WILEY | Hindawi

*Research Article*

# An Urban Bus Network Generation Algorithm Based on Particle Swarm Optimization and Force Field Properties

**Yiming Chen** [ID],[1] **Yanhong Gu** [ID],[1] **Zhengdong Huang,**[2] **Tianhong Zhao,**[2] **and Liangyuan Guo**[1]

[1]*College of Mathematics and Statistics, Shenzhen University, Shenzhen, China*
[2]*Research Institute for Smart Cities, School of Architecture and Urban Planning, Shenzhen University, Shenzhen, China*

Correspondence should be addressed to Yanhong Gu; yhgu@szu.edu.cn

Due to continuous urban sprawl, large-scale bus network design has become a major challenge in urban transport planning. The continuous increase in urban population and scale makes the factors considered in the urban route network design increasingly complex. Contemporary public transportation network design problems are based more on efficiency goals such as the accessibility and comfort of the transportation network, which increases the difficulty of analyzing the problem. Bus network design is not only an NP-hard (nondeterministic polynomial) problem but also a multivariable and multiobjective problem. This paper focuses on the bivariate and multiobjective bus network design problem of route generation and station selection. This paper proposes an algorithm called the Pseudo Force Field. By combining the idea of Particle Swarm Optimization (PSO) and the properties of the force field, a feasible route generation scheme is provided for the design of the bus network. The algorithm does not need to determine the end station and has a high degree of completion of the demand. This solves the problem of the selection of terminal stations in large-scale road network design. On this basis, the article combines Genetic Algorithm (GA) and Pareto frontier to provide a new route optimization algorithm and proves the effectiveness of the algorithm. The model has achieved theoretical results in the design of the bus route network in the megacity of Shenzhen, China.

## 1. Introduction

Public transportation is one of the important ways to achieve a balance between supply and demand of transportation, energy savings, and emission reduction, and it has been highly valued by cities at all levels worldwide. According to existing research, transport systems in some countries account for approximately 20% of total annual greenhouse gas emissions from the energy sector [1]. Motorized travel is the main contributor to the worsening global greenhouse effect, and a reasonable public transportation system is of great significance in reducing the global greenhouse effect and achieving global sustainable development [2]. On the other hand, the public transportation system is the main mode of travel for residents of large cities. According to a survey report on residents' traffic behavior and willingness conducted by the Shenzhen Municipal Government in 2019,

public transportation accounted for 48% of residents' motorized travel, reaching 61% during peak hours. In terms of travel experience, more than 40% of residents hope to have a better travel experience. The public transportation system plays an irreplaceable role in the travel of urban residents. A good public transportation system helps to improve the people's happiness index and promotes the economic development of the city, which is of great significance to the development of the city.

Among the existing travel modes, rail transit and bus transportation are the main forms of public transportation. The concept of the transit network design problem (TNDP) was first proposed by Baaj and Mahmassani in 1991 [3] to describe the public transport system network design problem. It is a typical TSP-type multiobjective NP-hard problem [4]. The common urban rail transit is mainly based on the subway, and bus transportation is realized by the

operation of the bus network. In contrast, bus transportation has more flexible routes and greater accessibility than rail transportation, but it is far inferior to rail transportation in terms of operating speed and load capacity. In cities with a variety of public transportation, the route network design method of bus transportation often accommodates some traffic demands that cannot be met by rail transportation. Therefore, the design of the bus transportation system is often more rigorous and meticulous than rail transportation, and more influencing factors need to be considered.

In the existing research, most of the route generation algorithms are studied based on economic benefit indicators such as route length and optimized based on them [5]. However, in the design of an urban route network, the priority of the route network's requirements for accessibility and demand completion has exceeded the traditional economic benefit index. Moreover, the problem of low demand completion will lead to a general decrease in the quality of the overall set of alternative routes, which may cause the results obtained by the algorithm to lose practical value. This is particularly prominent in large city network design. Therefore, a route generation method with the main purpose of demand fulfillment is of great significance to the design of an urban route network.

In this paper, a route generation algorithm is proposed. The prototype of the algorithm is inspired by the logistics distribution problem and has similarities with the Particle Swarm Optimization (PSO). This paper will construct a Pseudo Force Field model based on this idea and design a new route generation method. The main contributions of the research are as follows:

(1) The article explores a new way of route generation. This method combines the properties of the PSO algorithm and the physical force field to provide a better initial solution set for the Genetic Algorithm (GA). The combination of algorithm and GA is shown in Figure 1. The article will demonstrate the advantages of this algorithm compared with the traditional initial solution generation algorithm through the first experiment.

(2) Make a fundamental change to the optimization method of the GA and propose a new optimization method. Through this route generation method, the GA is transformed from the traditional route selection problem to the site ranking problem. Through experiment two, the article mainly proves that this optimization method can ensure the route network quality of the algorithm in the optimization process and expand the whole solution set space.

## 2. Literature Review

Route network analysis is often analyzed by transforming it into a topology network structure. Through mathematical topology techniques, the relative influence between different routes and the traffic efficiency of the route network is studied [6]. Rivera-Royero et al. study route network performance from 11 RNP concepts and develop a classification scheme to map possible relationships and boundaries between them [7]. Munir et al. provide a template for the analysis of demand type indicators by evaluating the effectiveness of travel demand management strategies [8]. Khan and Fatmi provide a metric for assessing the safety of traffic networks, filling a gap in this field [9]. Jiang et al. analyzed the impact of route network design on people's lives from multiple perspectives, including environmental pollution, traffic accidents, noise emissions, and so on [10]. The researchers analyzed the impact of route network design on urban operation through different angles and means. This further reflects the importance of route network design and provides a strong theoretical basis for route optimization evaluation.

In the early design of the route network, the economic benefits of the route were the main consideration in the research. Pentek et al. demonstrated a relatively classic economical route network design method through the study of forest route design [11]. In the optimization algorithm, two-layer planning is the main method of early route network design. This is a way to optimize the route network based on the idea of mathematical operations research. The research of Ben-Ayed et al. and Zhang and Gao is a typical example of two different periods of this kind of road network algorithm [12], [13]. With the introduction of bionic algorithms, traditional mathematical programming methods such as bilevel programming have been gradually replaced. Early bionic algorithms are cited, such as Martins and Pato and Pattnaik et al., who applied the tabu road algorithm and genetic algorithm, to the problem of bus network design [14], [15]. Ngamchai and Lovell improved the encoding method of the route algorithm, which greatly improved the problem-solving efficiency of bionic algorithms [16]. The abovementioned studies are all designed with basic route attributes such as route length as the research object, and there are very few algorithms for generating initial routes according to requirements. However, there are already relatively mature algorithms for the demand data extraction of urban traffic systems [17]. Due to the high volatility of demand data, the relevant fluctuation of demand data cannot be considered in road network design. Therefore, the current public transportation network design is based on deterministic transportation networks and deterministic travel demand [6]. Zhang et al. provide a study on route saturation optimization by combining the knowledge of the golden ratio to improve the genetic algorithm [18]. This research is more classical research on the problem of demand nature. Badia et al. further deepen the accessibility of the demand problem and add the transfer accessibility problem to the TNDP problem [19].

According to a summary of the existing research, the route network optimization algorithm has gradually matured in theory. However, there is still a large gap in the research on demand orientation, including route generation and solution. The algorithm proposed in this paper gives an example of the research on demand-based goal-oriented route network optimization. In fact, the algorithm idea has been adopted as the generation method of the initial route in the road network design by Sun et al. [20]. In addition, there
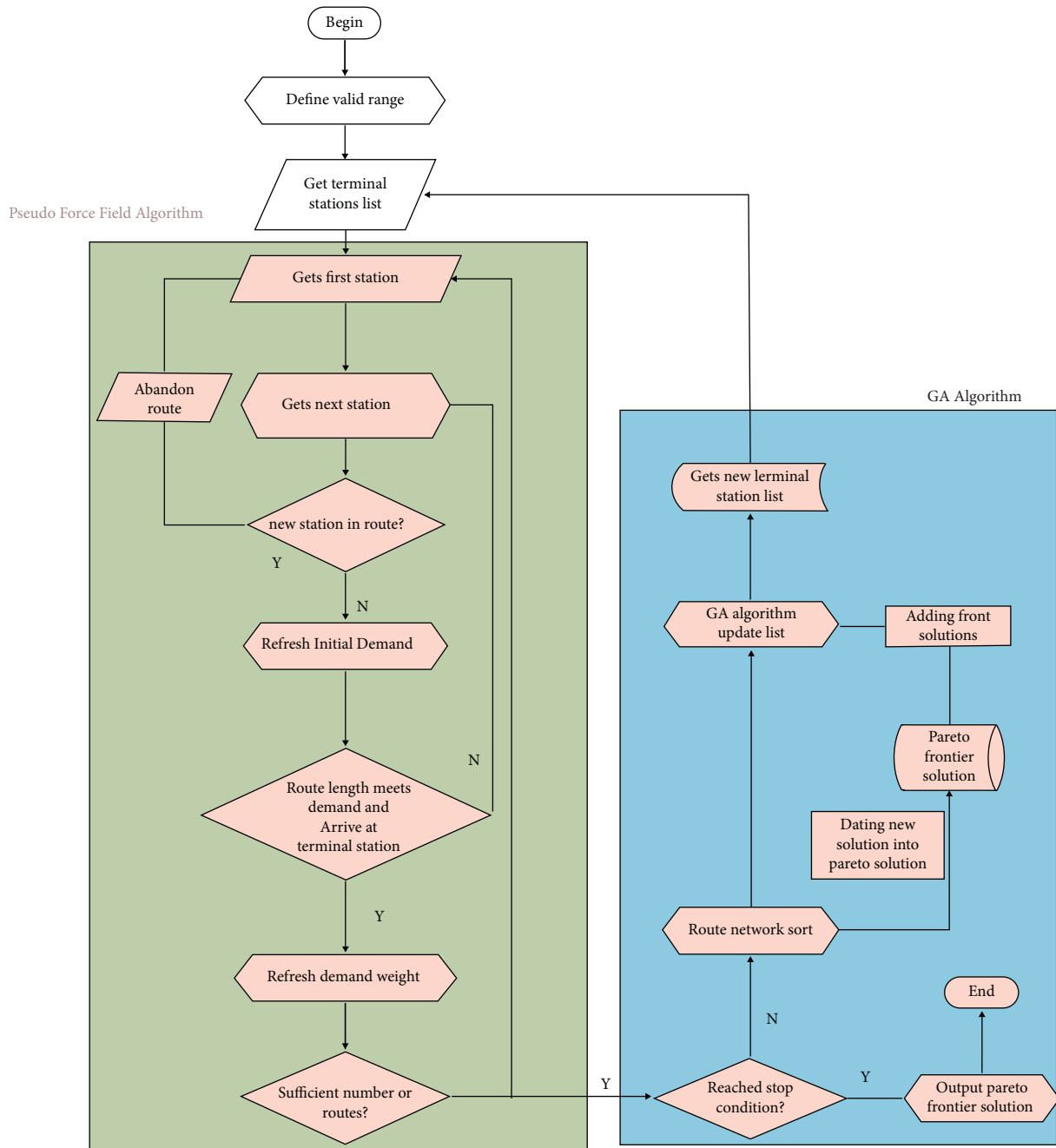
FIGURE 1: The entire process of route generation and optimization.

have been similar studies in other fields on the combined application of PSO and GA, and it has been proven to be feasible. Taking the research of Pandey et al. as an example [21], they combined PSO and GA to build a multiobjective model to solve the problem of increased power loss in the power supply system. In the design of public transport network, GA cannot solve the bivariate conflict between station selection and route generation, which is particularly prominent in large-scale route network design problems. This paper provides a feasible solution for the route generation of GA by combining the PSO idea and the force field properties. In the evaluation and discussion in this paper, the

fulfillment of requirements is the most important research objective.

## 3. Preliminary Preparation

*3.1. Assumptions.* In the design of this study, the following assumptions are made. These assumptions are valid in subsequent model representations and simulation calculations. Assumptions will better constrain the application scenarios of the model and improve the operational efficiency of the model.

(1) The model does not consider the problem of one-way and two-way traffic on different road sections, and this aspect is not constrained in the simulation calculation.

(2) Routes only consider major bus service lines, and each line must connect two terminal stops at both ends.

(3) There are no isolated bus stations in the station set; that is, each station connects at least two lines.

(4) The experimental background assumes that the influence of the subway on passenger transport is unchanged, and the influence of the route structure on the subway is not considered in the calculation.

TABLE 1: Notation and name in the paper.

| Set | Name |
| --- | --- |
| $I(A)$ | Unitization function |
| $E$ | Unity matrix |
| Ones$(X, Y)$ | All-one matrix |
| NEED$_{\text{level}}$ | Demand class matrix |
| NEED | Demand matrix |
| OD | Reachable route length matrix |
| $Vqx_{\text{point}}, Vqy_{\text{point}}$ | Point vector matrix |
| ODV$qx_{\text{point}}$, ODV$qy_{\text{point}}$ | Reachable point vector matrix |
| $k$ | Gravitational coefficient |
| Pointnum | Number of anchors |
| BGnode | Number of terminal stations |
| roadlen | Length of route; |

*3.2. Explanation of Symbols.* In the algorithm, the latitude and longitude grid is regarded as a two-dimensional plane space for calculation, where the longitude is the $x$ coordinate and the latitude is the $y$ coordinate. The following variables are defined as initial variables that do not change in the full-text discussion. They are the origins of other variables. The variables are shown in Table 1. In the formula of the article, "∘" stands for Hadamard product and "×" stands for matrix multiplication. "·" stands for number multiplication and vector multiplication.

First, some special matrices are explained. These matrices are used in subsequent formulations. They do not have practical meaning in the discussion of this article. Ones$(X, Y)$ represents a column matrix whose elements are all 1 and whose shape is $(X, Y)$. $I(A)$ is a matrix binary function:

$$I(A) = \begin{cases} 1, & A_{(i,j)} \neq 0, \\ 0, & A_{(i,j)} = 0. \end{cases} \quad (1)$$

The demand data are contained in the two NEED-type matrices corresponding to the NEED$_{\text{level}\,(i,j)}$ and NEED matrices in the above table. Each element in the matrix represents a requirement, and they are all square matrices with the same dimension as the number of stations. Among them, NEED$_{(i,j)}$ represents the demand from the $i$-th station to the $j$-th station. NEED$_{\text{level}\,(i,j)}$ represents the demand level corresponding to the station $i$ to station $j$ (The definition of the demand class will be detailed in the data processing section). The OD matrix contains information about the accessibility of the route, including three matrices OD$_{(i,j)}$, ODV$qx_{\text{point}}$, and ODV$qy_{\text{point}}$. OD is a square matrix with the same dimension as the number of stations, and the value of OD$_{(i,j)}$ represents the length of the route between two stations; that is, the matrix contains the length information of the route and reachability information. If the element is 0, it means that the two points are not reachable. For the two station vector matrices of $Vqx_{\text{point}}$ and $Vqy_{\text{point}}$, each element of the matrix satisfies the following:

$$Vqx_{\text{point}\,(i,j)} = \text{Point}_{(j,x)} - \text{Point}_{(i,x)},$$
$$Vqy_{\text{point}\,(i,j)} = \text{Point}_{(j,y)} - \text{Point}_{(i,y)}, \quad (2)$$

where Point$_{ix}$ and Point$_{iy}$ represent the x-coordinates of the $i$-th and $j$-th stations, respectively; Point$_{iy}$ and Point$_{jy}$ represent the y-coordinates of the $i$-th and $j$-th points, respectively. According to the above formula, the mathematical expressions for ODV$qx_{\text{point}}$ and ODV$qy_{\text{point}}$ are as follows:

$$\text{ODV}qx_{\text{point}\,(i,j)} = Vqx_{\text{point}\,(i,j)} \circ I(\text{OD}),$$
$$\text{ODV}qy_{\text{point}\,(i,j)} = Vqy_{\text{point}\,(i,j)} \circ I(\text{OD}). \quad (3)$$

## 4. Mathematical Model

*4.1. Pseudo Force Field Algorithm.* In the original PSO, the algorithm simulates the feeding habits of the bird flocks to find the optimal solution. Since the optimization process of PSO is similar to the path generation process, it is often applied to the line generation problem. In the route generation algorithm introduced in this paper, the idea of PSO in the optimization process is borrowed to deal with the relationship between route generation and demand changes. Specifically, the Pseudo Force Field algorithm will construct a fitness function according to different locations and changing demands during the route generation process and update the "speed" of route generation.

According to the expression in Sun et al. article, the pseudo force field algorithm is based on the basic form of electric field strength calculation [20]. Considering the basic properties of the electric field, there is the following formula for the magnitude of the electric field experienced at a point in the $R^2$ space:

$$E = \sum k \frac{Q}{r^2}. \quad (4)$$

The calculation of the force field for each station in the route network should not be affected by the demands of all other stations throughout the route network. When calculating the effective field force experienced by each calculated station, the set of stations affecting the calculated station needs to be determined. Taking the design of shorter routes as an example, stations that cannot be reached under the specified route length should be excluded from the station set [22]. For a route design from a station, the set of stations

FIGURE 2: Force analysis in equipotential map.

that have an impact on the calculated station is called the set of valid stations. It should be the set of all the stops that can be reached under the conditions of the specified route length. Therefore, the demand matrix formed by the set of stations that affect a calculated station should satisfy the following formula:

$$
\begin{aligned}
\text{NEED}^*_{\text{level}} = &\ T_{\text{roadlen}+[(\text{BGnode}/\text{Pointnum})]} \\
&\circ \text{NEED}_{\text{level}} \circ T^T_{\text{roadlen}+[(\text{BGnode}/\text{Pointnum})]},
\end{aligned}
\tag{5}
$$

where

$$
T_n = I\big(\big((\text{OD} \circ T_{n-1})^T + E\big) \\
\times \text{Ones}(\text{Pointnum}, \text{Pointnum})\big),
$$

$$
T_1 = \begin{cases} T_{1\,(i,j)} = 1, & \forall j,\ i = \text{point}, \\ T_{1\,(i,j)} = 0, & \forall j, i \neq \text{point}, \end{cases}
\tag{6}
$$

$$
\text{for} \quad i, j = 1, 2, 3 \ldots \text{Pointnum}.
$$

Point represents the first station of the route; $T$ is a column matrix of $(\text{Pointnum}, \text{Pointnum})$ shape; $\text{NEED}^*_{\text{level}}$ represents the Demand Class Matrix generated by all valid stations. In the parametric design, the setting of the size of the influence area will be slightly larger than the length of the route. Its purpose is to solve the problem that the route length may have reached the target during the route generation process, but the route has not reached the next

terminal station, resulting in the failure of route generation. In the simulation calculation, [BGnode/Pointnum] is used as the expansion volume of the station set, which is designed based on the density of terminal stations in all stations. If the rate of generation is slow, the value can be adjusted subjectively according to the actual situation without affecting the subsequent calculation. In the process of generating the route, each station is regarded as measuring 0 on $R^2$, and the charge of the station is its initial demand (the initial demand refers to the sum of the demand from this point as the starting station). The point charge between the points is on $R^2$, with the Euclidean distance as the distance between the two. Under the assumption that the difference between vehicles is not considered, the urban bus network will be regarded as a simple superimposed electric field, and the electric field force experienced at each point is the electric field strength at that point. The specific presentation in the route network is shown in Figure 2.

A vector diagram of the local field strength at a point is presented in Figure 2. At this point, the calculated point is subjected to the field strength of all valid points (or stations). The field strength acting at each point is affected by the charge at the effective point and the distance between it and the calculated point. The vector direction of the field force on the calculated point in the force field can be obtained by superimposing the vector of the field strength on the calculated point. Its formula is as follows:

$$
F_{\text{Point}i} = \left( \sum_{j=1}^{\text{pointnum}} k \cdot \frac{\text{NEED}^{**}_{\text{level}(j,1)} \cdot Vqx_{\text{point}(i,j)}}{\sqrt{Vqx_{\text{point}(i,j)}^2 + Vqy_{\text{point}(i,j)}^2}}, \sum_{j=1}^{\text{pointnum}} k \cdot \frac{\text{NEED}^{**}_{\text{level}(j,1)} \cdot Vqy_{\text{point}(i,j)}}{\sqrt{Vqx_{\text{point}(i,j)}^2 + Vqy_{\text{point}(i,j)}^2}} \right).
\tag{7}
$$

In the formula, $\text{NEED}^{**}_{\text{level}}$ represents the matrix formed by the initial demand (the definition of the actual meaning of this matrix will be explained in Part 4.2, formula (13)), which is a (pointnum, 1) matrix. $F_{\text{Point}i}$ represents the resultant force of point $i$ in the pseudo force field, which is a two-dimensional vector.

According to the force field vector of the calculated point, the most suitable approximate path direction at the point can be obtained at this time. It will direct the route to areas with more intensive demand to ensure that the route can complete more demand. In the algorithm, the direction of the force field vector will be used as an important criterion for choosing the next station of the route. That is, the vector
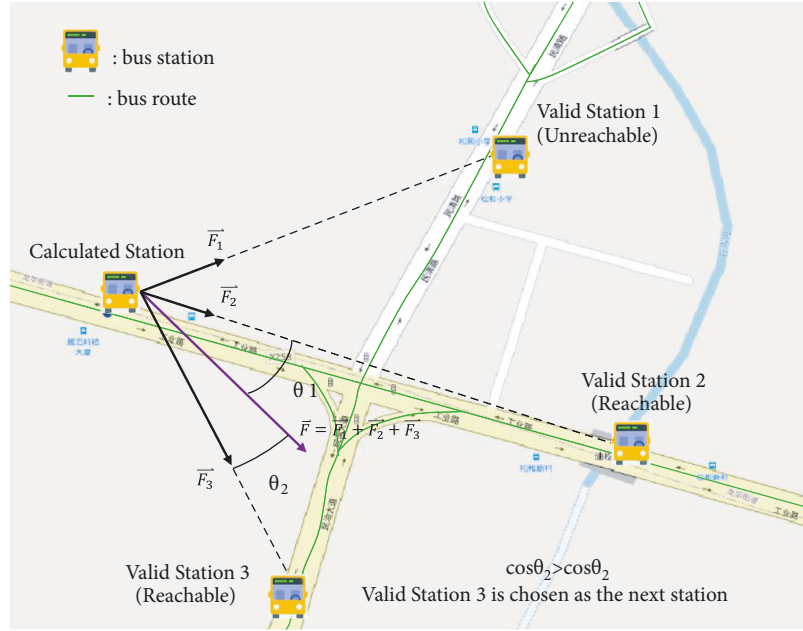
Figure 3: Station selection in pseudo force field.

here is the direction of the "speed" determined by the particle at that point according to the demand data. In the selection of the next station, the station with the smallest cosine of the angle between the vector of the reachable station and the calculated station and the force field vector of the calculated station itself will be used as the next station to be selected. The specific presentation is shown in Figure 3.

In Figure 3, the next reachable station should select the station with the largest cosine value of the angle between the two vectors. The vector direction between two points has nothing to do with the route; it is only related to the relative position of the two points in space. In the process of selecting the next reachable station, the station with the cosine value of the included angle less than 0 should be eliminated first to avoid the phenomenon of loopback or going back in the route. In route generation, reachable stations are selected by looping until the length of the route reaches the given demand and another starting and ending station appears in the set of route stations. Its basic mathematical formula is expressed as follows:

$$F(i, j) = \frac{F_{\text{point}i} \cdot \left(\text{ODV}qx_{\text{point}(i,j)}, \text{ODV}qy_{\text{point}(i,j)}\right)}{\left|F_{\text{point}i}\right| \cdot \left|\left(\text{ODV}qx_{\text{point}(i,j)}, \text{ODV}qy_{\text{point}(i,j)}\right)\right|},$$

$$\text{newpoint} = \text{MAX}\{F(i, j) | \text{for } i \in \text{Caluated Station}, \forall j\},$$

$$(8)$$

where

$$\text{MAX}\{A\} = \begin{cases} A_{(i,1)} = 0, & F(i, j) \neq \max(A), \\ A_{(i,1)} = 1, & F(i, j) = \max(A). \end{cases} \quad (9)$$

In the above formula, newpoint represents the selected reachable stations. Suppose its node ID as $i$, then matrix

newpoint is a (Pointnum, 1) shape which elements are 0 except $\text{newpoint}_{(i,1)}$ as 1. The route generation algorithm makes the route go to the demand-intensive area as much as possible under the constraint of the specified route length to solve the problem that the route avoidance demand produces invalid routes to optimize some demanding goals in the algorithm solution. In this algorithm, the size of the algorithm's "speed" does not change. The algorithm only uses the direction of "speed" for station selection, and the distance traveled each time is one station. The inertia index of the algorithm is 0. That is, the speed calculation at each station (or position) is completely determined by this point, regardless of the "speed" of the previous stations.

*4.2. Model Optimization.* Through the Pseudo Force Field algorithm, the route design problem can be transformed into stations or station groups ranking problem, and the optimal Pareto frontier can be obtained by combining the GA algorithm. However, in the process of generating the route, the algorithm often encounters many problems, such as loopback and interruption. For this, more fine-grained constraints must be placed on the model. In this work, the following three basic requirements are put forward for route generation:

(1) Generated routes without loopbacks or station duplication;

(2) The first and last points of the generated route must belong to the set of terminal stations;

(3) The route length is equal to or greater than the required length threshold.
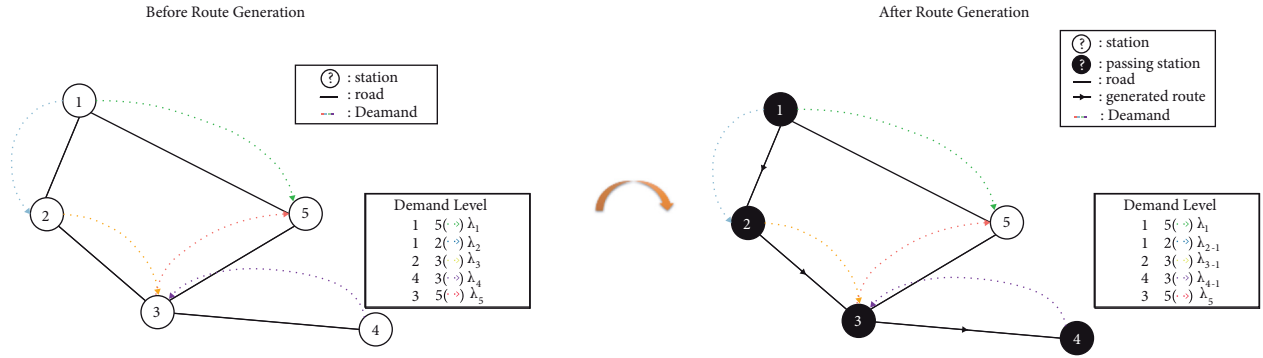
FIGURE 4: Demand cleaning 1.

If the generated route cannot meet the above constraints, it is considered that the route generation fails and the route needs to be discarded.

For the nonreordering that already has a set of stations or station groups, if the station is directly sorted, it is enough to loop through the station sequence. If sorting by station group, to generate several groups of routes with lengths $A_1$, $A_2$, $A_3$, ..., and the number of routes $a_1$, $a_2$, $a_3$, ..., it is necessary to assign tasks to the station groups. Since the number of stations in a station group is not necessarily the same, the number of routes undertaken by each station group should also vary. For a station group $E$, let the ratio of the number of sites it contains to the total number of stations be $E'$. For $n$ types of routes with different lengths, there is the following formula for the number of routes undertaken by the station group:

$$\text{worknum} = \sum_{i=1}^{n} [E' \cdot a_i]. \quad (10)$$

Each item of the series in the formula represents the corresponding number of routes undertaken by this station group for a certain type of route length and then assigns the number of routes to each station group. In generating routes, the choice of stops within a station group is random. (The random selection will be explained in detail in the data processing section later.) In the process of generating a route, regardless of whether there is a route output, each time a station is selected, the station will be removed from the station group. This method avoids the fact that multiple routes are generated at the same site at the same time, resulting in a high degree of route coincidence, which will cause the route generation effect to decrease. If the stations of a station group have completely taken the station group, but the route task has not been completed, record the remaining route task amount and temporarily skip the station group. After completing the tasks of all other station groups, retraverse the station groups and complete the original unfinished build tasks for each station group. In the classification, a certain station group may account for a very small proportion of the entire number of stations, so the station group cannot be allocated to line tasks according to the proportion. At the beginning of assigning tasks, if the number of routes undertaken by the station group is 0, then add 1 to the number of routes for this station. This may cause the number of routes to be inconsistent with the originally designed number of routes, but it guarantees the number of routes.

After a route is generated, the generated route will affect the demand data, so it is necessary to reduce the demand level of the corresponding demand satisfied by the route. The specific formula is as follows:

$$\text{NEED}(\text{new})_{\text{level}} = \text{NEED}_{\text{level}} - \text{Re}(\text{Road}, \text{NEED}_{\text{level}}), \quad (11)$$

where

$$\text{Re}(\text{Road}, \text{NEED}_{\text{level}}) = \begin{cases} \text{NEED}_{(i,j)} = 1, & \text{if} \quad i, j \text{ both in Road and NEED}_{\text{level}(i,j)} \neq 0, \\ \text{NEED}_{(j,i)} = 1, & \text{if} \quad i, j \text{ both in Road and NEED}_{\text{level}(j,i)} \neq 0, \\ \text{NEED}_{(i,j)} = 0, & f \quad i \text{ or } j \text{ not in Road.} \end{cases} \quad (12)$$

Assuming that the route can connect station $i$ and station $j$, reduce the demand level from station $i$ to station $j$ and station $j$ to station $i$ by 1 level (the basis for the reduction will be detailed in the data processing section below). If the demand level is 0, keep the original value level unchanged. The final output of the Re function is a matrix with the same shape as the initial parameters of $\text{NEED}_{\text{level}}$. $\text{NEED}(\text{new})_{\text{level}}$ and $\text{NEED}_{\text{level}}$ are the same variable. $\text{NEED}(\text{new})_{\text{level}}$ is a temporary representation of $\text{NEED}_{\text{level}}$ after the update. This part of the modification will affect matrix $\text{NEED}_{\text{level}}$. The meaning of which is more intuitively represented by Figure 4.

In the process of generating the route, the route is processed by the needs of the station to avoid the reverse situation and so on. Take a column matrix of Pointnum rows of $\text{NEED}_{\text{level}}^{**}$ to represent the tentative demand in the route generation process. Among them, $\text{NEED}_{\text{level}(i,1)}^{**}$ represents the sum of the demand level of the point with node ID $i$; that
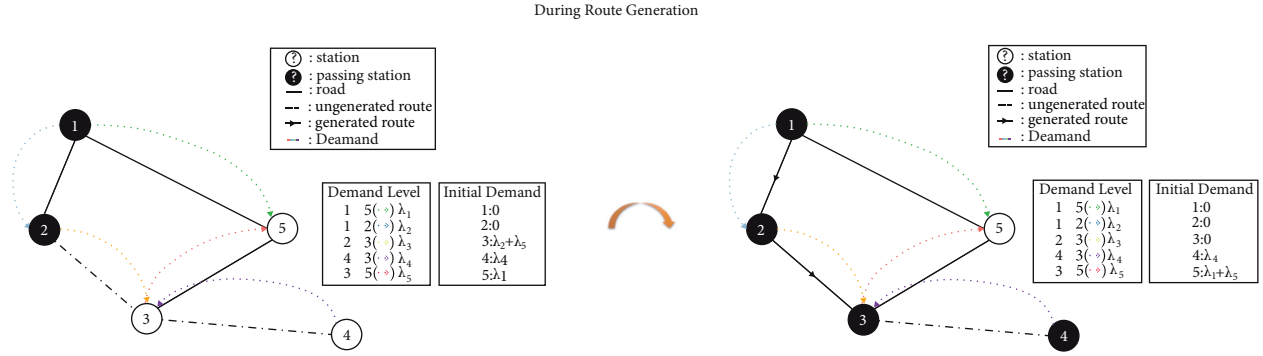
During Route Generation



FIGURE 5: Demand cleaning 2.

is, the charged amount of this point. $\text{NEED}^{**}_{\text{level}}$ only takes effect in this route generation and does not affect the original demand data.

In the process of route generation, $\text{NEED}^{**}_{\text{level}}$ should be changed as the route is extended. When the route selects a new station, demand should change as follows.

(1) The demand (or charge) in $\text{NEED}^{**}_{\text{level}}$ of the passed stations should be defined as 0. This ensures that routes do not loop back or be chosen repeatedly.

(2) When a new stop is selected for the route, all demand starting from the new stop should be "activated." In other words, for all the demands starting from this station, the target station of this demand needs to increase the corresponding demand (or charge) in $\text{NEED}^{**}_{\text{level}}$. The amount of increase is determined by the level of that requirement.

$$\text{NEED}^{**}_{\text{level}} = \text{NEED}^{*}_{\text{level}} \times \text{Ones}(\text{pointnum}, 1), \quad (13)$$

$$\begin{aligned} \text{NEED}(\text{new})^{**}_{\text{level}} &= \text{NEED}^{**}_{\text{level}} + \text{NEED}^{**}_{\text{level}} \\ &\quad \times I(C - \text{Ones}(\text{pointnum}, 1)), \end{aligned} \quad (14)$$

$$\text{NEED}(\text{new})^{**}_{\text{level}} = \text{NEED}^{**}_{\text{level}} \text{OD}, \quad (15)$$

where

$$C = \begin{cases} C_{(i,1)} = 0, & \text{for point } i \text{ is Selected Point,} \\ C_{(i,1)} = 1, & \text{for point } i \text{ isn't Selected Point,} \end{cases}$$

$$D = \begin{cases} D_{(i,1)} = 0, & \text{for point } i \text{ in Road,} \\ D_{(i,1)} = 1, & \text{for point } i \text{ not in Road.} \end{cases} \quad (16)$$

Formula (13) explains the relationship between $\text{NEED}^{**}_{\text{level}}$ and $\text{NEED}^{*}_{\text{level}}$. The definition of $\text{NEED}^{*}_{\text{level}}$ is already given in formula (5). In addition, formulas (14) and (15) show that $\text{NEED}^{**}_{\text{level}}$ is constantly changing during route generation. These changes will only take effect for this route generation, not for other routes. $\text{NEED}(\text{new})^{**}_{\text{level}}$ and $\text{NEED}^{**}_{\text{level}}$ are the same variable. $\text{NEED}(\text{new})^{**}_{\text{level}}$ is just a temporary representation of $\text{NEED}^{**}_{\text{level}}$ after making the previous two changes. $C, D$ are two matrices of shape of (pointnum, 1). In formula (14 and 15, formula (14) needs to be performed before

formula (15) in each processing. This guarantees that the demand (or charge) of each passed station is 0. The specific generation process is shown in Figure 5.

In the selection of the length of the line, the short line missions are given priority. On the one hand, during the task of generating short lines, the set of stations that are valid for the calculated station can be determined according to the scope of influence and will not be affected by the distant unreachable stations. On the other hand, short and small routes can clear the demand in small areas, reduce the complexity of demand data so that further long routes will not be affected by the demand in small areas, and better complete some cross-regional demand work, maximizing the realization of the value of long lines.

*4.3. Evaluating Indicator.* In modern bus networks, route network analysis should consider the operational efficiency and comfort of the network [23]. According to the actual situation, two indicators are used as measurement standards in the calculation, namely, the load degree of the network and the complexity of the route network.

The route index mainly involves the travel demand of passengers and the length of the route, which reflects the complete efficiency of the route to the demand and the comfort of passengers. In the calculation, the average load level of all routes in the entire route network is used as the rating index, which is the ratio of the passenger turnover of a route to the total vehicle mileage. Its mathematical formula is as follows:

$$\text{Weight}_{\text{need}(i,j)} = \frac{\text{NEED}_{(i,j)}}{N}, \quad (17)$$

$$\text{Weight}_{\text{road}} = \sum_{i \in P, j \in P} \frac{\text{Weight}_{\text{need}(i,j)}}{\text{Roadlen}}, \quad (18)$$

$$\text{Weight}_{\text{network}} = \sum \frac{\sum \text{Weight}_{\text{road}}}{\text{Roadnum}}. \quad (19)$$

In the formula, $N$ represents the number of routes that can satisfy the demand between stations $i$ and $j$; $P$ is the set of all stations that the route passes through; Roadlen represents the length of the route segment from the $i$-th station to the $j$-th station on the route; Roadnum represents the

number of routes; $Weight_{road}$ represents the load of a single route in the route network.

The complexity of the route network can intuitively reflect the load balance of all routes in the bus network. The indicator adopts the Lorentz curve model and is calculated based on the Gini coefficient formula that reflects fairness. Its specific formula is as follows:

$$H = \frac{2\sum_{i=1}^{Roadnum} i \cdot Weight_{road\,i}}{Roadnum \cdot \sum_{i=1}^{Roadnum} Weight_{road\,i}} - \frac{Roadnum + 1}{Roadnum}.$$

(20)

All routes are arranged in ascending order according to parameter $Weight_{road}$ in formula (18), $Weight_{road\,i}$ in formula (20) represents the value of the $i$-th route in the arrangement on parameter $Weight_{road}$.

The abovementioned $Weight_{network}$ in formula (19) and $H$ in formula (20) are the main bases for obtaining the Pareto frontier. The subsequent optimization process and results will revolve around the above two indicators.

*4.4. Genetic Algorithm Design and Pareto Frontier.* Due to the large number of terminal stations, the terminal stations can be classified into several station groups by K-Means clustering (the details of this part will be explained in detail in the Data Processing, Part 5.2.1). In the population structure, each group of bus networks is regarded as an individual, and the multiple groups of bus networks generated by the ordering of multiple groups of different terminal stations are regarded as a population. According to this situation, the fitness determination in the genetic algorithm will also be analyzed based on the score of each bus network individual on the two indicators.

Before detailing the GA, the application of the Delaunay algorithm to the Pareto Frontier needs to be supplemented. The Delaunay triangulation algorithm can construct several points on a two-dimensional plane into a nonconcave triangular network. According to this algorithm, all edges and edge points can be extracted through the triangular network. According to the basic principles of data envelopment analysis (DEA), it is not difficult to prove that the Pareto front must be composed of several edge points and edge segments of the triangulation. Therefore, the Pareto Frontier of the two-dimensional plane point can be obtained by reasonably sorting the edge points extracted by the triangulation network in the calculation. In the specific calculation, the minimum value of any index in the edge point set is selected as the first selection point. Set it as $M_1 = (x_1, y_1)$; then, in the solution set DM composed of frontier solutions, the correspondence between other frontier points and this point should conform to the following formula:

$$M_{i+1} = (x_{i+1}, y_{i+1}) \in DM,$$

(21)

where

$$\begin{cases} x_{i+1} > x_i, \\ y_{i+1} < y_i. \end{cases}$$
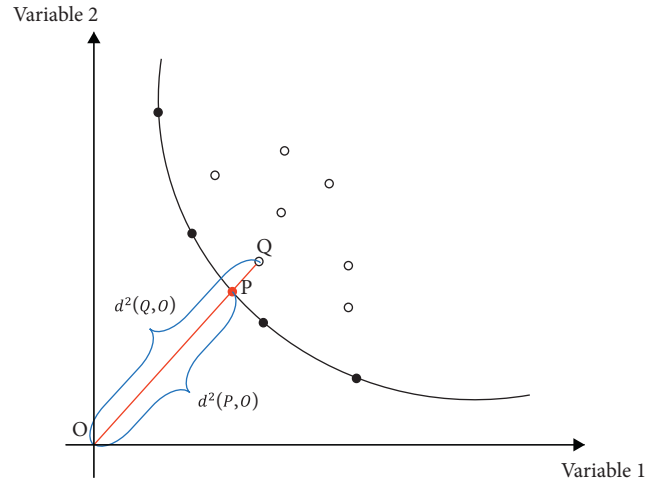
(22)



FIGURE 6: Fitness score in GA.

According to the above formula, all the selected edge points and edges are connected to form a polyline as the Pareto front. With Delaunay's nonconcave triangular network properties and the basic theory of DEA in the Pareto frontier, it is not difficult to prove that all other solutions of Pareto support can be realized for this set of Pareto frontier points, which has been confirmed in other fields [24].

In the model solution, several stations or station groups are randomly arranged and the corresponding bus network is generated as the initial population. According to the Pseudo Force Field algorithm, each station ranking represents a set of definite and unique bus networks; that is, the station ranking can form a one-to-one mapping relationship with the bus network. The two negative indicators of the network are combined with the Delaunay algorithm to obtain the Pareto frontier individuals in the population [25]. The two index values of each individual are used to map the point $Q$ on the two-dimensional coordinate system, connect the point to the origin, and the intersection P of the straight line and the Pareto frontier, with the formula:

$$Mark = \frac{d^2(P, O)}{d^2(Q, O)}.$$

(23)

Solve Mark as fitness in GA, which is shown in Figure 6.

In the processing of the optimization algorithm, the traditional sorting genetic algorithm is used for calculation. Because this paper does not study the performance of the optimization algorithm, very precise considerations are not adopted in the GA optimization process. In the subsequent calculation work, the GA algorithm will face the optimization of the sequence number sorting problem without repetition. The specific crossover and mutation processes are shown in Figures 7 and 8.

In addition, in the GA algorithm, individuals are selected in the form of roulette. This selection is made depending on the probability constructed according to the fractional proportions defined above. The specific GA process is shown in the following pseudocode.
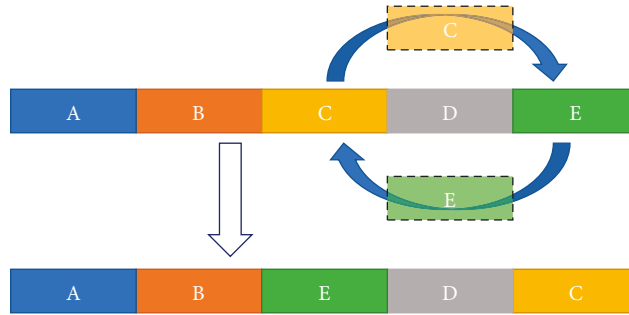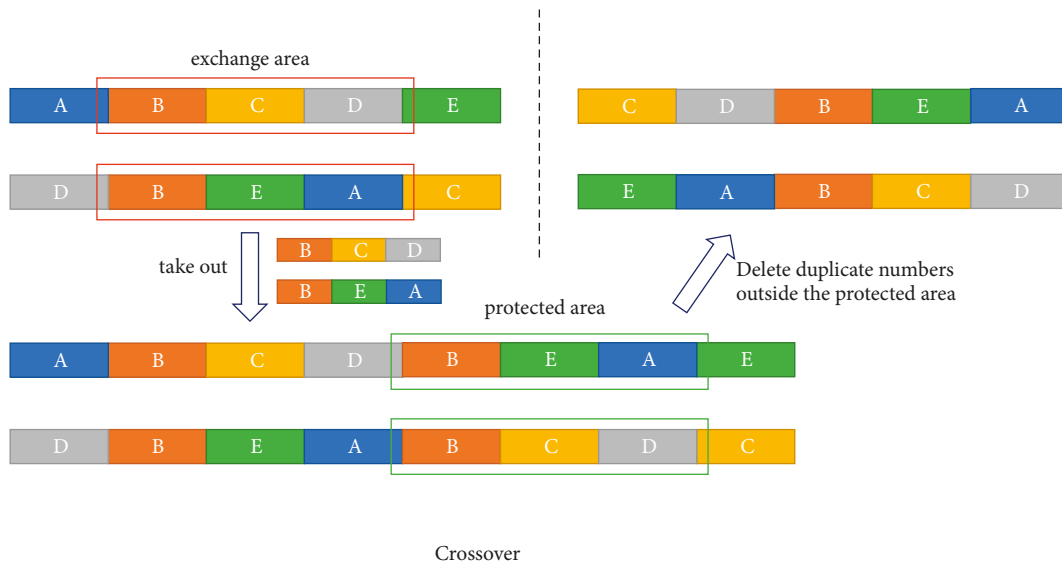
FIGURE 7: Mutation in GA.



Crossover

FIGURE 8: Crossover in GA.

Algorithm: GA
**Input**: Demand, terminal station, station
**Output**: route network
(1)     **def** Pseudo Force Field (terminal station sequence):
(2)       **output** route network
(3)     **def** Evaluation (route network):
(4)       **output** Pareto Front
(5)     **begin**
(6)       **initialize** terminal station sequence
(7)       $T \longleftarrow 0$
(8)       **while** ($T <$ cycle times):
(9)         $T = T + 1$
(10)        route network $\longleftarrow$ Pseudo Force Field (terminal station sequence)
(11)        Mark $\longleftarrow$ Evaluation (route network)
(12)        new sequence $\longleftarrow$ Inheritance, Crossover, Mutation by Mark
(13)        terminal station sequence $\longleftarrow$ new sequence
(14)      **end while**
(15)    **end**

ALGORITHM 1: Pseudocode for the optimization algorithm.

To improve the convergence speed of the calculation, an elite retention strategy is adopted in the algorithm. Each time a Pareto frontier solution is generated, the value of the frontier solution, including the station group ordering corresponding to the value, and the generated bus network will be retained to the next generation to participate in the

construction of the Delaunay triangulation and the competition for survival. If a new solution replaces or joins the original frontier solution, the solution becomes part of the frontier solution, and the replaced frontier solution is treated as a common solution, participating in algorithms such as selection and intersection. It can enrich the frontier solution set of the results and provide more data for analyzing excellent route network results.

## 5. Numerical Experiment

The results will mainly show the advantages of the algorithm from two aspects: the change in demand completion in route generation and the changes in demand completion in the optimization process. The experiment uses K-Shortest-Path (KSP) algorithm for comparison. The KSP algorithm is the most commonly used initial route generation algorithm. In the GA of a large route network, it is the main way of generating the initial route set. As a comparison, the effectiveness of the algorithm in this paper can be visually demonstrated. The approximate data of the computing environment of the two algorithms are shown in Table 2.

For the selection of route number, it needs to be set before the experiment. The design of these parameters will be affected by the actual situation. The city of Shenzhen has 2,406 simplified stations after sorting, with a total of approximately 900 conventional lines. For the theoretical route generation design of 512 stations, that is, approximately 1/5 of the area of Shenzhen, the theoretical estimate of the number of routes should be approximately 180. Using the period of maximum passenger flow among all periods of the regular day as the calculation data, the number of routes is set to be 5% more than the regular number. In the calculation, 200 routes will be generated.

*5.1. Performance of Route Generation.* In the first experiment, more than 180 routes were generated and recorded with random station ordering using the Pseudo Force Field algorithm. In the generated route group, the routes with the same start and end station pairs are excluded. In the KSP algorithm, the generated route starting and ending station pairs are used and the shortest route is generated. Again, the routes generated in this section follow the guidelines mentioned earlier. These guidelines are relisted in the following sections:

(1) The generated route does not have loopbacks. That is, there are no duplicate stations in the route;

(2) The generated route must use the terminal station as the start and end of the route;

(3) The route length must be longer than or equal to the length of the route task design;

(4) The route does not allow the same terminal stations.

Routes that do not meet the above requirements need to be eliminated during the generation process and are regarded as invalid routes. In the subsequent optimization experiments and generation comparison experiments, the above basic constraints need to be observed.

TABLE 2: Attribute parameters.

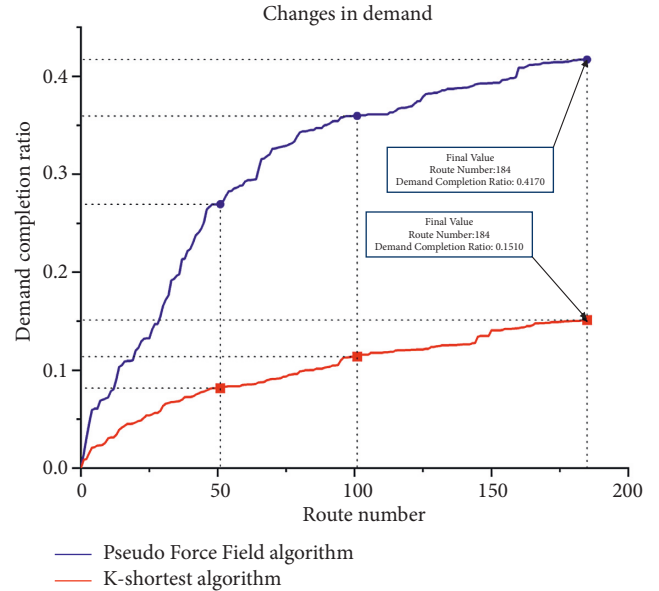| Parameters | Value |
| --- | --- |
| Point number | 512 |
| Number of terminal stations | 117 |
| Route number | 200 |



FIGURE 9: Comparison of KSP algorithm and pseudo force field algorithm.

*5.1.1. Comparison.* For the route algorithm proposed in this paper, routes of length 30 (according to the algorithm, it may be longer than 30) are generated in the calculation, and the number of routes keeps adding up. The route is assumed to fully complete all requirements traversed, regardless of route affordability. This means that the reduction of the demand level mentioned in Part 4.2 should be defined here as zero for passing demand. Moreover, each demand is defined as 1, and no demand grading is performed anymore. The variation between the route and demand completion of the two algorithms is shown in Figure 9.

As shown in Figure 9, the Pseudo Force Field algorithm has a better performance in demand completion. In the same process of generating 184 routes, the Pseudo Force Field algorithm can complete 41% of the demand. In contrast, the KSP algorithm fulfills approximately 15% of the demand. It means that within a reasonable number of routes, the Pseudo Force Field algorithm is more efficient for demand completion. Experiments with more routes were not performed and shown because they were considered to be in excess of reasonable numbers. Too many routes are meaningless for transit network design, even if it performs better in demand fulfillment.

Through the above experiments, the line generated by the Pseudo Force Field algorithm has a high completion efficiency for demand. This has an excellent performance in

large urban road networks. In the following experiments, we will further demonstrate the performance of the Pseudo Force Field algorithm in optimization.

*5.2. Performance of Optimization Process.* Based on the Pseudo Force Field algorithm, a set of route networks can be generated in sequence through an ordered station code. Therefore, the problem of bus network design will be generated from the original route and transformed into the problem of arranging the terminal stations. This experiment will demonstrate the stability of the generative algorithm to satisfy the demand during the optimization process through two indicators related to the demand.

*5.2.1. Data Processing.* There are 117 terminal stations included in the 512 stations in the selected range. Due to the limitation of computing power, it is not possible to directly optimize the ranking of the stations in the calculation. Therefore, the K-Means clustering method is used in the simulation calculation, and the starting and ending stations are divided into a limited set of station areas [26]. Each area contains several terminal stations, and the 117 station codes are converted into station group codes. Since the selection of stations within the station group is random in route generation, the station group code and the road network cannot form a meaningful mapping. To ensure the successful convergence of the algorithm, the station classification must satisfy the stability of the mapping between the ranking code of the same station group and the score generated route network. That is, the score fluctuation should be within a small range. It is not difficult to see that for a fixed set of terminal stations, the fewer station group categories, the higher the volatility of the score. Therefore, it is necessary to find a suitable number of station classifications through experiment.

Based on the above theoretical basis, the K-Means clustering method is used to continuously reduce the number of classification categories from 100 categories. Each time the same station group is sorted and encoded to generate 50 groups of route networks, the similarity between the route networks is calculated, and the stability of the classification method is analyzed from the degree of dispersion of the road network on the two index data [27]. Its specific calculation formula is as follows:

$$G_e\left(x_1, x_2, \ldots \ldots x_n\right) = \sum_{i=1}^{n} \sum_{j=1}^{n} d_e^2\left(x_i, x_j\right). \quad (24)$$

In the formula, the data $x$ represent the vector on the $R^2$ metric space distributed on the two indicators of route network complexity and route network load in this problem, and $x_i$ represents the value of the $i$-th data.

In addition, to account for the order of magnitude difference between the two indicators, the variance calculation uses the degree of change relative to the mean as the raw data. That is, the raw data are processed by the following formula:
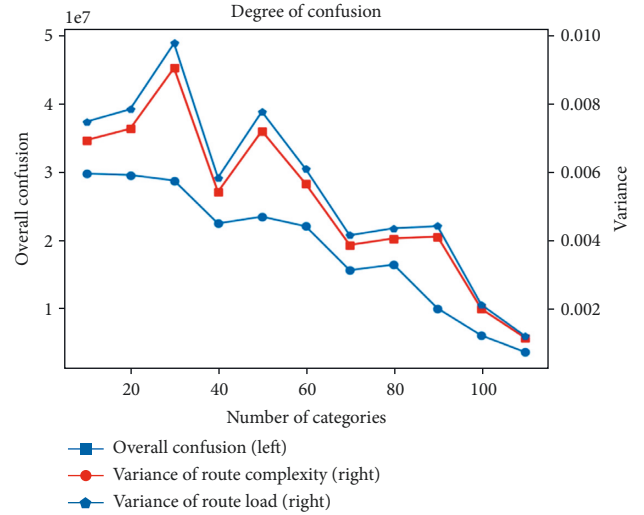


Figure 10: Index confusion under different classifications.

$$D = \frac{\sqrt{\sum_{i=1}^{n} \left(\mu_i - \overline{\mu}/\overline{\mu}\right)^2}}{n}, \quad (25)$$

where $\mu$ represents the original value and $n$ represents the number of route network groups.

According to the information entropy obtained by the above formula, combined with the respective variances of the two indicators, 48 route networks are generated for each classification method for data analysis. The specific data performance is shown in Figure 10.

When the overall data are in the 10–60 categories, there will be large numerical fluctuations in the overall degree of confusion and the variance of the load degree, and the reference as a basis for classification is poor. After the data are greater than 60 categories, the variance of the load degree, the variance of the complexity, and the overall degree of confusion all show a relatively stable or declining trend. Therefore, when choosing the number of clustering categories, more than 60 categories should be preferred for K-Means area two-dimensional clustering. In the simulation calculation, the K-Means clustering method is adopted to divide the terminal stations into 70 station groups, and the route is generated and solved.

According to Article 4 of "Technical Conditions for Safe Operation of Motor Vehicles" issued by China in 2004, the floor area for standing passengers in urban buses and trolleybuses shall be not less than 0.125 square meters per person. In summary, the standard bus design verification number is 45 people, and the bus operation standard is 5 min during the peak period. Therefore, according to the three bus scheduling standards formulated by Sheu [28], the bus dispatch frequency is set to be time-invariant, and real-time passenger demand data are considered to be collected through advanced intelligent transportation system technologies (such as automatic passenger counting systems), regardless of changing passengers arriving at the terminal and arriving at the originating station. In the calculation, it is assumed that each bus line theoretically takes 540 passengers for each one-hour demand, and the subsection clustering

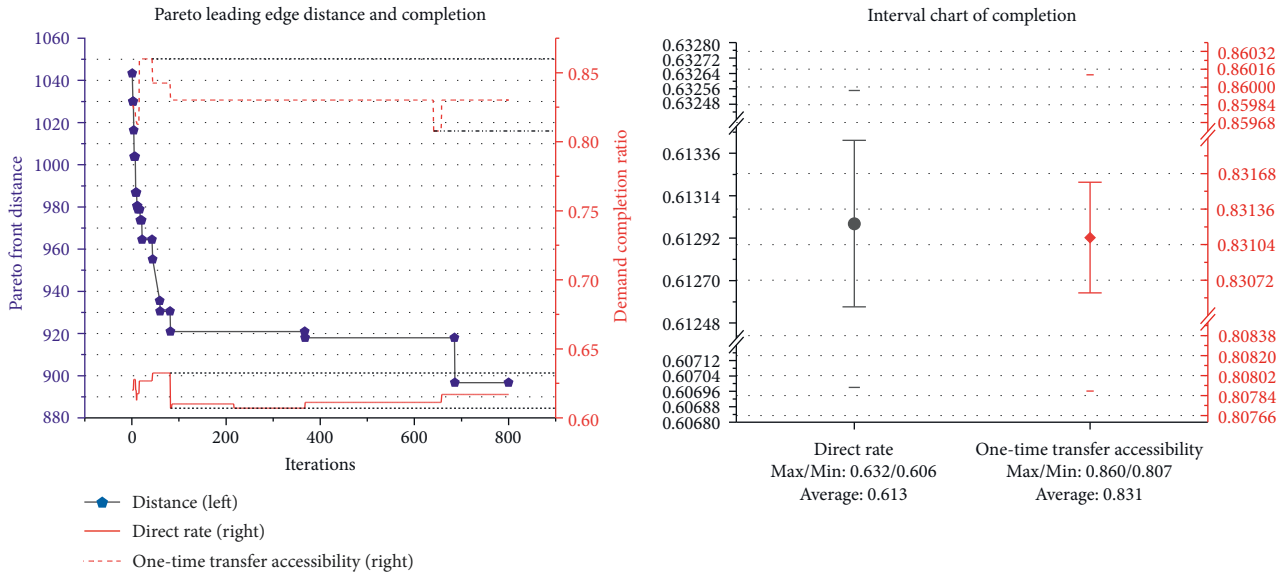| Parameters | Value |
| --- | --- |
| $k$ | 1.5 |
| Route kind 1 | Length: 10 number: 80 |
| Route kind 2 | Length: 20 number: 60 |
| Route kind 3 | Length: 30 number: 60 |
| Classification number of terminal stations | 70 |
| Number of individuals per generation | 50 |
| GA cycle numbers | 800 |
| Crossover length | 30 |
| Crossover probability | 0.8 |
| Mutation length | 5 |
| Mutation probability | 0.2 |



FIGURE 11: Achievable efficiency and index optimization.

method is used to classify all demands. In the processing of demand data, for a certain station pair, the maximum demand of the station pair in all periods on a regular day is selected as the original demand data to construct a demand matrix. In addition, with 540 as the dividing line, the raw demand data are graded; that is, in the case of legal transportation, each time a vehicle passes through, the weight of this station is reduced by one unit. Therefore, the demand level of each point where demand is not 0 can be expressed as follows:

$$\text{DemandDegree} = \left[\frac{\text{Demand}}{540}\right] + 1. \qquad (26)$$

The practical significance of this division is that the demand from the $i$-th station to the $j$-th station satisfies the required optimal number of routes.

*5.2.2. Result of Calculation.* In the calculation, multithreading can be used for the generation to speed up the calculation speed. According to actual conditions, a total of 200 lines of 10, 20, and 30 are generated in the calculation.

The proportion of routes is approximately 3 : 3 : 4, and the $k$ value is selected as 1.5 for testing. According to the above algorithm, the variables in Table 3 for the experiments are determined in the calculation.

In the experiment, an i7-11700k is used to calculate in the Windows 10 environment, and without GPU acceleration, the time to generate 50 sets of lines by multiprocessing is 5 mins. In the presentation of the results, the solution with the minimum distance from the origin in the frontier solutions of each generation of the population will be presented as the overall population level. The two indicators corresponding to the optimal route network are shown in Figures 11 and 12.

The coordinates on the left represent the distance between the Pareto frontier and the origin, in which the minimum value of the multiplication of the two indicators in the entire Pareto frontier is used as the numerical result. The right side represents two demand completion degrees, namely, the accessibility rate in the case of direct access and the accessibility rate in the case of one transfer. The image shows that in the process of solving the Pareto frontier, the algorithm ensures that the direct access rate is stable above
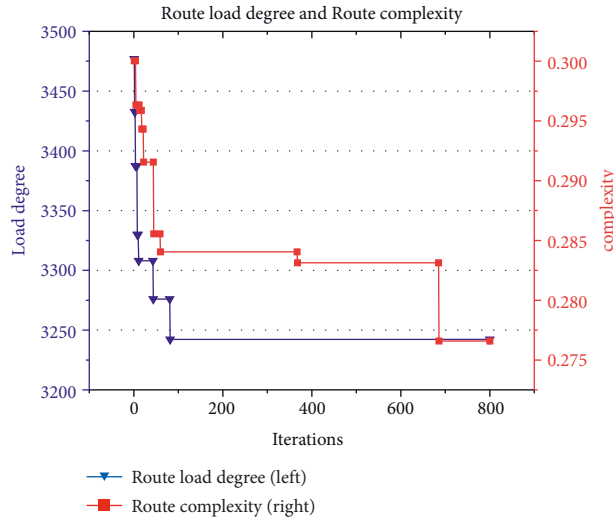
Figure 12: Changes in two indicators.

Table 4: Changes in experimental indicators.

| Iteration | Load degree | Complexity | Distance | Direct rate | One-time transfer |
|---|---|---|---|---|---|
| 1 | 3,477.01 | 0.3000 | 1,043.25 | 0.6200 | 0.8260 |
| 2 | 3,432.68 | | 1,029.95 | 0.6200 | 0.8260 |
| 4 | 3,387.33 | | 1,016.35 | 0.6279 | 0.8217 |
| 6 | | 0.2963 | 1,003.85 | 0.6279 | 0.8217 |
| 8 | 3,329.92 | | 986.83 | 0.6200 | 0.8146 |
| 11 | 3,308.07 | | 980.36 | 0.6176 | 0.8125 |
| 13 | | 0.2958 | 978.82 | 0.6176 | 0.8125 |
| 18 | 3,308.07 | | 973.63 | 0.6266 | 0.8601 |
| 22 | | 0.2915 | 964.53 | 0.6266 | 0.8601 |
| 44 | 3,275.94 | | 955.16 | 0.6325 | 0.8424 |
| 45 | | 0.2855 | 935.46 | 0.6325 | 0.8424 |
| 60 | | 0.2840 | 930.53 | 0.6325 | 0.8424 |
| 82 | 3,242.2190 | | 920.95 | 0.6069 | 0.8301 |
| 368 | | 0.2831 | 917.97 | 0.6112 | 0.8301 |
| 686 | | 0.2765 | 896.74 | 0.6170 | 0.8302 |
| 800 | 3,242.2190 | 0.2765 | 896.74 | 0.6170 | 0.8302 |

60% and the one-time transfer reachability rate is above 80%. The algorithm ensures the stability of the requirement completion degree and realizes the optimization of the target. It should be noted that the demand for this experiment is defined in the context of demand stratification. It means that completing the same demand multiple times can repeatedly reduce the remaining value of the demand. This resulted in data-level differences between the two experiments.

Figure 12 shows the performance of the two indicators on the GA. In Figure 12, the left ordinate represents the load of the corresponding optimal route network in each iteration, and the right side represents the complexity of the route network. Both index values steadily decrease in the iterative process. The main data changes in the experiments are shown in Table 4.

## 6. Conclusion

The above results and analysis suggest that the Pseudo Force Field has higher performance in demand completion than traditional route generation algorithms. On the other hand, the Pseudo Force Field algorithm gives researchers a unique optimization method and guarantees the quality of the route during the optimization process. Compared with the traditional route generation algorithm, the Pseudo Force Field has a larger solution space. For example, faced with a basic route network design problem with $k$ starting and ending stations. Without considering the constraints, in the traditional optimization method, the number of combinations of $n$ lines selected from $m$ lines is $C_m^n$. In contrast, there are $k^n$ cases where there can be duplicate station orderings and $k!$ cases where there can be duplicate station orderings (this paper uses nonrepeated sites for experiments). The explosive growth of the solution set space makes the advantages of

the Pseudo Force Field algorithm particularly prominent in large route networks. This means that the Pseudo Force Field algorithm can provide a route generation method with greater potential for better optimization methods in the future and further improve the route network optimization results.

It is not difficult to see in the research that the common sorting GA algorithm cannot satisfy the convergence of such a large-scale solution set space. In addition, the completion of the route still needs to be improved. In the experiment, the processing method of the demand has a great influence on the generation effect of the route. Due to the lack of processing methods for changing demand, the algorithm should be more precisely considered in follow-up work to further improve the generation effect.

## Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

## References

[1] S. Sim, J. Oh, and B. Jeong, "Measuring greenhouse gas emissions for the transport-ation sector in Korea," *Annals of Operations Research*, vol. 230, no. 1, pp. 129–151, 2015.

[2] X. Xu, A. Chen, and C. Yang, "A review of sustainable network design for road netw-orks," *KSCE Journal of Civil Engineering*, vol. 20, no. 3, pp. 1084–1098, 2016.

[3] M. H. Baaj and H. S. Mahmassani, "An AI-based approach for transit route system planning and design," *Journal of Advanced Transportation*, vol. 25, no. 2, pp. 187–209, 1991.

[4] P. R. Tong, W. Du, M. Li, J. Q. Huang, W. Q. Wang, and Z. Qin, "Last-mile school shuttle planning with crowdsensed student trajectories," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 293–306, 2021.

[5] K. Sohn, "Multiobjective optimization of a road diet network design," *Transportation Research Part A: Policy and Practice*, vol. 45, no. 6, pp. 499–511, 2011.

[6] R. Zanjirani Farahani, E. Miandoabchi, W. Y. Szeto, and H. Rashidi, "A Review of Urban Transportation Network Design Problems," Article ID 5735702, 2019, https://doi.org/10.1155/2019/5735702.

[7] D. Rivera-Royero, G. Galindo, M. Jaller, and J. Betancourt Reyes, "Road network performance: a re-view on relevant concepts," *Computers & Industrial Engineering*, vol. 165, Article ID 107927, 2022.

[8] T. Munir, H. Dia, and H. Ghaderi, "A systematic review of the role of road network pricing in shaping sustainable cities: lessons learned and opportunities for a post-pandemic world," *Sustainability*, vol. 13, no. 21, Article ID 12048, 2021.

[9] A. A. Khan and Z. Fatmi, "Strategies for prevention of road traffic injuries (RTIs) in Pakistan: situational analysis," *Journal of the College of Physicians and Surgeons-Pakistan: JCPSP*, vol. 24, no. 5, pp. 356–360, 2014.

[10] B. Jiang, S. Liang, Z. R. Peng et al., "Transport and public health in China: the road to a healthy future," *The Lancet*, vol. 390, no. 10104, pp. 1781–1791, 2017.

[11] T. Pentek, H. Nevecerel, D. Picman, and T. Porsinsky, "Forest road network in the Republic of Croatia - status and perspectives," *Croatian Journal of Forest Engineering*, vol. 28, no. 1, pp. 93–106, 2007.

[12] O. Ben-Ayed, D. E. Boyce, and C. E. Blair, "A general bilevel linear programming formu-lation of the network design problem," *Transportation Research Part B: Methodological*, vol. 22, no. 4, pp. 311–318, 1988.

[13] H. Zhang and Z. Gao, "Bilevel programming model and solution method for mixed transportation network design problem," *Journal of Systems Science and Complexity*, vol. 22, no. 3, pp. 446–459, 2009.

[14] C. L. Martins and M. V. Pato, "Search strategies for the feeder bus network design problem," *European Journal of Operational Research*, vol. 106, no. 2-3, pp. 425–440, 1998.

[15] S. B. Pattnaik, S. Mohan, and V. M. Tom, "Urban bus transit route network design using genetic algorithm," *Journal of Transportation Engineering*, vol. 124, no. 4, pp. 368–375, 1998.

[16] S. Ngamchai and D. J. Lovell, "Optimal time transfer in bus transit route network design using a genetic algorithm," *Journal of Transportation Engineering*, vol. 129, no. 5, pp. 510–521, 2003.

[17] D. Bauer, M. Tulic, and W. Scherrer, "Modelling travel time uncertainty in urban networks based on floating taxi data," *Eur. Transp. Res. Rev.*, vol. 11, no. 1, p. 46, 2019.

[18] L. Zhang, M. Zhang, W. Yang, and D.-G. Dong, "Golden ratio genetic algorithm based approach for modelling and analysis of the capacity expansion of urban road traffic network," *Compitational Intelligence and Neuroscience*, 2015.

[19] H. Badia, J. Argote-Cabanero, and C. F. Daganzo, "How network structure can boost and shape the demand for bus transit," *Transportation Research Part A: Policy and Practice*, vol. 103, pp. 83–94, 2017.

[20] J. Sun, J. Huang, Y. Chen, P. Wei, J. Jia, and C. Song, "Flexible bus route optimal scheduling model in response to dynamic demand[J]," *Journal of Beijing University of Technology*, vol. 47, no. 3, pp. 269–279, 2021.

[21] R. S. Pandey and S. R. Awasthi, "A multiobjective hybrid algorithm for optimal planning of distributed generation," *Arabian Journal for Science and Engineering*, vol. 45, no. 4, pp. 3035–3054, 2020.

[22] Y. X. Wang, J. Bi, C. R. Lu, and C. Ding, "Route guidance strategies for electric vehicles by considering stochastic charging demands in a time-varying road network," *Energies*, vol. 13, no. 9, p. 2287, 2020.

[23] C. Zhang, Z. Juan, Q. Luo, and G. Xiao, "Performance evaluation of public transit systems using a combined evaluation method," *Transport Policy*, vol. 45, pp. 156–167, 2016.

[24] Z. Miao, Y. Geng, and J. C. Sheng, "Efficient allocation of CO2 emissions in China: a zero sum gains data envelopment model," *Journal of Cleaner Production*, vol. 112, pp. 4144–4150, 2016.

[25] Y. S. Chen, D. J. Deng, and C. C. Teng, "Range-based localization algorithm for next generation wireless networks using radical centers," *IEEE Access*, vol. 4, pp. 2139–2153, 2016.

[26] C. A. Iliopoulou, C. P. Milioti, E. I. Vlahogian-ni, and K. L. Kepaptsoglou, "Identifying spatio-temporal patterns of bus bunching in urban networks," *Journal Of Intelligent Transportation Systems*, vol. 24, no. 4, pp. 365–382, 2020.

[27] A. Kolacz and P. Grzegorzewski, "Measures of dispersion for multidimensional data," *European Journal of Operational Research*, vol. 251, no. 3, pp. 930–937, 2016.

[28] J. B. Sheu, "A fuzzy clustering approach to real-time demand-responsive bus dispatching control," *Fuzzy Sets and Systems*, vol. 150, no. 3, pp. 437–455, 2005.