early warning mechanisms for gas emergencies. However, the characteristics of gas leakage incidents, the perfect early warning mechanism cannot prevent all incidents. Once an accident occurs, the first emergency response is particularly important, as the location of the duty point determines whether the maintenance personnel can arrive at the site within a certain time.

At present, most studies on the site selection of emergency sites are based on graph theory, where each demand area is represented by a point on the network, and then classical location models or improved mathematical models based on them are used to locate facilities. The consequences often lead to large errors, uneven division may occur in the method of region division. For a large region, it is unrealistic to represent the coverage of the whole region only with the center point or the center of gravity point of the region, which will result in some demand points not being covered. Later, with the development of GIS system, researchers combined GIS to divide the whole demand area into small grids [3-5], greatly accurately locating the demand points, which is more suitable for selecting fire stations and emergency demand stations. But it is not suitable for the selection of gas duty points, as the occurrence of gas events is based on the distribution of its pipe network. If the whole area is divided into small grids, there may be situations where there are no pipelines distributed in some areas and it will be classified as gas emergency demand points, which will cause significant

Manuscript received May 11, 2023; revised September 13, 2023.

This work was supported in part by the State Key Laboratory of Oil and Gas Reservoir Geology and Exploration, and supported by the National Social Science Foundation of China (22XGL019).

Xianfeng Ding is an associate professor in the School of Science, Southwestern Petroleum University, Chengdu 610500, PR China. (e-mail: fxxd@163.com).

Qian Yin is a postgraduate student in the School of Science, Southwest Petroleum University, Chengdu 610500, PR China. (e-mail: yq1569262926@163.com).

Kang Cen is a professor in the School of Civil Engineering and Surveying, Southwest Petroleum University, Chengdu, 610500, PR China. (e-mail: cenkangxt@126.com).

errors. Some researchers combined the data mining technology [6-8], combining the clustering algorithms with facility location selection, and dividing regions when some attributes are similar, and using the center point of the region as the facility location point. Unfortunately, these methods did not fully consider the randomness on the demand space. Su [9] used Gaussian mixture clustering algorithm to describe the distribution of spatial demand, and decomposed a large number of spatial random demands into a mixture of multiple Gaussian distributions for approximation. Han [10] established a coverage model based on spatiotemporal big data set, combined demand with traffic status, and integrated multi-source data such as real-time traffic data and road network data obtained through the OPEN API of Gaud Maps into the spatiotemporal database. For facilities location with a large amount of data, this problem is appropriate. But, for the location of emergency maintenance duty point of gas demands, the actual data is not as much as the first aid data, and the occurrence of gas events is related to the distribution of pipeline. The collected data does not obey the Gaussian distribution, so the Gaussian mixture model cannot be used to describe the randomness of event occurrence.

In the selection of emergency facilities, commonly used methods include coverage problem, P-median problem, and P-center problem. These methods usually aim to minimize the number of facilities, minimize emergency time, and minimize the total rescue distance. However, the actual situation is always complex and accompanied by many uncertain situations, such as demand, response time, cost uncertainty and other factors, the above single model cannot be used to obtain results. Therefore, many mathematicians began to study the location of emergency facilities in uncertain environments. Ghaffarinasab [11] assumed that the transportation service demand is a Bernoulli random variable, so, the original random problem has been transformed into a deterministic problem, and verifies the effectiveness of the model in single and multiple allocation settings. Dong [12] considered waste sorting rate and house price fluctuation coefficient as uncertain parameters, and proposed a bi-objective robust optimization model to replan the waste incineration facilities' location. Sener [13] established a stochastic optimization model under three uncertain conditions, including demand, transportation costs, and their correlation, to solve the problem of multiple distribution hub coverage. Andaryan [14] considered the randomness of demand in a certain range and established a mixed integer Linear programming model with the goal of minimizing cost. In practical applications, there are always certain limitations in terms of time. Bozorgi-Amiri [15] established a multi-objective dynamic stochastic programming model by considering both demand and cost uncertainties, as well as the impact of uncertain travel time parameters on facility location and logistics allocation. Yang [16] developed a two-stage robust approximation model based on the optimization problem of container yard templates when the arrival time of ships is uncertain. Maghfiroh [17] combined location routing problem with time window constraints to establish a random location model under various time constraints. Yu [18] set different emergency times for different levels of risk areas, and built the location model of maritime rescue airbase considering time satisfaction. Wang [19] comprehensively considered road closures and regional risk classification under COVID-19 in 2019, and established a two-level emergency vehicle routing problem with time window allocation. With the development of fuzzy mathematics [20-21], some scholars have begun to consider using fuzzy numbers to represent uncertain factors. For example, Yan [22] used triangular fuzzy numbers to represent the uncertainty of demands, converted triangular fuzzy numbers into fixed numbers using a quantization method, and established a multi-objective optimization mathematical model. Esra [23] used the fuzzy linear programming model to select the most suitable location for incineration power stations in central Anatolia, Türkiye. Shang [24] studied the fuzzy hierarchical multimodal transportation hub positioning problem in the cargo delivery system, and analyzed the uncertainty of the travel time and loading and unloading time of the cargo delivery system. It is worth mentioning that uncertain factors may lead to changes in coverage. Therefore, Zhang [25] established an uncertain position set coverage model and a maximum coverage model, and explored the opportunity maximum coverage model under coverage constraint α and demand constraint confidence level β.

To sum up, scholars mainly studied other emergency facilities, such as fire stations and ambulance stations, and paid little attention to the location planning of maintenance duty points in gas accidents. In addition, scholars describe uncertain factors in two forms: one is focusing on the spatial randomness of demand while ignoring the emergency time and the risk degree of demand consequences; the other is focusing on the uncertainty of demands and time while ignoring the randomness of demand spatial distribution and the weight value of demands. Therefore, this paper comprehensively considers the spatial randomness of demand distribution and other uncertain information to select the location of gas duty points.

Previous studies on the influence of the randomness of demand space are all based on a large number of actual data, and Gaussian distribution is used to describe the randomness of demand space. However, for small cities, large-scale data is unrealistic. Since gas events are " radiation" to a certain extent, they will have a certain impact on the surrounding environment. This paper is based on the actual situation, combined with the formula for the impact radius of gas pipeline leakage, and uses the Birch algorithm to redefine the demand areas. Because the Birch clustering algorithm clusters under the constraints of equilibrium factors and spatial threshold parameters, effectively describing the " radiation" of gas events. At the same time, this article combines the Birch algorithm and K-means algorithm to better describe the characteristics of gas event demands.

Due to the complex urban transportation system, vehicle travel time is uncertain. Through investigation and analysis, it is found that the variation form of vehicle travel time in a day follows the Poisson distribution. Therefore, this paper combines the probability theory to construct a time constraint function. Assuming the speed of the vehicle is constant, convert time into a coverage probability function based on distance. This function not only solves the time uncertainty problem, but also describes the spatial distribution of the whole demand.

Due to the different forms of development after the occurrence of gas emergencies, they can be divided into emergency repair events and emergency rescue events [26]. The consequences caused by the two are different. Emergency repair incidents refer to natural gas leaks caused by damage to natural gas pipelines, which are controllable. After a rescue event occurs, it may cause personal injury and threaten social security. Therefore, this paper sets different hazard levels based on the consequences of natural gas accidents and assigns corresponding weights to different demand areas.

Finally, based on uncertainty theory, a corresponding "fair" opportunity constrained location model was established, and the optimal facility location under different coverage risk coefficients θ was discussed.

II. NOTATIONS AND BASIC ASSUMPTIONS

A. notation introduced for the models

TABLE I
NOTATIONS USED IN THIS PAPER

symbol	introductions
I	set of demand points (number of potential gas pipeline hazards)
A	set of demand area
A_i	set of demand points ($i \in A$)
J	set of all candidate locations
C	assume an ideal total coverage of the whole area
C_i	time satisfaction of demand area i
θ	risk parameter
N_i	$N_i = \{j \mid d_{ij} \le D\}$ potential collection of facilities within a certain range of services
d_{ij}	distance from demand point i or demand area A_i to the duty point j
D	standard distance for facility service
t_{ideal}	time under ideal condition (the time it takes for the facility point to reach the demand area except the rest time during the morning and evening rush hours)
t_{over}	the time it takes for the facility point to reach the demand area during the morning and evening rush hours
ω_i	the weight of demand region i
R	coverage radius, which is the service distance in response time
$\Delta\xi$	relaxation radius
P	the number of duty stations planned to be established
x_i	1 if demand point i or area A_i is covered at least one facility j ; 0 otherwise
y_j	1 if a facility is located at potential location j ; 0 otherwise
z_{ij}	1 if demand point i is served by j , 0 otherwise

B. Basic assumptions of the paper

(1) A demand area has only one duty point serving it;
(2) Each demand region is independent of each other;
(3) The demand in each demand area is independent from each other;
(4) Assume that the vehicle speed remains constant.

III. ESTABLISHMENT OF SITE SELECTION MODEL OF DUTY POINT

A. Requirement space division based on Birch clustering algorithm

Birch algorithm [27] is a tree structure from top to bottom, including two basic concepts: clustering feature (CF) and clustering feature tree (CF-tree) (Fig. 1).

(1) Clustering Feature (CF): cluster $\{a_i\}(i = 1, 2, ..., N)$ contains N D-dimensional data points, its clustering feature is a triplet $CF=(N,LS,SS)$, where N is the number of data points in the cluster; LS is the linear sum of the N data points, equivalent to $\sum_{i=1}^{N} a_i$, and SS is the sum of squares of the N data points, equivalent to $\sum_{i=1}^{N} a_i^2$.

(2) Clustering feature tree (CF-Tree): a clustering feature tree is a balanced tree, which stores the features of hierarchical clustering clusters. Each node represents a cluster, and each child node (child cluster) contains a CF entry. The form of CF-Tree is determined by three parameters: the non-leaf node branching factor (B) limits the maximum number of sub nodes of each non leaf node; the leaf node branching factor (L) limits the maximum number of sub clusters per leaf node; The maximum radius threshold (T) of the sub cluster limits the maximum radius of the sub cluster.

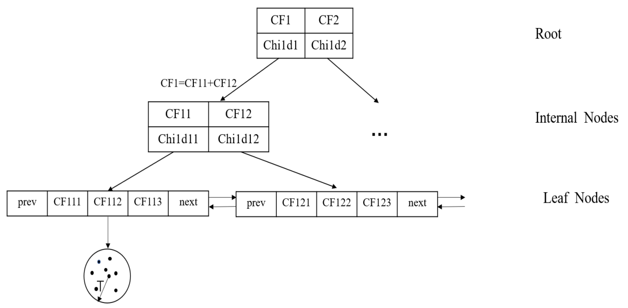

Fig. 1. CF-tree (B=2, L=3)

Birch clustering is the process of constructing a clustering feature tree with all samples, parameters B and T have great influence on the final clustering result, they determine the structure of CF-Tree. Since the selection of threshold is related to the distribution of data, the values of B and T are determined based on the actual situation in this paper. The initial threshold T is set for the first time according to the influence range of the gas accident, and then the optimal B is determined. Before using Birch and K-means algorithms to describe the spatial distribution of demands, this paper selects densely populated communities such as schools and hospitals as initial clustering centroids based on the identification criteria for high consequence areas. This not only reflects the characteristics of the problem itself but also improves the efficiency of the clustering algorithm.

B. Location Model

1) LSCP

Toregas [28] first proposed the location coverage model (LSCP) in 1971 in order to meet the requirement that facility points can provide services for demand points within a specified time or distance. Simply put, it means covering all demand points with the minimum number of facility points, the model expression is

$$\min \sum_{j \in J} y_j \tag{1}$$

$$s.t.\ \sum_{j\in N_i} y_j \geq 1, \forall i \in I \tag{2}$$

$$y_j \in \{0,1\}, \forall j \in J \tag{3}$$

Constraint (2) indicates that all demand points are within the service scope of at least one facility. The set coverage model lays more emphasis on the service "efficiency" of facility points to ensure that every point can be covered. In addition to considering the efficiency issue, it also needs to consider the issue of "fairness", so that all demand points can be provided with corresponding services by facility points under certain fair conditions. Consider the following P-median model for this purpose.

2) P-median

The P-Median model [29] optimizes the average "quality" of services when the "quantity" is guaranteed. Its mathematical expression is

$$\min \sum_{i\in I} \omega_i d_{ij} x_{ij} \tag{4}$$

$$s.t.\ \sum_{j\in J} x_{ij} = 1,\ \forall i \in I \tag{5}$$

$$x_{ij} \leq y_j, \forall i \in I, j \in J \tag{6}$$

$$\sum_{j\in J} y_j = P \tag{7}$$

$$x_{ij} \in \{0,1\}, \forall i \in I, j \in J \tag{8}$$

$$y_j \in \{0,1\}, \forall j \in J \tag{9}$$

Constraint (5) ensures that the demand point can only be covered by one facility point. Constraint (6) is the constraint of distribution rationality. Only when the duty point is established can the demand be allocated. Constraint (7) is to specify the number of facilities. "Fairness" and "benefit" are like fish and bear's paw, which cannot be both. Using the two models for site selection alone may be inconsistent with the reality. Therefore, this paper firstly uses the set coverage model to determine the minimum number of duty points to be established, and then, combined with the P-median model, chooses the facility location with the best average performance.

C. The model of location selection of duty points under uncertain factors

Combined with the characteristics of natural gas events, LSCP and P-medium all ignored the randomness of natural gas events and the uncertainty of response time. Therefore, based on the P-median model, this paper proposes the location problem of gas duty points under uncertain environment. The response time for different time periods varies depending on the urban traffic situation. Fig. 2 shows the fluctuation of travel time at a certain distance in a city within a day.

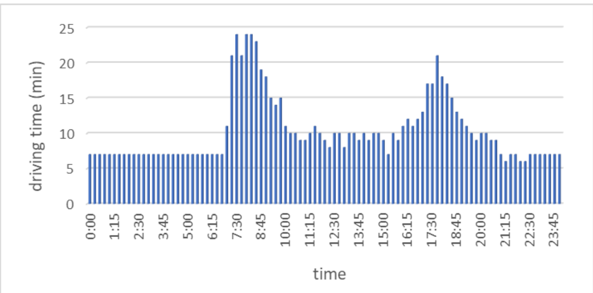

Fig. 2. Response time changes in different time periods

In Fig. 2, it may take a long time to arrive at another point during the morning peak (6:30-9:30) and evening peak (17:30-19:00) periods. The response time during the noon peak is relatively stable with small fluctuation, and the time difference is not obvious. Other time periods can be reached within about 10 minutes. Due to the strong correlation between arrival time and traffic conditions, the probability of natural gas accidents occurring at the location follows a Poisson distribution, and the specific parameter values of the distribution are related to time and location. A probability function needs to be established to describe the situation where a powerful maintenance personnel arrive at the rescue demand point within a certain period of time. Assume the relaxation interval of facility point coverage radius is $[R, R+\Delta\xi]$, and the probability of emergency repair personnel arriving at the demand area is q_{ij}, then the opportunity constrained location model is constructed. According to the requirements of gas companies, it is necessary to arrive at the scene within t_{ideal} minutes after the demand occurs. Assuming that the driving speed remains constant at v km/h, then $R = v \cdot t_{ideal}$, and the time taken by the driving in the morning and evening rush hours is t_{over}, then the value of $\Delta\xi$ can be $\Delta\xi = v \cdot (t_{over} - t_{ideal})$. When the distance between the duty point and the demand area is obviously less than the actual service capability upper limit R of the duty point, the demand area can be considered as covered. When the distance between the demand point and the upper limit of the service capacity of the duty point are less than $R+\Delta\xi$, it is necessary to determine the probability that the work point can cover the demand point.

$$G(d_{ij}) = \begin{cases} 1, d_{ij} \leq R \\ e^{-\lambda(d_{ij}-R)}, R < d_{ij} \leq R+\Delta\xi, \forall i \in A(i), j \in J \\ 0, d_{ij} > R+\Delta\xi \end{cases} \tag{10}$$

Since the randomness of events is also approximately subject to Poisson distribution, this function (10) can solve the probability coverage problem under the response time on the one hand, and also solve the error of the demand points represented by the demand center in the region division process on the other hand, thus better describing the spatial randomness of demands. Among them, $\Delta\xi$ is not only the relaxation of response time influenced by road traffic, but also the variation range of response time of demand points in the demand area.

The strong randomness of the gas events location indicates that the distance d_{ij} between the gas demand point i and the duty point j is uncertain, and the probability of random coverage is also random. So, this paper considers the whole demand space network structure, based on the probability function (10), establish an opportunity constrained location model in uncertain environments. and pre determines the threshold C to maximize the probability $\Phi(\sum_{i\in A}\sum_{j\in J} G(d_{ij}) \geq C)$ of the entire demand area being covered.

$$\max\ F = \Phi(\sum_{i\in A}\sum_{j\in J} G(d_{ij}) z_{ij} \geq C) \tag{11}$$

For the uncertain factors in the objective function, under the condition of achieving optimal average service quality, the objective function (11) is converted into the form (12), and the constraint (15) is added to obtain the probabilistic coverage model considering the randomness of demand space is

$$\min \sum_{j\in J}\sum_{i\in A}\omega_i d_{ij} z_{ij} \tag{12}$$

$$s.t. \quad \sum_{j\in J} z_{ij} \ge y_j, \quad \forall i \in I \tag{13}$$

$$\sum_{j} y_j \le P \tag{14}$$

$$\Phi(\sum_{j\in J}\sum_{i\in A} G(d_{ij}) z_{ij} \ge C) \ge \theta \tag{15}$$

$$z_{ij} = \begin{cases} 1, & G(d_{ij}) \ge C_i \\ 0, & \text{otherwise} \end{cases} \tag{16}$$

$$x_i, y_j \in \{0,1\}, \forall i \in A, j \in J \tag{17}$$

Constraint (12) is assuming that the facility point is at j, combined with function (10), for the covered demand area, if $z_{ij} = 1$, it indicates that the time to reach the demand point is within the satisfaction range C_i, then it means that the demand point is covered, otherwise it is 0. Due to the frequent occurrence of emergencies in reality, in order to enable decision-makers to better analyze the current situation, the site selection can be considered by setting the risk coefficient θ in constraint (15).

D. Assessment demands weight

Natural gas accidents have strong randomness and uncertainty, and theoretically, natural gas emergencies can occur at any location in the pipeline network distribution. In the actual calculation process, a large amount of data can make the final site selection decision very complex, and human factors are more obvious. Therefore, this paper evaluates the risk level of demand areas based on the spatial distribution of demand obtained from Birch clustering. According to the divided demand region, its center is taken to represent the region, and the central point of the region is expressed as the emergency demand point.

TABLE II
RISK LEVEL QUANTIFICATION TABLE

hazard analysis level	quantized value
low risk (minor hidden danger)	1
low risk (general hidden danger)	2
low risk (major hidden danger)	3
medium risk (minor hidden danger)	4
medium risk (general hidden danger)	5
medium risk (major hidden danger)	6
high risk (minor hidden danger)	7
high risk (general hidden danger)	8
high risk (major hidden danger)	9

According to the ABC three-level rating results of gas pipelines by gas companies, as well as other relevant information, pipeline hazards are divided into three types: high risk, medium risk, and low risk, each type is further divided into minor hazards, general hazards, and major hazards. dangers can be divided into three types: high risk, medium risk and low risk according to relevant information, and each type can be further divided into minor risk, general risk and major risk. The high consequence area refers to the area that will cause high harm when the gas accident occurs. The high-consequence area is further divided into first-level impact area, second-level impact area and third-level impact area. Therefore, this paper divides them into high risk (minor hidden danger), high risk (general hidden danger) and high risk (major hidden danger). The pipeline hazards are classified into 9 categories and quantified labels are set from low to high, as shown in Table II.

Regional risk level (SR) can be evaluated by the mean value of pipeline hidden danger risk assessment level $\overline{SR}$ of demand points in the region. Assuming that the demand area A_i contains n demand points, and the hidden danger risk level of sr_s demand point is $s(s \in A_i)$, then the hidden danger level of the whole demand area is

$$\overline{SR_i} = \frac{\sum_{s\in A_i} sr_s}{n} \tag{18}$$

The risk assessment value is normalized to obtain the corresponding weight value ω_i of the demand region

$$\omega_i = \frac{\overline{SR_i} - \min\left(\overline{SR_i}\right)}{\max\left(\overline{SR_i}\right) - \min\left(\overline{SR_i}\right)} \tag{19}$$

E. SAPSO

Based on the characteristics of gas events, this paper adopts the random simulation technology of particle swarm optimization (PSO) algorithm to deal with opportunity constraints in uncertain environments. The PSO algorithm is integrated with simulated annealing (SA) algorithm, which not only maintains the simplicity and easy implementation of PSO algorithm, but also maintains the probability jump ability of SA in the search process, effectively avoiding falling into the local minimum solution in the search process

The PSO algorithm, jointly proposed by Kennedy and Eberhart [30], is a random search algorithm based on group cooperation. Its speed and position update formulas are shown as follows:

$$\begin{aligned} v_{i,j}(k+1) = \chi[&v_{i,j}(k) + c_1 r_1 (p_{i,j}(k) - x_{i,j}(k)) \\ &+ c_2 r_2 (p_{g,j}(k) - x_{i,j}(k))] \end{aligned} \tag{20}$$

$$x_{i,j}(k+1) = x_{i,j}(k) + v_{i,j}(k+1) \quad (j = 1, \cdots, n) \tag{21}$$

Where, the compression factor $\chi = \dfrac{2}{|2 - C - \sqrt{C^2 - 4C}\,|}$, $C = c_1 + c_2, C > 4$, since the optimal position of the population is adopted in the velocity update formula (20), all particles will fly to the optimal position of the population. If the optimal position of the population is in the local minimum, all particles will tend to the local minimum solution. In order to improve the algorithm's ability to avoid falling into the local minimum solution, a position is selected from many p_i, which is denoting $\hat{p}_g$. Instead of p_g in equation (20), there is

$$\begin{aligned} v_{i,j}(k+1) = \chi[&v_{i,j}(k) + c_1 r_1 (p_{i,j}(k) - x_{i,j}(k)) \\ &+ c_2 r_2 (p'_{i,j}(k) - x_{i,j}(k))] \end{aligned} \tag{22}$$

p_i that performs well should be given a higher probability of selection. Using the mechanism of SA algorithm, we consider that p_i is a particular solution worse than p_g. We can calculate the jump probability $e^{-(f_{p1}-f_{pg})/t}$ of p_i relative to p_g at temperature t, where f is the value of the objective function. If the jump probability value is considered a suitable configuration of p_i, p_i is used instead of p_g probability. The jump probability can be calculated by equation (23)

$$e^{-(f_{p1}-f_{pg})/t} / \sum_{j=1}^{N} e^{-(f_{p1}-f_{pg})/t} \tag{23}$$

Where N is the population size.

The solution steps are as follows:

Step1: Randomly generate a population containing N particles, and randomly initialize the position and speed of particles;

Step2: Calculate the fitness of each particle i, record the position p_{best} of each particle and the global optimal position GB;

Step3: Conduct SA domain search for particles, and update the population according to formula (23);

Step4: Use roulette losing strategy to determine a globally optimal alternative value p'_g from all p_i;

Step5: Calculate the fitness value of each particle, update the optimal position p_{best} of each particle and the optimal position p_{best} of the population;

Step6: Perform annealing operation to determine whether the termination conditions are met. If so, stop searching and output results; otherwise, go to Step4.

This paper selects appropriate parameters to ensure the convergence of the PSO algorithm while removing speed boundary restrictions. This algorithm achieves good results in solving practical problems.

IV. INSTANCE

A. Description of the location problem of the duty point of the gas company

This experiment selects the main urban area of a city with a dense road network and good traffic conditions. When a gas event occurs, the emergency maintenance personnel are required to arrive at the scene within 25 minutes, and the average driving speed is about 45km/h. At present, the entire experiment is divided into three subregions, with three companies managing one area each, which has led to some shortcomings in emergency management: (1) Macro-control cannot be realized in emergency rescue. For example, in a gas accident that requires emergency rescue, nearby rescue personnel may not be able to carry out rescue work due to being outside the management scope of their own company, which may result in serious loss of life and property. (2) The independent operation of each company makes it difficult to share the facility resources, and the repeated purchase of facility resources leads to serious waste. Especially among the management bodies of adjacent gas companies, multiple operating companies coexist but are dispersed, which affects the efficiency of maintenance and also causes unnecessary economic losses. (3) The setting of multiple duty points and the excessive number of management personnel have resulted in long-term high management costs and difficulty in improving management efficiency. These drawbacks seriously reduce the management efficiency of each company and increase unnecessary operating costs.

Therefore, this paper considers regional integrated management without considering the corresponding management area of each company. For the whole area, according to the actual situation, the probability of multiple gas accidents occurring at the same time is basically small, so this paper considers the spatial area rather than individual demand point.

B. BIRCH clustering model

1) Determination of cluster radius threshold T value

The formula of potential influence radius of gas pipeline leakage is:

$$r = 0.099d\sqrt{p} \tag{24}$$

Where, d is the outer diameter of the pipeline (mm); p is the maximum allowable pressure (MAOP or MPa); r is the radius (m) of the affected area. Different pipelines, pipeline diameter is not uniform, pipeline pressure is also different, according to the relevant information provided by a city gas company, in the city, gas pipeline diameter by 32mm, 38mm, 57mm, 89mm, 108mm, 159mm and other 6 kinds. Pressure levels include low pressure (⩽ 0.4MPa), medium pressure (0.4MPa) and high pressure (⩾0.4MPa). Different outer diameters and pipeline pressures affect the radius of the affected area. As the distribution of pipelines is very complex, there may be multiple types of pipelines in the same area. In this paper, the average value $\overline{d} = 80.5\text{mm}$ of the external diameter of pipelines and the pressure $p = 0.4\text{MPa}$ are taken to calculate the influence radius of pipeline leakage, and then $r \approx 5\text{m}$. That is, after a pipeline leaks, the area within a radius of 5m can be considered as a rescue area.

According to the requirements of the gas company, the identification standard of high consequence area is shown in Table III.

TABLE III
IDENTIFICATION CRITERIA FOR HIGH-RISK AREAS

	Identification Criteria
city gas pipeline	The fourth level areas that the pipeline passes through (areas with high concentration of buildings, heavy traffic, and dense underground facilities on four floors and above (excluding basements))
	The third level areas that the pipeline passes through (sections with 100 or more households, including suburban residential areas, commercial areas, industrial areas, development areas, and densely populated areas that do not meet the conditions of the fourth level area)
	There are hospitals, schools, nurseries, sanatoriums, prisons, clinics and other building areas within 200m on both sides of the pipeline that are difficult to evacuate people
	There are areas within 200m on both sides of the pipeline where 20 or more people gather for at least 50 days within a year. For example, markets, temples, sports fields, squares, entertainment and leisure areas, theaters, campsites, village committees, quarries, brick and tile factories, tea shops, and other areas.

In Table III, the densely populated area within 200m of the pipeline is set as the high-consequence area. Considering the general consequence area and the influence range after gas leakage, assuming that the whole area is covered by a circle, a certain point is considered as the center of the circle and the area with a radius of $r' = 100r$ m is divided into one demand area.

In this paper, the demand point is located according to its longitude and latitude coordinates, and using latitude and longitude to calculate distance can result in significant deviation, so the actual distance (r') is converted into the longitude and latitude distance (d'). Supposing that the latitude and longitude coordinates of the two demand points are (X_1, Y_1) and (X_2, Y_2) respectively, then the actual distance between the two points is

$$D = R_{\text{earth}} \cdot ar\cos[\cos(Y_1)\cdot\cos(Y_2)\cdot\cos(X_1 - X_2) + \sin(Y_1)\cdot\sin(Y_2)] \quad (25)$$

Where R_{earth} represents the radius of the earth ($R_{\text{earth}} = 6371.0\text{km}$), and the distance between latitude and longitude is calculated by the distance of Manhattan.

$$d' = |(X_1 - X_2) + (Y_1 - Y_2)| \quad (26)$$

Combined with formula (25) and formula (26), the actual distance is two points of 500m, and the distance between latitude and longitude is about 0.0045 degrees. Therefore, the maximum cluster radius threshold $T = 0.0045$.

2) Determination of Branching factor B value

After obtaining the initial threshold *T*, determine the value of *B* and collect the relevant data of gas demands. When $T = 0.0045$ remains unchanged, the value of *B* starts from 2 and increases continuously until *B* reaches 50. This paper use the Calinski-Harabasz index to evaluate the quality of model parameters and clustering effectiveness. The higher the score of Calinski-Harabasz, the better the clustering model. Table IV shows the specific situation.

TABLE IV
BIRCH PARAMETER VALUES

node	branch factor	Calinski-Harabasz index	node	branch factor	Calinski-Harabasz index
15	33	438.1038	26	8	413.9375
16	21	438.6304	27	6	416.1803
17	21	456.9277	28	21	415.2652
18	22	454.6679	29	21	445.3929
19	21	449.5691	30	18	452.066
20	6	448.1119	31	18	448.2805
21	6	434.007	32	6	416.1803
22	14	431.8096	33	6	448.214
23	21	436.4823	34	15	449.5103
24	2	429.6466	35	15	464.4048
25	8	428.2124			

In Tabel VI, when $T = 0.0045$, the number of clusters is 17 and 35, the branching factor is 21 and 15, respectively, the variance criterion is larger than that of other clusters, and the specific value changes are shown in Fig. 3.

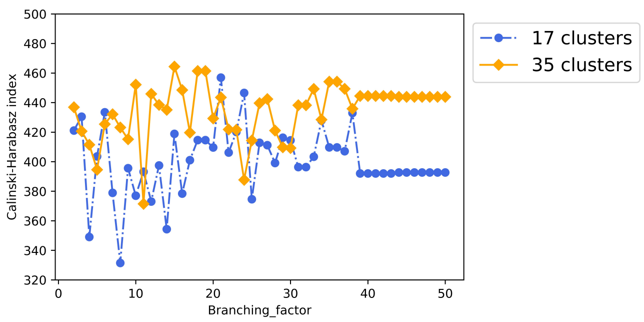

Fig. 3. Variance criterion coefficients under different branching factors

In Fig. 3, the variance criterion coefficients of 35 clustering results are basically greater than those of 17 clustering results, indicating that the clustering results are relatively good. When *B* is 15, the Calinski-Harabasz index is the highest, at 456.9277. Obviously, when the value of *B* is within [39, 50], the Calinski-Harabasz index remains relatively stable regardless of the number of clusters. Therefore, when *B*=39 and *T*=0.0045, by comparing and judging the Calinski-Harabasz index, the optimal number of clusters is selected as 35 (*N*=35), as shown in Fig. 4. And the final clustering results is shown in Fig. 5.

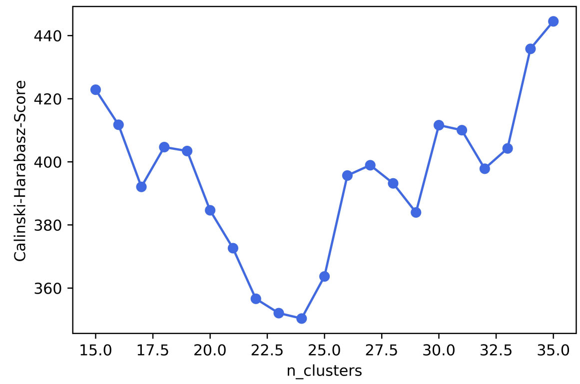

Fig. 4. Variance criterion coefficient of different cluster numbers with branching factor 39

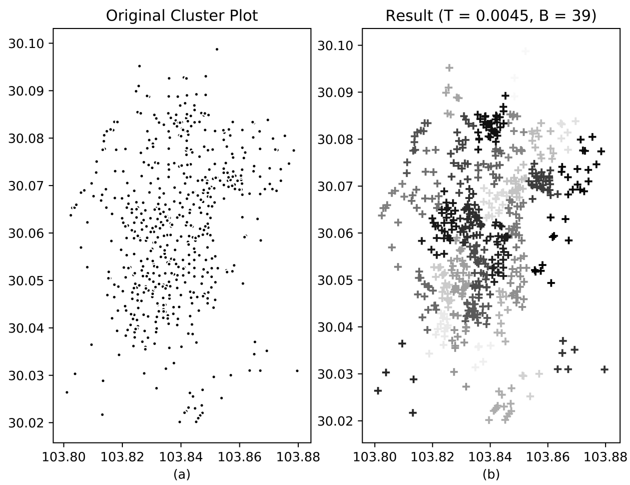

Fig. 5. Clustering result
Fig (a) on the left represents the distribution of all demand regions, and Fig (b) on the right represents the 35 divided regions.

C. Requirement area weight value

To calculate the response risk levels for different demand areas using formulas (18) and (19) based on the impact range after the occurrence of a gas event, as shown in Table V.

In order to simplify the calculation, if $\omega_i > 0.7$ is the high-risk area, if $0.4 < \omega_i \le 0.7$ is the medium risk area, and if $\omega_i \le 0.4$ is the low-risk area. Table VI shows the ideal arrival times corresponding to each risk level area. Fig. 6 shows the distribution of the entire demand area.

TABLE V
WEIGHT VALUE OF DEMAND AREAS

demand regions	A_1	A_2	A_3	A_4	A_5	A_6	A_7
$\overline{SR}$	8.565	4.235	1.182	7.167	7.515	1.400	8.781
ω	0.946	0.404	0.023	0.771	0.814	0.050	0.973
demand regions	A_8	A_9	A_{10}	A_{11}	A_{12}	A_{13}	A_{14}
$\overline{SR}$	2.077	8.629	5.000	5.538	7.786	5.263	8.600
ω	0.135	0.954	0.500	0.567	0.848	0.533	0.950
demand regions	A_{15}	A_{16}	A_{17}	A_{18}	A_{19}	A_{20}	A_{21}
$\overline{SR}$	8.636	4.667	8.000	1.182	2.667	1.200	7.867
ω	0.955	0.458	0.875	0.023	0.208	0.025	0.858
demand regions	A_{22}	A_{23}	A_{24}	A_{25}	A_{26}	A_{27}	A_{28}
$\overline{SR}$	5.143	7.364	1.905	4.800	1.400	5.000	7.214
ω	0.518	0.795	0.113	0.475	0.050	0.500	0.777
demand regions	A_{29}	A_{30}	A_{31}	A_{32}	A_{33}	A_{34}	A_{35}
$\overline{SR}$	5.762	6.000	2.214	2.111	3.500	1.333	1.667
ω	0.595	0.625	0.152	0.139	0.313	0.042	0.083

TABLE VI
DEMAND REGION GRADE DIVISION RESULTS AND DRIVING TIME SATISFACYION TABLE

risk grade	Number of regions	$[t_{ideal}, t_{over}]$
high-risk areas	13	[20min，25min]
medium risk areas	10	[15min，20min]
low risk areas	12	[10min，15min]

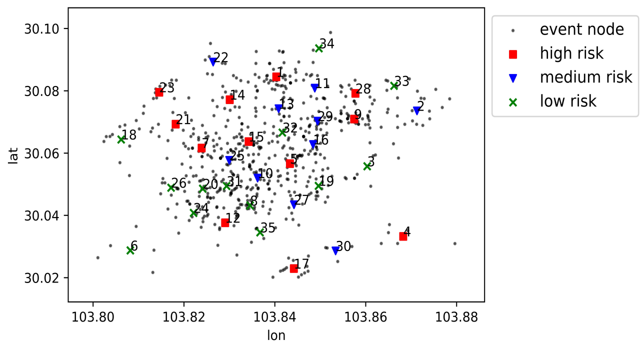

Fig. 6. Demand point distribution diagram

D. Interpretation of result

This paper comprehensively considers the spatial randomness description of requirements and the study of uncertain information. In order to better demonstrate the practicality of the location selection model established in this paper in uncertain environments, it is compared and analyzed with the P-median model and the mean model (MLP). The mean model uses mean to describe the uncertainty of distance and time, and treats the risk level of demand areas as medium risk. So, the uncertainty constraint (15) in the opportunity constrained location model (OCLP) is transformed into

$$\sum_{j \in J} \sum_{i \in A} G(\overline{d_{ij}}) z_{ij} \geq C \tag{27}$$

SAPSO algorithm was used to solve the three models separately. The population size was set as 35, the annealing rate as 0.95, and the number of terminated iterations as 200. Table VII shows the experimental results.

TABLE VII
COMPARISON OF SELECTED VALUES OF DIFFERENT MODELS

models	P	Locations	optimal value
P-median	1	32	46.40984
	2	21，28	31.90845
	3	9，21，28	25.74087
MLP	1	--	--
	2	--	--
	3	10，14，17	17.04700
OCLP	1	--	--
	2	16，20	11.92231
	3	3，18，20	10.83160

In Table VII, when the number of selected duty points is 3, the opportunity constraint model established in this paper obtains an optimal value of 10.83160, which is smaller than the optimal value obtained by the P-median model and the mean model, indicating that the location model established in this paper is superior to the two location models mentioned above.

In order to reduce the influence of other uncertain factors, different sizes of risk coefficient θ are set. Due to the large error between latitude and longitude distance and the actual distance, the Miller projection is used to convert latitude and longitude into actual coordinates, and the Manhattan distance is used to solve the problem. In the absence of a better location layout scheme, this experiment first obtains the optimal number of duty points to be selected through the location coverage model. Through calculation, when the number of construction duty points is 3, the global optimal solution can be achieved, as shown in Fig. 7. Therefore, this paper only considers situations where the number of duty points is 1, 2, and 3, the final site selection results obtained by setting different risk coefficients θ are shown in Table VIII.

TABLE VIII
SITE SELECTION UNDER DIFFERENT RISK GRADE COEFFICIENT AND NUMBER OF GUARD POINTS

θ	P	Locations	optimal value
0.1	1	--	--
	2	16，20	11.92231
	3	3，18，20	10.83160
0.2	1	--	--
	2	16，20	8.026620
	3	3，18，20	1.831600

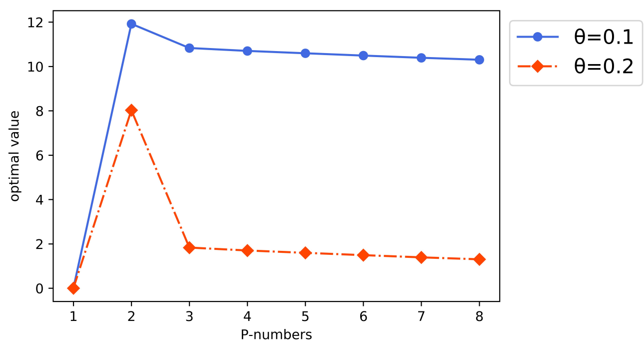

Fig. 7. Optimum value of the model corresponding to different number of site selection

In Table VIII, it can be seen that when the set number of work points is 1 and the risk parameter value is 0.1 or 0.2, OCLP model does not have the optimal solution. In addition, regardless of whether the risk parameter is 0.1 or 0.2, when the number of working points is determined, except for the optimal value of the objective function, the position of the working points remains unchanged. This indicates that the location of the site is tending towards stability. In Fig. 7, it can be seen that when the number of duty points is greater than or equal to 3, regardless of the value of the risk coefficient, the optimal value tends to be consistent, indicating that the optimal value for site selection converges, and the number of optimal duty points is 3.

V. CONCLUSION

This paper studies the location selection of maintenance duty points in gas companies, and proposes corresponding solutions for the spatial randomness of gas events and the uncertainty of response events. Regarding the spatial randomness of gas events, starting from the actual situation and starting from the "radiation" of gas events, spatial clustering is performed by setting two important parameters of the Birch algorithm, and K-means algorithm is applied to describe the distribution characteristics of gas event demand. On this basis, from the urban traffic conditions, combined with probability theory, time uncertainty is analyzed, and a chance constrained planning model based on spatial randomness and time uncertainty is established. According to the distribution of pipelines and the surrounding environment, the demand areas are divided into three risk levels: high, medium, and low, with each level corresponding to different response time requirements. Assign different weights to different risk level requirement areas to ensure the most effective protection for different risk areas as much as possible. In order to reduce the impact of other uncertain factors, a risk coefficient was set to discuss the site selection situation under different risk coefficients. The SA algorithm with global optimization ability and PSO algorithm with random search ability are used to solve the models under different risk coefficients, and the best protection point position is obtained. It is found that the final location of the duty points are the same, indicating that the model has a certain degree of stability. The OCLP model was compared with the P-median model and the mean model to verify the feasibility and practicality of the OCLP model. The experimental results indicate that the opportunity constrained model established in this paper is more suitable for solving the problem of selecting the working point location of gas enterprises. It is worth pointing out that this paper did not consider the impact of human resources, materials, and construction costs on site selection optimization layout, which can be further studied in the future.

REFERENCES

[1] BP. *BP Statistical Review of World Energy 2022. London, June, 2022.*

[2] Tian, B., et al., *Statistics and Regulations Analysis of urban Gas explosion Accidents from 2016 to 2020.* J. Journal of Safety and Environment. p. 1-9.

[3] Chen, Y.-C., et al., *An Analysis of the Optimal Facility Location of Tourism Industry in Plain Region by Utilizing GIS.* Sage Open, 2022. **12**(2).

[4] Lei, T.L., *Integrating GIS and location modeling: A relational approach.* Transactions in Gis, 2021. **25**(4): p. 1693-1715.

[5] Ulutas, A. and C.B. Karakus, *Location selection for a textile manufacturing facility with GIS based on hybrid MCDM approach.* Industria Textila, 2021. **72**(2): p. 126-132.

[6] Ping, W., C. Xianjun, and Z. Xuebin, *Research on Location of Logistics Distribution Center Based on K-Means Clustering Algorithm.* J. Security and Communication Networks. 2022. **2022**.

[7] Qing, L., et al., *Application of Clustering Algorithms in the Location of Electric Taxi Charging Stations.* J. Sustainability. 2022. **14**(13).

[8] Chen, J., et al., *Base Station Planning Based on Region Division and Mean Shift Clustering.* Mathematics, 2023. **11**(8).

[9] Su, Q., Yang,W. and Wang,Q.G, *Ambulance location planning considering the spatial randomness of demand. China.* J. Manag. Sci. 2019. 27(10): p. 110-119.

[10] Han, B., et al., *Site Selection of Fire Stations in Large Cities Based on Actual Spatiotemporal Demands: A Case Study of Nanjing City*. J. ISPRS International Journal of Geo-Information. 2021. Vol.7(No.8): p. 542.

[11] Ghaffarinasab, N., *Stochastic hub location problems with Bernoulli demands.* Computers & Operations Research, 2022. **145**.

[12] Dong, J., et al., *Overcapacity-driven regional waste incineration facility network planning with residential land value maximization involved: A case study of Shanghai, China.* Sustainable Cities and Society, 2023. **94**.

[13] Sener, N. and O. Feyzioglu, *Multiple allocation hub covering flow problem under uncertainty.* Annals of Operations Research, 2022.

[14] Andaryan, A.Z., K. Mousighichi, and N. Ghaffarinasab, *A heuristic approach to the stochastic capacitated single allocation hub location problem with Bernoulli demands.* J. European Journal of Operational Research. 2023.

[15] Bozorgi-Amiri, A. and M. Khorsi, *A dynamic multi-objective location–routing model for relief logistic planning under uncertainty on demand, travel time, and cost parameters.* J. The International Journal of Advanced Manufacturing Technology. 2016. **85**(5-8).

[16] Yang, L., T.S. Ng, and L.H. Lee, *A robust approximation for yard template optimization under uncertainty.* Transportation Research Part B-Methodological, 2022. **160**: p. 21-53.

[17] Maghfiroh, M.F.N., et al., *A Location Routing Problem with Time Windows Consideration: A Metaheuristics Approach.* J. Applied Sciences. 2023. **Vol.13**(No.843): p. 843.

[18] Yu, G., Zhao, Q.H, and Xi, M.H, *A method of location selection for Marine emergency rescue airbase considering time satisfaction.* J. Practice and Understanding of Mathematics. 2020. 50(23): p. 82-92.

[19] Wang, Y., et al., *Emergency logistics network optimization with time window assignment.* Expert systems with applications, 2023. **214**: p. 119145-119145.

[20] Yan, S., and Qi, J.P, *Study on multilevel emergency logistics facility location considering demand uncertainty*. J. Operations Research and Management Science. 2022. 31(09): p. 7-13.

[21] Wu, X., S. Liang, and G.D. Çaylı, *Characterizing idempotent nullnorms on a special class of bounded lattices.* J. Fuzzy Sets and Systems. 2022. Vol.427(No.C): p. 161-173

[22] Xinxing Wu, H.T., Zhiyi Zhu, Lantian Liu, Guanrong Chen, Miin-Shen Yang, *Nonlinear strict distance and similarity measures for intuitionistic fuzzy sets with applications to pattern classification and medical diagnosis.* J. Scientific reports. 2023. Vol.13(No.1): p. 13918.

[23] Ilbahar, E., C. Kahraman, and S. Cebi, *Location selection for waste-to-energy plants by using fuzzy linear programming.* Energy, 2021. **234**.

[24] Shang, X., et al., *A credibility-based fuzzy programming model for the hierarchical multimodal hub location problem with time uncertainty in cargo delivery systems.* International Journal of Machine Learning and Cybernetics, 2021. **12**(5): p. 1413-1426.

[25] Zhang, B., et al., *Sustainable multi-depot emergency facilities location-routing problem with uncertain information.* J. Applied Mathematics and Computation. 2018. **333**.

[26] Tan, Y.H, *Standardization and simulation of emergency repair for gas pipe network leakage*. R. 2018, Beijing University of Civil Engineering and Architecture.

[27] Shang, J.Z., An, W.P., and Guo, Y.D., I*mprovement and Analysis of BIRCH Algorithm based on thresholds.* J. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition). 2020. **32**(03): p. 487-494.

[28] Toregas, C., et al., *The Location of Emergency Service Facilities.* J. Operations Research. 1971. **19**(6).

[29] Hakimi, S.L., *Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph.* J Operations Research. 1964. **Vol.12**(No.3): p. 450-459.

[30] Kennedy, J. and R. Eberhart. *Particle swarm optimization*. in *Proceedings of ICNN'95* - International Conference on Neural Networks. 1995.