# Analysis of County Bus Operation Status Based on Bus GPS Data

Shuyue Ge, Xianghong Li*

Henan Polytechnic University of Energy Science and Engineering, Jiaozuo, Henan, 454003, China,

(*lixianghong001@126.com)

**Abstract.** This article uses bus GPS data to obtain and analyze the operating status of bus vehicles, identify bus arrival and departure information, and evaluate bus operations. Taking bus GPS data as an example, the operation status of bus No. 6 in Weishan County, Jining City was analyzed in detail. Through data preprocessing and trajectory line matching, high-precision operating data was obtained, and key operating indicators such as one-way time consumption, operating speed, and arrival interval were calculated. Research results show that these operating indicators can effectively reflect the operating efficiency and service quality of the bus system and reveal the characteristics of bus operations. The analysis of this article provides a scientific basis for optimizing the county's public transportation services and promotes the modernization and intelligent development of the county's public transportation system.

**Keywords:** Bus operation status, GPS data, track line matching, operational indicators

## 1 INTRODUCTION

With the acceleration of urbanization, the public transportation system in county towns has gradually become one of the main ways for residents to travel in their daily lives. However, due to resource and technological limitations, compared to large cities, the public transportation system in county towns faces many unique challenges and opportunities. County towns usually have small geographical areas, low population density, and relatively concentrated transportation demand on a few major routes. The planning and adjustment of public transportation routes often lack scientific basis, making it difficult to respond to the constantly changing travel needs of residents in a timely manner. Therefore, how to improve the efficiency and service quality of public transportation remains a major challenge for county traffic managers.

In recent years, the development of GPS technology has provided new possibilities for the collection and analysis of bus operation data. Through in-depth analysis of these data, scientific support can be provided for the optimization of the bus system. Although many scholars have conducted in-depth research on transportation in large cities using public transportation GPS data, there is relatively little research on small



G. Chen et al. (eds.), *Proceedings of the 2024 International Conference on Rail Transit and Transportation (ICRTT 2024)*, Advances in Engineering Research 254,
https://doi.org/10.2991/978-94-6463-610-9_14

county towns. For example, Lin et al.[1] proposed three algorithms for matching vehicles with the bus network based on bus GPS data, evaluated their ability to predict bus arrival information from stability, accuracy, and robustness, and selected the best prediction algorithm. Based on GPS data of all bus routes in Shanghai, Wang Ling et al.[2] analyzed the significant factors and influencing patterns that affect the reliability of bus route travel time, and constructed a logistic regression model to quantitatively analyze the correlation between influencing factors and bus route reliability. Zhang[3] proposed a set of evaluation indicators for public transportation efficiency and stability based on fine-grained GPS trajectory data, combining GPS data, IC card data, POI data, and bus network topology data from multiple sources. Sun et al.[4] considered the historical speed of vehicles and the road conditions, combined with real-time GPS data, to obtain the average driving speed of public transportation vehicles on different road sections; Collected data on Chongqing Bus Route 871 for instance verification.

In the above context, this article takes the No. 6 bus in Weishan County, Jining City as an example, based on bus GPS data, and uses Python for data processing and analysis. By spatially matching the data with the geographic information of bus route stops, the operating status of buses is obtained, a bus operation diagram is drawn, the arrival and departure information of buses is identified, and relevant operating indicators are calculated to conduct in-depth analysis of their bus operation status. This article aims to reveal the current situation and potential problems of the public transportation system in small county towns (such as Weishan County, Jining City), and provide scientific basis for further optimizing public transportation services.

## 2 DATA PREPROCESSING

Bus GPS data (i.e. trajectory data) records key information such as vehicle ID, GPS time, longitude and latitude, speed, and up and down directions, and is an important data source for analyzing the operation status of buses. Among them, data preprocessing is a key step to ensure the accuracy and reliability of bus operation analysis. By cleaning and integrating raw data, noise and outliers can be eliminated, data quality can be improved, and the effectiveness and accuracy of subsequent analysis can be ensured. The preprocessed high-quality data lays a solid foundation for accurately evaluating the operation status of public transportation and proposing optimization suggestions.

### 2.1 Sampling Interval

The sampling interval of GPS data is very necessary for subsequent research on the operation characteristics of public transportation. The higher the sampling interval frequency of public transportation trajectory data, the more accurate the description of its operation trajectory will be. If the data sampling interval is fine enough, the acceleration and deceleration status as well as the entry and exit behavior of public transportation vehicles can be clearly described; If the sampling interval is rough, the arrival and departure characteristics of public transportation vehicles cannot be accurately characterized. According to literature, it is recommended that the sampling interval of trajectory data for studying the operation status of public transportation should not exceed 30 seconds[5]. The sampling interval kernel density probability distribution of this dataset is shown in Figure 1. The sampling interval is concentrated around 1 second, belonging to high-frequency sampling with good continuity. The proportion of available data is high, and it has high accuracy and credibility, which can more accurately reflect the dynamic operation of buses.

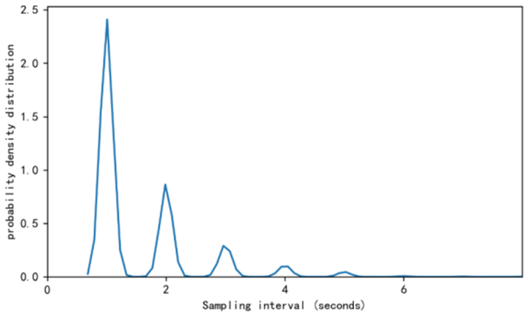



**Fig. 1.** Sampling interval kernel density probability distribution map.

## 2.2 Trajectory Matching

Trajectory route matching is a method of checking whether GPS points are on the path. Due to interference factors such as equipment and satellite signals, there is a certain degree of error between the GPS data recorded by the equipment and the actual location of the bus. In order to reduce the error of trajectory data, the driving trajectory of public transportation vehicles is matched with the GIS information of the route and corrected to eliminate outliers.

This study used the Kd-Tree method for map matching[6][7]. Kd-Tree can help efficiently locate the nearest road point to GPS points. By constructing a Kd-Tree for road networks, it is possible to quickly locate the nearest neighbors of each GPS point in complex road networks, thereby achieving accurate map matching. The specific steps are as follows:

(1) Project the GPS data of public transportation to the nearest location of the bus route using the project method.

Set the original $G=(g_{1,}g_2,...,g_i,...g_n)$. Using the project function to retrieve the closest position point (i.e. matching point) between each point $g_i$ and the starting point $l$ of the line with a radius of r, obtain the distance $D=(d_1,d_2,...,d_i,...d_n)$ between each matching point and the starting point $o$ of the line, and output the distance set. The rules for distance retrieval can represent:

$$D = \underset{\forall[d^{(j)}\in l]}{argmindist}\left(d_i^{(j_n)},o\right) \tag{1}$$

In the formula: $d_i^{(j_n)}$ is the distance to be searched on the bus route $l$; $dist(d_i^{(j_n)}, o)$ is the Euclidean distance from the starting point $o$ of $d_i^{(j_n)}$.

(2) Use the interpolation method to obtain the geometric shape of the nearest point as the result of map matching.

Based on the Euclidean distance $dist(d_i^{(j_n)}, o)$ between the original point $G=(g_{1,}g_2,...,g_i,...g_n)$ and the matching point output by (1) and the starting point $o$, interpolate the generated distance set $D=(d_1,d_2,...,d_i,..d_n)$ to obtain the matching point coordinates $P=(p_1,p_2,...,p_i,...,p_n)$ and output them. Input the distance set $D=(d_1,d_2,...,d_i,...d_n)$, retrieve the matching point $P=(p_1,p_2,...,p_i,...,p_n)$ corresponding to each point $g_i$ on the line $l$, and output the coordinates of the matching points.

(3) Draw a probability density distribution map of the distance values between the original data points and the matching data points, and screen out outliers at long distances. As shown in Figure 2, the distance values between the original point and the matching point are mainly concentrated within 20 meters. Therefore, data with a distance value exceeding 20 meters are considered as outliers and removed to make the data more reliable.

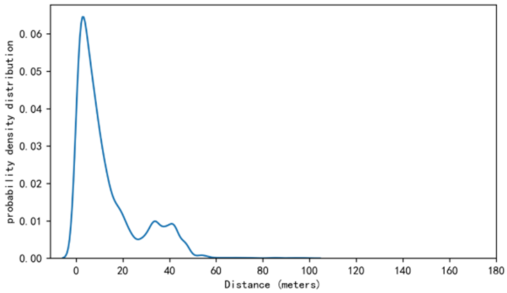



**Fig. 2.** Probability density distribution map of distance values between original data points and matching data points.

## 2.3 Run Chart Drawing

After completing the sampling interval adjustment and map matching, we will use the processed data to draw a bus operation map. In GPS data, take the data time as the x-coordinate and the location of the bus as the y-coordinate. Connect the GPS points of the vehicles with lines on the graph to obtain the vehicle's operation chart. From the timetable, it is possible to accurately determine when the bus departs, arrives at the station, where there is a delay on the bus route, and the speed of the vehicle during operation. The graph shown in Figure 3 shows the daily operation of a bus, and the slope of the curve in the graph represents the operating speed of the bus at this time.

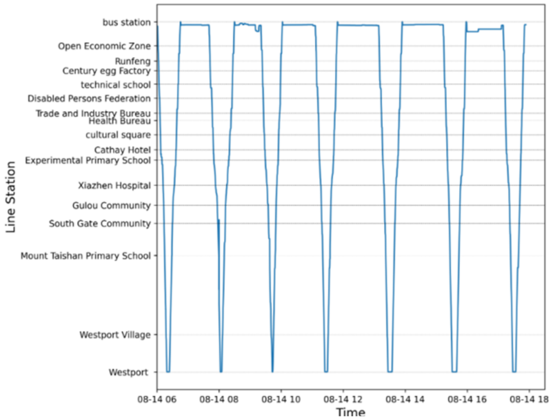



**Fig. 3.** The daily operation chart of a certain bus on Weishan Route 6.

### 2.4 Principle of Bus Arrival and Departure Information Recognition

The accurate identification of bus arrival and departure information is an important foundation for evaluating the operation status of public transportation. Using the speed changes and dwell time characteristics in GPS data, combined with the geographical location of bus stops, automatically identify the arrival and departure information of buses. The recognition principle is that when a vehicle arrives near a station, it undergoes a process of deceleration upon arrival, stopping, starting acceleration, and then leaving the station. If a buffer zone is set within a certain distance range around the station, as long as the intersection of the bus schedule and the gray area is taken, the starting and ending positions of the curve can be identified as the arrival and departure times. At the same time, the arrival and departure times of the vehicle can be obtained by looking at the abscissa position of the departure station[8][9].

## 3 BUS OPERATION INDICATORS

After data preprocessing is completed, the key to analyzing the operation status of public transportation lies in selecting and calculating appropriate operational indicators. These indicators not only objectively reflect the operational efficiency and service quality of the public transportation system, but also provide data support for improvement and optimization. This study selected one-way travel time, operating speed, and single station interval as the main operating indicators to analyze the reliability of public transportation itinerary[10].

### 3.1 One Way Time Consumption

The one-way travel time represents the total running time of the bus from the starting station to the destination station, including the stopping time. The one-way travel time can directly reflect the overall operational efficiency and service quality of the bus route. Shorter one-way travel time usually means higher operational efficiency and faster service speed, which is beneficial for improving the travel experience of passengers. The one-way travel time can be expressed by the following formula:

$$T = T_a - T_d \tag{2}$$

In the formula: $T$ is the one-way travel time of the bus; $T_a$ is the arrival time of the bus; $T_d$ is the departure time of the bus.

### 3.2 Operating Speed

The operating speed represents the average speed of a bus during its operation. Used to measure the overall efficiency of the public transportation system and whether it can reach its destination on time. A higher operating speed means that buses can complete routes in a shorter time, reduce waiting and boarding time for passengers, improve bus turnover, and serve more passengers. The operational speed can be expressed using the following formula:

$$\mathrm{V_{avg}} = \frac{\mathrm{D}}{\mathrm{T}} \tag{3}$$

In the formula: $V_{avg}$ represents the average speed; D represents the total distance of the bus route; T represents one-way time consumption.

### 3.3 Arrival Interval

The arrival interval refers to the time interval between two consecutive buses passing through the same station. This indicator directly reflects the frequency of departure between two consecutive buses on the same route. A shorter single stop interval usually means that the bus system provides more frequent services, allowing passengers to take the next bus without waiting for too long, thereby improving the service frequency and accessibility of the bus system. The arrival interval can be expressed using the following formula:

$$\mathrm{I} = \mathrm{T_2} - \mathrm{T_1} \tag{4}$$

In the formula: $I$ represents the arrival interval (interval time); $I_1$ represents the time when the first bus passes through a certain station; $I_2$ represents the time when the second bus following closely passes through the same station.

## 4 CASE ANALYSIS

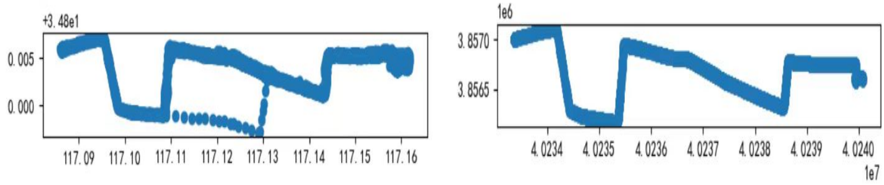
This article uses GPS data of public transportation vehicles on Route 6 in Weishan County on August 14, 2023 for research. The data for this date is chosen because it is a working day and can represent the bus operation on regular working days, avoiding interference from holidays or special events on the data. In addition, the weather conditions on that day were normal, with no extreme weather or special traffic control situations, ensuring the reliability and representativeness of the data. The data volume of this study is about 20000 pieces, including detailed information such as the vehicle ID, route name, longitude and latitude, operating speed, recording time, and up and down directions of public transportation vehicles. The specific data format is shown in Table 1[5]. Statistical analysis of data is conducted using Python's numerical analysis library Pandas and geospatial analysis library GeoPandas, including data preprocessing, map matching, characterization of bus operation status, and analysis of bus operation indicators.

**Table 1.** Weishan County Bus Data Format.

| Field Name | Example | Explain |
|---|---|---|
| License number | 9156 | License plate |
| Driver ID | 11813 | Driver identification |
| GPS time | August 14, 2023 08:00:00 | Data collection time |
| Line name | 3006 | The name of the route where the vehicle is traveling |
| Longitude | 117.152 | East longitude |
| Latitude | 34.806 | North latitude |
| Up and down | Up | Up and down directions |
| Site Name | Area for development | The name of the station where the vehicle has arrived |

### 4.1 Analysis of Site Matching Results

According to the trajectory data processing steps, preprocess the GPS data of the No. 6 bus in Weishan County. The results before and after map matching are shown in Figure 4. The distribution of data points before matching is uneven, with obvious errors and deviations, especially in some road sections where abnormal jumps and deviations occur, reflecting the noise and outliers in the original GPS data. After being processed by the Kd-Tree method, the matched data points have a more coherent and consistent distribution, with significantly smoother trajectories, and are more in line with the actual geographical location of bus routes.




(a)Before matching (b)After matching

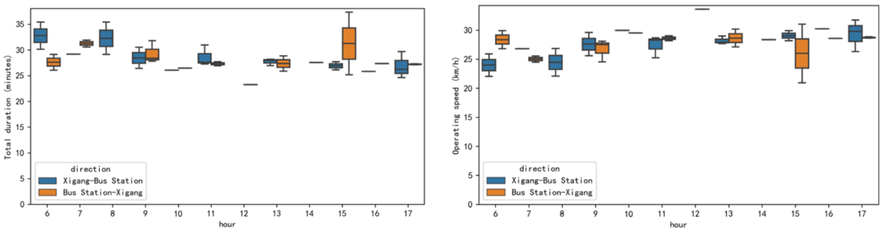
**Fig. 4.** Schematic diagram of the results before and after matching the bus trajectory data with the route.

## 4.2 Analysis of Operational Indicators

Based on the bus timetable, calculate the one-way travel time, operating speed, and arrival time interval of the bus vehicles.

The one-way travel time is the time taken by the bus from the starting point to the destination. The time distribution box diagram is shown in Figure 5 (a). From the figure, it can be seen that there are certain differences in one-way travel time between different time periods. During the morning rush hour (6-9 o'clock), the one-way travel time at 6 o'clock is concentrated between 20-25 minutes, and it takes about 20 minutes from the west port to the bus station at 7 o'clock. The direction from the bus station to the west port is slightly shorter, which is 15-20 minutes, and it stabilizes at around 20 minutes in both directions at 8 o'clock. During non peak hours (10-15 o'clock), the one-way travel time from 10 o'clock to 11 o'clock is reduced to about 15 minutes, increased to 20 minutes at 12 o'clock, approached 25 minutes from West Port to the bus station at 13 o'clock, and further reduced to about 15 minutes at 14 o'clock and 15 o'clock. The one-way travel time during the evening peak period (16-17 o'clock) has increased again to about 20 minutes. Overall, the morning and evening peak hours (such as 6, 7, 16, and 17) take relatively longer, reflecting the impact of traffic congestion during peak hours. During non peak hours, such as noon and early afternoon, the one-way travel time is relatively short, indicating that the traffic conditions during this period are relatively smooth. The one-way travel time in the two directions is roughly similar, but there are slight fluctuations in certain hours, mostly distributed in different directional traffic flows.

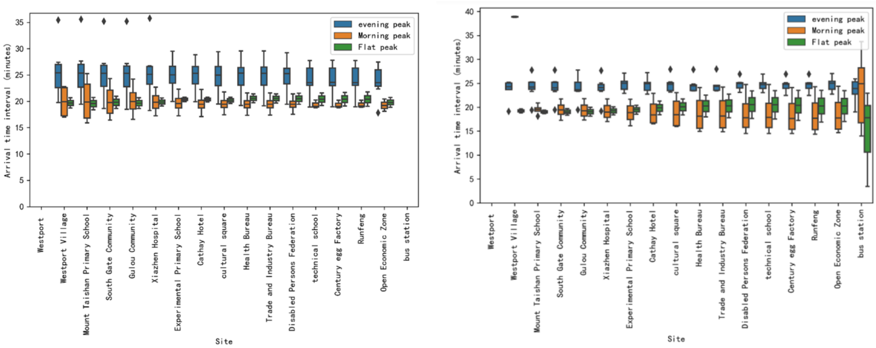
The operating speed box diagram is shown in Figure 5 (b), and there are certain differences in operating speeds during different time periods. Usually, the speed is lower during morning and evening peak hours. During morning peak hours, the operating speed in the direction from West Port to Bus Station at 6 o'clock is 25-27 km/h, while the operating speed in the direction from Bus Station to West Port is slightly higher, at around 30 km/h; During peak hours in the evening, the operating speed significantly decreases, both between 20-25 km/h, reflecting traffic congestion during peak hours. The operating speed during non peak hours is relatively high, both at 27-30km/h, indicating that the traffic conditions are relatively smooth during this period. Therefore, by optimizing route design and improving scheduling flexibility, the operational efficiency of public transportation during peak hours can be improved to a certain extent.




(a)One way time consumption. (b)Operating speed.

**Fig. 5.** Box plot of one-way time consumption and operational speed results.

To make statistics on the arrival time interval, it is necessary to distinguish the direction of arrival and mark whether the arrival record belongs to the morning peak, evening peak, or off peak period. According to the traffic operation characteristics of Jining City, the morning rush hour is from 6:00 to 9:00, the evening rush hour is from 16:00 to 18:00, and the rest of the time is during off peak hours. The box diagram of the arrival time interval is shown in Figure 6.




(a)West Port Bus Station Direction. (b)Bus Station - West Port Direction.

**Fig. 6.** Box diagram of arrival time interval.

Most stations in the direction of Xigang Bus Station have an arrival time interval between 20 and 30 minutes during evening and morning peak hours, while the arrival time interval during peak hours is relatively short and stable, about 10 to 20 minutes. Some stations, such as Xigang Village and Xigang, have significant outliers during the evening peak hours, with time intervals exceeding 30 minutes. On the whole, the change trend of the time interval between the bus station and the west port is similar to the upward trend. The arrival time interval in the evening peak and the morning peak is mostly between 20 to 30 minutes and 15 to 25 minutes. The time interval of some stations such as the west port in the evening peak is close to 40 minutes, and some stations such as Mount Taishan Primary School and Nanmenkou Community in the morning peak have more outliers. The arrival time interval during peak hours is between 10 and 20 minutes, which is generally stable. Overall, the punctuality rate of buses during peak hours needs to be improved, especially focusing on stations with large fluctuations in arrival time intervals. By optimizing routes and increasing transportation capacity, passengers can improve their travel experience. Meanwhile, despite performing well during peak hours, further optimization is still needed to ensure that buses maintain a high on-time rate and stable arrival time intervals throughout all time periods.

## 5 CONCLUSION AND PROSPECT

Using GPS data of public transportation to analyze the operational status of urban public transportation and evaluate the reliability indicators of public transportation routes. An experimental demonstration was conducted on the bus trajectory data of Weishan County on August 14, 2023, and the operational status of a single route was analyzed. The specific conclusion is as follows:

(1) One-way travel time: The analysis results show that during peak hours (morning and evening), one-way travel time significantly increases, mainly influenced by factors such as traffic congestion and frequent boarding and alighting of passengers. The one-way travel time during off peak hours is shorter, and the bus operation is smoother.

(2) Operating speed: The operating speed during peak hours is relatively low, especially on heavily congested road sections. By optimizing route design and improving scheduling flexibility, operational speed can be improved to a certain extent.

(3) Arrival interval: The arrival time interval of buses fluctuates greatly during peak hours, and some stations have obvious outliers, which affect the on-time rate and performance during off peak hours, but there is still room for improvement.

Based on the analysis of the above conclusions, the following optimization suggestions are proposed. Firstly, increase the frequency of departure during peak hours to reduce passenger waiting time and alleviate traffic congestion. Secondly, optimize bus routes, choose alternative routes with less congestion or establish dedicated bus lanes to improve operational speed during peak hours. Establish a real-time monitoring and dynamic scheduling system, adjust departure time and interval in a timely manner, ensure high-frequency operation during peak hours, and flexible scheduling during non peak hours.

The next step of the work will combine weekly and monthly trajectory data to refine the study of the situation where the up and down directions of bus routes do not overlap, evaluate it, and provide suggestions to improve the efficiency of public transportation in county towns.

## REFERENCES

1. Lin, Wei-Hua, and Jian Zeng. "Experimental study of real-time bus arrival time prediction with GPS data." Transportation Research Record 1666.1 (1999): 101-109.
2. WANG Ling, WANG Yanli, QI Xinzhou, et al. Analysis of Reliability Factors of Conventional Bus Travel Time Based on GPS. Journal of Wuhan University of Technology (Transportation Science and Engineering), 2022, 46(1): 1R5.
3. Zhang, Hui, et al. "Analysis of Spatial-Temporal Characteristics of Operations in Public Transport Networks Based on Multisource Data." Journal of Advanced Transportation 2021.1 (2021): 6937228.
4. Sun, Dihua, et al. "Predicting bus arrival time on the basis of global positioning system data." Transportation Research Record 2034.1 (2007): 62-72.
5. Deng Hongxing, Bai Yiyang. Bus Operation Status Recognition Considering Trajectory Data. Journal of Chongqing University of Technology (Natural Science), 2023, 37(2): 233-240.
6. Anwar, Muhammad, et al. "Subgraph Query Matching in Multi-Graphs Based on Node Embedding." Mathematics 10.24 (2022): 4830.
7. Zou, Qian, et al. "Research on Real-Time Anomaly Detection Method of Bus Trajectory Based on Flink." Electronics 12.18 (2023): 3897.

8. Ou, Yuming. "AI for real-time bus travel time prediction in traffic congestion management." Humanity Driven AI: Productivity, Well-being, Sustainability and Partnership (2022): 63-84.
9. Skhosana, Menzi, and Absalom E. Ezugwu. "A Real-Time Machine Learning-Based Public Transport Bus-Passenger Information System." International Journal of Sustainable Development & Planning 16.7 (2021).
10. Byon, Young-Ji, et al. "Bunching and headway adherence approach to public transport with GPS." International Journal of Civil Engineering 16 (2018): 647-658.