sustainability

MDPI

Article

Revealing Freight Vehicle Trip Chains and Travel Behavior: Insights from Heavy Duty Vehicle GPS Data

Bo Yu [1,2], Gaofeng Gu [3], Yuandong Liu [4,*] and Yi Li [4]

[1] College of Computer Science, Chongqing University, Shapingba District, Shazheng Road #174, Chongqing 400044, China; boyu_cq@hotmail.com
[2] Information Center, Chongqing Transport Planning Institute, No. 339 Longshan Avenue, Yubei District, Chongqing 401120, China
[3] College of Transportation Engineering, DaLian Marinetime University, No. 1, Linghai Road, Dalian 116026, China; ggu@dlmu.edu.cn
[4] School of Transportation Engineering, Chang'an University, Xi'an 710064, China; liyi@chd.edu.cn
* Correspondence: yuandong_liu@hotmail.com

Abstract

High-quality, well-structured trip chain data are essential for analyzing the daily activity patterns, travel behaviors, and logistical decisions of commercial vehicles, as well as for supporting sustainability-oriented freight management and low-carbon urban logistics. This study introduces a novel methodology for analyzing truck travel patterns using extensive GPS data, focusing on identifying freight trip chains and enhancing urban freight systems. A road-constrained clustering approach was developed to accurately identify vehicle stops and truck stop locations, addressing limitations in previous studies that struggled with misclassification. A trip chain reconstruction methodology was formulated, key characteristics were extracted and clustering techniques were applied to categorize trucks based on their travel behavior. A case study in Chongqing demonstrates that the proposed method outperforms traditional clustering algorithms, reducing misclassification rates in stop location identification. The findings reveal consistent trip chain patterns and distinct travel behaviors within truck groups. This research presents a data-driven framework that provides a foundation for optimizing logistics, fleet management, and low-carbon freight system planning. By enhancing the accuracy of trip chain analysis, this methodology contributes to the design of energy-efficient and sustainable urban freight systems, helping reduce emissions and foster eco-friendly logistics solutions.

Keywords: GPS data; truck trip chains; travel patterns; heavy-duty vehicle; sustainable urban freight

Check for updates

Academic Editor: Carlos Llopis-Albert

Received: 23 November 2025
Revised: 22 January 2026
Accepted: 23 January 2026
Published: 28 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

A well-functioning freight system is essential for efficient supply chain operation and timely delivery of goods. The organizational layouts of freight locations—serving manufacturing, distribution, and transportation—form the backbone of urban freight networks. Large trucks facilitate critical interactions between metropolitan regions and businesses, connecting markets, logistics hubs, and industrial centers. To align with sustainability goals, freight networks must evolve by incorporating low-carbon technologies, such as electric trucks and alternative fuels, reducing emissions, and enhancing energy efficiency. As cities grow, optimizing these networks for sustainability becomes essential, helping to reduce carbon emissions, improve air quality, and alleviate traffic congestion, all of

which contribute to urban well-being. However, the data supporting the evaluation of the dynamics and interaction patterns of different freight locations, as well as a comprehensive analysis of the Heavy Duty Vehicle (HDV) behavior, are still largely unexplored. Addressing this data gap is critically important beyond methodological advancement. The freight transportation sector, and heavy-duty trucking in particular, is a major and growing contributor to global greenhouse gas emissions and urban air pollutions. Current operational patterns and emissions trajectories of HDV fleets are significantly out of sync with international net-zero climate targets, creating an urgent need for data-driven insights to inform effective policy and planning [1]. A precise understanding of freight vehicle behavior—including trip chains, dwell times, and facility interactions—is fundamental to designing targeted interventions such as zero-emission zones, logistics consolidation schemes, and infrastructure for alternative fuels. Therefore, this study aims to contribute to this urgent research priority by developing and validating a framework to reconstruct and analyze HDV activity patterns from GPS data, thereby providing a scalable tool to support evidence-based urban freight management and sustainability policy.

A truck trip chain or tour is a sequence of trips that includes the origins, destinations, and intermediate stops a vehicle makes during a journey. For commercial vehicles, these trip chains often consist of diverse purposes, such as goods pick-ups, deliveries, and non-freight-related activities like rest or refueling. By connecting various business establishments, truck trip chains create spatial interactions between various facility locations, linking different parts of one or more supply chains. Thus, analyzing the trip chain patterns of commercial vehicles can yield valuable insights into supply chain agents' decision-making and behavior. In addition, trip chain data are also a fundamental component of freight modeling [2,3]. Trip chains have a behavioral foundation; the interconnected trips within a tour are considered collectively to reflect their logistical interactions [4]. Freight tours, driven by economic decisions aimed at minimizing logistics costs, are particularly suited as the analytical unit for freight movement studies. This makes trip chain analysis essential for understanding the activity patterns, logistical decisions, and travel behaviors of commercial vehicles.

Therefore, high-quality, well-structured trip chain data are needed for freight analysis and modeling. Traditionally, trip chain data has been collected through driver surveys [5–7], where truck drivers document trip details, such as stops and purposes. While this method provides rich and accurate data, it is labor-intensive, costly, and often lacks comprehensive and timely coverage, limiting its utility for freight modeling. The advent of GPS technology has revolutionized trip chain data collection. Although GPS data does not directly provide behavioral information, it provides detailed vehicle trajectory information, enabling the reconstruction of trips and trip chains from these trajectories and the inference and extract behavioral knowledge using a variety of techniques [8–12]. Extensive studies have utilized GPS data for various research purposes, including trip end identification [13,14], trip purpose inferring [9,15–19], commercial vehicle travel pattern analysis [9,20], and freight modeling [4,21,22], and, of course, truck trip chain mining.

Compared to passenger trip chains, which typically follow simpler home-based or work-based patterns, commercial vehicle trip chains are generally more complex. Their identification requires determining the start and end of a tour, for which no unified definition exists. The literature offers various perspectives on freight tour definitions. Some studies define the end of a tour as a return to a base location [22–24], while others consider a new tour to begin after a delivery stop is followed by a pick-up stop [25,26]. One study explores various definitions of freight vehicle trip chain/tour chains [23]. The study evaluated three commonly used definitions, base-based trip chains, trip purpose-based trip chains and capacity-based trip chains. It compares the extracted tour chains according to different definitions and demonstrates that tour chain types are highly dependent on

the tour chain definition. These differing definitions reflect varied analytical priorities: base-driven methods emphasize activity regularity, while others focus on trip purposes or vehicle capacity usage. For freight modeling and prediction purposes, base-driven methods are widely favored due to their ability to capture the regularity of truck activities, particularly since GPS data typically lacks trip purpose and vehicle capacity information.

In general, the reconstruction of trip chains typically involves three steps: (1) identifying truck stops from raw GPS data; (2) determining trip ends; and (3) identifying distinct stop locations. Generally, speed threshold method is used to identify truck stops [19,27]. A truck stop is indicated when the vehicle speed drops below a predefined threshold (e.g., 5 km/h). Once stops are identified, they are classified based on their purpose. Numerous studies have focused on inferring trip purposes from GPS data. Most studies categorize freight-related purposes into two groups: (1) freight-related, such as delivery, unloading, and loading, and (2) non-freight-related activities, such as rest, refueling, and dining. For truck behavior analysis, non-freight-related stops are typically excluded, as freight-related activities are more relevant for understanding truck travel patterns and the relationships between different trip ends. Once trip ends are identified, truck trips and trip chains can be reconstructed based on the behavioral information from these stops.

The third step, identifying distinct truck stop locations, has received relatively little attention. While truck stops refer to any stopping event by a truck, trip stop locations specifically refer to sites such as business establishments, company-owned parking lots, logistics depots, and freight hubs where trucks stop for similar purposes. Accurately identifying these stop locations is crucial for reconstructing trip chains for both individual trucks and groups of trucks. Correct identification ensures that stops at the same location are grouped together rather than treated as distinct stops, improving the accuracy of trip chain reconstruction. For grouped trucks, identifying shared stop locations also facilitates the evaluation of collective behaviors, such as trip chain similarities, and provides insights into stop location characteristics, such as base classifications. This is particularly important as freight transport patterns vary across different facility types and locations [28].

Despite its importance, relatively few studies have focused on identifying stop locations. Two common methods are used in the literature to determine when vehicles stop near each other. One approach employs spatial constrained methods, such as grid-based methods or the Voronoi method, to identify and characterize truck stops within a region [10,29]. Spatial constrained methods divide a space into different regions, which can result in an establishment being split across multiple regions. This method tends to favor identifying collective behaviors rather than distinguishing individual locations. Another widely used approach involves clustering techniques [24,30–33], such as the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) method, which is valued for its ability to detect clusters of varying shapes and for its robustness to noise [34]. We tested the DBSCAN algorithm using practical data and found that it cannot effectively differentiate distinct stop locations in areas with high truck stop densities as the spatial constrained is not taken into consideration.

This study therefore aims to uncover trip chains from the GPS data, revealing the patterns of activity that groups of trucks display, and offering a strong data base for improving freight planning tactics. For this purpose, this study introduces an integrated method that takes the advantage of both the spatial-constrained method and the DBSCAN method, namely, a roadway-constrained method to identify trip ends and stop locations from GPS data, reconstructing trip chain information for individual vehicles, and classifying trucks based on travel behavior. This study contributes to the literature in three key areas:

Development of the Roadway-Constrained Method: A novel, road-constrained method is proposed to accurately identify truck stop locations, addressing the limita-

tions of previous studies that relied on traditional clustering algorithms, which struggled to effectively capture distinct truck stop locations.

Introduction of an Iterative Procedure for Trip Chain Identification: An iterative procedure is introduced to determine truck trip chains, offering empirical evidence that supports the planning of sustainable freight systems.

Identification and Classification of Consistent Truck Travel Behaviors: This study identifies consistent truck travel behaviors and classifies trucks into distinct groups based on trip chain properties, providing deeper insights into truck behavior. By classifying trucks according to these properties, this study enhances the understanding of freight dynamics and offers valuable implications for sustainable freight management.

The proposed framework can accurately extract trip chain data from GPS trajectories and profile trucks into distinct behavioral groups for analysis. By addressing the complexities of trip chain identification and leveraging GPS data, this study contributes to the advancement of freight modeling and the understanding of commercial vehicle travel behavior.

This paper is organized as follows: Section 2 describes the dataset used for this study and defines truck trip chain. Section 3 introduces the trip chain mining methodology and truck classification methods. Section 4 summarizes the results of this study. Finally, Section 5 concludes this study.

2. Data Description and Trip Chain Representation

2.1. Data Description

GPS trajectory data: GPS data for HDVs in this study were collected from Chongqing, China. The dataset consists of over 2 million GPS pings recorded from more than 108,000 heavy trucks between 1 September and 7 September 2023. GPS signals were sampled at intervals of approximately 30 s, capturing key information such as timestamp, latitude, longitude, speed, and direction angle. Due to signal loss and device malfunctions, the dataset includes instances of missing data. To ensure data quality, a preprocessing step was implemented to handle the missing GPS signals.

Road network data: road network data were utilized in the analysis to establish road-constrained regions for accurately identifying single stop locations. For reproducibility, this study employed open-source road network data from OpenStreetMap. This integration of road network data enhances the robustness of the stop identification procedure.

Freight-related Area of Interest (AOI) data: Freight-related AOIs refer to specific regions or areas of significant interest, typically associated with key industrial activities. In this study, AOI data for Chongqing was sourced from Baidu Map (https://map.baidu.com/ (accessed on 3 June 2023)), identifying a total of 7402 freight-related AOIs within the city. Note that the coverage of AOIs provided by Baidu Map is limited. As a result, this dataset was used for validation purposes in this study.

2.2. Trip Chain Representation

Trip chain, or tours for trucks or HDVs, are defined in different ways in the literature. Previous studies have explored three definitions of trip chains based on the regularity of activities: the base-driven approach, the purpose-driven approach, and the capacity-driven approach [23]. We adopted the base-driven method, consistent with the approaches taken by the previous researchers [10]. Within this framework, each truck's journey is conceptualized as a chain of activities originating from a base location and proceeding through multiple intermediate stops. Consequently, a trip chain of truck t can be represented as a sequential series of activities: $TC_t^i = (b_t, s_1, s_2, ...s_n)$, where i denotes the trip chain index, t

refers to the truck number, and b_t represents the base of truck t. Figure 1 illustrates three distinct trip chains for a t:

$$\mathrm{TC}_1^1 : \ base\ 1,\ stop\ 1,\ stop\ 2,\ stop\ 3$$

$$TC_1^2 : \ base\ 1,\ stop\ 4,\ stop\ 5$$

$$TC_2^1 : \ base\ 2,\ stop\ 1,\ stop\ 2$$

TC_1^1 represents the first trip chain of truck 1 and TC_1^2 represents the second trip chain performed by truck 1. Note that a truck can perform different trip chains. TC_2^1 is then the first trip chain performed by truck 2. TC_1^1 & TC_1^2 both originate from base 1 but have different stop sequences. TC_1^1 & TC_2^1, on the other hand, originate from different bases but share the same intermediate stops. As demonstrated in Figure 1, a truck's trip chain is composed of a base and multiple intermediate stops. Different trucks could share the same stops and bases. Also, the base of one trip chain can be the stop of another trip chain, and vice versa. Therefore, to identify truck trip chains, it is necessary to determine bases and stops, which requires identifying truck stops and grouping these stops together in a meaningful way to determine the stop location.

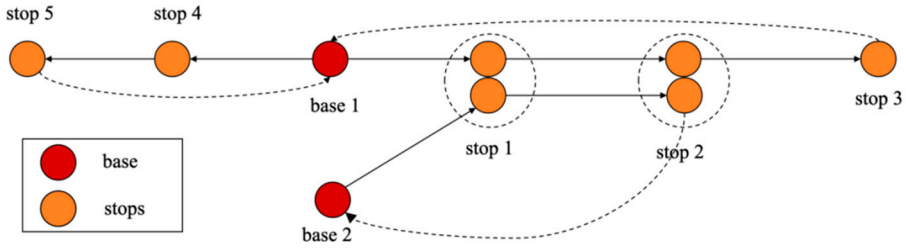


Figure 1. Trip Chain Illustration: The arrow represents the truck's movement direction; the dashed arrow indicates the truck returning to the base; the dashed circle denotes a repeated stop.

3. Materials & Methodology

The proposed approach consists of the following four major steps. (1) Truck stops identification and clustering. (2) Trip chain extraction. (3) Truck trip chain feature selection. (4) Truck profiling and classification. Figure 2 outlines the overall framework, and the detailed procedure is described in the following sections.

3.1. Truck Stop Identification

A heuristic-based approach was developed to identify truck stop activities using the available GPS data, which included each truck's spot speed. Following methods established in prior study [27], a speed threshold of 5 km/h was applied to determine truck stop behavior. Additionally, consecutive stops that were spatially and temporally close were merged to account for stop-and-go patterns, such as those occurring at parking facilities. The merging rules were defined based on empirical data:

- If the distance between two consecutive stops was smaller than a predefined distance threshold and the temporal difference was less than a specified temporal threshold, the two stops were considered to be the same stop.
- If the arrival time between two consecutive stops was shorter than a defined time difference threshold and the average speed between the two stops was below a specified speed threshold, the two stops were also regarded as the same stop.

The first rule identifies truck stops that are both spatially and temporally close, suggesting that they should be merged together. The second rule, which is applied after

the first, further refines the process by identifying temporally connected stops with very low-speed movement between stops. This could indicate trucks waiting in line but not fully stationary or exhibiting stop-and-go behavior. A combination of these thresholds was tested (Section 4.1) to select the best set of parameters for accurate stop identification.

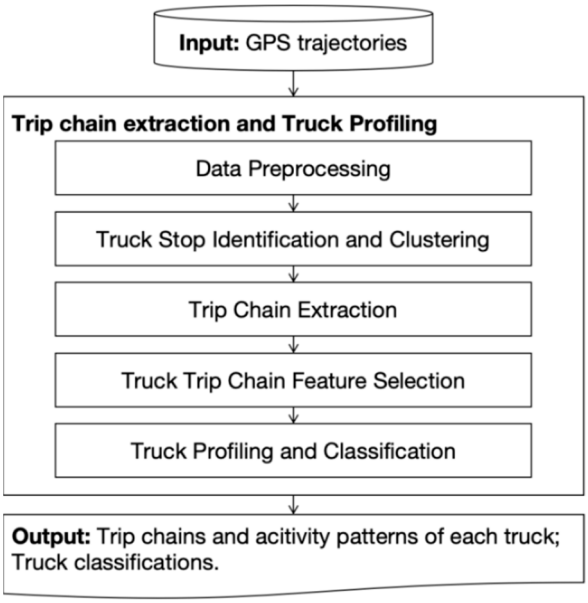


Figure 2. Flowchart of this study.

3.2. Truck Stop Clustering

Then, we aimed to identify the stop locations. Stop locations refer to places such as business establishments, company-owned truck parking lots, logistics depots, freight hubs, and ports where trucks stop for similar trip purposes. Accurate identification of these locations is essential for constructing vehicle trip chains. Identifying stop location requires grouping vehicle stops in a meaningful way. Previous studies generally use two different approaches: Spatial clustering techniques, such as DBSCAN [34], or the Spatial Constrained Approach [10,29]. Upon examining freight-related AOIs, truck stops, and the roadway network, it was found that the majority of business establishments in China are located within regions delineated by higher levels of roadways. It is uncommon for high-level roadways to traverse through business establishments (detailed explanation in Section 3.2.1). Based on this observation, we extended the spatial constrained method and developed a roadway-constrained spatial clustering method, described in the following sections.

The roadway-constrained clustering algorithm consists of three main steps: (1) defining roadway-constrained zones by using OSM road network data to create zones based on major roads and assigning truck stops to these zones, (2) associating related zones by grouping adjacent roadC zones likely belonging to the same business cluster, and (3) identifying stop locations by applying the DBSCAN clustering algorithm within each roadC group to accurately refine truck stop locations. Detailed procedures for each step are provided in the following sections.

3.2.1. Roadway Constrained Area

In this study, we introduced roadway-constrained zones (referred to as roadC zones) to capture the spatial distribution of truck stops within the road network. These zones were generated by dividing the study area based on the road network data. Specifically, the roadC zones are polygonal regions that are delineated by high-level major roads, which serve as the boundaries of each zone. Each truck stop was assigned to a corresponding roadC zone based on its proximity to these road network boundaries. This approach

ensured that the spatial units of analysis aligned with the structure of the urban road network and the distribution of truck stops within these high-traffic areas. The road network data was obtained from OpenStreetMap, which provides a classification system for differentiating roads based on their function and importance. These classifications are typically organized into nine categories, including motorways, primary roads, and secondary roads, among others. Using these classifications, higher-level roadways were selected to define the boundaries of roadway-constrained zones.

A detailed examination of roadway classifications and business establishment boundaries in Chongqing was conducted. A sample of 7402 freight-related AOIs, which included manufacturing factories/business establishments in Chongqing, was evaluated, and their boundaries were compared against various levels of roadways. The analysis revealed that six high-level roadway classes—motorway, trunk, primary, secondary, tertiary, and residential—rarely traverse within the boundaries of establishments (see Table 1 for classification details). Figure 3a provides a comparison between the boundaries of the AOIs and the OSM roadway network. The AOIs were consistently confined within the boundaries defined by the selected roadway network level, demonstrating alignment in the spatial context. Figure 3b,c show a comparison. Figure 3b illustrates six of the 7402 AOIs (depicted with green line patterns and labeled ① to ⑥) alongside the OSM roadway network. The gray lines represent roadways filtered to the six selected classes, while the red dashed lines indicate the remaining roadway classes. The figure demonstrates that the selected roadway classes effectively enclose establishments within distinct roadway-defined zones, while roads within the establishments are generally lower-level roadways (depicted in red dashed lines).

Table 1. Description of truck characteristics.

Features	Description
Temporal Variables	
Avg dwell time	Average stop duration at each intermediate stop
Dwell time variation	The variation of stop duration at each intermediate stop
Spatial Variables	
Average stop distance	Average distance from truck stops to its base
Trip chain radius	the maximum distance from the truck stops to its base
Stop distance variation	The variation of distance from trucks stops to its base.
Travel Attributes	
Daily trips	Number of daily trips completed by the truck. Trip is defined as a journey from one stop to the next one.
Unique trip chains	Total number of unique trip chains of each truck.
Daily trip chains	Average number of trip chains completed every day by each truck.
Average stops per chain	Average number of intermediate stops of the trip chains of each truck.

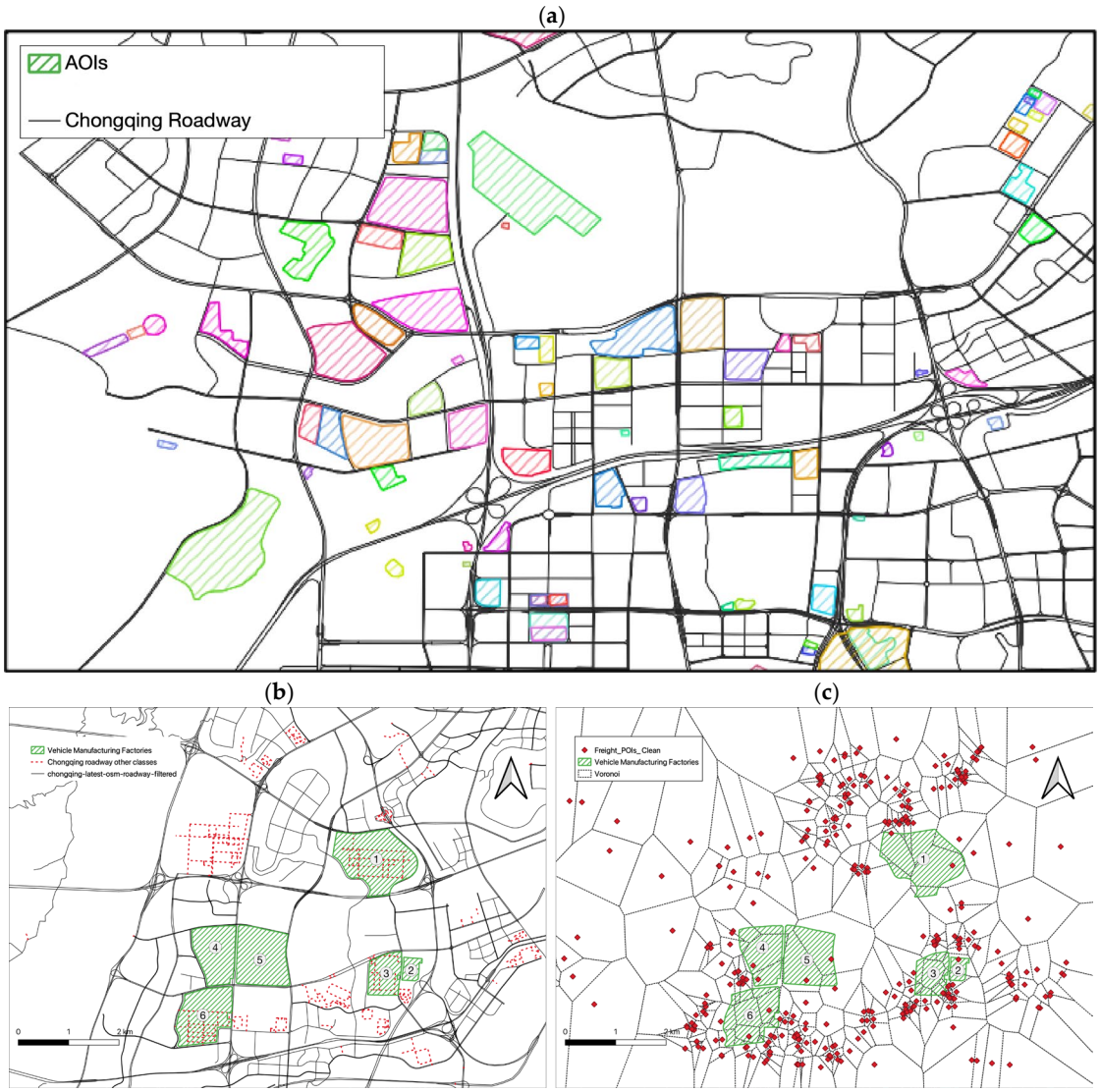


Figure 3. Examples of six establishments and OSM roadway network with different classifications. (**a**) OSM roadway network. (**b**) The Voronoi method (POI comes from Amap data). (**c**) OSM roadway network & AOIs. The green areas in the figure represent established vehicle manufacturing factories, while the red dots indicate freight-related POIs.

In contrast, Figure 3b showcases a spatial division using the Voronoi method used to define stop locations in other studies [29]. The Voronoi method is a computational geometry technique that partitions a plane into regions based on proximity to POIs. The Voronoi method, apparently, performs poorly in detecting the boundaries of POIs. The city of Chongqing was therefore divided into roadway-constrained zones (roadC zones) using the filtered road network. A unique ID r (r∈R) was assigned to each zone.

3.2.2. RoadC Zone Clusters

While relatively uncommon, certain freight hubs, ports, and industrial parks extend across multiple roadC zones. We compared the AOIs with RoadC zones and found that out of 7402 observed AOIs, 440 spanned more than one RoadC zone. Additionally, heavy trucks may park or rest in nearby areas outside company boundaries due to spatial constraints or road conditions. To resolve this issue, a road zone clustering algorithm was developed based on the functional connections between roadC zones. The algorithm uses the DBSCAN (a density-based clustering algorithm) clustering method to group truck stops based on

their spatial density and functional connectivity. Two criteria guided the grouping of roadC zones. For a truck t, the cluster of its j th stop s_t^j was $C_{s_t^j}$.

Criteria 1 Functional Connectivity: for all truck stops s_t^j in the same cluster k ($C_{s_t^j} = k$), the corresponding roadC zones $R_{s_t^j}$ that each truck stop s_t^j located in are functional connected (simplified as connected).

Criteria 2 Transitive Connectivity: If roadC zone r_l is connected with roadC zone r_m, and r_m is connected to roadC zone r_n, then r_l, r_m, and r_n are grouped as one cluster.

Following the two criteria, the connected roadC zones can be identified and grouped together to form a roadC zone cluster. Figure 4 provides an example of connected roadC zones using the trajectory of a truck with plate number 'A6****' from 1 September 2023 to 7 September 2023. Different trajectory colors represent different dates, and dotted points within the dashed circle indicate the truck's stop points, identified as belonging to the same cluster. These stops span two separate roadC zones: 3107 and 3114. Following the criteria, these zones were determined to be connected and grouped into one roadC cluster.

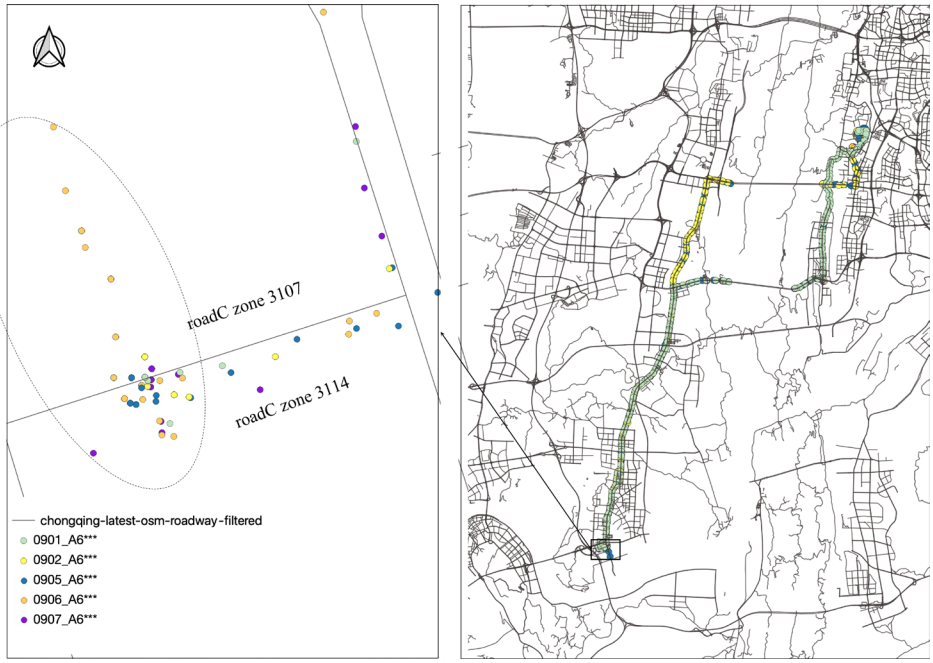


Figure 4. roadC zone clustering illustration. The dashed circle in the left panel indicates stops that have been identified as belonging to the same cluster using the traditional DBSCAN method. The dotted points, each in a different color, represent truck stops on different dates.

The detailed workflow is as below:

Step 1 Initialization: Each roadC region is initially treated as a separate cluster.

Step 2 Cluster Truck Stops: Then for each truck t, the DBSCAN algorithm is applied to its stops s_t^i to get the cluster $C_{s_t^i}$ for each stop.

Step 3 Group RoadC Zones: For all stops s_t^j of truck t within the same cluster c ($C_{s_t^j} = c$), the corresponding roadC zones $R_{s_t^j}$ that these stops locate on are grouped together.

Step 4 Iterate Across Trucks: Steps 1–3 are repeated until all trucks are processed.

3.2.3. Stops Clustering and Stop Location Identification

After defining roadway clusters, truck stops within each roadC cluster are grouped using the DBSCAN algorithm, which requires selecting two parameters: minPoints and eps (distance threshold). Various parameter combinations were tested, and after comparing clustering results against the 7402 AOIs (see Section 4.1 for more details), the optimal

parameters were determined to be minPoints = 3 and eps = 90 m. This combination resulted in an acceptable False Alarm Rate (FAR), effectively balancing clustering accuracy and false alarms.

3.3. Trip Chain Identification

According to the trip chain definition in Section 2, the base location refers to a place where trucks return after completing a series of delivery or pickup activities. Two criteria were employed to define the base:

Most Frequently Visited Location: Trucks often visit multiple stops during different trip chains but consistently return to a specific location, designated as the base.

Long Stop Durations Location: A stop with a dwell time exceeding a certain threshold is considered the start of a new trip chain and thus defined as the base. Previous studies [23,35] suggested a threshold of 240 min, which demarcates operational stops (e.g., rest, loading, unloading) from non-operational stops (e.g., overnight stays).

We analyzed the empirical data in Chongqing and adopted the broken power law method to determine an appropriate dwell time threshold. This mathematical model identifies a "break point" where a power law relationship changes its behavior; it was previous adopted in a study to identify the time threshold for differing temporary stops with freight-related stops [31]. By fitting truck dwell times to a broken power law, two thresholds were identified (Figure 5): 300 min, indicating stops related to rest or non-operational activities, marking the base or start of a new trip chain, and 1000 min, representing long-term stays where trucks are parked for extended periods without daily use.

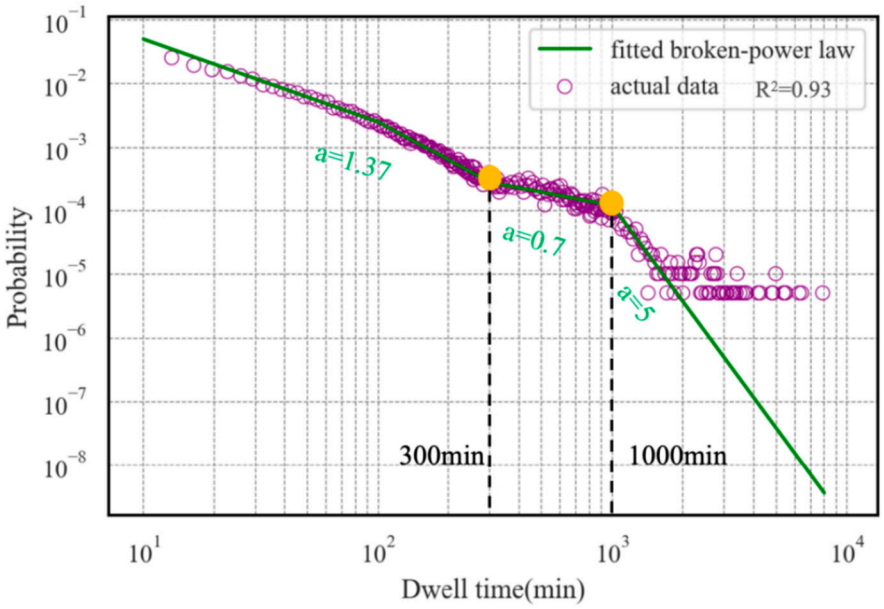


Figure 5. Probability Distribution of Dwell Time for HDVs at Stop Locations. The three segmented green lines represent fitted power law segments. The scaling exponents α for the three segments were separately 1.37, 0.7, and 5, respectively. The two dots in yellow indicate the breakpoints of the distribution at 300 min and 1000 min.

With the above definition, the trip chain TC_t^i for each truck t was identified through the following steps:

Identify Base Location: For each truck t, identify the most frequently visited location $l (l \in L)$ and set it as the base b_t.

Sort Stops: sort all stops of truck t in chromonic order and iterate over these stops.

Define Trip Chains: Assume the existing trip chain is TC_t^i. For a truck stop s_t^j, if $s_t^j = b_t$, or the stop dwell time at s_t^j is longer than 300 min, start a new trip chain $TC_t^{i+1} = \left(s_t^j,\right)$. Otherwise, append the stop to the existing trip chain $TC_t^i = (b_t, \ldots s_t^{j-1}, s_t^j)$.

Iterate: repeat step 2 to step 4 until all stops are processed.

Using these steps, the trip chains for each truck are identified, enabling a detailed analysis of truck travel behavior.

3.4. Truck Profiling

We aimed to profile trucks by analyzing trip chains. We incorporated trip chain characteristics into consideration and classified the trucks based on their traveling patterns. To identify typical truck travel patterns, clustering analysis was conducted using selected features derived from trip chain characteristics, including the following:

- Temporal Variables: Average dwell time, dwell time variation.
- Spatial Variables: Average stop distances, trip chain radius.
- Travel Attributes: Number of intermediate stops, trip chain frequency.

Note that most of the travel attributes were defined based on the trip chains of each truck because a trip chain could contain multiple sub-trips and reflect the hidden travel structure. Table 1 list these attributes by their categories.

We tested different clustering algorithms including K-Means [36], Agglomerative Clustering [37], Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [34], Gaussian Mixture Models (GMMs) [38], Mean Shift [39], and Spectral Clustering [40] to provide a comprehensive comparison of clustering performance. Further details of this comparison and evaluation strategy are provided in Section 4.4.1. We finally adopted the K-Means clustering algorithm to uncover distinct truck travel patterns and provide a foundation for improving freight logistics.

4. Results & Analysis

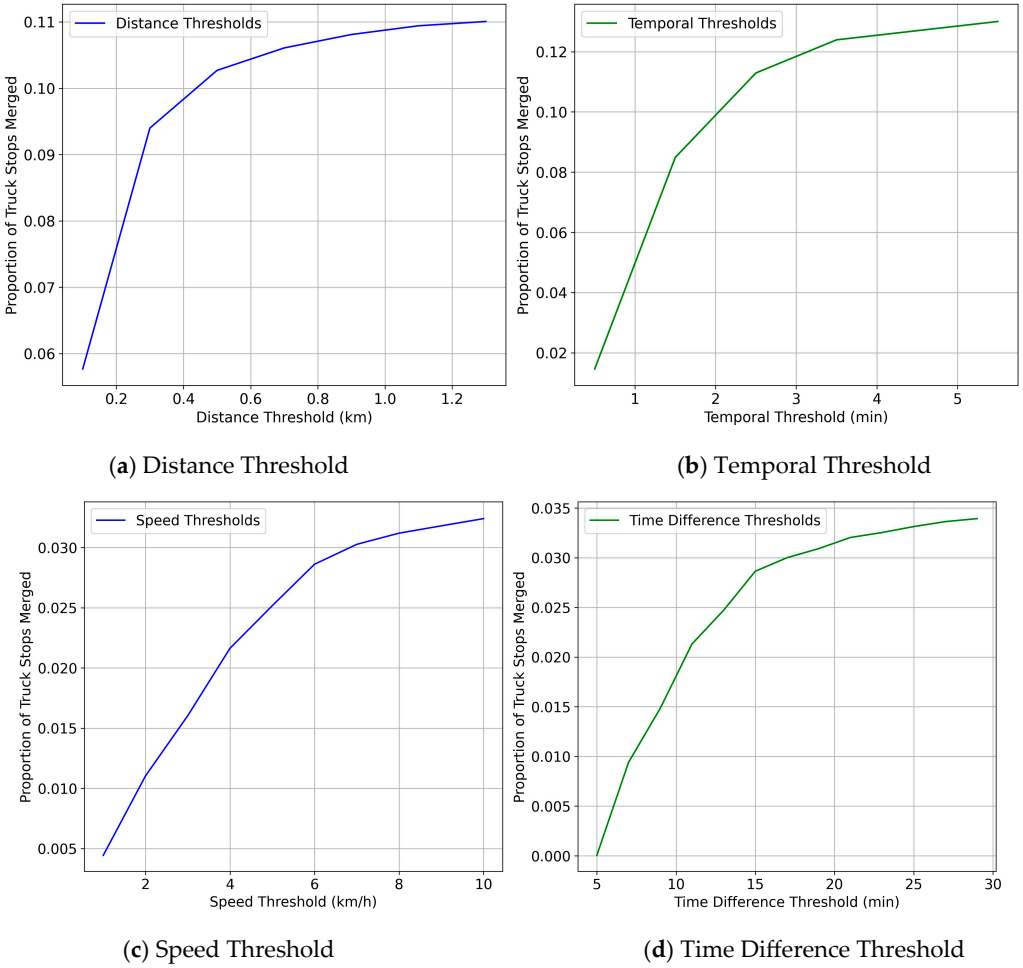
In this section, we apply the proposed method to the heavy-duty vehicle GPS data described in Section 2.1.

4.1. Parameter Selection

Merging nearby stops required a set of parameters to determine when two stops should be considered as one. Consecutive stops that were spatially and temporally close were merged to account for stop-and-go patterns, such as those occurring at parking facilities. Four parameters needed to be tested separately: the distance threshold, temporal threshold, time difference threshold, and speed threshold. The distance threshold was tested within a range of 100 m to 1300 m, with increments of 200 m. The temporal threshold ranged from 0.5 min to 5.5 min, the time difference threshold from 5 min to 35 min, and the speed threshold from 3 km/h to 10 km/h.

Figure 6 illustrates the sensitivity analysis results, where the y-axis represents the share of truck stops after merging. Figure 6a shows that the distance threshold followed an elbow pattern. When the distance was between 100 m and 500 m, the share of stops decreased sharply with an increase in the distance threshold, but this reduction slowed after the threshold exceeded 500 m. Figure 6b shows the sensitivity to the time threshold. The elbow range was observed between 1.5 and 3.5 min, indicating that selecting a time threshold within this range was acceptable. The speed and time difference thresholds were tested after stops were initially merged based on the first rule. The proportion of stops further merged was then calculated using these parameters. As shown in Figure 6c, the share of stops decreased clearly within the speed range of 2 km/h to 6 km/h, suggesting that 6 km/h

may be the optimal speed threshold. Finally, Figure 6d also shows an elbow value of 15 m to be the optimal time difference threshold. From the figure and based on the changing point pattern, the best parameters were set to 500 m for distance threshold, 2.5 min for temporal threshold, 6 km/h for speed threshold and 15 min for time difference threshold.



(**a**) Distance Threshold (**b**) Temporal Threshold

(**c**) Speed Threshold (**d**) Time Difference Threshold

Figure 6. The proportion of trucks stops remaining after stop merging for different threshold values.

To assess the robustness of these parameters, we conducted a sensitivity analysis on key parameter thresholds (speed, time, and distance) using the elbow method for threshold selection. The speed and time thresholds had minimal impact on stop merging, with only 3–4% of stops affected, indicating stable clustering results. However, varying the distance and temporal thresholds (ranging from 300 to 500 m and 1.5 to 2.5 min) caused slight variations in merged stops. Therefore, we tested different combinations of these thresholds and found that they had minimal impact on the final results. Further inspection revealed that even if some stops were not merged initially, the majority of them were grouped into the same cluster in subsequent clustering stages. Overall, these findings demonstrate that the framework remains stable and reliable across different threshold settings.

4.2. Clustering Results and Validation

As noted in Section 2.1, the raw dataset contained missing records. Trucks with missing data exceeding one hour were excluded from the analysis. After data cleaning and preprocessing, 94,345 unique trucks remained in the dataset. Chongqing City was divided into 3748 roadC regions, which were then processed using the roadC clustering algorithm. The final result included 3271 roadC clusters. Then, the stops of each truck were processed with the proposed road constrained clustering algorithm.

Previous studies often relied on land use data and satellite imagery for method validation, manually comparing clustering results with satellite images. However, land use data provides only general geographic information for polygonal areas with similar land use pattern and cannot differentiate between individual establishments or factories, limiting its application for detailed validation. Similarly, satellite imagery lacks precise boundaries for distinguishing between different factories or establishments. In this study, we address these limitations by using both AOI (Area of Interest) data and satellite images to validate the accuracy of the proposed road-constrained clustering methods. Specifically, we utilized 7402 freight-related AOIs to assess the accuracy of our identification results and employed satellite imagery to visually verify. Figure 7 illustrates an example of the clustering results along with the AOIs. In this figure, each colored dot signifies a distinct cluster. The areas enclosed by bold white boxes denote individual AOIs.

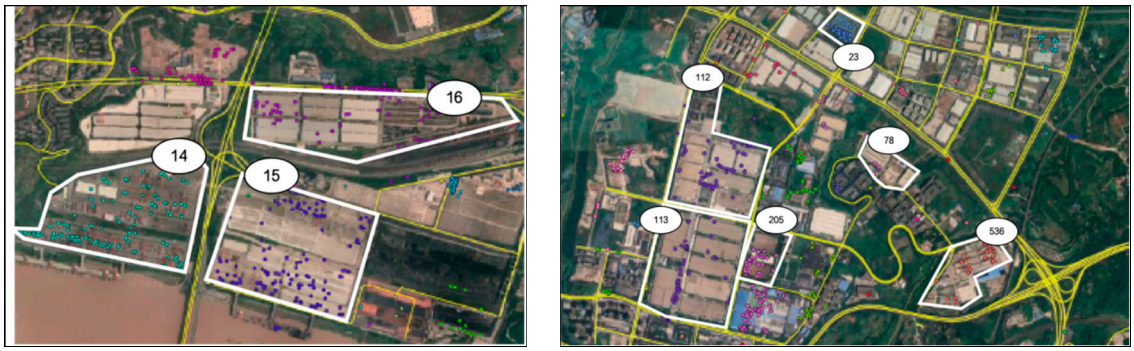


Figure 7. Illustration of stop clusters. Different number represents the ID of the AOI; Each colored dot signifies a distinct cluster.

We compared the clustering results with the AOI boundaries and introduced two metrics to evaluate the results:

$$FAR_C = n_c/NC \quad (1)$$

$$FAR_A = n_a/NA \quad (2)$$

where n_c denotes the number of AOI grouping events where different AOIs are grouped into one cluster. n_c indicates the total number of clusters. n_a indicates the number of AOI split events where AOIs are split into more than one cluster. NA represents the total number of AOIs. FAR_C and FAR_A represent misclassification rates from two different perspectives. FAR_C captures the rate at which distinct AOIs are incorrectly grouped into a single cluster, while FAR_A measures the rate at which AOIs are erroneously split into multiple parts. These two types of misclassifications affect downstream trip chain analysis in different ways. When FAR_C occurs, vehicle stops at physically and functionally distinct locations are incorrectly merged into the same location, leading to the loss of important distinctions between activity types (e.g., a warehouse vs. a fuel stop). This misclassification distorts the topological structure of the trip chain, affecting key elements such as the number of visited locations, inter-location transitions, and inferred relationships between activities. As a result, FAR_C can significantly alter the overall flow and meaning of trip chains, making it more disruptive to downstream analysis. In contrast, FAR_A causes fragmentation of stops that actually belong to the same location, creating multiple pseudo-locations. This misclassification increases the apparent number of nodes in the trip chain and may lead to inflated trip chain complexity. While the effects of FAR_A are more localized and primarily influence node-level statistics (such as visit frequency and dwell time), they do not fundamentally alter the overall sequence of activities.

Consequently, FAR_C typically has a more significant impact on the semantic interpretation and connectivity of trip chains, whereas FAR_A mainly results in overestimated complexity without distorting the broader structure.

To evaluate the performance of the proposed method, we compared it with the benchmark DBSCAN method. The comparison results are shown in Figure 8. From the figure, we observe that as the distance threshold increased, the FAR_A value (dotted line) decreased, while the FAR_C value (solid line) increased. Our goal was to achieve low values for both FAR_A and FAR_C, ideally balancing both to minimize errors in classification.

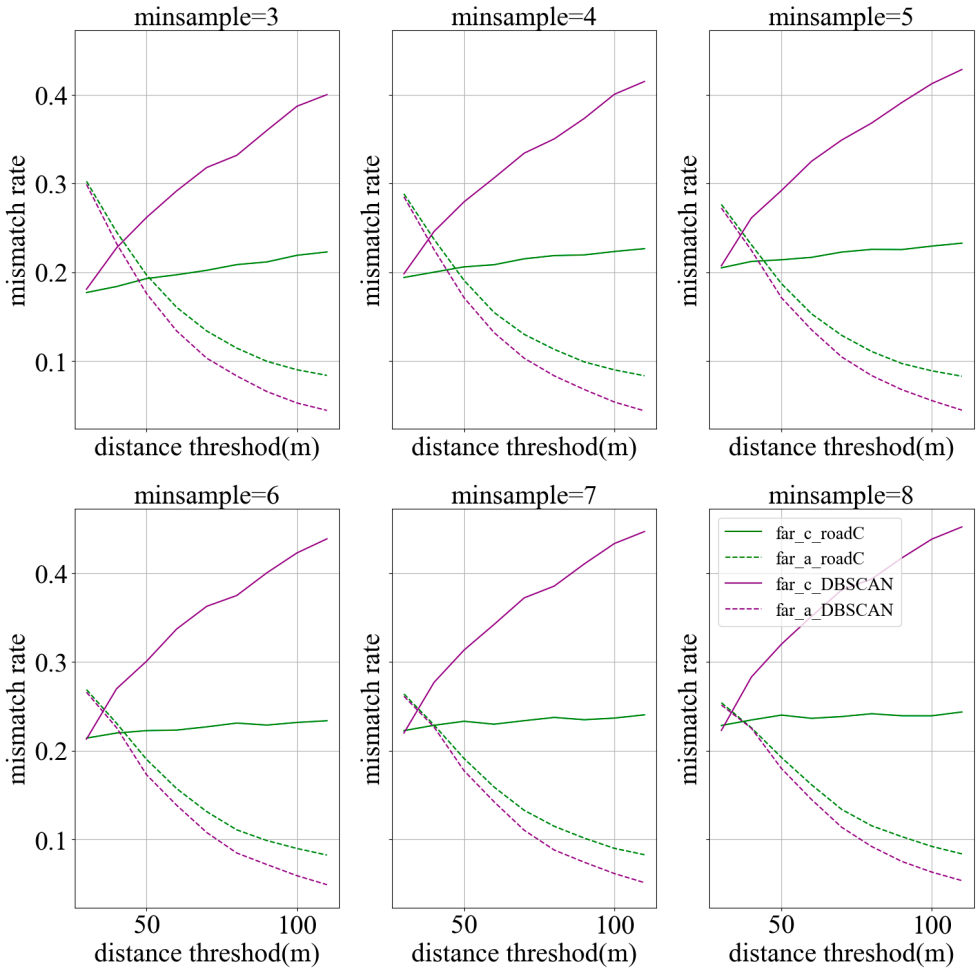


Figure 8. FAR_A and FAR_C for the proposed algorithm and traditional DBSCAN algorithm.

When comparing the proposed method with DBSCAN, we see that the proposed method was particularly effective in reducing the FAR_C value. Specifically, the solid purple line (representing DBSCAN) increases more sharply as the distance threshold grows, while the solid green line (representing the proposed method) shows a more gradual increase. This suggests that the proposed method is more robust at preventing different AOIs (such as establishments and hubs) from being mistakenly grouped into the same cluster. In contrast, the FAR_A values between the two methods did not show significant differences.

AOI data were used in this study as a reference for validating the clustering results of truck stop locations. This choice offers an interpretable and operationally relevant benchmark; however, several limitations should be noted. First, AOIs are typically defined based on land-use or administrative criteria, which may not align with the actual operational boundaries of freight activities. Vehicles operating within the AOI may be performing different activities, but these activities will be grouped together as serving the same stop,

leading to potential misclassification. Another issue is the incomplete nature of AOI data, particularly for smaller or non-central facilities that are not represented as discrete AOIs. The unclear boundaries of these smaller facilities can complicate the validation of stops and their associated activity patterns. This ambiguity could lead to misclassifications, where distinct stops are merged or different stops are grouped together. In contrast, larger facilities with clearer boundaries are more likely to be validated accurately, which could result in a somewhat overly optimistic view of the trip chain accuracy.

4.3. Trip Chain Identification and Truck Profile

Using the proposed methods, a comprehensive list of trip chains was identified. Figure 9 summarizes the distribution of intermediate stops within these trip chains and their corresponding shares. Notably, over 60% of trip chains include only one intermediate stop, representing direct trips where trucks move directly from origin to destination without additional stops. This observation aligns with findings from previous studies [31].

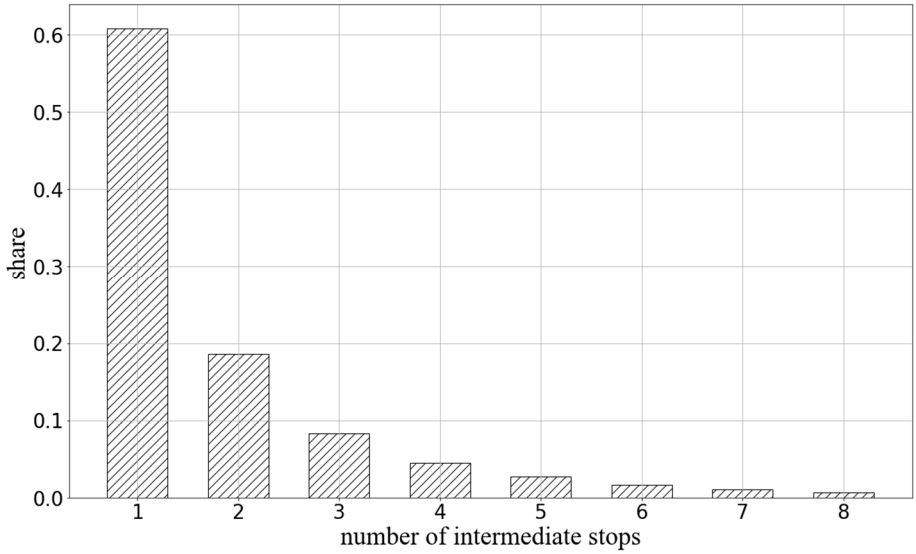


Figure 9. Number of intermediate stops.

The clustering results obtained using the k-means algorithm demonstrated better differentiation among vehicle classes. Based on these results, vehicles were categorized into six distinct classes, as outlined in Table 2. Each column details the statistics for each of the six categories. This analysis highlights variations in truck travel patterns.

Table 2. Classification Results.

Category	All	Short -Distance Trucks	Medium-Short -Distance Trucks		Medium-Long -Distance Trucks		Long-Distance Trucks
		C1	C2	C3	C4	C5	C6
avgStpTime (min)	63.6	45.2	52.7	42.9	122.0	57.7	76.0
avgStpDist (km)	43.7	11.0	16.0	21.8	43.3	43.9	219.7
Radius (km)	76.8	15.3	33.6	35.9	85.7	95.2	331.0
dailyStops	1.5	3.6	3.1	0.9	0.8	2.3	0.9
avgStpPerChain	2.4	1.1	1.7	1.7	1.9	5.2	3.0
dailyTripChains	0.7	3.2	2.0	0.5	0.5	0.5	0.3
timeEachChains IsPerformed	1.2	7.4	1.7	1.1	0.9	1.0	0.9
Vehicle Counts	94,345	1145	15,523	38,550	16,572	15,118	7438

Category 1: Short-distance single direct trucks

These vehicles operate over short travel distances with the highest trip chain similarity, indicating that these vehicles typically serve fixed destinations. They have the highest daily travel frequency, averaging over 7 trips per day. These trucks have minimal intermediate stops, averaging only one per trip, indicating involvement in full truckload transportation serving a single freight stop. These trucks primarily serve fixed destinations, with trips mapped to concentrated areas around vehicle manufacturing hubs. Analysis of origins and destinations relative to nearby POIs reveals that these trucks predominantly serve the vehicle manufacturing industry, a key sector in Chongqing City. Chongqing's automobile manufacturing industry demonstrates industrial concentration, with clusters of auto parts factories and logistics centers around vehicle manufacturing bases. These short-distance trucks frequently perform multiple round trips between these bases, supporting the region's industrial ecosystem.

Category 2 & 3: Medium-short-distance trucks

Chongqing is a municipality with over 20 sub-cities/regions within its jurisdiction. These trucks operate primarily within each city/region in Chongqing, with an average trip chain radius of about 33 km. The similarity of the trip chains is about 0.4, indicating both variability and similarity in destinations. Trips typically involve 1 to 2 intermediate stops. These trucks rarely perform long distance trips; they mainly serve the target industry and surrounding areas. The key difference between Category 2 & 3 is that Category 2 trucks have a higher trip frequency, with trucks active throughout the week, while Category 3 trucks have lower utilization rates, with trucks operating on specific days and remaining at their base for the rest of the week, indicating underutilization.

Category 4 & 5: Medium-long-distance trucks

These trucks mainly connect different regions/cities in Chongqing, covering a broader service radius of around 90 km per trip chain. These trucks have smaller trip chain similarity compared to shorter-distance categories, reflecting greater variability in served destinations.

The difference between C4 and C5 is that trucks in C4 have a longer average stop time at its intermediate stops, averaging 2 h, the highest among all groups. These trucks also typically involve 2 stops per trip chain, balancing between frequent stops and longer durations. Truck in C5, on the other hand, has an average stop time of 57 min. They also have relatively longer trip chains, with over 5 intermediate stops of each trip chain. Digging into the data reveals that these trucks typically have multiple bases, frequently visiting a secondary base in addition to their primary base.

Category 6: Long-distance intercity trucks

Category 6 is named long-distance trucks because it is characterized by the longest travel distance among all the six categories. These trucks primarily serve intercity and cross-regional routes, starting from Chongqing and reaching major destinations such as Chengdu, Beijing-Tianjin, Shanghai, and Guangzhou. On average, these trucks have 3 intermediate stops per trip. These trucks have low trip chain similarity, reflecting diverse destinations and operational flexibility. These trucks usually originate from logistics hubs or transshipment centers.

Figure 10 presents density plots for several selected features of trucks, offering further insights into truck behavior patterns. Figure 10a illustrates the arrival times (hour of the day) at intermediate stops. The data reveals that truck operations are minimal during nighttime hours (12:00 a.m. to 5:00 a.m.). Arrival times exhibit two peaks: one around 10:00 a.m. and another around 3:00 p.m., indicating distinct activity periods.

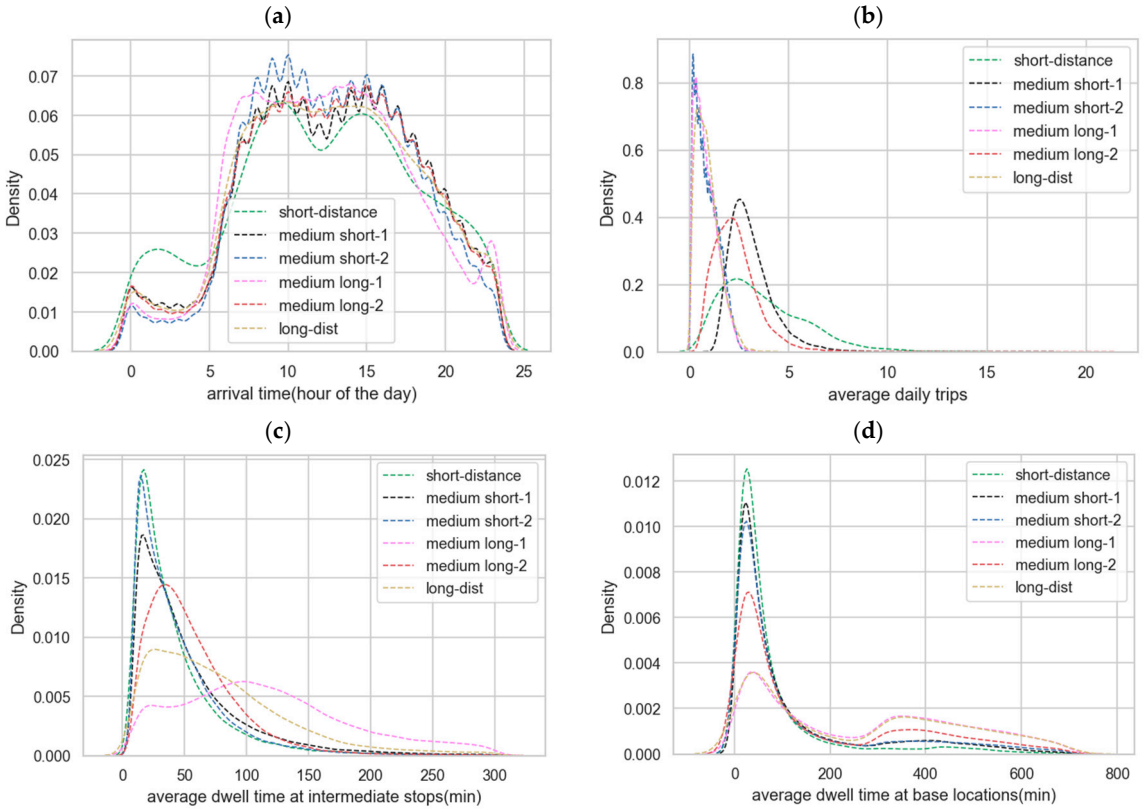


Figure 10. Density plots for truck features. (**a**) represents the distribution of the arrival time by truck categories. (**b**) shows the average daily trips by truck categories; (**c**) represents the average dwell time at intermediate stops by truck categories; (**d**) illustrates the average dwell time at base locations by truck categories.

Figure 10b depicts the average number of daily trips made by each group of trucks, where a trip is defined as travel between two consecutive stops. Short-distance trucks (C1), medium-short distance trucks (C2), and medium-long distance trucks (C5) exhibit a higher number of daily trips compared to other groups. Notably, trucks in group C5 also demonstrate the highest average number of intermediate stops per trip chain. This suggests that these trucks may be engaged in more complex operations, possibly connecting multiple locations or handling intricate logistical tasks within a single trip chain.

Figure 10c highlights the average dwell time at intermediate stops, where the C4 truck group exhibits a unique pattern. For this group, dwell times have two peaks: one around 15 min and another at approximately 100 min. This distribution differs significantly from other groups, indicating distinct operational behaviors.

Figure 10d displays the average dwell time at base locations, showing two peaks across all truck groups. The first peak occurs around 50 min, likely reflecting typical loading or unloading activities at the base location during daytime. The second, significantly longer peak suggests activities such as overnight parking or extended rest periods.

4.4. Robustness Analysis

We conducted a robustness analysis to evaluate the performance of different clustering approaches for identifying truck groups based on travel attributes, using widely recognized indicators. We also acknowledge the potential impact of data loss on the clustering results. Sensitivity analyses were performed to assess how varying levels of missing data affect the results.

4.4.1. Clustering Approach Comparison

The performance of various clustering algorithms was assessed using three widely recognized evaluation metrics: Silhouette Score [41], Davies-Bouldin Index [42], and Calinski-Harabasz Index [43], as summarized in Table 3. Silhouette Score measures how similar each point is to its own cluster compared to other clusters, with values ranging from -1 to 1. Higher values indicate better clustering, with values close to 1 showing well-separated clusters, while values near 0 or negative suggest poor clustering. Davies-Bouldin Index assesses the average similarity between clusters, with lower values indicating better separation. Calinski-Harabasz Index calculates the ratio of between-cluster dispersion to within-cluster dispersion, where higher values reflect better-defined and more separated clusters. The algorithms compared include K-Means, Agglomerative Clustering, DBSCAN, Gaussian Mixture Model (GMM), Mean Shift, and Spectral Clustering.

Table 3. Comparison among clustering methods.

Clustering Method	Silhouette Score	Davies-Bouldin Index	Calinski-Harabasz Index
K-Means	0.79	0.29	5742.04
Agglomerative Clustering	0.79	0.29	5742.04
DBSCAN	0.39	1.16	422.74
GMM	0.79	0.29	5742.04
Mean Shift	0.60	0.52	574.89
Spectral Clustering	0.79	0.29	5742.04

The results show that K-Means, Agglomerative Clustering, GMM, and Spectral Clustering all performed equally well, each achieving a Silhouette Score of 0.79, indicating strong clustering with well-separated and compact clusters. They also achieved a Davies-Bouldin Index of 0.29, the lowest among all methods, signifying good separation, and a Calinski-Harabasz Index of 5742.04, the highest, indicating very good separation and compactness. The next best method was Mean Shift, which showed a slightly lower Silhouette Score of 0.60, suggesting less distinct clusters but still adequate separation. However, Mean Shift had a higher Davies-Bouldin Index and Calinski-Harabasz Index, indicating more overlap between clusters and weaker clustering performance. DBSCAN performed the worst in all three metrics.

Given these results, any of K-Means, Agglomerative Clustering, GMM, or Spectral Clustering could be used. However, K-Means was ultimately selected for its consistent performance, efficiency, and scalability. Compared to Agglomerative Clustering, GMM, and Spectral Clustering, K-Means is more computationally efficient, scales better with larger datasets, and is easier to implement and interpret.

4.4.2. Analysis of Clustering Stability Under Missing Data

To evaluate the impact of potential data missingness and ensure the robustness of our clustering results, we conducted a systematic stability analysis. This procedure assessed whether the identified vehicle behavior patterns remained consistent when only a subset of the data was available. We employed a sub-sampling approach where the original dataset was randomly partitioned into subsets ranging from 10% to 90% of the total population. For each sub-sample, the K-means clustering algorithm was re-applied independently. To quantify the stability, we compared the cluster assignments of the sub-sampled data against the baseline assignments derived from the full dataset.

A critical challenge in clustering stability is the label switching problem, where the same behavioral group may be assigned different cluster IDs across different runs. To

overcome this, we utilized the Adjusted Rand Index (ARI) as the primary evaluation metric [44]. The ARI is a statistical metric used to measure the similarity between two cluster groupings of the same data [44]. It is widely used in machine learning for stability analysis because it focuses on the clustering structure, not on arbitrary cluster labels. ARI compares every pair of data points, checking whether they are in the same cluster in both groupings. A match occurs when points are either in the same or different clusters in both groupings. The adjusted part accounts for chance agreements, ensuring that the score reflects meaningful similarity. ARI ranges from −1 (completely different) to 1 (perfect match), with 0 indicating random clustering.

Figure 11 summarizes the ARI scores across different levels of missing data. As shown, the ARI scores remained consistently high across all sampling levels, demonstrating the robustness of the clustering results. Even when only 10% of the data was available, the ARI score was still above 0.86, and for samples exceeding 50%, the scores consistently exceeded 0.95. These results suggest that the clustering method was resilient to the absence of data and that the truck behavior clusters were not overly sensitive to the specific data points included in the analysis. The stability of the ARI scores, even with progressively larger amounts of missing data, indicates that the variables selected for the clustering process captured strong, underlying patterns in truck travel behavior. This highlights the robustness of the identified truck categories, ensuring that they reflected generalizable, empirical groupings of truck travel behavior rather than being contingent on specific data points.

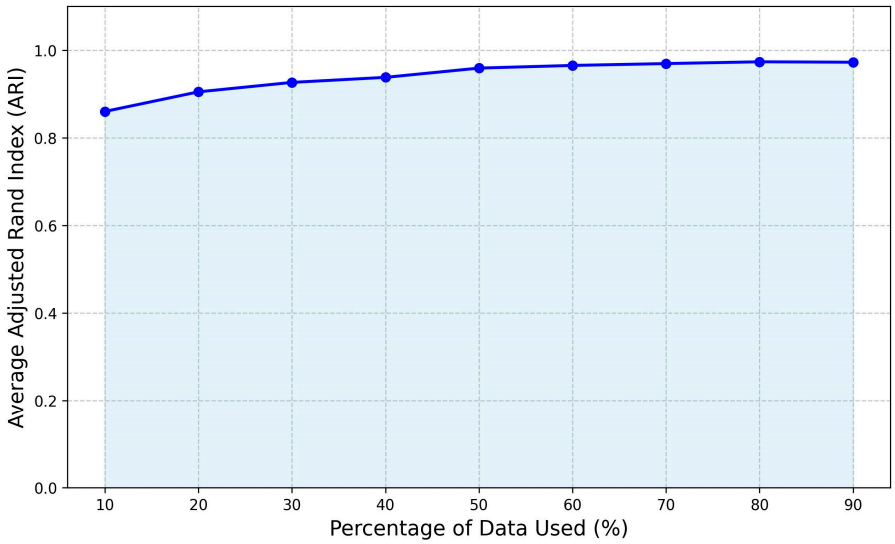


Figure 11. Clustering stability across varying levels of missing data.

4.5. Trucks and Industries

In this section, we analyze the potential industries served by each group of trucks. Each truck was assigned an industry category according to data provider, covering a total of 19 distinct industry classifications. Figure 12 illustrates the proportion of each industry served by trucks across different classes. Each cell in the figure indicates the share of trucks within a specific truck group that served the corresponding industry. This section presents a descriptive analysis of the industries served by different truck categories based on the proportions observed in the data, without making direct claims about their operational significance.

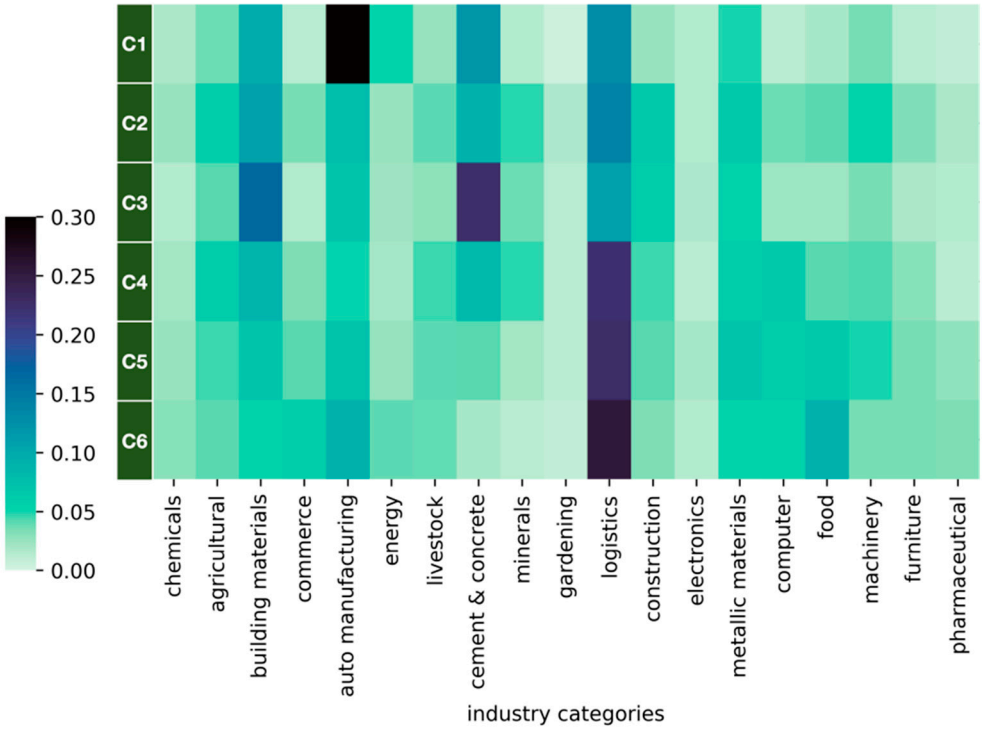


Figure 12. Truck categories by industries.

Short-Distance Trucks: A significant majority of short-distance trucks served the automobile manufacturing industry, a key sector in Chongqing characterized by a high level of industrial complementarity. This reflects the localized nature of automobile production and the proximity of related supply chain operations.

Medium-to-Short-Distance Trucks: A high proportion of trucks in this category transport building materials like cement and concrete, supporting construction projects. Other industries, such as agriculture, minerals, and machinery, also make up significant shares.

Medium-to-Long-Distance Trucks and Long-Distance Trucks: A significantly high proportion of trucks in these categories serve logistics companies, handling goods that require transportation over longer and even intercity distances. For example, over a quarter of intercity trucks are dedicated to logistics companies, underscoring the importance of freight and logistics hubs in facilitating intercity transportation. Conversely, very few long-distance trucks are utilized for building materials, such as cement or concrete, as these are typically transported shorter distances due to cost and practicality constraints.

Several observations were summarized from industry-specific insights:

Logistics Sector: Trucks serving logistics companies constitute the largest proportion across all categories, highlighting the central role of logistics in truck utilization.

Commerce, Automobile Manufacturing, and Food Industries: The share of intercity trucks serving these industries is relatively higher compared to other categories, reflecting their demand for longer-range transportation networks.

Other Industries: Certain industries, like automobile manufacturing, require both short-distance and long-distance truck services, indicating a complex supply chain with localized production and intercity distribution components.

4.6. Transferability Discussion

This section discusses the transferability of the proposed framework across diverse geographic and operational contexts. The methodology comprises three critical components, each designed for generalization to other cities, regions, or freight systems. While the overarching structural framework is broadly adaptable, certain parameters are intended to

be calibrated to account for local geographic contingencies and specific industrial layouts. These three pillars are summarized as follows:

Stop Location Identification Component: The methodology used for stop identification, based on GPS data, is broadly applicable and can be generalized to other urban areas with similar data sources. While the Road-Constrained clustering technique is established for regions with established road hierarchies, its direct implementation may face challenges in areas with inconsistent network structures. In such contexts, the framework maintains its robustness by cross-referencing Area of Interest data with OpenStreetMap attributes. This comparative approach facilitates the assessment of classification suitability and ensures that the most representative road hierarchy is selected for the local environment.

Trip Chain Identification Component: The methodologies for base location detection and power law-based dwell thresholding are designed for deployment with empirical data across diverse geographic contexts, ensuring the framework's broad adaptability. Specifically, the trip chain identification logic offers high flexibility, allowing for seamless customization to align with the unique operational characteristics of different urban freight systems.

Vehicle Behavior Analysis Component: While the empirical findings reflect the local context of Chongqing, the underlying analytical framework—specifically the feature engineering and clustering techniques—is highly transferable. Sensitivity tests demonstrate that the method remains reliable despite data loss, ensuring its efficacy in capturing truck travel behaviors across different data qualities.

In summary, while the framework's performance may fluctuate across different regions due to variations in road network structures and data quality, its core methodologies remain robust. Although specific numerical thresholds derived in this study may require localized calibration, the underlying threshold selection protocols are universally applicable. Successful implementation in alternative urban contexts necessitates high-fidelity GPS trajectories and accurate road network data, such as those from OpenStreetMap. Furthermore, integrating AOI data is essential for cross-regional validation. By accounting for these local contingencies and ensuring data integrity, this framework provides a scalable and adaptable solution for urban freight management.

5. Discussion and Conclusions

In recent years, the need for sustainable urban freight systems has become increasingly urgent as cities grapple with growing congestion, environmental challenges, and the need for more efficient logistics. However, data to support these goals is often lacking. To address these challenges, this study proposes a method for mining the trip chain patterns of heavy-duty vehicles (HDVs) using GPS data. This approach offers significant potential to enhance sustainable urban freight management by providing a clearer understanding of truck movement patterns.

In this study, we proposed a method to mine the trip chain patterns of heavy-duty vehicles using GPS data. A road-constrained clustering approach was developed to identify truck stop locations, addressing the limitations of traditional clustering methods. This technique ensures more accurate differentiation of stop locations, thereby enhancing the reconstruction of truck trip chains. Results from a comparison of over 7000 AOIs show that the proposed method is more effective at preventing different AOIs (such as establishments and hubs) from being mistakenly classified as the same stop location. Based on these identified stop locations, a procedure was designed to identify base locations for each truck and to extract trip chains, creating a robust database to support freight planning and modeling.

Key trip chain characteristics of HDVs were extracted and analyzed using GPS data from Chongqing. The results indicate that, on average, heavy trucks in Chongqing spend 63 min at intermediate stops, with an average service radius of approximately 76 km. Based on trip chain characteristics, HDVs were classified into four main categories: short-distance, medium-short distance, medium-long distance, and long-distance trucks. The analysis revealed that the service range plays a significant role in classification, with trucks of similar service ranges exhibiting comparable trip chain patterns. Long-distance trucks primarily serve intercity and cross-regional routes. Further analysis of industry data showed that over one-quarter of long-distance trucks are engaged in logistics services. Medium-short and medium-long distance trucks are primarily involved in intra-city movements. Medium-short distance trucks typically serve nearby destinations, with industries such as construction and building materials being two of the main users. In contrast, medium-long distance trucks connect regions or cities within Chongqing, with a broader service radius of approximately 90 km per trip chain.

Understanding these patterns reveals important opportunities to optimize urban freight systems by improving the efficiency of truck movements, reducing unnecessary stops, and streamlining routes. For example, identifying long-distance and medium-short distance truck behaviors helps pinpoint the most congested areas and timing for freight transport. These insights can be leveraged to reduce traffic congestion and promote the use of more environmentally friendly routes, leading to a decrease in fuel consumption and emissions. By making urban freight operations more efficient, this study contributes to the development of sustainable transport systems that minimize environmental impact while maintaining the flow of goods essential to the economy.

However, this study has several limitations. First, the analysis was based on short-term data collected over just seven days, which does not capture seasonal variations or long-term evolutionary patterns in freight behavior. Additionally, the empirical study was conducted in Chongqing; comparing data from multiple cities could offer a more comprehensive understanding of freight behavior, highlighting both common trends and regional differences that may arise due to varying geographic, economic, or infrastructural factors.

Several future research directions can be explored. First, studies could expand on the trip chain data from this study by incorporating additional factors that influence the selection of trip chain patterns. Investigating the drivers behind these patterns (e.g., operational constraints and geographic or economic factors) could provide deeper insights into truck route planning. Identifying these underlying influences could improve the understanding of how trucks optimize their trips within complex logistics networks. Additionally, future research could explore how trip chain destinations are selected. Analyzing the impact of industries, supply chains, and logistics hubs on truck destinations would enhance understanding of freight flow dynamics, informing more accurate predictive models and improving sustainable urban freight management. Finally, while this study excludes non-freight-related stops, these stops can significantly influence long-distance freight operations by affecting travel schedules and delivery performance. Future research could integrate non-freight stops into trip chain analysis, exploring their interaction with freight activities. Classifying these stops may reveal their impact on operational efficiency and help refine models for predicting truck movement, ultimately optimizing logistics networks and improving freight flow management.

Author Contributions: Conceptualization, B.Y. and Y.L. (Yuandong Liu); Methodology, Y.L. (Yuandong Liu); Validation, B.Y., Y.L. (Yuandong Liu) and G.G.; Formal Analysis, B.Y. and Y.L. (Yuandong Liu); Investigation, G.G. and Y.L. (Yuandong Liu); Resources, Y.L. (Yi Li); Data Curation, Y.L. (Yuandong Liu); Writing—Original Draft Preparation, B.Y.; Writing—Review & Editing, Y.L. (Yuandong Liu); Visualization, G.G.; Supervision, Y.L. (Yi Li); Project Administration, B.Y.; Funding Acquisi-

tion, B.Y. and Y.L. (Yuandong Liu). All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Fundamental Research Funds for the Central Universities, CHD (No. 300102344102) and Natural Science Basic Research Program of Shaanxi Province [Grant S2025-JC-QN-0569].

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data used in this study involve sensitive information and cannot be made publicly available due to privacy and confidentiality restrictions. Data can be accessed from the authors upon reasonable request and subject to approval and necessary data-use agreements.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Dua, R.; Almutairi, S. A perspective on emerging policy and economics research priorities for enabling low-carbon trucking. *Energy Res. Soc. Sci.* **2025**, *124*, 104025. [CrossRef]
2. Hunt, J.D.; Stefan, K. Tour-based microsimulation of urban commercial movements. *Transp. Res. Part B Methodol.* **2007**, *41*, 981–1013. [CrossRef]
3. Sakai, T.; Alho, A.R.; Bhavathrathan, B.; Dalla Chiara, G.; Gopalakrishnan, R.; Jing, P.; Hyodo, T.; Cheah, L.; Ben-Akiva, M. SimMobility Freight: An agent-based urban freight simulator for evaluating logistics solutions. *Transp. Res. Part E Logist. Transp. Rev.* **2020**, *141*, 102017. [CrossRef]
4. Khan, M.; Machemehl, R. Analyzing tour chaining patterns of urban commercial vehicles. *Transp. Res. Part A Policy Pract.* **2017**, *102*, 84–97. [CrossRef]
5. Oka, H.; Hagino, Y.; Kenmochi, T.; Tani, R.; Nishi, R.; Endo, K.; Fukuda, D. Predicting travel pattern changes of freight trucks in the Tokyo Metropolitan area based on the latest large-scale urban freight survey and route choice modeling. *Transp. Res. Part E Logist. Transp. Rev.* **2019**, *129*, 305–324. [CrossRef]
6. Allen, J.; Piecyk, M.; Piotrowska, M.; McLeod, F.; Cherrett, T.; Ghali, K.; Nguyen, T.; Bektas, T.; Bates, O.; Friday, A. Understanding the impact of e-commerce on last-mile light goods vehicle activity in urban areas: The case of London. *Transp. Res. Part D Transp. Environ.* **2018**, *61*, 325–338. [CrossRef]
7. Toilier, F.; Serouge, M.; Routhier, J.-L.; Patier, D.; Gardrat, M. How can urban goods movements be surveyed in a megacity? The case of the Paris region. *Transp. Res. Procedia* **2016**, *12*, 570–583. [CrossRef]
8. Joubert, J.W.; Meintjes, S. Repeatability & reproducibility: Implications of using GPS data for freight activity chains. *Transp. Res. Part B Methodol.* **2015**, *76*, 81–92.
9. Siripirote, T.; Sumalee, A.; Ho, H. Statistical estimation of freight activity analytics from Global Positioning System data of trucks. *Transp. Res. Part E Logist. Transp. Rev.* **2020**, *140*, 101986. [CrossRef]
10. Gao, Z.; Janssens, D.; Jia, B.; Wets, G.; Yang, Y. Identifying business activity-travel patterns based on GPS data. *Transp. Res. Part C Emerg. Technol.* **2021**, *128*, 103136.
11. Qian, X.; Ukkusuri, S.V. Spatial variation of the urban taxi ridership using GPS data. *Appl. Geogr.* **2015**, *59*, 31–42. [CrossRef]
12. Liu, D.; Kan, Z.; Lee, J. The proposal of a 15-minute city composite index through integrating GPS trajectory data-inferred urban function attraction based on the Bayesian framework. *Appl. Geogr.* **2024**, *173*, 103451. [CrossRef]
13. Yang, X.; Sun, Z.; Ban, X.J.; Holguín-Veras, J. Urban freight delivery stop identification with GPS data. *Transp. Res. Rec.* **2014**, *2411*, 55–61. [CrossRef]
14. Gong, L.; Sato, H.; Yamamoto, T.; Miwa, T.; Morikawa, T. Identification of activity stop locations in GPS trajectories by density-based clustering method combined with support vector machines. *J. Mod. Transp.* **2015**, *23*, 202–213. [CrossRef]
15. Bohte, W.; Maat, K. Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transp. Res. Part C Emerg. Technol.* **2009**, *17*, 285–297. [CrossRef]
16. Feng, T.; Timmermans, H.J. Detecting activity type from GPS traces using spatial and temporal information. *Eur. J. Transp. Infrastruct. Res.* **2015**, *15*, 662–674. [CrossRef]
17. Chen, C.; Jiao, S.; Zhang, S.; Liu, W.; Feng, L.; Wang, Y. TripImputor: Real-time imputing taxi trip purpose leveraging multi-sourced urban data. *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 3292–3304. [CrossRef]

18. Liao, C.; Chen, C.; Guo, S.; Wang, Z.; Liu, Y.; Xu, K.; Zhang, D. Wheels know why you travel: Predicting trip purpose via a dual-attention graph embedding network. In Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, New York, NY, USA, 11–15 September 2022; Volume 6, pp. 1–22. Available online: https://dl.acm.org/doi/epdf/10.1145/3517239 (accessed on 3 May 2023).
19. Gingerich, K.; Maoh, H.; Anderson, W. Classifying the purpose of stopped truck events: An application of entropy to GPS data. *Transp. Res. Part C Emerg. Technol.* **2016**, *64*, 17–27. [CrossRef]
20. Zhao, Y.; Cheng, S.; Lu, F. Spatiotemporal interaction pattern of the Beijing agricultural product circulation. *J. Geogr. Sci.* **2023**, *33*, 1075–1094. [CrossRef]
21. Hess, S.; Quddus, M.; Rieser-Schüssler, N.; Daly, A. Developing advanced route choice models for heavy goods vehicles using GPS data. *Transp. Res. Part E Logist. Transp. Rev.* **2015**, *77*, 29–44. [CrossRef]
22. Ruan, M.; Lin, J.J.; Kawamura, K. Modeling urban commercial vehicle daily tour chaining. *Transp. Res. Part E Logist. Transp. Rev.* **2012**, *48*, 1169–1184. [CrossRef]
23. Romano Alho, A.; Sakai, T.; Chua, M.H.; Jeong, K.; Jing, P.; Ben-Akiva, M. Exploring algorithms for revealing freight vehicle tours, tour-types, and tour-chain-types from GPS vehicle traces and stop activity data. *J. Big Data Anal. Transp.* **2019**, *1*, 175–190. [CrossRef]
24. Duan, M.; Qi, G.; Guan, W.; Guo, R. Comprehending and analyzing multiday trip-chaining patterns of freight vehicles using a multiscale method with prolonged trajectory data. *J. Transp. Eng. Part A Syst.* **2020**, *146*, 04020070. [CrossRef]
25. Alho, A.R.; e Silva, J.d.A. Analyzing the relation between land-use/urban freight operations and the need for dedicated infrastructure/enforcement—Application to the city of Lisbon. *Res. Transp. Bus. Manag.* **2014**, *11*, 85–97. [CrossRef]
26. Jing, P.; Zhang, Y.; Jeong, K.; Alho, A.; Ben-Akiva, M. Modeling daily tour-chaining pattern choice of urban heavy commercial vehicles. In Proceedings of the 98th Annual meeting of transportation research board, Washington, DC, USA, 13–17 January 2019.
27. Sarti, L.; Bravi, L.; Sambo, F.; Taccari, L.; Simoncini, M.; Salti, S.; Lori, A. Stop purpose classification from GPS data of commercial vehicle fleets. In Proceedings of the 2017 IEEE International Conference on Data Mining Workshops (ICDMW), New Orleans, LA, USA, 18–21 November 2017; pp. 280–287.
28. Allen, J.; Browne, M.; Cherrett, T. Investigating relationships between road freight transport, facility location, logistics management and urban form. *J. Transp. Geogr.* **2012**, *24*, 45–57. [CrossRef]
29. Yang, Y.; Jia, B.; Yan, X.-Y.; Li, J.; Yang, Z.; Gao, Z. Identifying intercity freight trip ends of heavy trucks from GPS data. *Transp. Res. Part E Logist. Transp. Rev.* **2022**, *157*, 102590. [CrossRef]
30. Patel, V.; Maleki, M.; Kargar, M.; Chen, J.; Maoh, H. A cluster-driven classification approach to truck stop location identification using passive GPS data. *J. Geogr. Syst.* **2022**, *24*, 657–677. [CrossRef]
31. Yang, Y.; Jia, B.; Yan, X.-Y.; Jiang, R.; Ji, H.; Gao, Z. Identifying intracity freight trip ends from heavy truck GPS trajectories. *Transp. Res. Part C Emerg. Technol.* **2022**, *136*, 103564. [CrossRef]
32. Yang, Y.; Cai, J.; Yang, H.; Zhang, J.; Zhao, X. TAD: A trajectory clustering algorithm based on spatial-temporal density analysis. *Expert Syst. Appl.* **2020**, *139*, 112846. [CrossRef]
33. Gong, L.; Yamamoto, T.; Morikawa, T. Identification of activity stop locations in GPS trajectories by DBSCAN-TE method combined with support vector machines. *Transp. Res. Procedia* **2018**, *32*, 146–154. [CrossRef]
34. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the KDD'96: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2 August 1996; pp. 226–231.
35. You, L.; Zhao, F.; Cheah, L.; Jeong, K.; Zegras, C.; Ben-Akiva, M. Future mobility sensing: An intelligent mobility data collection and visualization platform. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4 November 2018; pp. 2653–2658.
36. MacQueen, J. Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 18–21 June 1965; pp. 281–297.
37. Nielsen, F.; Nielsen, F. Hierarchical clustering. In *Introduction to HPC with MPI for Data Science*; Springer: Cham, Switzerland, 2016; pp. 195–211.
38. Wolfe, J.H. Pattern clustering by multivariate mixture analysis. *Multivar. Behav. Res.* **1970**, *5*, 329–350. [CrossRef]
39. Comaniciu, D.; Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619. [CrossRef]
40. Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905. [CrossRef]
41. Rousseeuw, P.J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
42. Davies, D.L.; Bouldin, D.W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *PAMI-1*, 224–227. [CrossRef]

43. Caliński, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.-Theory Methods* **1974**, *3*, 1–27. [CrossRef]
44. Hubert, L.; Arabie, P. Comparing partitions. *J. Classif.* **1985**, *2*, 193–218. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.