Analysis of Bus Line Operation Reliability Based on Copula Function

Zhang Xiaoliang 1,2,* and Jia Limin 1

1 School of Traffic and Transportation, Beijing Jiaotong University, Beijing 100044, China; lmjia@bjtu.edu.cn
2 Research Institute of Highway, Ministry of Transport, Beijing 100088, China
* Correspondence: 16114226@bjtu.edu.cn; Tel.: +86-010-62079527

Abstract: To promote the development level of urban sustainability, more and more cities have been paying attention to the improvement of public transportation. City managers intend to attract people from private cars to public transport by improving the service level of urban public transport. The operational reliability of bus lines plays a crucial role in maintaining high-level service of public transportation. Previous studies have focused on the service level of the whole line by investigating the overall stability of departure punctuality, line running time, and punctuality. This study aims to clarify the relationship between the whole line and sites in terms of the influence on the operational reliability of bus lines. This study proposes a bus route reliability evaluation method, based on copulas connect function, and the actual line data are taken as a case study. The results show that the method reveals the relationship between the overall line and the interstation paths in the context of operational reliability. This study finds the interstation path, which is more critical to the whole line. This method can provide a good reference for the diagnosis and evaluation of line operation and the optimization of bus routes. The results can support the sustainable development of public transportation.

Keywords: bus line operation; reliability; correlation; copula function

1. Introduction

It has been realized that developing public transport is a promising and sustainable way to alleviate traffic congestion and improve the urban environment [1]. Correspondingly, many efforts have been made to develop public transportation [2]. Among these efforts, service quality improvements to public transport takes a dominant position. The operational reliability of bus lines is an important factor for the public to consider when choosing a trip. At the same time, it is also one of the most important indicators for industry departments to improve the service level and enterprises to improve the bus operation efficiency.

The main reason that determines the sustainable development of public transportation is its reliability. The reliability of bus services is an important reflection of the quality of bus service, and is also a key factor for residents to consider when choosing transportation modes [3,4]. Improving the reliability of public transportation is conducive to improving the service quality of public transportation, effectively enhancing the satisfaction of public transportation travelers, and promoting the sustainable development of public transportation.

The concept of public transport service reliability was first proposed by Sterman in 1976 [5]. Later, bus service reliability was defined as the ability to provide stable service within a period of time [6–9]. A survey report of bus service quality for Seoul Metropolitan City in 2003 provided a more comprehensive definition of bus service reliability, which included the accurate operation of public transportation systems, the probability of...
vehicles arriving at a station on time, the delay rate, and so on [10]. Some other reports noted that reliability of the public transportation system is the ability to complete specified tasks under certain operating conditions and times [11–15]. Additionally, time reliability of a public transportation system is divided into vehicle running time reliability and passenger waiting time reliability [16–18].

In terms of influencing factors, the Transit Capacity and Quality of Service Manual (TCQSM) system lists factors that affect reliability, which include traffic conditions, road construction and maintenance, vehicle and maintenance quality, availability of vehicles and drivers, bus priority, completion of the whole schedule, passenger demand balance, driver technology, use of wheelchair elevators and slopes, line length, and number of stations [19]. Abkowitz and Engelstein [20] pointed out that factors that may affect the operational reliability of bus lines include line characteristics (line length, number of signalized intersections, degree of parking on the road, distance between stations, etc.), operation conditions (traffic volume, service frequency, passenger activities), drivers (departure delay, driving behavior) [21,22], etc. Due to the influence and restriction of many factors on the reliability of public transport operations, research on the reliability of public transport has been a difficult problem for operators and researchers all over the world.

In terms of evaluation indexes, the TCQSM from the United States sets up evaluation indexes related to reliability at both the station and line levels when evaluating the quality of public transport services [19]. It is known that the reliability of adjacent stations is almost the same, so the line-level indexes, such as punctuality, has received a great deal of attention. In the following research, a number of indexes have been taken into account [23], such as the time coefficient of variation, running time ratio [24,25], time reliability [26], headway [27–29], and other indicators. Moreover, Chen et al. put forward three performance parameters to evaluate the reliability of public transport services, namely, on-time rate index (PIR), departure index (DIS), and uniformity index (EIS) [30]. Bruinsma et al. studied passenger departure times and section operation times [31].

In accordance with the above indexes, many reliability evaluation methods have been put forward, see [32–38]. Statistical methods play an important role in these methods. It is concluded that the bus travel time for one line obeys a normal distribution, lognormal distribution, gamma distribution, Weibull distribution, etc. Therefore, Zhao et al. proposed the system engineering theory [39] and Song studied bus lines as a series system and then solved the problem using Monte Carlo simulations [40]. Sharov and Boisjoly et al. studied the multi-level evaluation model and algorithm of operation reliability of conventional public transportation networks [40,41].

In summary, there has not been a consensus definition regarding bus operation reliability. From different perspectives, different indexes of bus service reliability have been provided. From the perspective of passengers, the adopted indexes are mainly about the smoothness of headway, punctual arrival, and waiting times. From the operating point of view, they are concerned with the execution rate of the running time. Most of the research perspectives involve variations of the operational time variance between stations, but there are few studies on bus service reliability from the perspective of the relationship between station and station adjacent paths.

To fill this research gap, this paper takes the conventional bus operation system on a single line as the research object, and defines the connotation of the operation reliability of bus lines with regard to system theory. The functional cause of operational reliability of a single bus line is defined to be composed of routes between stations in the middle of the bus line, and the reliability evaluation method under the influence of the correlation of the components is studied. Through the study of the relationship between the routes between stations and the total reliability of the bus line, a more reasonable basis will be determined for the optimization, adjustment, and service improvement of the bus line.

Along these lines, this paper evaluates and analyzes the reliability of bus routes and the relationship between routes at various stations using the Copula function. As a multivariable correlation analysis function [42], it can well express the correlation of different
variables and has the advantage of system reliability analysis model construction. The other sections of this paper are mainly arranged as follows: in the second part, bus line operation reliability evaluation index is presented. The third part covers the general idea and bus line reliability evaluation model based on the Copula function. The fourth part covers the actual case analysis (including data and calculation steps). Conclusions and prospects are made in the last section.

2. Evaluation Index of Bus Line Operation Reliability

2.1. Definition of Bus Line Operation Reliability

This paper takes one-way operation of a single bus line as the research object. The operational reliability of the bus line is defined from the view of the passenger. The main transport function of a bus line is to transport passengers from the starting point to the end point, within a certain period of time, under the specified route conditions.

Specific description: Suppose a one-way single bus line is \( G \), the running direction of the line is \( N_1 \rightarrow N_{m+1} \), and it is composed of \( m+1 \) \((m \geq 1)\) bus stops. The bus stops are denoted by \( N_i \) \((i = 1, 2, ..., m, m \geq 1, m \) is the number of stations. \( L_j \) \((j = 2, ..., m)\) represents path between two stations.

In the operation of one-way bus lines, the situation of alternative lines is not considered; that is, only the operation of a single line is considered. In order to realize the function of one-way operation of the bus line, it is necessary to ensure the normal operation (that is, not invalid) of the bus vehicles between the stations on the bus line.

A functional structure diagram of the bus line is shown in Figure 1.

![Figure 1](image)

Figure 1. Schematic diagram of functional structure of bus line series system, \( G \).

In the operation process of the bus line system, \( G (N_1 \rightarrow N_{m+1}) \), any failure of inter-station path operation will lead to failure of the whole bus line operation. Therefore, the reliability of the one-way bus route, \( G \), can be represented by the failure probability of all the paths between \( m \) stations. The specific calculation method is as follows:

The operating reliability of one-way single bus route \( G (N_i \rightarrow N_{m+1}) \) is denoted by \( R_d(t) \), and the operating reliability of path \( L_i \) between each station is expressed by \( R_{L_i}(t) \). In a previous study, it was assumed that the bus routes run independently between each station, the reliability of bus routes (reliability estimation of series system) is:

\[
R_d(t) = \prod_{j=1}^{m} R_{L_j}(t)
\]  

In practice, the operation of routes between bus stops may affect each other. At the same time, bus operation is affected by many factors. It is a difficult problem to describe the relationship between inter-station path and the reliability of the line. This paper will adopt a new method to address this problem, which will be discussed in Section 3.

2.2. Operational Reliability Evaluation Index of Bus Lines

The operational reliability of a single bus line is related to the operation of vehicles. When considering the reliability evaluation index, it is mainly based on the following principles:
1. Reflect the operation reliability characteristics of bus lines, which is convenient to analyze and find existing problems;
2. Should be able to reflect the passengers for the bus to provide operational capacity requirements, that is, expectations;
3. It should be operable and easy to calculate based on the given data. After comprehensive consideration of the above factors, one-way bus line operation is regarded as a system based on the perspective of system theory, and each inter-station path is regarded as a subsystem.

Accuracy and regularity are the key indicators to measure the reliability of public transportation. The higher the accuracy and regularity, the more reliable the public transport system is. In a mathematical expression, accuracy and regularity are usually expressed by average time standard deviation and coefficient of variation. So the evaluation indexes of reliability in the operation of bus lines are selected, as listed in Table 1.

Table 1. The evaluation indexes of reliability in the operation of bus lines.

| The Evaluation Object | Time | Index |
|-----------------------|------|-------|
| a line                | day/hour | average running time (\(T\)) |
|                       |       | on-time rate (\(\eta\)) |
| inter-station path    | day/hour | average running time (\(T_{Lj}\)) |
|                       |       | average running speed (\(V_j\)) |
|                       |       | average time standard deviation (\(\delta_T\)) |
|                       |       | coefficient of variation (\(CV_T\)) |

The coefficient of variation, \(CV_T\), is \(\frac{\delta_T}{\bar{T}}\), where \(\delta_T\) is the standard deviation of the line running time or the running time between stations; \(\bar{T}\) is the average of the line running time or the station running time. The coefficient of variation is expressed in the form of relative numbers, which can eliminate the influence of different comparison results between measurement units and mean values, facilitating the analysis of the running time of routes and stations, judge the degree of dispersion of data distribution, and reflect the balance and stability of the bus line operation process.

Accompanied with the above indicators, a series of threshold ranges should be assigned to judge the reliability or unreliability of bus line operations and inter-station route operations. These threshold ranges are affected by many factors, such as the benefit and cost of bus, road traffic conditions, passenger demands and passenger expectations, etc., and the ranges may be different for different routes and different inter-station routes at different periods.

Broadly, the reliability index is the probability of a system during its normal life. So a one-way bus line operating life is the non-failure bus running time. That is generally within the range of the average running time, expressed in \(T\). The distribution function corresponding to \(T\) is: \(F(t) = P[T > t], t \geq 0\). That's the life distribution \(F(t)\), that is, the probability that the bus line is normal (not invalid) before time \(t\); that is, the reliability or reliability function of the bus line at time \(t\) is:

\[
R_d(t) = P\{T > t\} = 1 - F_d(t) = F_\bar{D}(t), t \geq 0
\]  

(2)

On the contrary, when the entire bus line running time exceeds the normal average running time range, it is a fault. Fault distribution is \(F_d(t)\).

\(T_{Lj}\) is the inter-station bus travel time when the running time of vehicles between two adjacent stations falls within the variation range of normal running time between stations, and the corresponding distribution function of \(T_{Lj}\) is:

\[
F_{Lj}(t) = P\{T_{Lj} > t\}, t \geq 0
\]  

(3)
Life distribution, \( F_{L_j}(t) \), that is, the probability that inter-station paths are normal (not failing) before time \( t \), that is, the reliability or the reliability function of inter-station paths at time \( t \) is:

\[
R_{L_j}(t) = P\{ T_{L_j} > t \} = 1 - F_{L_j}(t) = F_{L_j}(t), \quad t \geq 0
\]  

(4)

On the contrary, the operation time of the inter-station path exceeds the normal variation range of the inter-station operation time, which is a fault.

Under the specified conditions, the failure probability of a bus line in unit time after time \( t \) is called the failure probability density function of time \( t \), which is denoted as \( f(t) \); then, \( f(t) \) is:

\[
f(t) = \lim_{\Delta t \to 0} \frac{P\{ t < T \leq t + \Delta t \}}{\Delta t}
\]  

(5)

Similarly, the failure probability of the inter-station path in unit time after time \( t \) under specified conditions is called the failure probability density function of the line at time \( t \), denoted as \( f_{L_j}(t) \); then, \( f_{L_j}(t) \) is:

\[
f_{L_j}(t) = \lim_{\Delta t \to 0} \frac{P\{ t < T_{L_j} \leq t + \Delta t \}}{\Delta t}
\]  

(6)

The conditional probability of a bus line failure in a unit time or running path at time \( t \) is called the failure rate of the line, at time \( t \), which is abbreviated as the failure rate, denoted as:

\[
\lambda(t) = f(t)/R_0(t)
\]  

(7)

In the same way, the conditional probability of the failure of the route between stations in the unit time or running path, at time \( t \), is called the failure rate of the line at time \( t \), denoted as:

\[
\lambda_{L_j}(t) = f_{L_j}(t)/R_{L_j}(t)
\]  

(8)

3. General Idea and Methods

3.1. Research Framework

The main research ideas of this paper are shown in Figure 2. Firstly, bus route \( G \) is described by the system theory method, and a series system is established. Next, the evaluation indexes of line operation reliability are determined, including the line indexes and the inter-station path indexes. Then, the Copula function model is determined. In addition, the parameter values are determined using the estimation algorithm. Finally, the line reliability and inter-station path dependency are analyzed using actual data.
3.2. Reasons for Choosing the Copula Function

As a multivariable correlation analysis function [41], the Copula function could well express the correlation of different variables and has the advantage of system reliability analysis model construction. To characterize the interaction between bus routes and station routes, the Copula function is used in this paper. The Copula function was derived from the Sklar theorem [43], which is able to construct multivariate joint distribution functions with arbitrary marginal distributions and has strong flexibility and adaptability. The Copula function connects the joint distribution function, \( H(x_1, x_2, \ldots, x_n) \), of two or more random variables, \( X_1, X_2, \ldots, X_n \), with their marginal distributions, \( F_1, F_2, \ldots, F_n \), so that \( H(x_1, x_2, \ldots, x_n) = C(F_1(x_1), F_2(x_2), \ldots, F_n(x_n)) \) holds.

Therefore, the Copula function can be used to describe the interaction of paths between stations when there is an interaction relationship between the paths between stations. Suppose that bus line \( G \) is connected in series by \( m \) inter-station paths, \( L_{ij} \), and the life of the \( i \)-th inter-station path is \( T_{L_{ij}} \), reliability is \( R_{L_{ij}}(t) \). If the life distribution function is \( F_{L_{ij}}(t) \), \( i = 1, 2, 3, \ldots, n \), then the lifetime of \( G \) is \( T = (T_{L_{i1}}, T_{L_{i2}}, \ldots, T_{L_{in}} > t) \). The joint distribution function is:

\[
H(t_1, t_2, \ldots, t_n) = P(T_{L_{i1}} \leq t_1, T_{L_{i2}} \leq t_2, \ldots, T_{L_{in}} \leq t_n)
\]

(9)

According to Sklar’s theorem, there exists an \( n \)-dimensional copula \( C \) such that:

\[
H(t_1, t_2, \ldots, t_n) = C(F_{L_{i1}}(t_1), F_{L_{i2}}(t_2), \ldots, F_{L_{in}}(t_n))
\]

(10)

Because \( F_{L_{ij}}(t) \) is a continuous function, \( C(F_{L_{i1}}(t_1), F_{L_{i2}}(t_2), \ldots, F_{L_{in}}(t_n)) \) is unique. Then, the reliability calculation of bus line \( G \) is as follows:
\[ R_y(t) = P\{ \min (T_{L1}, T_{L2}, ..., T_{Ln} > t) \} \]
\[ = P(T_{L1} \leq t_1, T_{L2} \leq t_2, ..., T_{Ln} \leq t_n) \]
\[ = 1 - \sum_{i=1}^{n} P(T_{Li} \leq t) + \sum_{1 \leq i < j \leq n} P(T_{Li} \leq t, T_{Lj} \leq t) + \sum_{1 \leq i < j < k \leq n} P(T_{Li} \leq t, T_{Lj} \leq t, T_{Lk} \leq t) + \ldots + (-1)^n \sum_{1 \leq i_1 < i_2 < \ldots < i_n \leq n} P(T_{L_{i_1}} \leq t, T_{L_{i_2}} \leq t, \ldots, T_{L_{i_n}} \leq t) \]
\[ = 1 - \sum_{i=1}^{n} F_{Li}(t) + \sum_{1 \leq i < j \leq n} (-1)^j \times C(F_{L_1}(t), F_{L_2}(t), ..., F_{L_j}(t), 1, 1, ..., 1) \]

where \( 2 \leq j \leq n \).

### 3.3. Copula Function

There are two types of popular Copula functions, elliptic Copula functions and Archimedean Copula functions. Elliptic copula functions could be further divided into integral normal Copula functions and t-copula functions, in which the Gauss copula function could be extended to the N-element model.

\[ G_{R}^{\phi}(u_1, u_2, ..., u_n; R) = \int_{-\infty}^{\varphi^{-1}(u_1)} \ldots \int_{-\infty}^{\varphi^{-1}(u_n)} \frac{1}{2\pi^{n/2}|R|^{1/2}} \exp \left\{ -\frac{1}{2} (x)^{T} R^{-1} (x) \right\} dx_1 dx_2 ... dx_n \]  

(12)

\( \varphi(.) \) and \( \varphi^{-1}(.) \) denotes the standard normal distribution function and its inverse function, respectively, \( R \) is the \( n \)-order correlation coefficient matrix, \( X = (x_1, x_2, ..., x_n) \) corresponds to \( n \) analysis variables, which is suitable for the same group of multivariable in system reliability analysis.

### 3.4. Selection and Estimation of Copula Function Model

Because the relationship between bus routes is very complex, it is very important to choose a copula function; it is necessary to analyze the correlation of indicators. This paper chooses a copula function based on empirical distribution analysis. The specific steps are as follows:

Let \( (x_1, y_1), (x_2, y_2), ..., (x_n, y_n) \) be a set of samples selected from the population \((X, Y)\). The population \((X, Y)\) has a joint distribution function \( H(x, y) \), and the corresponding copula is a copula function. The empirical joint distribution function is \( \hat{H}(x, y) \) for the empirical copula, \( \hat{C}(u, v) \) indicates that the copula function model is selected according to the scatter diagram of \((X, Y)\).

Before using the copula function to estimate the reliability of bus lines, the parameters of the copula function should be adjusted \((\theta)\) to make an estimate. At present, the most commonly used estimation method is maximum likelihood estimation. Suppose \( n \)-dimensional random vector, \( x_{1t}, x_{2t}, x_{3t}, ..., x_{nt} \), the edge distribution is \( F_1(x_1), F_2(x_2), ..., F_n(x_n) \), the joint density function is \( f(x_1, x_2, ..., x_n) \), and the joint distribution function is \( F(x_1, x_2, ..., x_n) \). Hypothesis \( u_i = F_i(x_i), \delta_i \) is the edge distribution parameter, where \( i = 1, 2, ..., n, \theta \) is a copula function parameter, then:

\[ f(x_1, x_2, ..., x_n) = C(u_1, u_2, ..., u_n) \prod_{i=1}^{n} f_i(x_i) \]  

(13)

Log likelihood function is:
\[
\ln L(x_1,x_2,...,x_n; \delta_1, \delta_2,...,\delta_n, \theta) = \sum_{i=1}^T \left( \sum_{t=1}^n \ln f_i(x_{it}; \delta_1, \delta_2,...,\delta_n, \theta) \right) + \ln C(F_1(x_{i1}; \delta_1), F_2(x_{i2}; \delta_2),..., F_n(x_{in}; \delta_n); \delta_1, \delta_2,...,\delta_n, \theta)
\]

(14)

According to the extremum theorem:

\[
(\hat{\delta}_1, \hat{\delta}_2,..., \hat{\delta}_n, \theta) = \arg\max L(x_1,x_2,...,x_n; \delta_1, \delta_2,...,\delta_n, \theta), \quad \text{then}
\]

\[
\frac{\partial L(x_1,x_2,...,x_n; \delta_1, \delta_2,...,\delta_n, \theta)}{\partial \delta_1} = 0
\]

(15)

\[
\frac{\partial L(x_1,x_2,...,x_n; \delta_1, \delta_2,...,\delta_n, \theta)}{\partial \delta_2} = 0
\]

(16)

\[
\frac{\partial L(x_1,x_2,...,x_n; \delta_1, \delta_2,...,\delta_n, \theta)}{\partial \delta_n} = 0
\]

(17)

\[
\frac{\partial L(x_1,x_2,...,x_n; \delta_1, \delta_2,...,\delta_n, \theta)}{\partial \theta} = 0
\]

(18)

4. Data and Model Calculations

4.1. Data Introduction

1. Raw data

Taking the bus route 430 in Beijing as an example, GPS data and vehicle scheduling data from July 2020 were selected for the case analysis. Starting from Dongshagezhuang bus station to Andingmen, the whole journey is about 21.4 km. There are 25 stations in this direction, and 24 inter-station routes in the middle, as shown in Figure 3.

Figure 3. The 430 bus route diagram.

The corresponding relationship between number and station is shown in Table 2.

| NO. | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   | 10  | 11  | 12  | 13  | 14  | 15  | 16  | 17  | 18  | 19  | 20  | 21  | 22  | 23  | 24  | 25  |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|     | Dongshagezhuang bus station | Dongshagezhuang | Pingxi Wangfu intersection south | South East banner  | South Station of Tiantong- yuan | Dongsanqiu South Station | Tiantong- yuan Tai pingzhuang | Tiantong Xiyuan North | Tiantong Xiyuan South | Lishuiqiao Station | Lishuiqiao North Station | Dayangfang |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |

The original GPS data are shown in the following table (Table 3).
Table 3. GPS data sample.

| CAR_CODE | LINE_CODE | UP_OR_DOWN | GPS_TIME      | LONGITUDE | LATITUDE | RUN_STATUS | PARK_STATUS | NEXTSTATTION_CODE | UPDATETIME |
|----------|-----------|------------|---------------|-----------|----------|------------|-------------|-------------------|------------|
| 27376    | 44        | 0          | 2020-07-06 7:30 | 116.3824333333 | 39.9477733333 | 1          | 1           | 0                 | 2020-07-06 7:31 |
| 22137    | 1         | 1          | 2020-07-06 7:30 | 116.3034833333 | 40.162125    | 1          | 0           | 0                 | 2020-07-06 7:31 |

1. Explains: CAR_CODE is the vehicle number; LINE_CODE is a line number; UP_OR_DOWN is the direction of the bus line, 0 for up and 1 for down; LONGITUDE and LATITUDE are in the WGS-84 coordinate system (World Geodetic System-1984 Coordinate System); RUN_STATUS is the state of operation of the vehicle, 1 for operation and 0 for non-operation; PARK_STATUS is the entry and exit status of the parking lot, 1 for in and 0 for out.

An example of raw data for a trip plan is shown in Table 4.

Table 4. Sample table of trip plan data.

| Date       | Line id | License Plate | Trip Index | Actual Departure Time | Actual Time of Arrival | Direction | Stardown | Planned Time of Arrival | Originating Terminal | Terminal       |
|------------|---------|---------------|------------|-----------------------|------------------------|-----------|----------|------------------------|----------------------|------------------|
| 2020-07-30 | 10919   | AN8721        | 2          | 04:06:32              | 05:40:35               | 1         | 04:06:00 | 2:05:41:00             | North of Jingzhang Road | Deshengmen       |
| 2020-07-30 | 10919   | AN8723        | 3          | 04:12:26              | 05:46:55               | 1         | 04:12:00 | 04:57:00               | North of Jingzhang Road | Deshengmen       |
| 2020-07-30 | 10919   | AW7333        | 7          | 04:18:40              | 05:51:02               | 1         | 04:18:00 | 05:53:00               | North of Jingzhang Road | Deshengmen       |
| 2020-07-30 | 918     | AV0516        | 1          | 04:30:00              | 06:20:00               | 1         | 04:30:00 | 06:20:00               | Pinggu new bus station | Dongzhimen Bus Hub |
| 2020-07-30 | 918     | AW1319        | 8          | 04:21:11              | 06:20:00               | 1         | 04:20:00 | 06:20:00               | Pinggu new bus station | Dongzhimen Bus Hub |

2. Data Processing

The raw data cannot be directly used due to the data completeness, noise, consistency, and other problems. In order to improve the data quality, it is necessary to preprocess the raw collected data, including missing data completion, error data deletion, duplicate data deletion, and GPS drift data processing. Taking GPS drift data processing as an example, the characteristics of GPS drift data are that the position of individual GPS data deviates seriously from the whole GPS data series.

\[
\Delta S_m = V_h (t_k - t_{k-1}) + 2r
\]  
(19)

where \(\Delta S_m\) is the current GPS positioning point relative to the previous GPS positioning point displacement; \(t_k\) is the current positioning point at positioning time; \(t_{k-1}\) is the previous positioning time; \(r\) is the range of errors of the GPS receiver, and \(V_h\) is the maximum speed limit of the road network of the target city, and the \(V_h\) of Beijing is 80 km/h.

\[
\Delta S = R \times \arccos (\sin (y_{k-1}) \sin (y_k) + \cos (y_{k-1}) \cos (y_k) \cos (x_{k-1} - x_k))
\]  
(20)

where \(\Delta S\) is the shortest spherical distance between two fixed points; \(R\) is the radius of the Earth; \(x_{k-1}\) and \(x_k\) are the longitude coordinates of GPS fixed points; \(y_{k-1}\) and \(y_k\) are the latitude coordinates of GPS fixed points. When \(\Delta S > \Delta S_{mr}\), a GPS position point is identified as a drift point.

3. Index calculation

According to Table 1 (in the second part), the line reliability evaluation index values are calculated through these original data. The specific steps are as follows:

- inter-station path average running time (\(T\)).

Step 1: Select the real-time bus dynamic data according to the route number, get the one-day arrival time of all the vehicles in a line, save the results as bus route arrival time table.
Step 2: According to the vehicle number and arrival time, get the arrival time of each vehicle.

Step 3: Each station is traversed, and the running time of each station is obtained by subtracting the time when the vehicles arrive at the station in turn.

Step 4: Calculate the average running time interval of each station path, that is, the average running time between stations of the line within the statistical period.

- Inter-station path average running speed ($V_j$)

Step 1: Get average running time ($\bar{T}$) above.

Step 2: Match the average running time of up and down bus routes with the distance between bus stations.

Step 3: Calculate the ratio

$$V_j = \frac{\sum l_i}{\sum T}$$

where $l_i$ is the operational mileage of the inter-station path.

- Coefficient of variation ($CV_T$)

$$CV_T = \frac{\delta_T}{\bar{T}},$$

where $\delta_T$ is the running time between stations:

$$\delta_T = \frac{\sqrt{\sum_{i=1}^{n} (\bar{T}_i - \bar{T})^2}}{n}$$

where:

- $\bar{T}_i$ is the inter-station path running time;
- $\bar{T}$ is the average inter-station path average running time within a statistical time period;
- $n$ is the total of inter-station paths.

Through statistical analysis of one month’s data of the line, the line reliability evaluation index results are obtained, as shown in Table 5 below.

**Table 5. Calculation results of bus line reliability evaluation indexes.**

| Index | $L_1$ | $L_2$ | $L_3$ | $L_4$ | $L_5$ | $L_6$ | $L_7$ | $L_8$ | $L_9$ | $L_{10}$ | $L_{11}$ | $L_{12}$ |
|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|
| $T_{kj}$ | 1.3   | 4.0   | 5.0   | 6.5   | 4.4   | 3.1   | 2.2   | 2.4   | 3.7   | 1.2    | 6.1    | 4.1    |
| $\delta_T$ | 1.2   | 2.0   | 2.3   | 2.1   | 1.9   | 2.8   | 2.4   | 2.8   | 1.3   | 1.6    | 1.6    | 2.0    |
| $CV_T$ | 0.9   | 0.5   | 0.5   | 0.3   | 0.4   | 0.9   | 1.1   | 1.2   | 0.4   | 1.3    | 0.3    | 0.5    |
| $V_j$ (km/h) | 11.3  | 23.2  | 17.4  | 24.7  | 13.4  | 16.3  | 13.0  | 22.9  | 13.5  | 27.2   | 14.9   | 19.5   |

4.2. Determine the Edge Distribution Model

Many distribution functions have been used for bus line running times, and include symmetric distributions, such as normal distribution and logistic distribution, asymmetric distribution, such as lognormal distribution and gamma distribution, and log logistic distribution. For headway distribution, exponential distribution and gamma distribution are usually used.

Based on past experience, assuming that bus 430 runs from Dongshagezhuang bus station to Andingmen, and the bus operation between stations follows a normal distribution, the probability distribution model, i.e., density function, is as follows:
\[ f(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right], -\infty < x < +\infty \] (23)

where \( \mu \) — the mean value of random variable \( x \);
\( \sigma \) — the standard deviation of the random variable \( x \).

Using the maximum likelihood estimation method, the parameters of the density function of the bus line with a confidence of 95%, \( \mu_{lj} \) and \( \delta_{lj} \), as in Table 6.

| Index | \( \mu_{lj} \) | \( \delta_{lj} \) |
|-------|----------------|-----------------|
| \( L1 \) | 1.3 | 4.0 |
| \( L2 \) | 6.0 | 5.0 |
| \( L3 \) | 4.4 | 2.3 |
| \( L4 \) | 3.1 | 1.9 |
| \( L5 \) | 2.2 | 2.8 |
| \( L6 \) | 2.4 | 2.4 |
| \( L7 \) | 3.7 | 1.3 |
| \( L8 \) | 1.2 | 1.6 |
| \( L9 \) | 6.1 | 1.6 |
| \( L10 \) | 4.1 | 2.0 |

4.3. Determine Copula Function Model

Suppose that bus line 430 is composed of 24 inter-station paths, \( L_i \), in series, and the life of the \( i \)-th inter station path is \( T_{lj} \), reliability is \( R_{lj} \). The life distribution function is \( F_{lj}(t), i = 1,2,3,...,24 \), then the lifetime of \( G \) is \( T = \min(T_{l1}, T_{l2}, \ldots, T_{l24} > t) \), and its joint distribution function \( H(t_1,t_2,...,t_n) = P(T_{l1} \leq t_1, T_{l2} \leq t_2, \ldots, T_{ln} \leq t_n) \). There is an n-dimensional copula, \( C \), such that \( H(t_1,t_2,...,t_n) = C(F_{l1}(t_1),F_{l2}(t_2),...,F_{l24}(t_n)) \). Because \( F_{l1}(t) \) is continuous, \( C(F_{l1}(t_1),F_{l2}(t_2),...,F_{l24}(t_n)) \) is unique. Then, the reliability of bus line \( G \) is calculated:

\[ R_g(t) = 1 - \sum_{i=1}^{24} F_{l24}(t) + \sum_{1 \leq i_1 \leq 1 \leq \ldots \leq i_n \leq 24} C(F_{l1}(t_1),F_{l2}(t_2),...,F_{l24}(t_n),t_1,t_2,\ldots,t_n) \] (24)

Based on the previous empirical analysis method, the copula function is selected, and the reliability of the bus line running time conforms to the normal distribution. Therefore, the reliability estimation function of the whole bus line is as follows:

\[ G_R^{G_n}(u_1,u_2,...,u_n; R) = \int_{-\infty}^{\phi^{-1}(u_1)} \ldots \int_{-\infty}^{\phi^{-1}(u_n)} \frac{1}{\sqrt{2\pi n^2|R|^{1/2}}} \exp \left[ -\frac{1}{2} \left( x^T R^{-1} x \right) \right] dx_1 dx_2 \ldots dx_n \] (25)

\( \phi(\cdot) \) and \( \phi^{-1}(\cdot) \) denotes the standard normal distribution function and its inverse function, respectively. \( R \) is the correlation coefficient matrix of order 24.

According to the theorem, \( X \) is a normal distribution random variable, \( X_1, X_2, \ldots, X_n \), the mean values of \( N \) are \( u_1,u_2,\ldots,u_n \). The variances were \( \delta_1,\delta_2,\ldots,\delta_n \). If the covariance matrix is \( R \), then the random variable \( U_i = \phi(\frac{X_i - u_i}{\delta_i}) \), \( i \in I \), \( C(u_1,u_2,\ldots,u_n) \) is the copula function, which is called a normal Gauss copula function with covariance matrix \( R \), and the linear correlation matrix \( R \) is symmetric, so the positive definite matrix is \( \text{diag} R = 1 \).

Through the maximum likelihood estimation method, the following conclusions are obtained

\[ \hat{u} = \frac{1}{24} \sum_{i=1}^{24} X^{(i)} \] (26)

\[ \sum = \frac{1}{24} \sum_{i=1}^{24} (x^{(i)} - u)(x^{(i)} - u)^T \] (27)

The reliability estimates of No.1–No.24 inter-station paths using historical data are shown in Figure 4.
Figure 4. Reliability estimation distribution map of inter-station route operation.

It can be seen from the Figure 4 that the average distribution of reliability between stations is above 0.85; only the fifth section is below 0.86, and the overall situation is relatively stable. By analyzing the three-dimensional diagram of the operation time distribution of the routes between stations, as shown in Figure 5, it can be seen that the overall trend of the operation time of the routes between stations tends to be consistent, which can also be understood as the correlation between the routes between stations. The previous operation affects the operation of the adjacent routes between stations, and has an obvious series effect.

Figure 5. Average operation time distribution of route between stations.

Through maximum likelihood estimation, taking a 95% confidence interval, the reliability estimation value of the whole line is 0.86. Through the analysis of Figure 4, the reliability estimates of the route between stations are all above 0.85, which indicates that the reliability of the whole line is obviously related to the reliability of the route between stations, and the maximum value of the reliability of the route between stations is 0.95,
The minimum value is 0.85. This shows that the reliability of the whole line is most affected by the path between stations with the smallest reliability, that is, the path number 5 has the greatest influence. The path with number 5 just happens to be from the south of the Tiantongyuan north station and to Dongsanqi south station, which may be due to the large passenger flow or the large number of social vehicles. Therefore, a reliability improvement project can be carried out for this section in the later stage to improve the reliability of the whole line. As shown in Figure 6, managers should find and improve the inter-station sections with a low reliability in the line during the whole day.

Figure 6. Route map.

5. Results and Discussion

Analyzing the relationship of route reliability between stations is the core of analyzing the key nodes of the whole route operation reliability and the relationship between the routes of each station. Through the copula function-based reliability estimation of public transportation lines, the reliability index relationship between different routes can be analyzed.

Through the above theoretical analysis and practical case analysis, the operation reliability of the one-way bus line and the relationship with the path reliability between stations can be obtained. Assuming that the operation state of the paths between stations obeys a normal distribution, then the copula function in the whole line operation reliability relation function is called the normal Gauss copula function with covariance matrix R.

Through the actual line analysis, the typical inter-station path, which affects the operation reliability of the whole line, can also be obtained, and the response reasons can be determined. The operational reliability of the bus line between stations is affected by the operation of adjacent shifts and the passenger flow and road conditions, which provides the basis for improving the reliability of the whole line or the optimization adjustment of the line in the later period.

Through the analysis of the reliability of the bus line between stations, it is found that the reliability of the whole line is most affected by the lowest reliability between stations, and further proves the correctness of the assumption that the whole line is a series system. However, the operation reliability of the same line in different driving directions may vary greatly. Therefore, when calculating the bus operation reliability, the same line needs to be calculated according to the two driving directions of the bus, and the reverse direction of inter-station path reliability is also different.

In addition, traditional research takes route operation between stations as an independent subsystem, but this paper assumes that there is interaction relationship between the route operation between stations of all lines. It is difficult to make a conclusive
comparison in this regard, because few previous studies have discussed the relationship between stations. The traditional research method of station reliability is based on a statistical method to calculate the distribution of the punctuality rate of the arrival time of each station. Variable sites are identified by the distribution of site reliability. Therefore, it is considered that the station may be near an unstable section of the line. This statistical result is essentially consistent with the critical path result identified in this paper. The advantage of this paper is that it can also see the relationships and values between this section and the reliability of the whole bus line; traditional statistical methods however do not. In the calculation process, most of the traditional research conclusions are adopted, that is, the assumption that the bus lines and inter station routes are in line with the normal distribution; in practice, they may not be in line with the normal distribution, which needs further verification and to overcome one of the limitations of current computing.

Because the research on the reliability of the structure based on the copula theory is still in the primary stage, because of the complexity of bus line operations, there is uncertainty in this field, so it is a feasible exploration to establish bus line reliability evaluations using the copula function. However, the parameter estimation of multidimensional the copula function is involved. Due to the increase in dimensions, the calculation time and difficulty may be increased, which is an aspect that needs to be improved in follow-up research.

6. Conclusions

The operation reliability of bus lines plays a crucial role in maintaining high-level service of public transport. This study aims to clarify the relationship between whole line and sites on influencing the operation reliability of bus lines. This study proposes a bus routes reliability evaluation method based on copula connection functions, and the actual line data are taken as a case study. The results show that the method reveals the relationship between the overall line and the inter-station paths in terms of operation reliability. This study finds the interstation path that is more critical of the whole line. Previous studies focused on the service level of the whole line by investigating the overall stability of the departure punctuality, line running time, and punctuality [44].

The establishment of the copula function to evaluate and analyze the operational reliability of bus lines will lay a foundation for later expansion to the overall operational reliability of a bus line network. At the same time, it plays an important role in the operation and optimization of a single line. The method can provide a good reference for diagnosis and evaluation of line operation and the optimization of bus routes. The results can support the sustainable development of public transportation. It can make improvements to parts of a route, according to the relevant conclusions. In addition, it can also be used to optimize the operation scheme of the entire bus line, so as to achieve the effect of improving the punctuality rate of bus operations, service level and public satisfaction [45]. The next step is to attract more people to choose public transport and finally to provide support for green and sustainable development.

Author Contributions: Conceptualization, validation, data curation, writing—original draft preparation, Z.X.; methodology and formal analysis, writing—review and editing and supervision, J.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National key R&D plan (Project No: 2017YFC0840206), program of China and the Science and Technology Project of the Beijing Traffic Commission under Grant 201905-ZHJ/C2, Central Pub-lic-interest Scientific Institution Basal Research Fund (2019–0111 and 2021–9017a).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting reported results can be found in the article.
Acknowledgments: This research has been funded by National key R&D plan (Project No: 2017YFC08040206), program of China and the Science and Technology Project of the Beijing Traffic Commission under Grant 201905-ZHJ2, Central Public-interest Scientific Institution Basal Research Fund (2019-0111 and 2021-9017a).

Conflicts of Interest: Authors declare no conflict of interests.

References
1. Beaudoin, J.; Farzin, Y. Public transit investment and sustainable transportation: A review of studies of transit’s impact on traffic congestion and air quality. Res. Transport. Econ. 2015, 52, 15–22.
2. Firew, T. Analysis of Service Reliability of Public Transportation in the Helsinki CapitalRegion: The Case of Bus Line 550. Master’s Thesis, Aalto University, Espoo, Finland, 5 March 2016.
3. Paudel, J. Bus ridership and service reliability: The case of public transportation in Western Massachusetts. Transp. Policy 2021, 100, 98–107.
4. Frederico, C.S.; Pereira, A.L.S.; Marte, C.L.; Yoshioka, L.R. Mobile application for bus operations controlled by passengers: A user experience design project (UX). Case Stud. Transp. Policy 2021, 9, 172–180.
5. Sterman, B.P.; Schofer, J.L. Factors affecting reliability of urban bus service. Transp. Eng. J. 1976, 102, 147–159.
6. Chapman, R.A. Bus Reliability—Definition and Measurement; Research Report no.18; Transportation Operations Research Group, University of Newcastle upon Tyne: Newcastle upon Tyne, UK, 1976.
7. Polus, A. Modeling and measurements of bus service reliability. Transp. Res. 1978, 12, 253–256.
8. Sorvatini, J.A.; Liu, R.; Sinha, S. Assessing Bus Transport Reliability Using Micro-Simulation. Transp. Plan. Technol. 2008, 31, 303–324. Available online: http://www.tandfonline.com/doi/abs/10.1080/03081060802086512 (accessed on 31 August 2016).
9. Ceder, A. Public Transit Planning and Operation Theory, Modelling and Practice; Butterworth Heinemann: Haifa, Israel, 2016.
10. Park, J; Kho, S.Y. A new method to determine level of service criteria of headway. In Proceedings of the IAROR Conference 2007 Proceedings, Hannover, Germany, 28–30 March 2007.
11. Strathman, J.G.; Dueker, K.J.; Kimpel, T.; Gerhart, R.; Turner, K.; Taylor, P.; Callas, S.; Griffin, D.; Hooper, J. Automated bus dispatching, operations control and service reliability: Analysis of Tri-Met baseline service data. Transp. Res. Record 1999, 1666, 28–36.
12. Turnquist, M. A model for investigating the effects of service frequency and reliability on bus passenger waiting times. Transp. Res. Record 1978, 663, 70–73.
13. Herbert, S.L. The reliability of transit service: An historical perspective. J. Urban Technol. 2005, 12, 99–118.
14. Oort, N.V.; Nes, R.V. Reliability of urban public transport and strategic and tactical planning, a first analysis. In Proceedings of the IAROR Conference 2007 Proceedings, Hannover, Germany, 28–30 March 2007.
15. Dai, S.; Chen, Y.; Wei, Z. Analysis on System Reliability of Complex Public Transit Network. J. Wuhan Univ. Technol. Transp. Sci. Eng. 2007, 3, 412–414.
16. Jang, Y. Study and Case Evaluation on Operational Reliability of Conventional Public Transportation in Large Cities. Master’s Thesis, Beijing Jiaotong University, Beijing, China, 2008.
17. Chen, X.; Yu, L.; Zhang, Y.; Guo, J. Analyzing urban bus service reliability at the stop, route, and network levels. Transp. Res. Part A Policy Pract. 2009, 43, 722–734.
18. Song, X.; Yu, L. Analysis of factors affecting bus operation reliability and improvement measures. Integr. Transp. 2019, 3, 59–62.
19. Transportation Research Board. Transit Capacity and Quality of Service Manual, 2nd ed.; National Research Council; Transportation Research Board: Washington, DC, USA, 2003.
20. Akbikovitz, M.D.; Engelstein, I. Factors affecting running time on transit routes. Transp. Res. Part A 1983, 17, 107–113.
21. Strathman, J.G.; Hopper, J.R. Empirical analysis of bus transit on-time performance. Transp. Res. Part A 1993, 27, 93–100.
22. Chioni, E.; Illopoulou, C.; Milioti, C.; Kepaptsoglou, K. Factors affecting bus bunching at the stop level: A geographically weighted regression approach. Int. J. Transp. Sci. Technol. 2020, 9, 207–217.
23. Diab, E.; Badami, M.; El-Genidy, A. Bus Transit Service Reliability and Improvement Strategies: Integrating the Perspectives of Passengers and Transit Agencies in North America. Transp. Rev. 2015, 35, 292–328.
24. Dakic, I.; Yang, K.; Menendez, M.; Chow, J.Y.J. On the design of an optimal flexible bus dispatching system with modular bus units: Using the three-dimensional macroscopic fundamental diagram. Transp. Res. B: Methodol. 2021, 148, 38–59.
25. Van Oort, N. Service Reliability and Urban Public Transport Design. Ph.D. Thesis, Delft University of Technology, Delft, The Netherlands, 2011.
26. Abrantes, P.; Wardman, M. Meta-analysis of UK values of travel time: An update. Transp. Res. Part A Policy Pract. 2011, 45, 1–17.
27. Arhin, S.A.; Noel, E.C.; Dairo, O. Bus Stop On-Time Arrival Performance and Criteria in a Dense Urban Area. Int. J. Traffic Transp. Eng. 2014, 3, 233–238. Available online: http://article.sapub.org/10.5923.ijtte.20140306.01.html (accessed on 5 September 2016).
28. Junskik, P.; Seung-Youn, K. A New Method to Determine Level-of-Service Criteria for Headway Adherence; Transportation Research Record: Washington, DC, USA, 2005; Volume 1927, pp. 277–286. Available online: https://www.doc88.com/p-4903409240677.html (accessed on 23 July 2021).
29. Hammerle, M.; Haynes, M.; McNeil, S. Use of automated vehicle location and passenger count data to evaluate bus operations-experience of the Chicagou Transit Authority. *Transp. Res. Record* 2005, 1903, 27–34.
30. Alonso-González, J.M.; van Oort, N.; Cats, O.; Hoogendoorn-Lanser, S.; Hoogendoorn, S. Value of time and reliability for urban-pooled on-demand services. *Transp. Res. C: Emerg. Technol.* 2020, 115, 102621.
31. Rietveld, P.; Bruinsma, F.R.; van Vuuren, D.J. Coping with unreliability in public transport chains: A case study for Netherlands. *Transp. Res. A Policy Pract.* 2001, 35, 539–559.
32. Esmailpour, J.; Aghabayk, K.; Vajari, A.M.; De Gruyter, C. Importance—Performance Analysis (IPA) of bus service attributes: A case study in a developing country. *Transp. Res. A Policy Pract.* 2020, 142, 129–150.
33. Hammerle, M.; Haynes, M.; McNeil, S. Use of automatic vehicle location and passenger count data to evaluate bus operations for the Chicago transit authority. In Proceedings of the TRB 2005 Annual Meeting, Washington, DC, USA, 9–13 January 2005.
34. Camus, R.; Longo, G. Transit reliability level-of-service estimation based on automatic vehicle location data. In Proceedings of the TRB 2005 Annual Meeting, Washington, DC, USA, 9–13 January 2005.
35. Sperstad, I.B.; Solvang, H.E.; Jakobsen, H.S.; Gjerde, O. Data set for power system reliability analysis using a four-area test network. *Data Br.* 2020, 33, 106495.
36. Yanchun, Y.; Chunfu, S.; Yusu, G. Study on travel time reliability model of road network based on real-time data. *Mod. Transp. Technol.* 2006, 2, 70–73.
37. Zhang, W.; Xu, W. Simulation-based robust optimization for the schedule of single-direction bus transit route: The design of experiment. *Transp. Res. E Logist. Transp. Rev.* 2017, 106, 203–230.
38. Ishaq, R.; Cats, O. Designing bus rapid transit systems: Lessons on service reliability and operations. *Case Stud. Transp. Policy* 2020, 8, 946–953.
39. Zhao, H.; Song, R. Study on reliability of Transit Operation. *J. Highw. Transp. Res. Dev.* 2005, 10, 132–135.
40. Gao, S.; Zhang, S.; Zhang, Y.; Gao, Y. Operational reliability evaluation and prediction of rolling bearing based on isometric mapping and NoCuSa-LSSVMReliab. *Eng. Syst. Saf.* 2020, 201, 106968.
41. Sharov, M.; Mikhailov, A. Urban Transport System Reliability Indicators. *Transp. Res. Procedia* 2017, 20, 591–595.
42. Boisjoly, G.; Grisé, E.; Maguire, M.; Veillette, M.-P.; Deboosere, R.; Berrebi, E.; El-Geneidy, A. Invest in the ride: A 14-year longitudinal analysis of the determinants of public transport ridership in 25 North American cities. *Transp. Res. Part A* 2018, 116, 434–445.
43. Eryilmaz, S. Multivariate copula based dynamic reliability modeling with application to weighted-k-out-of-n systems of dependent components. *Struct. Saf.* 2014, 51, 23–28.
44. Liang, S.; He, S.; Zhang, H.; Ma, M. Optimal holding time calculation algorithm to improve the reliability of high frequency bus route considering the bus capacity constraint. *Reliab. Eng. Syst. Saf.* 2021, 212, 107632.
45. Diab, E.; El-Geneidy, A. Bus Public Transport Planning and Operations. In *International Encyclopedia of Transportation*; Vickerman, R., Ed.; Elsevier: Amsterdam, The Netherlands, 2021; pp. 326–332.