

Article

Identification of Truck Travel Endpoints and Traffic Demand Prediction Based on Trajectory Data

Xiaoyun Cheng ¹, Yichen Yao ^{2,*}, Jian Zou ², Tingjia Zhang ³ and Jitong Shao ²¹ College of Transportation Engineering, Chang'an University, Xi'an 710064, China; cxy@chd.edu.cn² Department of College of Transportation Engineering, Chang'an University, Xi'an 710064, China; zou_jian98@163.com (J.Z.); 2024901923@chd.edu.cn (J.S.)³ Chang'an Dublin International College of Transportation, Chang'an University, Xi'an 710064, China; 2023902678@chd.edu.cn

* Correspondence: 2023234102@chd.edu.cn; Tel.: +86-133-8914-9666

Abstract

The rapid expansion of the logistics industry has led to sustained growth in freight demand across regions. Analyzing the factors that influence truck travel demand and extracting loading/unloading locations from truck trajectory data can offer a theoretical foundation and policy support for freight logistics planning and management. Accordingly, this study is structured around two core tasks: identifying truck trip endpoints and forecasting truck travel demand, aiming to thoroughly examine urban freight travel patterns and their influencing factors. First, truck GPS data is integrated with a non-parametric Lorenz curve fitting method, road network data, and Areas of Interest (AOI) to precisely identify truck route endpoints. Subsequently, a Multi-Feature Dynamic GCN-LSTM model is constructed to predict truck travel demand by comprehensively considering multi-dimensional regional characteristics. Using Xi'an, China, as a case study, the results show that the proposed framework provides a more precise spatial distribution of truck trips and effectively forecasts truck travel demand in the city. This research has significant implications for enhancing urban freight efficiency, optimising the layout and facilities of urban transport networks, developing evidence-based transport policies, and promoting logistics development.

Keywords: Trip Demand Prediction; trip endpoint identification; Graph Convolution Neural Networks; long short-term memory

1. Introduction

Freight transportation serves as a vital foundation and prerequisite for regional economic development. The rapid rise of the logistics industry has driven regional socioeconomic growth, facilitating a smoother and more efficient flow of goods. Simultaneously, with the deepening of globalisation, the demand for goods circulation continues to rise. Consequently, the freight industry has experienced sustained growth worldwide, becoming a crucial link in ensuring smooth logistics. Modern urban freight transportation is a vital component in supplying daily necessities, industrial raw materials, and construction materials [1], representing key logistical activities that meet the production and living needs of urban residents [2]. Within urban freight systems, trucks undertake substantial transportation tasks by connecting factories, enterprises, logistics warehouses, and port terminals [3]. They thereby play a significant role in urban socioeconomic development.

Growing economies and higher purchasing power are diversifying market demands for services and products. Combined with ongoing economic restructuring, this is leading



Academic Editor: Sara Moridpour

Received: 10 December 2025

Revised: 17 February 2026

Accepted: 17 February 2026

Published: 27 February 2026

Copyright: © 2025 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

to continuous shifts in freight product mixes, necessitating more scientific management to unlock the road transport industry's potential and ensure its long-term growth. Vehicle GPS data, with its inherent temporal and spatial attributes, has been widely applied in studies of residential travel patterns and urban vehicle operations. However, its application in in-depth research on highway freight transportation systems remains relatively limited. Traditional freight demand forecasting often relies on regression analysis. Yet, due to the complex mechanisms governing freight demand generation, such methods frequently fall short in comprehensive transportation modeling [4]. The “four-stage method”, adapted from passenger transportation models, subsequently emerged as a framework for freight demand forecasting. Nevertheless, fundamental differences exist between freight and passenger transportation—in terms of stakeholders, decision-making processes, demand generation, and allocation principles—making direct application of passenger-based methods unsuitable. Moreover, existing freight research primarily focuses on network optimization and technological advancements. Demand analysis often depends on large-scale, coarse-grained forecasting models based on land use and economic indicators. These approaches struggle to capture the true characteristics of freight demand, which is marked by significant spatiotemporal variability and complex, dynamic product structures. To address these gaps, this study employs truck GPS data as its foundation. By identifying trip origins and destinations, we extract loading and unloading locations from truck trajectories to explore the spatial distribution characteristics of truck travel demand. Furthermore, we analyze the factors influencing this demand. This work aims to provide a theoretical foundation and policy basis for demand forecasting in the highway freight industry and for freight logistics planning and management.

2. Research Review

2.1. Freight Trip Identification Method

The growing availability and accuracy of trajectory data have led to a wider application of truck trajectory data in transportation research. Massive GPS trajectory datasets now provide a solid foundation for in-depth freight data analysis. Traditionally, obtaining truck trip data depended on time-consuming and labor-intensive manual surveys [5]. With the proliferation of GPS technology, however, GPS-based methods for identifying truck stops have become mainstream, effectively displacing traditional approaches [6]. A key limitation remains: GPS data typically include only geographic locations and vehicle status, lacking detailed contextual information about the journeys [7]. Therefore, a critical first step in GPS-based freight research is to accurately identify the endpoints of truck trips, which is essential for extracting complete and meaningful freight travel information.

The most common approach for extracting truck trips from GPS trajectories is to identify trip endpoints (origins and destinations) based on predefined thresholds for time, speed, and distance [8,9]. To improve the accuracy of endpoint identification, machine learning methods have been introduced. For instance, Random Forest [10], SVM [11] and PAM-CART [12]. Notably, the primary features used in these machine learning models also revolve around time, speed, and similar threshold parameters.

However, methods for determining time and distance thresholds can involve subjectivity in specific contexts, while machine learning models often suffer from limited interpretability and high dependency on data quality [13]. In practice, trucks typically dwell for extended periods at depots or bases and make shorter stops en route [14], with travel patterns varying considerably across different cities. To address these challenges, Yang et al. [15] proposed a multi-level time threshold method to distinguish between different types of trip endpoints. Their approach first identifies vehicle stops using a speed threshold and then applies tiered time thresholds to differentiate temporary pauses from

final trip endpoints. A complete freight trip may include various stop activities, such as loading/unloading, traffic-induced waiting, and driver breaks at service areas [16]. They further utilized POI data and road network GIS data to filter genuine trip endpoints for freight vehicles.

2.2. Travel Demand Prediction

Early research on travel demand prediction primarily relied on classical time series analysis methods, such as the Autoregressive Integrated Moving Average (ARIMA) model [17]. While effective in capturing periodic demand fluctuations under assumptions of stationarity and linearity, these methods fundamentally lack the capacity to model complex nonlinear features and spatial dependencies.

With advances in computational power and data availability, classical machine learning methods like Support Vector Regression (SVR) [18] and Gradient Boosting Decision Trees (XGBoost) [19] were introduced to the field. These approaches can handle nonlinear relationships through kernel functions or ensemble mechanisms and allow for rich feature engineering, demonstrating strong performance in data-rich scenarios.

In recent years, deep learning has opened new pathways for travel demand forecasting. Methods such as Graph Convolutional Networks (GCN) [20] and Graph Attention Networks (GAT) [21] explicitly model transportation networks as graph structures. Landmark studies like STGCN [22] and ASTGCN [23] leverage spectral or spatial graph convolutions to capture complex spatial dependencies explicitly. The field is now shifting from relatively isolated spatial or temporal models toward deeply integrated spatio-temporal frameworks. Multi-graph fusion has emerged as a key direction, where approaches such as Sumin et al. [24] integrate one or more regional features by constructing and fusing multiple graph views—such as those based on geographical distance, functional similarity, and real-time traffic connectivity—to better capture intricate inter-regional interactions.

Truck travel demand prediction faces unique challenges beyond those of passenger demand. Truck activities exhibit higher spatiotemporal sparsity and heterogeneity, and their demand is susceptible to disruptions from supply chain fluctuations, specific logistics orders, and traffic control policies. Consequently, research on truck demand prediction continues to evolve, with methodologies broadly categorized into traditional statistical models and data-driven machine learning/deep learning models. In the domain of traditional statistical research, Luo et al. [25] optimized and constructed a freight demand forecasting model by integrating characteristics of truck GPS data into the traditional four-stage method. Utilizing trip chain information, YU et al. [26] combined it with an improved maximum entropy model to predict freight demand across different periods of the year. Regarding machine learning and deep learning techniques, AHN et al. [27] integrated truck visit frequency data with a Support Vector Regression (SVR) model to predict urban construction material transport demand and estimate associated costs. Based on truck trajectory data, Li et al. [28] employed four models: KNN, CART, SVM, and LSTM, to predict truck travel paths, with results showing that all four achieved prediction accuracy above 70%. Similarly, Shao et al. [29] constructed four models: LSTM, GRU, Bi-LSTM, and Bi-GRU, using truck GPS trajectories to forecast truck traffic flow on Zhengzhou's expressways. Their findings demonstrated that all models attained high prediction accuracy, exceeding 90%.

Methods for identifying truck trip endpoints based on GPS data have achieved considerable success. The primary approach involves extracting trip endpoints by setting thresholds for time, speed, and distance. Some researchers have also attempted to improve recognition accuracy through machine learning methods. However, challenges remain, including the inherent subjectivity in determining these thresholds and the high

data quality requirements of machine learning models. In forecasting truck trip demand, emerging machine learning and deep learning techniques show significant advantages in capturing complex nonlinear relationships. Nevertheless, these methods often overlook the spatial heterogeneity of travel demand and rarely comprehensively integrate regional characteristics.

In summary, this study employs a non-parametric iterative method based on the Lorenz curve to determine an optimal parking duration threshold from truck trajectory data. Recognition accuracy is further improved by incorporating road network and AOI data. We construct a Multi-Feature Dynamic GCN-LSTM model that comprehensively accounts for regional characteristics, including land use, population and economy, truck-related points of interest (POI), and road networks. The model integrates multiple regional similarity graphs and introduces a sentinel attention mechanism to mitigate overfitting. Spatial dependencies are extracted by processing regional features through GCNs, while temporal dependencies are captured via LSTM networks. Through a graph weighting mechanism, the model automatically learns the weights of different feature maps, thereby integrating diverse spatial feature information. This framework ultimately enables precise prediction of truck travel demand. The research contributes to enhancing urban freight transportation efficiency and optimizing the layout of urban transportation networks and related facilities. It also provides theoretical support for evidence-based transportation policy formulation.

3. Methodology

3.1. Lorenz Curve-Based Identification of Truck Trip Endpoints

A typical truck trip usually starts at a loading location, and after loading is completed, the truck travels to an unloading location to deliver the cargo. Once unloading is finished, the trip ends. During a freight trip, in addition to loading/unloading stops, trucks may also make other stops due to traffic congestion, rest breaks at service areas, refueling, and so forth (as shown in Figure 1). Therefore, it is necessary to identify and retain effective truck stops as trip endpoints.

To identify truck trip endpoints, this study introduces the Lorenz-curve principle and proposes a multi-level time-threshold identification algorithm based on the concentration of the dwell-time distribution. The Lorenz curve, originally proposed by Max O. Lorenz (1905) [30], is used to characterize the degree of concentration in a distribution. In this study, it is adapted to analyze truck stopping behavior: the x-axis represents the cumulative percentage of stop-point frequency, and the y-axis represents the cumulative percentage of total parking duration, forming a Lorenz curve that reflects the concentration of dwell times. The extent to which the curve deviates from the line of equality reveals the concentration pattern of stop durations. Based on this property, we iteratively extract key inflection points from the Lorenz curve to derive multi-level time thresholds, enabling the identification of stops with different characteristics and providing a basis for subsequent endpoint filtering.

To identify truck trip endpoints (i.e., departure and arrival points), this study employs the Lorenz curve principle and proposes a multi-level time threshold method. First, the complete set of truck trajectory points is obtained by extracting all points from the GPS dataset (as shown in Figure 2). Each point is denoted as: $p_i = (lng_i, lat_i, t_i)$, indicating the truck's location (lng_i, lat_i) at time t_i . The trajectory dataset P for a single truck over one week is chronologically sorted as:

$$P = \{p_1, p_2, \dots, p_n\} \quad (1)$$

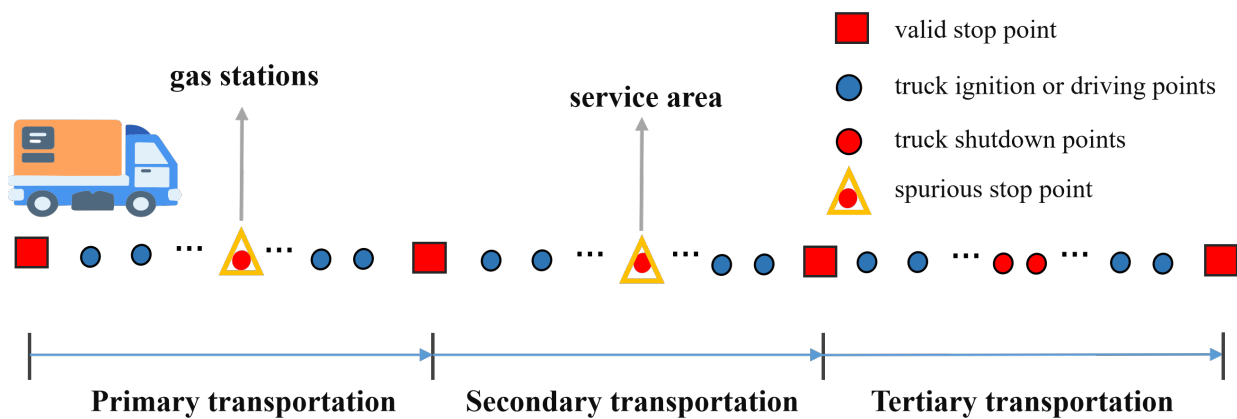


Figure 1. Schematic diagram of the truck's journey.

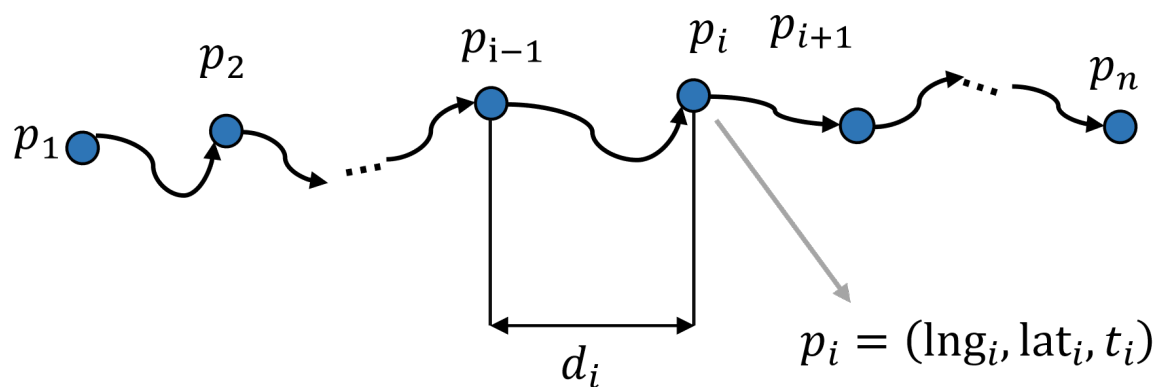


Figure 2. Schematic diagram of GPS trajectory of truck.

As shown in Figure 3, stationary points are extracted through the following steps: First, by calculating the distance and velocity between consecutive GPS points, adjacent trajectory points with zero velocity are clustered into stationary points. Each stationary point must contain at least two consecutive trajectory points. All stationary points are then sorted chronologically, denoted as:

$$\text{Stop} = \{\text{stop}_1, \text{stop}_2, \dots, \text{stop}_n\}. \quad (2)$$

For the set of stop points in S_k (with the initialization $S_0 = \text{Stop}$), the stop points are sorted in ascending order by dwell time Δt :

$$\Delta t_{(1)} \leq \Delta t_{(2)} \leq \dots \leq \Delta t_{(n)}. \quad (3)$$

Define the cumulative percentage of stop-point frequency (x-axis) and the cumulative percentage of parking duration (y-axis) as follows.

The cumulative frequency percentage up to the k -th stop point is:

$$F_k = \frac{k}{n} \times 100\%, \quad k = 1, 2, \dots, n. \quad (4)$$

The cumulative parking-duration percentage is:

$$L_k = \frac{\sum_{i=1}^k \Delta t_{(i)}}{\sum_{i=1}^n \Delta t_{(i)}} \times 100\%, \quad k = 1, 2, \dots, n. \quad (5)$$

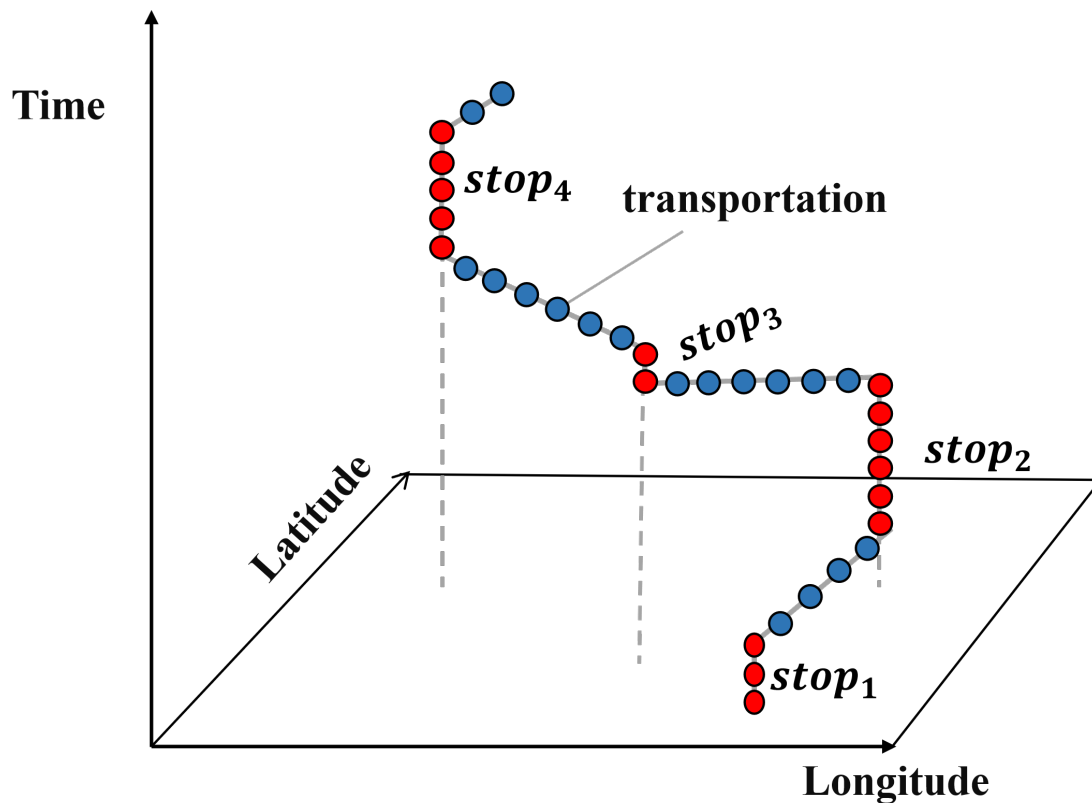


Figure 3. Schematic Diagram of Truck Stop Points.

The point set $\{(F_k, L_k)\}$ forms the Lorenz curve $L = f(F)$, as shown in Figure 4. Based on the constructed Lorenz curve, an iterative threshold identification algorithm is adopted to determine multi-level time thresholds. The k -th iteration consists of the following steps:

- (1) Sort the stop points in S_k in ascending order of duration and plot the Lorenz curve L_k ;
- (2) Compute the tangent line to L_k at the endpoint $(1, 1)$, and denote by F_k^* the horizontal coordinate of the intersection between this tangent line and the horizontal axis, which represents the minimum proportion of the core group required to contribute the vast majority of the total parking duration;
- (3) Locate, in the ordered sequence, the first stop point whose cumulative frequency percentage exceeds F_k^* , and define its duration as the current-level threshold T_k ; and
- (4) Separate from S_k all stop points with durations no less than T_k to form the k -th level, while the remaining stop points constitute the set S_{k+1} for the next iteration.

The iteration terminates when the remaining set reaches an approximately balanced distribution, i.e., when the Lorenz curve becomes close to a straight line, indicating that the distribution of stop durations has become nearly uniform. Through this sequential process, a set of time thresholds in descending order can be obtained: $T = \{T_1, T_2, \dots, T_n\}$, where $T_1 > T_2 > \dots > T_n$. These thresholds naturally correspond to different types of freight activities, such as long-term warehousing, standard distribution, and temporary stops.

To further exclude non-business-related extended stops, a spatial filtering mechanism is introduced. First, a road-network buffer zone is established based on the average urban road width and GPS positioning error. Stop points falling within this buffer are removed to effectively avoid interference from traffic congestion and toll-booth queues. Second, AOI data from OpenStreetMap is integrated to exclude stop points associated with facilities such as highway service areas, auto repair shops, and gas stations. Through this spatiotemporal joint filtering process, a clean set of truck OD pairs is ultimately obtained.

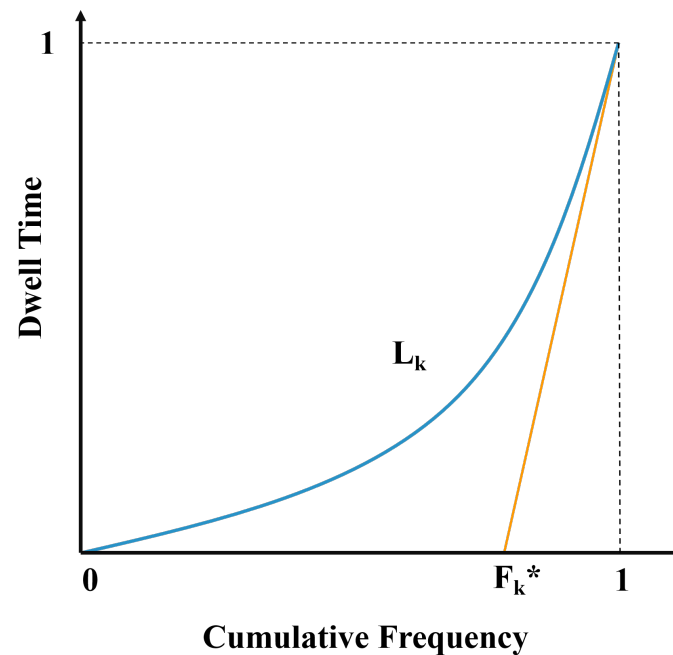


Figure 4. The Lorenz curve.

3.2. Truck Travel Demand Prediction Based on Multi-Feature Dynamic GCN-LSTM

In this study, we consider a geographic area divided into N study regions. The truck travel demand across all regions at time step t is represented by a vector $X_t \in \mathbb{R}^N$, which is forecast based on historical observations. Spatial relationships between regions are captured by a prior adjacency matrix $G_p \in \mathbb{R}^{N \times N}$, encoding pairwise similarities or distances. To improve the demand prediction for each region, we incorporate multi-dimensional regional features—such as industrial area size, commercial area size, and population density—as supplementary inputs. Assume there are K distinct types of regional features, each denoted as $F_k \in \mathbb{R}^{N \times d_k}$, where d_k is the dimensionality of the k -th feature. Our goal is to learn a predictive model f_{model} that estimates the future truck travel demand X_t using: A historical sequence of demand data $\{X_1, X_2, \dots, X_{t-1}\}$; the set of regional features $\{F_1, F_2, \dots, F_K\}$; and the inter-region adjacency matrix G_p . This can be formally expressed as:

$$f_{model}(X_{\tau=1}^{t-1}; G_p; F_{k=1}^K) \rightarrow X_t \quad (6)$$

Model Construction

We propose a Multi-Feature Dynamic GCN-LSTM model, designed to dynamically capture and predict truck travel demand by integrating multiple regional features through dynamic graph convolutional techniques. The core innovation of the model is its combination of a dynamic GCN and an LSTM to jointly model spatial dependencies and temporal dynamics. As shown in Figure 5, the overall architecture consists of three key components: multi-feature dynamic graph construction, graph convolutional spatial encoding, LSTM-based temporal modeling and final prediction.

(1) Multi-Feature Dynamic Graph Construction

In the dynamic graph construction stage, this study draws on the attention-based graph construction idea proposed by Sumin et al. [24]. Specifically, the features of each region are first projected into a latent (hidden) representation space, and an adjacency matrix is then generated by computing pairwise similarities between regions. In this way, the spatial dependency among regions can be explicitly characterized, allowing the learned graph to reflect inter-regional interactions and influence strength. Furthermore, to endow

the graph structure with temporal dynamics, a demand-context vector z_t is introduced at each time step t and concatenated with the static regional features to form a time-dependent node representation:

$$\tilde{F}_{k,t} = [F_k \parallel z_t] \in \mathbb{R}^{N \times (d_k + d_z)} \quad (7)$$

On this basis, latent node representations are computed as:

$$U_{k1,t} = \text{ReLU}(\tilde{F}_{k,t} W_{k1}) \in \mathbb{R}^{N \times D} \quad (8)$$

$$U_{k2,t} = \text{ReLU}(\tilde{F}_{k,t} W_{k2}) \in \mathbb{R}^{N \times D} \quad (9)$$

where W_{k1} and W_{k2} are learnable weight matrices and D is the hidden dimension. The inter-region similarity matrix is defined as:

$$e_{k,t} = \frac{U_{k1,t} U_{k2,t}^\top}{\sqrt{D}} \in \mathbb{R}^{N \times N} \quad (10)$$

A preliminary attention-normalized adjacency matrix is obtained via row-wise softmax:

$$A_{k,t}[i, j] = \frac{\exp(e_{k,t}[i, j])}{\sum_{o=1}^N \exp(e_{k,t}[i, o])} \quad (11)$$

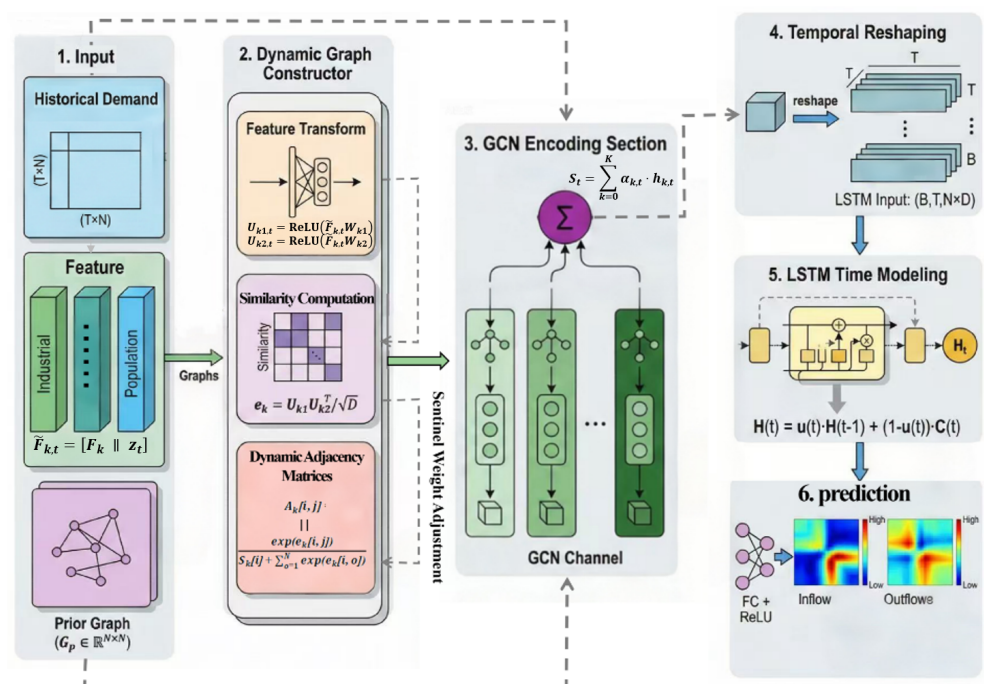


Figure 5. Multi-Feature Dynamic GCN-LSTM.

Since $\tilde{F}_{k,t}$ incorporates the demand-context vector z_t , the adjacency matrix is updated over time, thereby forming a dynamic graph structure.

To alleviate the curse of dimensionality and the risk of overfitting introduced by high-dimensional regional features, this study incorporates a Sentinel mechanism into the attention-based graph construction framework to adaptively regulate node-level information propagation strength. By learning the effectiveness and reliability of each region on the current feature graph, this mechanism suppresses the interference of noisy or weakly relevant regions in spatial dependency modeling, thereby improving the robustness of adjacency estimation and the predictive performance. Specifically, the Sentinel mechanism

learns a node-wise sentinel weight vector $S_{k,t} \in \mathbb{R}^N$, whose i -th element $S_{k,t}[i]$ reflects the suppression strength of node i . The sentinel weights are defined as:

$$S_{k,t} = \sigma(\tilde{F}_{k,t} W_s) \in \mathbb{R}^N \quad (12)$$

After obtaining $S_{k,t}$, it is incorporated into the attention normalization process so that the normalization term accounts not only for the similarity contributions from neighboring nodes but also for a node-wise suppression strength. This leads to the sentinel-augmented adjacency matrix:

$$A_{k,t}[i, j] = \frac{\exp(e_{k,t}[i, j])}{S_{k,t}[i] + \sum_{o=1}^N \exp(e_{k,t}[i, o])} \quad (13)$$

This mechanism enables the graph construction process to adaptively adjust inter regional influence strengths across different time steps, achieving “multi feature dynamic” modeling.

(2) Graph Convolutional Spatial Encoding and Dynamic Fusion

In the multi-feature dynamic graph convolutional encoder, regional attributes are used as model inputs, and a corresponding graph channel is constructed for each feature type. For the k -th feature, the adjacency matrix $A_{k,t}$ at time step t is generated using the aforementioned attention-based dynamic graph construction method, so that the spatial dependencies captured by each feature graph can be adaptively updated with changing demand states. The graph convolution on each channel is formulated as:

$$h_{k,t} = \text{GCN}(x_t, A_{k,t}), \quad k = 0, 1, \dots, K \quad (14)$$

where $h_{k,t} \in \mathbb{R}^{N \times D}$ denotes the spatial feature representation of N regions in channel k , and D is the hidden feature dimension.

To effectively integrate spatial information from different feature graphs, we employ a weighted fusion scheme. Specifically, the graph-convolutional representations from each channel are aggregated with learnable weights. The weight $\alpha_{k,t}$ for each feature graph is learned via a softmax function:

$$\alpha_{k,t} = \frac{\exp(\beta_{k,t})}{\sum_{m=0}^K \exp(\beta_{m,t})} \quad (15)$$

Finally, the unified spatial representation at time step t is obtained by weighted aggregation across all channels:

$$S_t = \sum_{k=0}^K \alpha_{k,t} \cdot h_{k,t} \quad (16)$$

where $\beta_{k,t}$ is a learnable importance score for each graph, and S_t denotes the final fused spatial feature representation.

(3) LSTM-based temporal modeling and final prediction

To capture temporal patterns, we employ an LSTM network to model the dynamic variations of freight demand. At each time step t , the fused spatial representation S_t produced by the multi-feature graph convolutional encoder is used as the input to the LSTM, which learns the temporal dependencies of freight demand. The input tensor to the LSTM has the following shape:

$$\text{LSTMInput} = [B, T, N \times D] \quad (17)$$

where B denotes the batch size, T is the number of time steps, N is the number of regions, and D is the feature dimension.

For each time step t , the LSTM combines the hidden state from the previous time step H_{t-1} with the current input features to compute the current hidden state H_t , which is updated as:

$$H_t = u_t \cdot H_{t-1} + (1 - u_t) \cdot C_t \quad (18)$$

Here, u_t is the update gate that controls how much information from H_{t-1} is retained; C_t is the candidate memory cell at time t ; and H_t is the hidden state at time t , capturing the temporal representation learned from the input sequence up to time t .

After temporal modeling, the final demand prediction is produced through a fully connected layer (FC) based on the learned temporal representation:

$$\hat{X}_{t+1} = FC(H_t) \quad (19)$$

where H_t is the hidden state at time step t , which encodes the spatio-temporal information by recursively integrating the fused spatial representations $\{S_\tau\}_{\tau=1}^t$.

4. Comparison and Analysis of Experimental Results

4.1. Research Area Scope and Dataset Description

This study focuses on all districts and counties of Xi'an City, Shaanxi Province, located within the geographical bounds of [108.0044433, 33.822271666, 109.75459166, 34.698108333]. The study area comprises 11 urban districts and 2 counties. The dataset consists of truck GPS data, AOI data, POI data, road network data, population and economic data, and land use classification data, with detailed descriptions provided in Table 1.

Table 1. Data description.

Data Type	Key Fields	Date	Brief Description
Truck GPS Data	Longitude, Latitude, UTC Time, System Time, GPS Speed (km/h), Province, City, District	1–7 November 2021	The GPS points are sampled at 30-second intervals. The dataset contains approximately 56,000 trucks with 138 million trajectory records.
AOI Data	AOI Name, Address, Type	2021	Vector data of Xi'an's AOIs in WGS-84 coordinate system. AOIs are polygon/region-based points of interest.
POI Data	Name, Category, Address, Administrative Division, Coordinates	2021	POIs are classified into 6 categories: Corporations, Shopping Services, Dining Services, Transportation Facilities, Commercial Buildings, and Automotive Services.
Road Network Data	Road ID, Road Type, Road Name, One-Way Flag	2021	Vector road network data of Xi'an in WGS-84 coordinate system. Roads are categorized into four types: expressways, arterial roads, secondary roads, and branch roads.
Population & Economic Data	Population Density, GDP Density	2021	Raster data with 1 km spatial resolution. Units are persons/km ² and 10,000 yuan/km ² , respectively.
Land Use Data	Industrial, Transportation, Medical, Residential, Commercial, Educational & Research Land	2021	Data in CGCS2000 coordinate system. Spatial distributions of land use types are vectorized into 500 m × 500 m grid cells.

4.2. Generation and Distribution of Truck Trip Origin-Destination Patterns

This study extracts OD patterns for freight vehicles based on GPS trajectory data from 56,049 trucks in Xi'an over one week (totaling 138 million records). Initial stop identification

was performed by calculating the distance, time interval, and average speed between adjacent trajectory points, yielding 1,120,171 initial stop points. Each truck generated an average of 19.97 stop points. The spatial distribution heatmap of initial stop points revealed distinct clustering patterns across the entire study area, as shown in Figure 6.

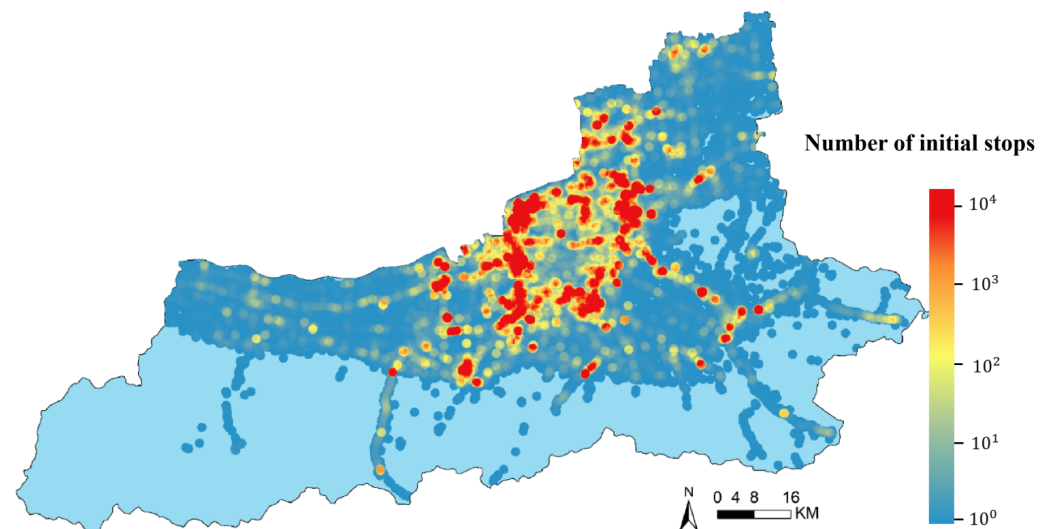


Figure 6. Heatmap of Initial Truck Stops.

To distinguish between valid stops and spurious stops, a nonparametric iterative method based on the Lorenz curve was employed to determine seven time thresholds: 779 min, 270 min, 104 min, 43 min, 14 min, 4 min, and 1 min, as shown in the Figure 7.

Using the seventh-level time thresholds as breakpoints, the distribution of parking durations at all stopping points was statistically analyzed. The results are shown in Table 2.

Table 2. Statistics of Truck Thresholds at Different Times.

Dwell Time (min)	Number of Stops	Cumulative Stops	Percentage	Cumulative Percentage
(0, 1)	283,617	283,617	25.32%	25.32%
[1, 4)	69,787	353,404	6.23%	31.55%
[4, 14)	324,959	678,363	29.01%	60.56%
[14, 43)	197,642	876,005	17.64%	78.20%
[43, 104)	101,051	977,056	9.02%	87.22%
[104, 270)	58,262	1,035,318	5.20%	92.43%
[270, 779)	47,739	1,083,057	4.26%	96.69%
≥ 779	37,114	1,120,171	3.31%	100.00%

Statistical analysis reveals that 60.56% of initial stops last within 14 min, implying that trucks frequently make brief pauses during freight trips—primarily for reasons such as waiting at traffic signals or queuing in congestion. Previous studies have applied stop duration thresholds of 10 min and 26 min to define truck trip endpoints. To balance the risk of excluding meaningful stops (if the threshold is too high) and retaining excessive invalid stops (if the threshold is too low), this study selects 14 min from the seven candidate thresholds as the criterion for identifying potential freight-related stops.

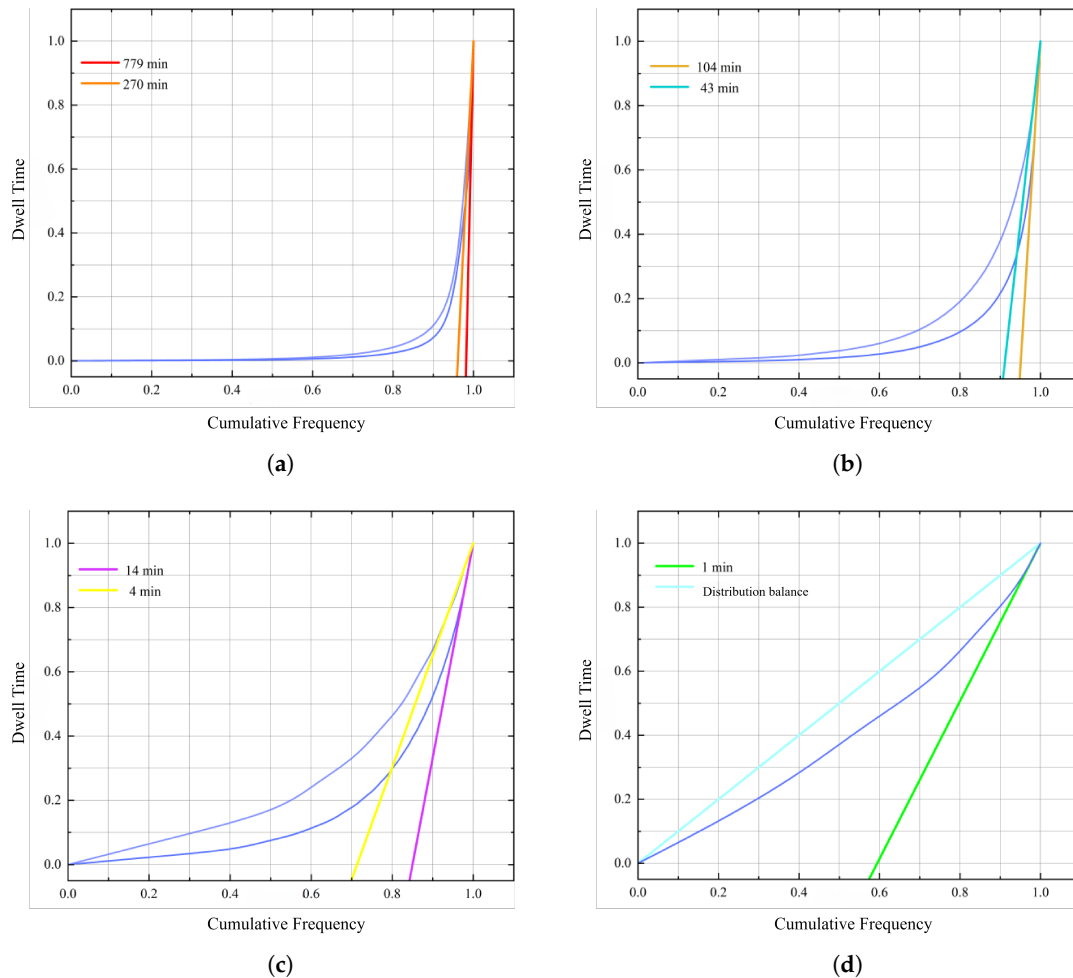


Figure 7. Time threshold diagram. (a) First- and second-level time thresholds. (b) Third- and fourth-level time thresholds. (c) Fifth- and sixth-level time thresholds. (d) Seventh-level time threshold. The blue line shows the Lorenz curve based on the cumulative distribution of parking durations.

The cumulative distribution curve of parking duration reveals a significant inflection point at 14 min (Figure 8). Prior to this threshold, the curve exhibits a steeper slope, which noticeably flattens afterward. This indicates that 14 min serves as a critical transition point in parking behavior patterns.

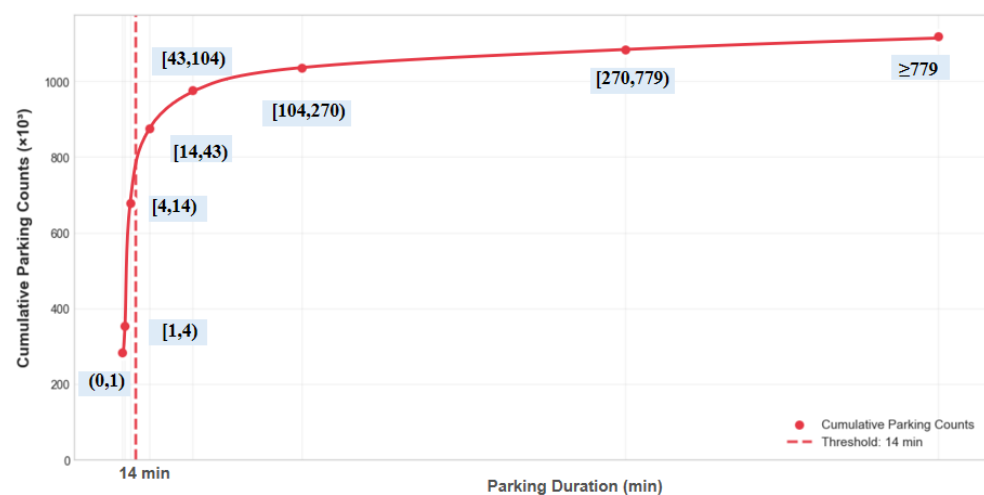


Figure 8. Cumulative distribution of truck parking duration.

According to China's Code for Design of Urban Road Engineering [31], and in conjunction with relevant studies [15,32], considering the influence of bicycle lanes, sidewalks, and green belts along urban roads, this study sets buffer distances of 40 m for expressways, 35 m for arterial roads, and 30 m for sub-arterial roads. Truck stop points falling within these road-network buffers were screened out and removed (Figure 9), eliminating 85,217 falsely identified stops caused by traffic congestion (19.3% of the total).

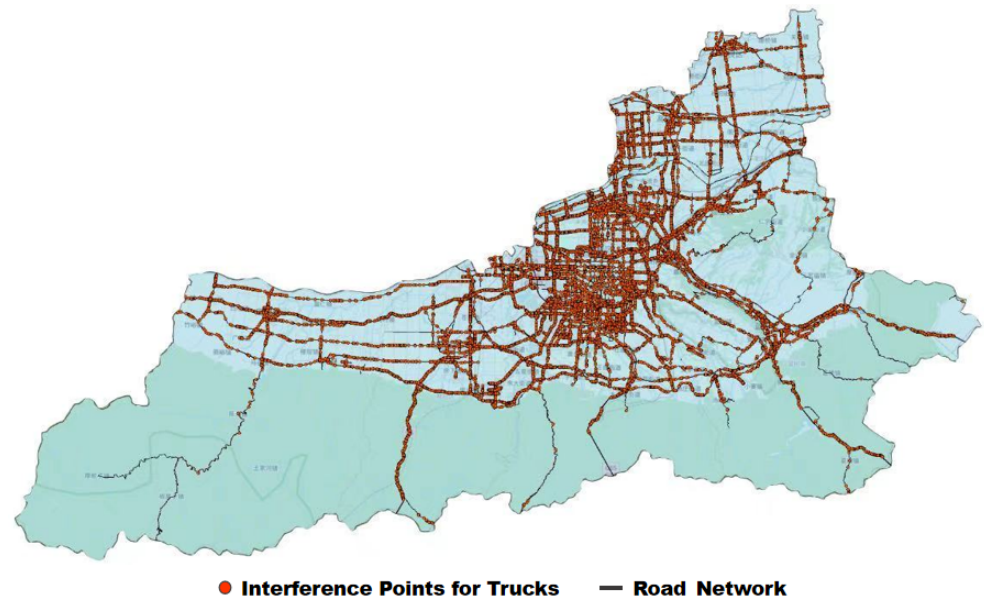


Figure 9. Road network-based filtering of interference stop points.

Based on high-precision AOI data, 15,329 stop points (accounting for 3.5%) located in areas such as expressway service areas, parking lots, and maintenance stations were excluded, as shown in Figure 10. A total of 341,262 stop points (77.2% of the total) within the 14–779 min duration range were retained.

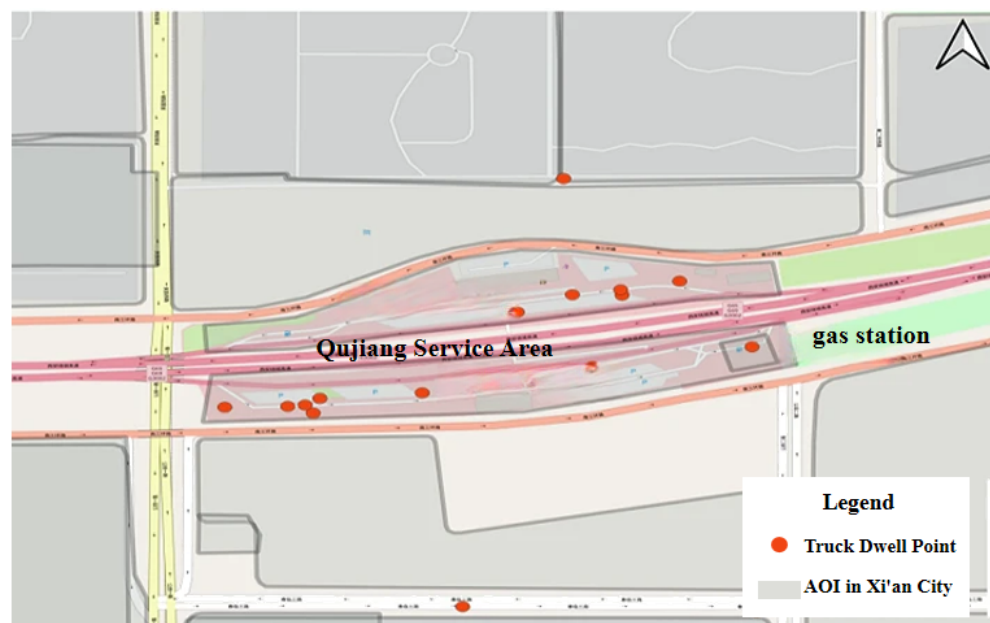


Figure 10. AOI-Based interference points in truck stops.

To validate the effectiveness of the method, this study employs a spatial sampling verification approach based on high-resolution satellite imagery and freight-related points

of interest (POI) data. The verification framework follows the methodological principles proposed by Gingerich et al. [7], combining systematic spatial sampling with manual interpretation to evaluate the accuracy of the stop-identification algorithm. Within the study area, complete daily GPS trajectories of 500 randomly selected trucks and the trip endpoints identified by the algorithm were used as validation samples. Through manual interpretation, 1151 true freight-related stop points were annotated. Dense sampling was applied around the 14-min interval (a critical parking-duration threshold) with sparse sampling in the range of 20 to 43 min, ultimately defining 17 candidate thresholds (10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 24, 28, 32, 36, 40, and 43 min). For each sampled trip endpoint, we manually located it on high-resolution satellite imagery and, considering the surrounding scene (e.g., logistics parks, factories, parking lots) and spatial proximity to freight POIs, determined whether it represented a genuine freight-activity point. The identification outcomes were classified into three categories: correctly identified true freight activity points (N_A), falsely identified non-freight points (N_M), and missed true freight points (N_E). We then quantified overall performance using the macro-accuracy metric:

$$\text{Mac} = \frac{N_A}{N_A + N_M + N_E}. \quad (20)$$

The validation results for different dwell-time thresholds are presented in Table 3.

Table 3. Sampling-based validation results for different dwell-time thresholds.

Sampled Threshold (min)	N_A	N_M	N_E	Mac
10	1090	268	61	0.768
11	1084	189	67	0.809
12	1074	143	77	0.830
13	1069	97	82	0.857
14	1067	60	84	0.881
15	1050	57	101	0.869
16	1035	54	116	0.859
17	1028	50	123	0.856
18	1016	50	135	0.846
19	1001	47	150	0.836
20	984	46	167	0.822
24	934	43	217	0.782
28	895	39	256	0.752
32	850	38	301	0.715
36	760	33	391	0.642
40	724	29	427	0.614
43	689	25	462	0.586

The results show that when the threshold is set to 14 min, $\text{Mac} = 0.881$, which is the highest among the 17 candidate thresholds and forms a clear peak (Figure 11). Further analysis indicates that when the threshold is lower than 14 min (e.g., 13 min), N_M increases markedly, whereas when the threshold is higher than 14 min (e.g., 15 min or 20 min), N_E rises substantially. The 14 min threshold therefore lies at the “turning point” between the increasing trends of false positives and false negatives, achieving the best trade-off between the two, and is thus selected as the criterion for identifying potential effective stops.

Based on this method, a total of 341,262 valid stops were extracted from 441,808 potential stops. Using the identified weekly truck trip endpoints, kernel density analysis was applied to visualize the spatial distribution of truck travel demand. The density values were classified into seven levels using the natural breaks method. Nine Level 1 hotspots, as shown in Figure 12, exhibited clear spatial patterns: they were predominantly clustered around

major road network nodes and expressway interchanges, showing a high degree of spatial overlap with logistics parks and industrial zones. This distribution pattern confirms that the freight network in Xi'an is centered on the Ring Expressway as its primary corridor, with interchanges naturally evolving into regional logistics hubs. Simultaneously, the hotspots observed in peripheral districts such as Huiyi and Gaoling reflect the spatiotemporal effects of “decentralized manufacturing operations.”

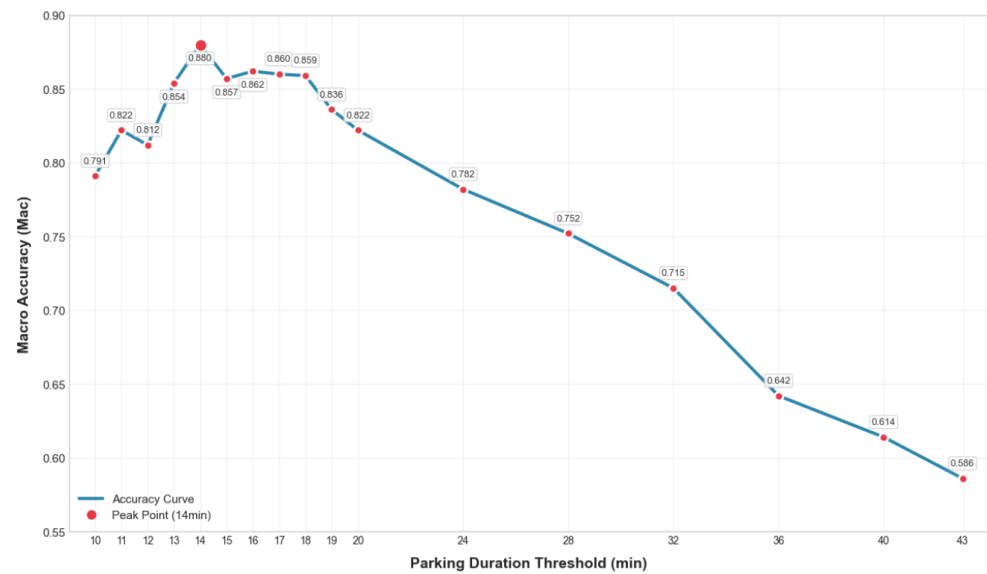


Figure 11. Comparison of sampling-based validation results across candidate dwell-time thresholds.

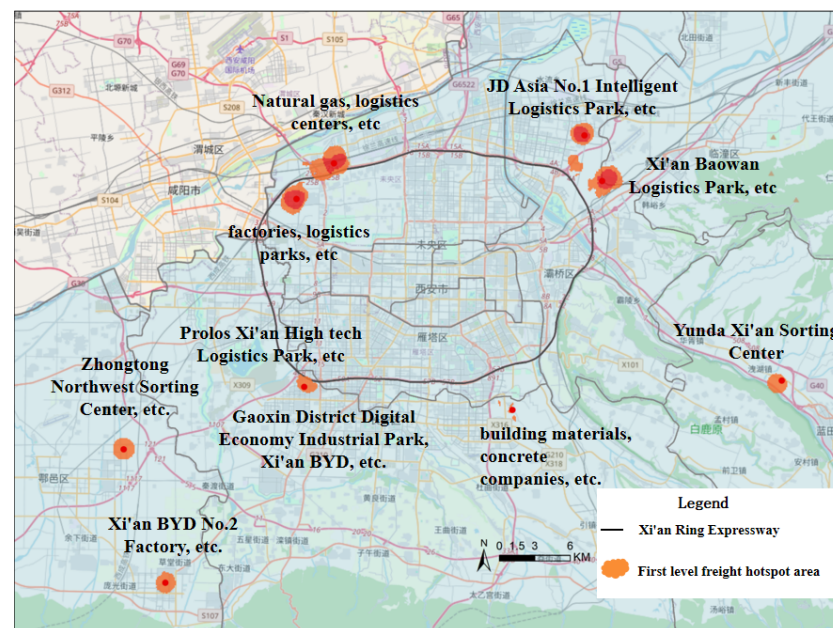


Figure 12. First-level hotspot area for truck travel demand within one week.

4.3. Truck Travel Demand Forecasting

In this study, the truck travel demand forecasting area was selected based on the observed concentration of demand around Xi'an's ring expressway, with suburban areas such as Zhouzhi County and Lantian County showing relatively low truck traffic. The chosen forecasting area includes Beilin, Xincheng, Lianhu, Weiyang, Yanta, and Baqiao Districts. The study area was then divided into 1 km × 1 km grids, as illustrated in the Figure 13.

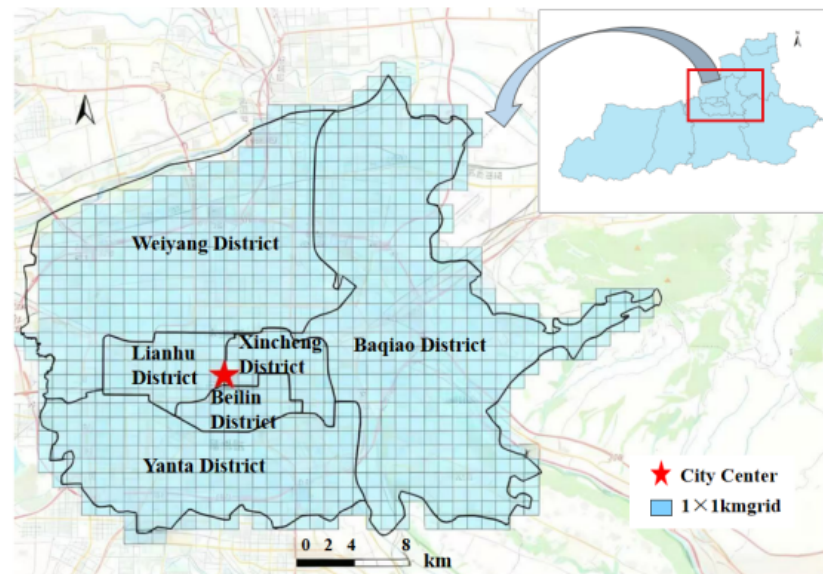


Figure 13. Study area.

Truck travel demand from 1 November to 7 November 2021 was used as the dependent variable for prediction. Based on the classic built environment theory and the characteristics of freight transportation, we constructed 16 initial regional feature variables across four core dimensions: land use type, population & economy, road network data, and built environment. The specific contents of the variables are shown in Table 4.

Table 4. Selection of Influencing Factors.

Factor	Variable	Variable Calculation
Land Use Type	Industrial Land Area	km^2/km^2
	Commercial Land Area	km^2/km^2
	Residential	km^2/km^2
	Public Administration & Service	km^2/km^2
Population & Economy	Population Density	$\text{persons}/\text{km}^2$
	Economy GDP	$10^4 \text{ CNY}/\text{km}^2$
Road Network Data	Expressway Density	km/km^2
	Arterial Road	km/km^2
	Secondary Arterial Road	km/km^2
	Branch Road Density	km/km^2
Built Environment	Corporations & Enterprises	count/km^2
	Shopping Services	count/km^2
	Food & Beverage Services	count/km^2
	Transportation Facilities	count/km^2
	Automotive Services	count/km^2
	Commercial & Enterprises	count/km^2

Following multicollinearity diagnostics and variable screening, six features significantly correlated with truck travel demand were ultimately selected. The significance of each feature was evaluated using the p -value ($\text{Prob}(t\text{-Statistic})$) from the t -test, where a p -value < 0.05 indicates a significant regression relationship between the independent variable and truck travel demand. Table 5 presents the six significant variables and their corresponding p -values.

Table 5. Significant variables related to truck travel demand.

No.	Significant Variable	<i>t</i> -Statistic	$p > t $
1	Industrial Land Area	11.368	0.001
2	Commercial Land Area	−7.598	0.023
3	Population Density	5.598	0.007
4	Expressway Density	7.273	0.005
5	Branch Road Density	8.874	0.004
6	Companies & Enterprises	2.930	0.001

During training, the model was configured with an input feature dimension of 64, a hidden dimension of 128, and comprised 2 LSTM layers and 2 GCN layers. The output dimension of the model, representing predicted demand, was set to 1. Regional features were normalized using min-max scaling. The dataset was partitioned into 80% for training and 20% for validation, and subsequently input into the model. To comprehensively evaluate model performance, we employed a complementary set of metrics. Mean Absolute Error (MAE) measures the average absolute deviation between predictions and true values. Root Mean Square Error (RMSE) is more sensitive to larger errors and reflects the stability of the prediction. The Mean Absolute Percentage Error (MAPE) provides a relative error perspective, facilitating cross-scenario comparison. The Coefficient of Determination (R^2) assesses the proportion of variance explained by the model. Together, these metrics capture prediction accuracy from different angles to form a holistic evaluation framework.

To assess how much individual features contributed to improved prediction performance, we conducted feature impact studies by adding or removing features. The performance of different feature combinations is presented in Table 6.

Table 6. Model performance with different feature combinations.

No. of Features	Feature Combination	MAE	RMSE	MAPE
0	No external features	7.891	13.043	12.11%
5	All features excluding Industrial Land Area	6.668	11.057	11.01%
	All features excluding Commercial Land Area	6.532	11.012	10.63%
	All features excluding Population Density	6.526	10.237	10.15%
	All features excluding Expressway Density	6.530	10.243	10.17%
	All features excluding Branch Road Density	6.553	10.537	10.45%
	All features excluding Companies & Enterprises	6.557	10.674	10.47%
6	All features	6.523	10.125	10.01%

Based on the results in the table, incorporating regional features significantly reduces the MAE, indicating that the inclusion of such features improves the model's accuracy in predicting truck travel demand. In particular, industrial land area exerts a substantial influence on model performance, which aligns with the strong known correlation between truck travel demand and industrial land use.

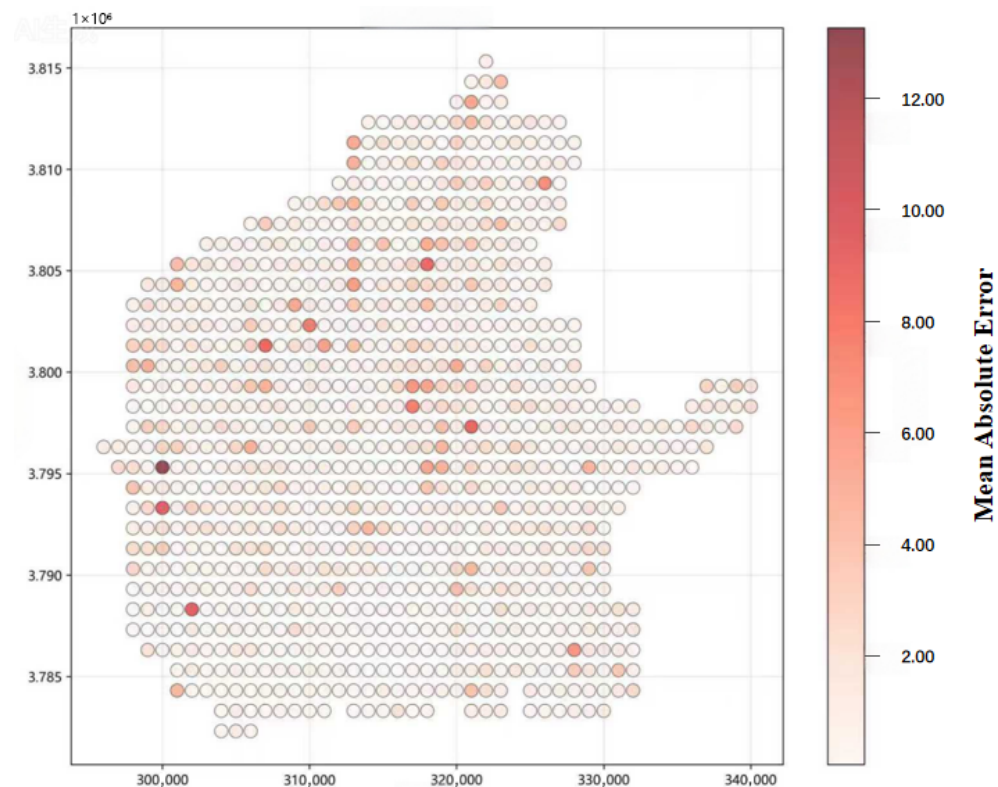
To comprehensively evaluate the model performance, this study selects STGCN, Graph WaveNet, STGAT, TCN, LSTM, and GCN-LSTM as comparative models. The corresponding prediction results are presented in Table 7.

The prediction results in the table reveal the following:

(1) In terms of MAE, the Multi-Feature Dynamic GCN-LSTM model achieved an MAE value of 6.523, significantly lower than the other models. This model demonstrated higher accuracy in predicting truck travel demand, effectively capturing complex patterns and nonlinear relationships within the data. Figure 14 illustrates the spatial distribution error of the predicted values.

Table 7. Overall and peak-hour performance comparison of truck demand prediction models.

Model	MAE	RMSE	R ²	MAPE	Peak-Hour MAE	Peak-Hour RMSE
STGAT	11.617	25.465	0.677	21.67%	1.917	2.621
STGCN	9.890	16.322	0.711	16.59%	1.127	3.013
GraphWaveNet	8.824	14.546	0.768	14.86%	1.141	2.539
TCN	8.023	13.732	0.802	12.47%	1.108	2.203
GCN-LSTM	7.891	13.043	0.813	12.11%	1.112	2.164
Multi-Feature Dynamic GCN-LSTM	6.523	10.125	0.897	10.01%	1.096	2.164

**Figure 14.** Spatial error distribution.

(2) Compared to other models, our model achieves a significant reduction in RMSE while maintaining high accuracy and good robustness ($RMSE/MAE = 1.55$), outperforming most comparative models. MAPE of 10.01% indicates that the model maintains stable relative accuracy across regions with varying demand levels. Figure 15 presents normalized truck travel demand values with random sampling to demonstrate the prediction performance of the Multi-Feature Dynamic GCN-LSTM model.

(3) In terms of the R^2 , the Multi-Feature Dynamic GCN-LSTM model attained the highest performance among all compared models, achieving a value of 0.897. This high level of fit reflects the model's superior capability to capture underlying trends and patterns in the data. Furthermore, given that freight demand is more prone to abrupt fluctuations and spatial spillover effects during peak hours, we conducted predictions specifically for these periods. The experimental results demonstrate that our model achieves lower MAE and RMSE during peak hours.

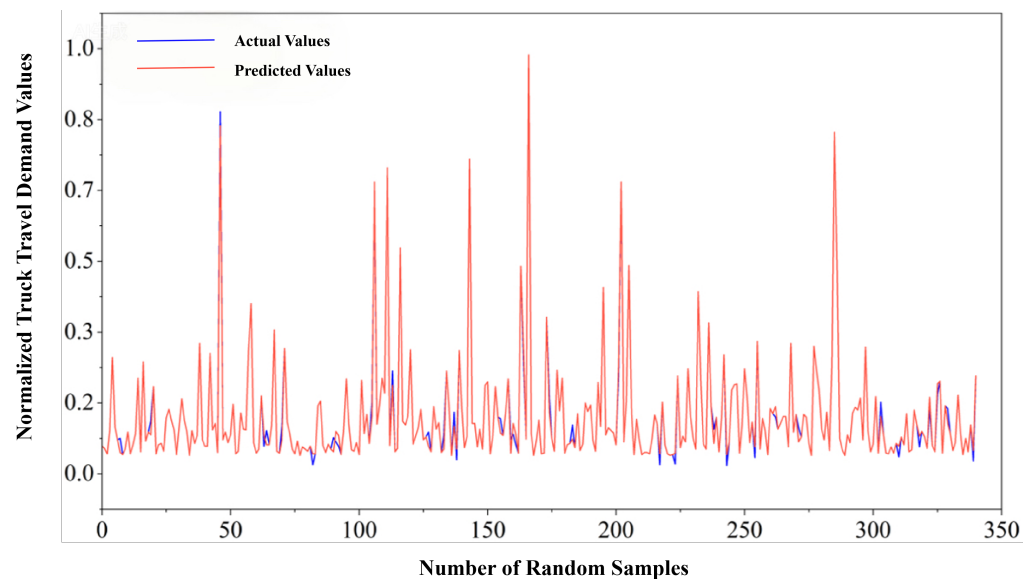


Figure 15. Prediction performance of truck travel demand.

In summary, the comparison of six models for predicting truck travel demand reveals that the Multi-Feature Dynamic GCN-LSTM model consistently outperformed the others across multiple metrics—including mean absolute error, accuracy, and goodness-of-fit. These results highlight its distinct advantages in handling spatio-temporal data such as truck travel demand, enabling more accurate identification of complex relationships and dynamic variations, and ultimately demonstrating stronger predictive capability.

5. Conclusions

This study, based on city-scale truck trajectory data, develops and validates a two-stage analytical framework of “truck trip endpoint identification—truck travel demand prediction,” aiming to provide transferable and scalable methodological support for fine-grained characterization of urban freight activities and management decision-making. The main findings and contributions of the research are reflected in the following aspects:

(1) Construction of the framework for truck endpoint identification and demand prediction under the operational conditions of the specified study area during the data collection period, potential truck stop locations are effectively identified by introducing a Lorenz curve-based non-parametric iterative method and reasonably setting a dwell-time threshold. By further integrating road network and area of interest (AOI) data, non-freight-related stops are filtered out, enabling the precise extraction of truck trip endpoints. On this basis, a Multi-Feature Dynamic GCN-LSTM model is constructed to predict truck travel demand. This model comprehensively incorporates land use characteristics, socioeconomic attributes, road network density, and truck-related points of interest (POIs). Comparative results indicate that the proposed model outperforms existing baseline models in key performance metrics, including mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2), demonstrating superior predictive capabilities. (2) Governance pathways and contributions toward sustainable urban freight from an environmental sustainability perspective, endpoint identification and demand prediction provide a foundation for optimizing freight routes, reducing empty trips and inefficient detours, thereby lowering fuel consumption and greenhouse gas emissions. For core hub areas with highly concentrated freight activities, the identification results can support the formulation of time-based loading/unloading strategies and the planning of truck staging/buffer facilities, mitigating concentrated exhaust emissions at the source. For freight corridors experiencing sustained growth, the

predictive model offers forward-looking evidence for planning low-carbon infrastructure, such as priority lanes for new-energy trucks and the layout of charging facilities.

From an economic and systemic sustainability perspective, the demand prediction results provide data-driven decision-making support for targeted investments in freight infrastructure (e.g., hub capacity expansion and prioritization of road upgrades), avoiding inefficient allocation of public resources. This enhances the resilience and long-term operational efficiency of urban logistics systems and contributes to building an economical, efficient, low-carbon, and resilient modern freight system.

At the same time, it is important to note the limitations of this study. Due to data availability constraints, this research utilizes one week of continuous truck trajectory data, which may not fully capture cyclical and seasonal patterns. Additionally, this study does not directly quantify the transportation externalities around freight hotspots, such as contributions to congestion, passenger-freight conflicts, safety risks, or emission impacts. The current spatial filtering of truck endpoints relies on a fixed buffer radius based on road design standards, which may not adequately capture roadside environments and local traffic characteristics. Future research could expand the temporal and spatial scope of the data and integrate multi-source data, such as floating car speed data, accident records, and emission models, to conduct quantitative analyses of the congestion, safety, and environmental impacts of freight activities. This would further establish a complete data-driven freight governance chain from “identification—prediction—evaluation—intervention,” while incorporating high-definition maps and other multi-source data to determine more flexible buffer radii.

In summary, this study enhances the accuracy of characterizing truck travel behavior through methodological innovation and proposes management application pathways with sustainability implications. It provides a reusable analytical framework for the governance of future urban freight systems, applicable across different time periods and cities. By transforming data-driven insights into integrated decision-making support for environmentally friendly, socially inclusive, and economically efficient policies, this framework contributes to the achievement of sustainable urban transportation and logistics goals.

Author Contributions: Conceptualization, X.C. and Y.Y.; methodology, J.Z., T.Z. and J.S.; software, Y.Y. and T.Z.; data collection, J.Z., X.C. and Y.Y.; formal analysis, Y.Y. and T.Z.; investigation, X.C.; writing—original draft preparation, Y.Y.; writing—review and editing, X.C. and J.S.; visualization, J.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data supporting the reported results are available upon request from the authors.

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

References

1. Behrends, S. Recent Developments in Urban Logistics Research—A Review of the Proceedings of the International Conference on City Logistics 2009–2013*. *Transp. Res. Procedia* **2016**, *12*, 278–287.
2. Zhang, R.; Li, L.; Jian, W. Urban Freight Vehicle Type Choice Model and Application. *J. Transp. Syst. Eng. Inf. Technol.* **2018**, *18*, 135–141. <https://doi.org/10.16097/j.cnki.1009-6744.2018.04.021>.
3. Demir, E.; Bektaş, T.; Laporte, G. A review of recent research on green road freight transportation. *Eur. J. Oper. Res.* **2014**, *237*, 775–793.
4. Chen Xiaohong, Liu Han, Y.Z. Research on the demand model and application of heavy freight transportation. *Transp. Port Navig.* **2023**, *10*, 253–263.

5. Allen, J.; Piecyk, M.; Piotrowska, M.; McLeod, F.; Cherrett, T.; Ghali, K.; Nguyen, T.; Bektas, T.; Bates, O.; Friday, A.; et al. 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.
6. Ma, X.; Wang, Y.; McCormack, E.; Wang, Y. Understanding freight trip-chaining behavior using a spatial data-mining approach with GPS data. *Transp. Res. Rec.* **2016**, *2596*, 44–54.
7. 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.
8. 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.
9. Thakur, A.; Pinjari, A.; Zanjani, A.; Short, J.; Mysore, V.; Tabatabaee, S. Development of Algorithms to Convert Large Streams of Truck GPS Data into Truck Trips. *Transp. Res. Rec. J. Transp. Res. Board* **2015**, *2529*, 66–73. <https://doi.org/10.3141/2529-07>.
10. 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.
11. 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.
12. Liu Li, Z.J. City freight stopping behavior recognition model based on truck GPS data. *Transp. Port Navig.* **2023**, *10*, 45–50. <https://doi.org/10.16487/j.cnki.issn2095-7491.2023.03.011>.
13. Gupta, N.; Mujumdar, S.; Patel, H.; Masuda, S.; Panwar, N.; Bandyopadhyay, S.; Mehta, S.; Guttula, S.; Afzal, S.; Sharma Mittal, R.; et al. Data quality for machine learning tasks. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, Online, 14–18 August 2021; pp. 4040–4041.
14. Hughes, S.; Moreno, S.; Yushimito, W.F.; Huerta-Cánepa, G. Evaluation of machine learning methodologies to predict stop delivery times from GPS data. *Transp. Res. Part C Emerg. Technol.* **2019**, *109*, 289–304.
15. 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.
16. Seya, H.; Zhang, J.; Chikaraishi, M.; Jiang, Y. Decisions on truck parking place and time on expressways: An analysis using digital tachograph data. *Transportation* **2020**, *47*, 555–583.
17. Williams, B.; Durvasula, P.; Brown, D. Urban Freeway Traffic Flow Prediction: Application of Seasonal Autoregressive Integrated Moving Average and Exponential Smoothing Models. *Transp. Res. Rec.* **1998**, *1644*, 132–141. <https://doi.org/10.3141/1644-14>.
18. Wu, C.H.; Ho, J.M.; Lee, D. Travel-time prediction with support vector regression. *IEEE Trans. Intell. Transp. Syst.* **2004**, *5*, 276–281. <https://doi.org/10.1109/TITS.2004.837813>.
19. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. <https://doi.org/10.1145/2939672.2939785>.
20. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2017**, arXiv:cs.LG/1609.02907.
21. Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Liò, P.; Bengio, Y. Graph Attention Networks. *arXiv* **2018**, arXiv:stat.ML/1710.10903.
22. Yu, B.; Yin, H.; Zhu, Z. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence. International Joint Conferences on Artificial Intelligence Organization, Stockholm, Sweden, 13–19 July 2018; pp. 3634–3640. <https://doi.org/10.24963/ijcai.2018/505>.
23. Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wan, H. Attention Based Spatial-Temporal Graph Convolutional Networks for Traffic Flow Forecasting. *Proc. AAAI Conf. Artif. Intell.* **2019**, *33*, 922–929. <https://doi.org/10.1609/aaai.v33i01.3301922>.
24. Han, S.; An, J.; Park, Y.; Kim, S.; Jang, K.; Lee, D. Multiple Areal Feature Aware Transportation Demand Prediction. *arXiv* **2024**, arXiv:cs.AI/2408.12890.
25. Luo, T. Research on Freight Traffic Demand Prediction Based on Truck GPS Data. Master's Thesis, Wuhan University, Wuhan, China, 2022.
26. Yu, H.; Jiang, R.; He, Z.; Zheng, Z.; Li, L.; Liu, R.; Chen, X. Automated vehicle-involved traffic flow studies: A survey of assumptions, models, speculations, and perspectives. *Transp. Res. Part C Emerg. Technol.* **2021**, *127*, 103101.
27. Ahn, S.; Han, S.; Al-Hussein, M. Improvement of transportation cost estimation for prefabricated construction using geo-fence-based large-scale GPS data feature extraction and support vector regression. *Adv. Eng. Inform.* **2020**, *43*, 101012.
28. Li, X. Analysis and Prediction of Truck Travel Path Preferences Based on Deep Learning. Master's Thesis, Southwest Jiaotong University, Chengdu, China, 2020.
29. Wang, S.; Shao, C.; Dong, C.; Huang, S.; Zheng, Y. Recurrent neural network model for predicting truck traffic flow demand based on GPS trajectory data. *J. Beijing Jiaotong Univ.* **2021**, *45*, 15–23.
30. Lorenz, M.O. Methods of Measuring the Concentration of Wealth. *Publ. Am. Stat. Assoc.* **1905**, *9*, 209–219.

31. CJJ37-2012; Code for Design of Urban Road Engineering. Ministry of Housing and Urban-Rural Development of the People's Republic of China: Beijing, China, 2012.
32. Wang, J. Travel Characteristics Analysis of Freight Vehicles Using GPS Data. Master's Thesis, Beijing Jiaotong University, Beijing, China, 2022. (In Chinese)

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.