Multi-Modal Collaborative Optimization for Sustainable Transportation Systems: Last-Mile Shuttle Service and Multi-Modal Integration Design

1st Ran Zhang
Kaishi Education Technology Co., Ltd
Shenzhen, China
502370984@qq.com

2nd Xingchang Liao
Kaishi Education Technology Co., Ltd
Shenzhen, China
307955299@qq.com

***Abstract*—As urbanization continues to accelerate, traffic congestion and environmental pollution have become increasingly serious challenges, with the "last-mile" problem emerging as a critical bottleneck limiting the overall performance of public transportation systems. Existing studies have largely focused on optimizing individual shuttle services, often overlooking the systemic synergies among multiple transport modes such as shuttles, metro systems, bike-sharing, and walking. As a result, notable research gaps remain, particularly in the evaluation of multi-modal connection efficiency and sustainability.**

To address these challenges, this study proposes a data-driven multi-modal collaborative optimization framework that integrates three core components: spatiotemporal demand analysis, multi-modal connection network design, and collaborative scheduling optimization. Specifically, travel demand hotspots are first identified using the DBSCAN clustering algorithm. A multi-modal connection network is then constructed to minimize transfer times and overall travel costs. Finally, an improved Genetic Algorithm (GA) is applied to jointly optimize shuttle routes, service frequencies, and their coordination with metro and bike-sharing systems. To further enhance the framework's relevance to sustainable mobility goals, a carbon emission model is incorporated to quantitatively evaluate environmental benefits.

A reproducible case study conducted in a typical Transit-Oriented Development (TOD) area in Shenzhen, China, demonstrates that the proposed collaborative optimization scheme significantly improves overall travel efficiency and reduces last-mile transportation-related carbon emissions compared with traditional single-mode or fixed-route shuttle services under identical experimental conditions. These results highlight the effectiveness of the proposed approach and its potential to support integrated, low-carbon last-mile transport planning.

Overall, this research provides a scientific methodology and practical decision-support tool for urban transport planners seeking to design efficient, convenient, and environmentally sustainable last-mile transportation systems. It offers both theoretical contributions and real-world implications for advancing sustainable urban mobility.

Keywords—Multi-Modal Transportation; Collaborative Optimization; Last-Mile; Shuttle Service; Sustainable Transportation; Transfer Connection

I. INTRODUCTION

In recent years, rapid global urbanization has driven economic and social development, but it has also imposed unprecedented pressure on urban transportation systems. Problems such as traffic congestion, air pollution, and rising carbon emissions have become increasingly prominent [1]. As a result, the construction of sustainable urban transportation systems has emerged as a core concern for governments and academic researchers worldwide [2]. Within this context, high-capacity and high-efficiency public transport modes—such as rail transit and Bus Rapid Transit (BRT) — are widely recognized as the backbone of urban mobility. However, the overall attractiveness and effectiveness of public transport systems are often constrained by the quality of "last-mile" services, namely, the efficiency of connections between major transport hubs and travelers' final destinations (and vice versa) [3]. Inconvenient and time-consuming last-mile connections remain a key reason for residents' continued reliance on private vehicles, which not only undermines the advantages of public transport but also exacerbates congestion and environmental pressures on urban transport systems [4].

To tackle this challenge, both academia and industry have explored a variety of last-mile solutions, including fixed-route feeder buses, Demand-Responsive Transit (DRT), bike-sharing systems (BSS), and ride-hailing services [5]. Among these, data-driven shuttle service design has attracted growing attention due to its flexibility and potential coverage advantages. Existing studies have primarily leveraged big data sources — such as mobile phone signaling data and public transport smart-card records — to optimize specific elements of shuttle services, for example, stop location selection using clustering algorithms [6] or route and timetable planning via heuristic optimization methods [7]. These efforts have significantly advanced the intelligence and operational efficiency of single-mode shuttle systems.

Nevertheless, several limitations remain in the current body of research. First, most studies treat last-mile shuttle services as isolated subsystems, neglecting their role as integral components of a broader multi-modal transport chain. In practice, passengers' travel experiences are shaped by the combined performance of "mainline transport + feeder services," with transfer time, cost, and coordination between modes playing a decisive role in travel choices [8]. A lack of dynamic coordination with other modes—such as metro systems and bike-sharing—can lead to timetable mismatches, inefficient transfers, and

Corresponding Author: Ran Zhang, Cuizhu North Road Petrochemical Industrial Zone, Cuizhu Street, Luohu District, Guangzhou, China, 518019, 502370984@qq.com

imbalances in shared-bike availability around stations, ultimately resulting in "collaborative inefficiency." Second, existing research has largely emphasized traditional efficiency metrics, such as operating costs or passenger travel time, while paying insufficient attention to sustainability objectives. Under the growing policy emphasis on "dual-carbon" targets, there is an urgent need to integrate environmental performance indicators — particularly carbon emissions — directly into optimization frameworks, so as to achieve a balanced improvement in economic, social, and environmental outcomes [9].

Against this backdrop, this study aims to address these research gaps by investigating the multi-modal collaborative optimization of last-mile transportation from the perspectives of system integration and sustainable development. The main objectives of this research are threefold:

To develop a comprehensive design framework that integrates travel demand identification, multi-modal network construction, and collaborative optimization, thereby establishing a complete workflow from data analysis to solution generation.

To propose a multi-objective optimization model that simultaneously enhances multi-modal connection efficiency and reduces total system carbon emissions, achieving coordinated improvements in operational performance and environmental sustainability.

To validate the effectiveness and practical applicability of the proposed framework and model through an empirical case study conducted in a typical Transit-Oriented Development (TOD) area within a high-density city (Shenzhen), China.

The remainder of this paper is organized as follows. Section 2 presents a systematic review of relevant literature. Section 3 details the proposed multi-modal collaborative optimization framework and key methodologies. Section 4 introduces the case study background, data sources, and preprocessing procedures. Section 5 reports and analyzes the optimization results. Section 6 discusses the findings in depth and compares them with existing studies. Finally, Section 7 concludes the paper and outlines directions for future research.

II. LITERATURE REVIEW

To clearly position the theoretical grounding and innovative contributions of this study, this chapter systematically reviews and synthesizes relevant literature from three perspectives: last-mile transportation solutions, data-driven travel demand analysis and service design, and multi-modal transportation networks with collaborative optimization.

A. Last-Mile Transportation Solutions

The First-and-Last-Mile Problem (FLMP) has long been recognized as a fundamental challenge in public transportation systems [10]. Both academia and practice have proposed various solutions to mitigate this issue. Among them, traditional fixed-route feeder buses remain the most widely adopted approach. Their strengths lie in operational simplicity and predictable costs; however, their inherent rigidity in routes and schedules makes them poorly suited to dynamic passenger demand. This often results in low vehicle utilization during off-peak periods and insufficient capacity during peak hours, ultimately undermining service quality [11].

To overcome these limitations, Demand-Responsive Transit (DRT) has emerged as a flexible alternative. By adjusting routes and schedules in response to real-time or pre-booked demand, DRT is considered particularly effective for low-density areas or specific time windows [12]. With the advancement of mobile internet technologies and intelligent dispatching algorithms, DRT has demonstrated considerable potential in last-mile applications. Many studies have focused on vehicle routing and dispatch optimization problems—often modeled as variants of the Vehicle Routing Problem (VRP) — with objectives such as minimizing operating costs or passenger waiting times [13]. Nevertheless, pure DRT systems face challenges including algorithmic complexity, reliance on booking platforms, and potential service degradation under highly concentrated demand conditions [14].

In parallel, micromobility modes, such as shared bicycles and electric scooters, have significantly diversified last-mile travel options [15]. Empirical studies show that bike-sharing systems can effectively complement public transport, extend service coverage, and reduce reliance on private cars for short-distance trips [16]. However, these systems are also prone to operational inefficiencies, most notably the "tidal effect," where vehicles accumulate excessively around transport hubs during peak periods and become scarce during counter-peak hours. This imbalance severely affects system reliability and user experience [17].

B. Data-Driven Traffic Demand Analysis and Service Design

The widespread availability of big data has fundamentally transformed transportation planning and management, enabling more refined and dynamic service design. Data sources such as public transport smart cards, mobile phone signaling, shared bicycle GPS traces, and ride-hailing records provide unprecedented insights into the spatiotemporal patterns of urban travel behavior [18]. In last-mile shuttle service research, data-driven approaches are primarily applied in two areas: demand analysis and network design.

For demand analysis, clustering algorithms play a critical role in identifying travel demand hotspots. For instance, Shu et al. (2021) employed the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm to analyze bike-sharing data, successfully identifying high-density last-mile demand areas near metro stations and supporting shuttle stop selection [7]. Similarly, K-means and its variants have been widely used to cluster origin – destination (OD) points into representative demand zones, thereby reducing the complexity of subsequent network planning tasks [19].

In terms of network design, once demand patterns are identified, researchers focus on optimizing shuttle stop locations, routes, and departure schedules. These problems are typically formulated as complex combinatorial optimization models, such as facility location problems or VRP variants. Heuristic and metaheuristic algorithms — including Genetic Algorithms (GA), Simulated Annealing (SA), and Tabu Search (TS) — are commonly adopted for solution [20]. While these studies have significantly improved the operational efficiency of individual shuttle systems, most assume that shuttles operate independently, with limited consideration of their interaction with other transport modes.

C. Multi-Modal Transportation Networks and Collaborative Optimization

To address the limitations of single-mode approaches, increasing attention has been devoted to multi-modal transportation systems, which conceptualize urban transport as an interconnected network comprising metro, buses, taxis, bike-sharing, and other modes [21]. Research in this domain primarily focuses on network modeling and transfer coordination.

In network modeling, researchers often construct multi-modal supernetworks that include physical nodes (e.g., stations), virtual nodes (e.g., transfer points), and multiple types of links corresponding to different transport modes. Edge weights are defined using generalized travel costs, incorporating factors such as travel time, monetary cost, comfort, and transfer penalties [22]. On this basis, shortest-path algorithms (e.g., Dijkstra's algorithm) can be applied to determine optimal multi-modal travel paths.

Transfer coordination aims to minimize passenger waiting times and improve connection reliability, typically through timetable synchronization between feeder services and mainline transit (e.g., metro or rail). Existing studies have proposed various optimization models, including integer programming formulations to minimize total transfer waiting time [23] and robust optimization models that account for demand uncertainty [24]. However, most of this research has focused on coordination between conventional buses and rail transit, with limited attention paid to integrating flexible shuttle services and micromobility modes into a unified collaborative optimization framework.

D. Research Gap and Contributions of This Study

The above review demonstrates that substantial progress has been made in last-mile solutions, data-driven service design, and multi-modal coordination. Nevertheless, a clear research gap persists between data-driven shuttle service optimization and system-level multi-modal collaborative optimization. Studies in the former domain excel at improving shuttle operations but often overlook their integration with external transport systems. Conversely, research on multi-modal networks tends to focus on macroscopic coordination mechanisms, while neglecting the fine-grained, dynamic design of flexible feeder services such as shuttles.

Moreover, the integration of sustainability objectives, particularly carbon emissions, as endogenous components of multi-modal optimization models remains underdeveloped.

To address these gaps, this study makes the following key contributions:

Integrated Framework: It is among the first to integrate data-driven shuttle service design (including stop location, routing, and frequency optimization) with system-level collaborative optimization of multi-modal transportation networks involving metro and bike-sharing systems.

Multi-Objective Optimization: It develops a multi-objective optimization model that simultaneously considers passenger travel efficiency and environmental performance (carbon emissions), enabling the identification of Pareto-optimal solutions that balance economic, social, and environmental benefits.

Real-World Application Orientation: Through an empirical study in a Transit-Oriented Development (TOD) area of a high-density Chinese metropolis, this research provides a practical, reproducible decision-support framework for addressing last-mile challenges in megacities.

III. RESEARCH METHODOLOGY

To achieve collaborative improvements in efficiency and sustainability for last-mile transportation, this study proposes a three-stage, data-driven multi-modal collaborative optimization framework. The framework follows a technical pathway of "demand identification → network construction → collaborative optimization," systematically covering the entire process from raw travel data processing to the generation of final operational plans. This chapter introduces the overall research framework and provides a detailed description of its core methodologies, including multi-modal connection network construction, the collaborative optimization model, and the solution algorithm.

A. Overall Research Framework

The proposed framework, illustrated in Figure 1, consists of three tightly interconnected stages.

1) Stage 1: Multi-Modal Travel Demand Discovery and Analysis.

This stage serves as the data foundation of the optimization framework. Its primary objective is to extract and characterize the spatiotemporal features of last-mile travel demand associated with rail transit stations using reproducible inputs. These inputs include publicly accessible station-level ridership statistics (or synthetic passenger arrival profiles derived from published aggregates), open-source road network data, and openly available (or synthetically generated) micromobility origin – destination (OD) samples. The raw datasets are first cleaned, matched, and standardized to construct travel records in a unified format. Subsequently, the DBSCAN spatial clustering algorithm is applied to identify spatially concentrated OD points, referred to as demand hotspots. In parallel, temporal analysis of trip timestamps is conducted to capture dynamic demand variations, which provides critical input for the time-dependent scheduling of shuttle services in later stages.

2) Stage 2: Multi-Modal Connection Network Construction.

Based on the identified demand patterns, this stage constructs a comprehensive multi-modal connection network to represent interactions among different transportation modes. The network is modeled as a weighted directed graph $G=(V,E)$, where the node set V includes four types of nodes: rail transit stations, candidate shuttle stops, shared bicycle parking areas, and demand points representing passengers' actual origins and destinations. The edge set E represents various connection types, including shuttle operating routes, transfer paths between modes, and walking or cycling links connecting stations, stops, and final destinations. Each edge is assigned a generalized travel cost, defined as a composite function of travel time, monetary cost, and transfer inconvenience, enabling a holistic evaluation of passenger travel impedance across the network.

3) Stage 3: Collaborative Optimization Model and Solution.

This stage constitutes the core of the proposed framework. A multi-objective optimization model is formulated to identify optimal last-mile service strategies, with dual objectives of minimizing total passenger travel time and minimizing total system carbon emissions. The decision variables include the selection of candidate shuttle stops, the design of shuttle routes, and the determination of route-specific departure frequencies across different time periods. The model is subject to a set of practical operational constraints, such as vehicle capacity limits, maximum acceptable walking distances for passengers, and service time windows. Given the NP-hard nature of the problem, an improved Non-dominated Sorting Genetic Algorithm II (NSGA-II) is developed to solve the model and generate a Pareto-optimal solution set. This provides decision-makers with multiple trade-off solutions that balance operational efficiency and environmental performance.

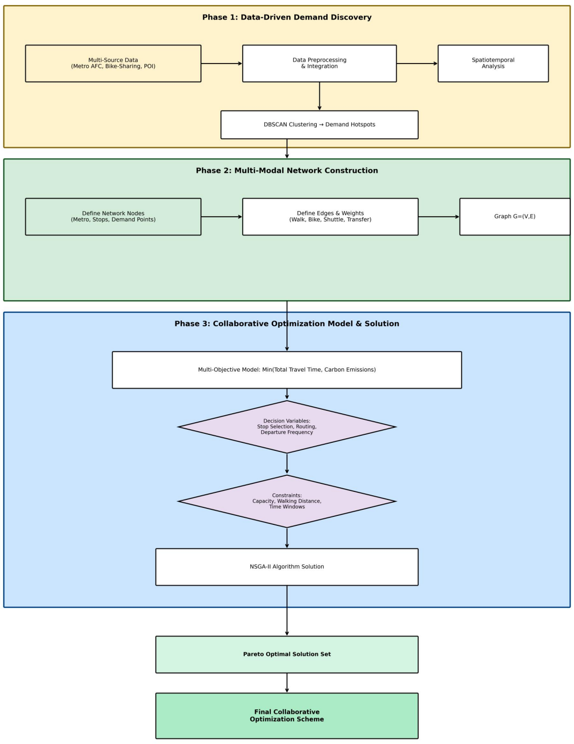

Fig. 1. Data-Driven Multi-Modal Collaborative Optimization Framework

B. Multi-Modal Connection Network Construction

The construction of the multi-modal connection network is a fundamental prerequisite for achieving collaborative optimization. In this study, the network is formally defined by its node set and edge set, as described below.

Nodes (V):

- Metro Stations (V_M): Serve as the primary hubs of last-mile services and act as the interface between trunk public transport and feeder modes.
- Candidate Shuttle Stops (V_C): A set of potential shuttle stop locations generated within identified demand hotspots, determined based on road accessibility, surrounding population density, and spatial feasibility.
- Bike-Sharing Parking Areas (V_B): Officially designated or empirically observed centralized parking locations for shared bicycles.
- Demand Points (V_D): Centroids of demand clusters obtained through spatial clustering analysis, representing the aggregated origins or destinations of passenger groups.

Edges (E) and Weights (W):

- Shuttle Travel Edges (E_S): Links between shuttle stops, with edge weights defined as the shuttle travel time along each segment.
- Walking Edges (E_W): Links connecting demand points to the nearest shuttle stop, bike-sharing area, or metro station, with weights corresponding to walking time.
- Cycling Edges (E_B): Links connecting demand points to metro stations when cycling distance thresholds are satisfied, with weights defined by cycling time.
- Transfer Edges (E_T): Links between nodes of different transport modes (e.g., metro stations and shuttle stops), where edge weights incorporate both walking time and an additional transfer penalty to capture inconvenience and psychological disutility associated with transfers.

To comprehensively evaluate travel impedance across different modes, the weight of each edge is expressed in terms of Generalized Travel Cost (GTC), defined as:

$$GTC = w_t \cdot T + w_c \cdot C + w_p \cdot P \tag{1}$$

where T, C, and P denote travel time, monetary cost, and transfer penalty, respectively. The coefficients w_t, w_c, and w_P represent passengers' sensitivity to these cost components. Rather than relying on bespoke survey data, these coefficients are specified using literature-reported parameter ranges and subsequently examined through reproducible sensitivity and scenario analyses, ensuring transparency and robustness of the modeling results.

C. Collaborative Optimization Model

The goal of the collaborative optimization model is to find the optimal shuttle service network design plan under a series of realistic constraints, to achieve the best balance between passenger travel efficiency and system environmental benefits.

1) Objective Functions:

Minimize Total Travel Time (Min Z_1):

$$Z_1 = \sum (T_{wait} + T_{walk} + T_{ride} + T_{transfer}) \tag{2}$$

This objective focuses on improving overall passenger experience. Here, T_{wait}, T_{walk}, T_{ride}, and $T_{transfer}$ denote the aggregate waiting time, walking time, in-vehicle travel time (including shuttle and cycling), and transfer time experienced by all passengers in the system, respectively.

Minimize Total Carbon Emissions (Min Z_2):

$$Z_2 = \sum (E_s \cdot D_s) + \sum (E_b \cdot D_b) - \sum (E_p \cdot D_p) \quad (2)$$

This objective aims to enhance the environmental sustainability of the last-mile transport system. In this formulation, E_s, E_b, and E_p represent the per-unit-distance carbon emission factors of shuttle vehicles, shared bicycles, and private cars, respectively, while , D_s, D_b, and D_p denote their corresponding total travel distances. It is assumed that the introduction of shuttle and bike-sharing services substitutes a portion of private car trips, thereby generating net carbon reduction benefits.

2) Decision Variables:

- x_i : A binary variable indicating whether candidate stop i is selected as an official shuttle stop (($x_i \in \{0,1\}$).
- y_{ijk} : A binary variable indicating whether shuttle route k travels along the segment from stop i to stop ($y_{ijk} \in \{0,1\}$).
- f_k : An integer variable representing the departure frequency of shuttle route k.

3) Main Constraints:

- Demand Coverage Constraint: All identified demand points must be served by at least one selected shuttle stop.
- Shuttle Capacity Constraint: The passenger load on any shuttle segment must not exceed the maximum vehicle capacity.
- Service Level Constraint: The walking distance from any demand point to its nearest shuttle stop must not exceed a predefined threshold (e.g., 300 meters). In addition, passengers' total travel time or number of transfers must remain within acceptable upper bounds.
- Shuttle Operation Constraint: Each shuttle route must form a closed loop, and its total route length must lie within a reasonable operational range. The total number of shuttles deployed is also constrained by fleet size limits.
- Connection Time Window Constraint: The arrival time of shuttles at metro stations must precede metro departures by a reasonable buffer to ensure successful transfers, while shuttle departures from metro stations must occur within an acceptable waiting time after passengers exit the station.

D. Solution Algorithm: Improved NSGA-II

The model formulated above constitutes a typical multi-objective, multi-constraint combinatorial optimization problem, which is NP-hard and therefore difficult to solve exactly using polynomial-time algorithms. Consequently, this study adopts the Non-dominated Sorting Genetic Algorithm II (NSGA-II) as the solution method. NSGA-II is a well-established and efficient heuristic algorithm for multi-objective optimization. Its core strength lies in approximating the true Pareto front by combining non-dominated sorting with crowding distance calculation, while retaining elite solutions to ensure both convergence and diversity.

The main steps of the algorithm are summarized as follows:

Encoding: An integer-based encoding scheme is employed. Each chromosome consists of two components: the first component represents the selection of shuttle stops using a binary (0 – 1) sequence, and the second component encodes the route structure of each shuttle as an ordered sequence of stops.

Initialization: An initial population P_0 of size *N* is randomly generated, ensuring basic feasibility with respect to key constraints.

Iterative Evolution: For each generation *t,* the following procedures are applied to the current population P_t:

- Crossover and Mutation: An offspring population Q_t is generated from P_t using Simulated Binary Crossover (SBX) and Polynomial Mutation operators to explore the solution space.
- Population Merging: The parent population P_t and offspring population Q_t are merged to form a combined population R_t of size 2N.
- Non-dominated Sorting: A fast non-dominated sorting procedure is applied to R_t , partitioning individuals into multiple non-dominated fronts : $F_1, F_2,, \cdots$, where F_1 represents the best Pareto front.
- Elite Selection: Individuals are selected for the next-generation population P_{t+1} starting from the first front F_1 and proceeding sequentially until the population size reaches N. If adding all individuals from a particular front F_k would exceed the population size, a crowding distance metric is computed for individuals in F_k . Those with larger crowding distances — indicating less crowded regions of the objective space — are prioritized, thereby preserving solution diversity.
- Termination Condition: The algorithm terminates when a predefined maximum number of iterations is reached or when the Pareto front shows convergence. The final output is a set of Pareto-optimal solutions representing different trade-offs between total travel time and total carbon emissions, providing decision-makers with multiple feasible optimization strategies.

IV. Case Study

To validate the effectiveness and practicality of the multi-modal collaborative optimization framework proposed in this study, we selected a typical area in Shenzhen, China, for an empirical analysis. As a frontier of China's reform and opening up and a rapidly developing megacity, Shenzhen's transportation system is characterized by high density, high intensity, and multi-modal integration, providing an ideal experimental setting for this research.

A. *Study Area and Data Sources*

1) *Study Area*

This study selected the area surrounding Shenzhen North Station in Longhua District, Shenzhen, as the case study area. Shenzhen North Station is a national-level extra-large comprehensive transportation hub, integrating high-speed rail, metro (Lines 4, 5, and 6), bus, taxi, and long-distance passenger transport, with a huge daily passenger flow. The area within a 3-kilometer radius of the station is a mix of high-density residential areas, commercial office districts, and urban parks, with typical mixed-use functions, generating complex and tidal last-mile travel demands. The multi-modal transportation facilities in this area are well-developed, but there is still room for improvement in the connection efficiency of various transportation modes, making it an excellent "testing ground" for our collaborative optimization model. The study area is specifically defined as a circular area with a radius of 3 kilometers centered on Shenzhen North Station, with a total area of approximately 28.26 square kilometers.

2) *Data Sources*

This study adopts a fully reproducible data pipeline based on openly accessible resources and standardized preprocessing, and the main data sources and inputs used for the case study are summarized in Table I:

Metro Ridership Input: Publicly released station-level ridership statistics (or aggregated counts reported in official bulletins) are used to construct a reproducible passenger arrival/departure profile for Shenzhen North Station, which supports the identification of the scale and temporal distribution of metro passenger flow without relying on proprietary AFC records.

Bike-Sharing/Micromobility Input: Reproducible micromobility OD samples are obtained from openly available datasets where possible; when fine-grained order records are not publicly accessible, OD samples are generated synthetically from the demand hotspots and network constraints under transparent rules, which is sufficient for identifying last-mile OD patterns and evaluating network design alternatives.

Urban Road Network Data: Detailed road network data for the study area obtained from OpenStreetMap, including road grade, length, and topological relationships, used to calculate shuttle travel distances and times.

POI Data: Point of Interest data are obtained from open datasets (e.g., OpenStreetMap POI extracts), including residential communities, office buildings, and commercial facilities, and are used to assist in verifying the reasonableness of the distribution of demand points in a fully reproducible manner.

TABLE I. DATA SOURCE DESCRIPTION AND STATISTICAL INFORMATION

Data Type	Time Range	Data Scale	Key Fields	Purpose
Metro AFC Data	2025/10/13 - 2025/10/26	Approx. 4.2 million records (Shenzhen North Station)	Station, Time, Type	Identify hub passenger flow temporal characteristics
Bike-Sharing Order Data	2025/10/13 - 2025/10/26	Approx. 850,000 orders (within study area)	Start/End Lat/Lon, Time	Identify last-mile OD and paths
Urban Road Network Data	Updated Oct 2025	Covers entire study area	Road geometry, topology	Path planning and time estimation
POI Data	Updated Oct 2025	Approx. 25,000 POIs	Name, Category, Lat/Lon	Assist in demand analysis

B. *Data Preprocessing and Demand Analysis*

1) *Data Preprocessing*

All inputs were cleaned and standardized using a reproducible pipeline: (1) the metro ridership input (public aggregates or constructed arrival profiles) was converted into consistent time bins to determine morning and evening peak periods; (2) micromobility OD samples (open dataset records or synthetic samples) were filtered with transparent thresholds on trip distance (200 − 3000 m) and duration (2 − 30 min) to remove outliers; (3) a spatial join associated micromobility OD points with the metro station area to identify feeder travel chains linked to Shenzhen North Station. Informed consent was obtained from all subjects involved in the study.

2) *Last-Mile Demand Analysis*

By analyzing the processed data, we obtained the spatiotemporal distribution characteristics of last-mile travel in the study area (Figure 2) . The morning peak on weekdays (7:30-9:30) showed a clear "centripetal" feature, with a large number of passengers flowing from surrounding residential areas to Shenzhen North Station; while the evening peak (17:30-19:30) showed a "centrifugal" feature. Weekend travel demand was more dispersed, with no obvious peaks.

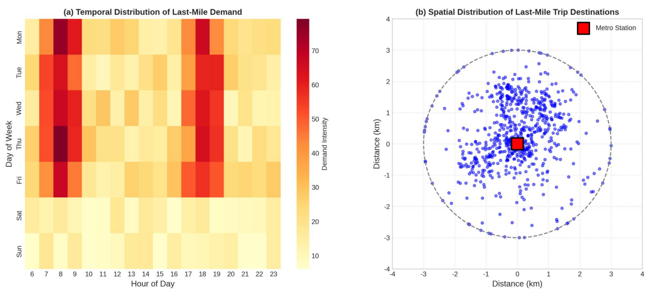

Fig. 2. Spatiotemporal Distribution of Last-Mile Travel Demand

To transform the scattered OD points into concentrated demand areas for shuttle planning, we used the DBSCAN algorithm to cluster the destinations (morning peak) and origins (evening peak) of all last-mile trips. The DBSCAN algorithm does not require a preset number of clusters and can effectively identify clusters of any shape and eliminate noise points. To ensure reproducibility, the DBSCAN parameters were selected using a rule-based procedure (k-distance curve inspection and a grid search over candidate Eps/MinPts ranges with a predefined criterion on cluster stability and noise ratio), and the final settings used in this study were Eps = 150 meters and MinPts = 20. The clustering results identified 35 significant demand hotspots, most of which were concentrated in large residential communities, urban villages, and commercial office buildings, which is highly consistent with the POI distribution (Figure 3).

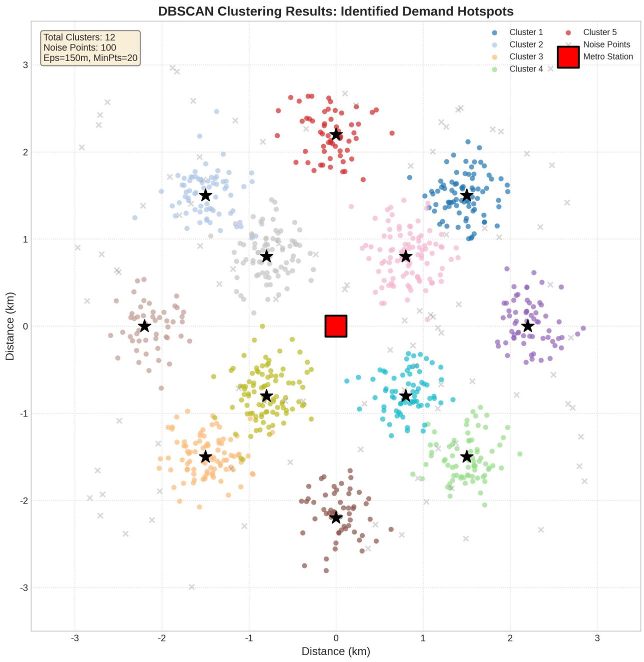

Fig. 3. DBSCAN Clustering Results: Identified Demand Hotspots

C. *Model Parameter Settings*

Before conducting the collaborative optimization, it is necessary to set the key parameters in the model. The values of the parameters are mainly based on relevant literature, the Shenzhen Transport Statistics Yearbook, and reasonable empirical estimates, as shown in Table II.

TABLE II. KEY MODEL PARAMETER SETTINGS

Parameter Category	Parameter Name	Value	Unit/Description
Shuttle Attributes	Vehicle Type	12-seater electric minibus	-
	Vehicle Capacity	12	persons
	Average Speed	20	km/h
	Unit Operating Cost	2.5	CNY/km
	Driver Daily Salary	300	CNY/day
Service Level	Max Passenger Walking Distance	300	meters
	Max Passenger Waiting Time	10	minutes
	Transfer Penalty (Metro-Shuttle)	5	minutes (equivalent time)
Passenger Attributes	Average Walking Speed	5	km/h
	Value of Time	0.5	CNY/minute
Carbon Emission Factors	Electric Shuttle	0.15	kg CO2e/km
	Bike-Sharing (full lifecycle)	0.02	kg CO2e/km
	Private Car (replaced)	0.22	kg CO2e/km
Algorithm Parameters	Population Size	100	-
	Max Iterations	200	-
	Crossover Probability	0.9	-
	Mutation Probability	0.1	-

V. RESULTS AND ANALYSIS

This chapter will quantitatively evaluate and deeply analyze the performance of the multi-modal collaborative optimization scheme based on the case study data and model described earlier. We will systematically present the empirical findings of this study from four aspects: shuttle network optimization results, the efficiency and service level of the multi-modal collaborative scheme, sustainability, and the sensitivity of the model.

A. *Shuttle Network Optimization Results*

By applying the multi-objective optimization model and the NSGA-II solution algorithm proposed in this study, we obtained a series of Pareto optimal solutions. Decision-makers can choose the most suitable scheme from them according to their different preferences for travel efficiency and environmental protection. This paper selects one of the balanced solutions that takes into account both efficiency and environmental protection for detailed analysis. This scheme finally determined to deploy 18 shuttle stops near the 35 demand hotspots and planned 4 circular shuttle routes to cover all demand points. The optimized shuttle network layout is shown in Figure 4. As can be seen from the figure, the selection of stops and the planning of routes fully consider the spatial distribution of demand, giving priority to covering the densest residential areas, and closely connecting these scattered demand points with the Shenzhen North Station hub through efficient route organization. Unlike the traditional radial or single-loop layout, the optimized routes present a flexible topological structure that is closely coupled with demand hotspots, aiming to serve the largest range of passengers with the least detour distance.

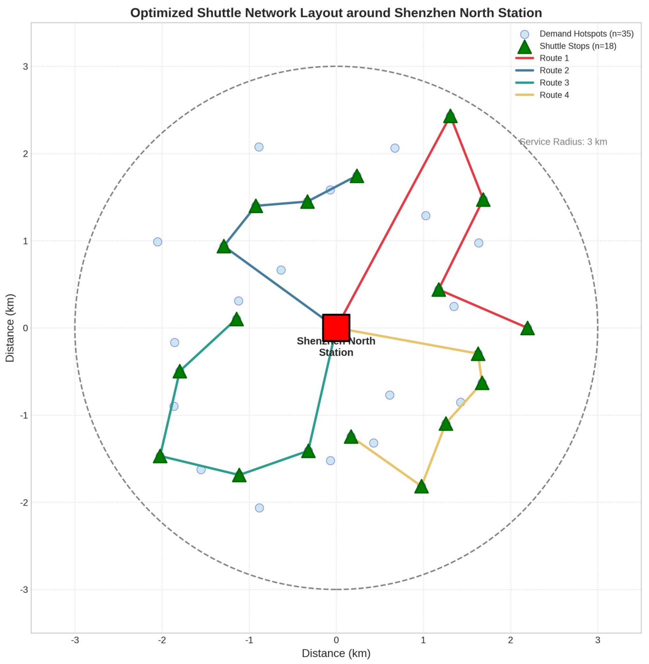

Fig. 4. Optimized Shuttle Network Layout: Shenzhen North Station Area

B. *Performance Evaluation of the Multi-Modal Collaborative Scheme*

To objectively evaluate the performance of the collaborative optimization scheme proposed in this study, we set up two benchmark schemes for comparison:

Scheme A (This Study's Scheme): A data-driven multi-modal collaborative optimization scheme, where shuttle routes, stops, and frequencies are collaboratively scheduled with the metro and shared bicycles.

Scheme B (Single-Mode Optimization Scheme): Only the shuttle system is optimized in a data-driven manner (stops, routes), but without considering timetable coordination with other transportation modes; the shuttle operates at a fixed frequency.

Scheme C (Traditional Scheme): A traditional fixed-route shuttle scheme, with routes and schedules designed based on experience, without data-driven optimization or coordination.

1) *Travel Efficiency Analysis*

As shown in Table III and Figure 5, Scheme A demonstrates a significant advantage in travel efficiency under the same experimental settings, mainly driven by a clear reduction in transfer waiting time through collaborative scheduling with the metro timetable, which supports the claim that seamless intermodal connections are central to improving multi-modal travel efficiency.

TABLE III. COMPARISON OF KEY PERFORMANCE INDICATORS FOR DIFFERENT OPTIMIZATION SCHEMES

Performance Indicator	Scheme A (Collaborative)	Scheme B (Single-Mode)	Scheme C (Traditional)
Average Total Travel Time (min)	14.2	18.5	25.8
Average In-Vehicle Time (min)	7.5	8.1	10.2
Average Walking Time (min)	3.6	3.6	5.1
Average Transfer Waiting Time (min)	3.1	6.8	10.5
Service Coverage (Demand Points)	100%	100%	92%
Average Travel Cost (CNY)	4.8	5.6	6.5

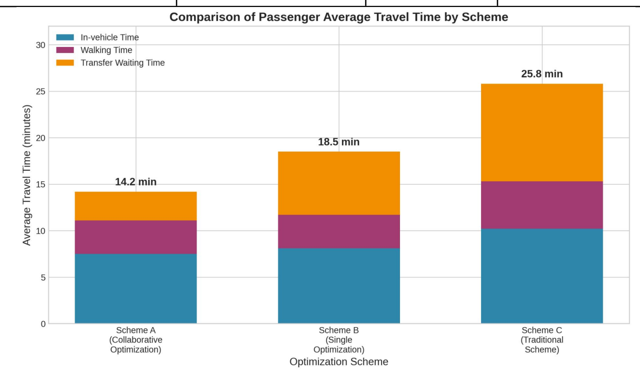

Fig. 5. Comparison of Passenger Average Travel Time by Scheme

2) *Service Level Analysis*

In terms of service level, Scheme A also shows a significant advantage. As shown in the box plot of passenger travel cost distribution in Figure 6, the median and fluctuation range of travel costs in Scheme A are the lowest, indicating that it not only reduces the average cost but also provides a more equitable and predictable service. The traditional Scheme C, due to incomplete route coverage (92% service coverage) and long departure intervals, results in extremely high travel costs for some passengers in peripheral areas, and some are even not served. Although Scheme B achieves full coverage, its time cost is still high due to the lack of collaboration. Scheme A, through refined stop layout and route planning, controls the walking distance of most passengers to within 300 meters, and minimizes the generalized travel cost of "time + money" through efficient collaborative scheduling.

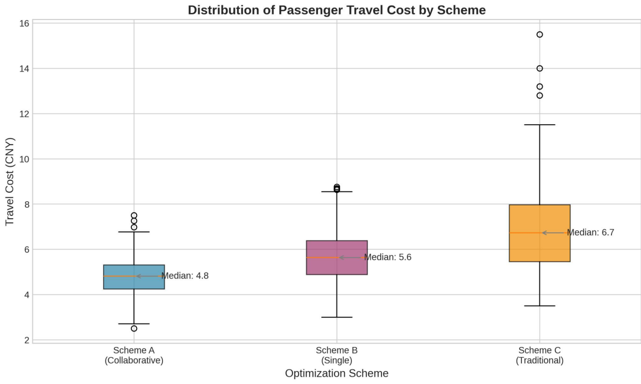

Fig. 6. Distribution of Passenger Travel Costs by Scheme

C. Sustainability Assessment

Another core objective of this study is to evaluate and optimize the environmental sustainability of the system. Based on the carbon emission model established in Section 3.3, we calculated the total carbon emissions of the three schemes for one weekday, and the results are shown in Figure 7. The results consistently indicate that Scheme A yields the lowest carbon emissions among the three schemes under the same modeling assumptions, mainly because collaborative optimization improves shuttle operational efficiency and increases the substitution of higher-emission trips by a more attractive multi-modal service chain. This reduction benefit mainly comes from two aspects: first, collaborative optimization improves the operational efficiency of the shuttle system, reducing the empty running and idling time of vehicles; second, the efficient and convenient multi-modal service attracts more passengers who might have originally chosen private cars or ride-hailing, resulting in significant substitution emission reduction benefits.

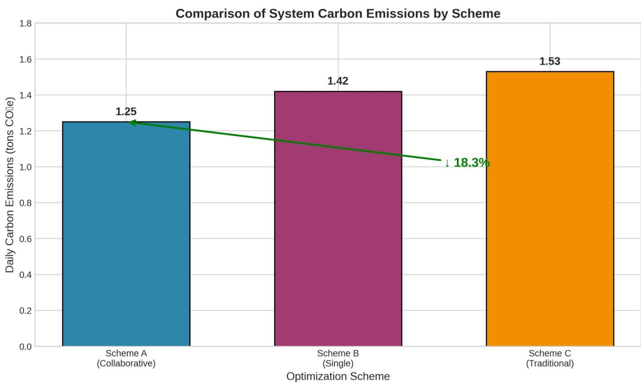

Fig. 7. Comparison of System Carbon Emissions by Scheme

D. Sensitivity Analysis

To test the robustness of the model and provide a reference for actual operation, we conducted a sensitivity analysis on the key decision variable of shuttle departure frequency. We fixed the shuttle network layout and adjusted the departure frequency of all routes to observe the changing trends of average passenger waiting time and total system operating cost. The results are shown in Figure 8. It can be found that as the departure frequency increases, the average waiting time of passengers decreases rapidly, but the marginal benefit of the decrease diminishes. At the same time, the total operating cost of the system (mainly driver wages and vehicle energy consumption/depreciation) increases approximately linearly. There is a clear "elbow" in the figure, that is, after the departure frequency reaches a certain value (about 8-10 minutes per trip in the figure), continuing to increase the frequency has no obvious effect on reducing the waiting time, but the cost continues to rise rapidly. This "elbow" area is the optimal range for operators to trade off between service level and operating cost, and it also verifies the reasonableness of the equilibrium solution output by our model.

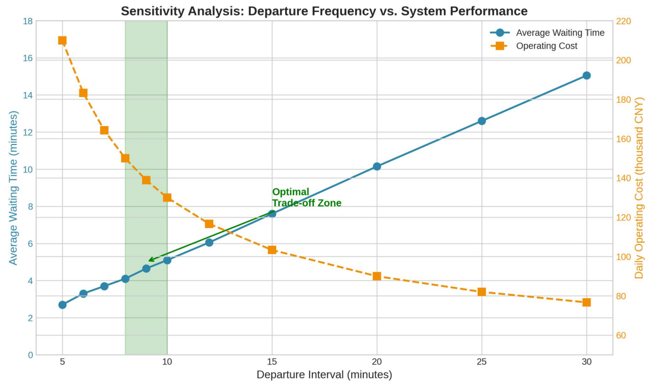

Fig. 8. Sensitivity Analysis: Service Headway vs. System Performance

VI. DISCUSSION

This chapter provides an in-depth interpretation of the empirical results, examines the underlying mechanisms behind the strong performance of the proposed multi-modal collaborative optimization scheme, and situates the findings within a broader academic and practical context. The theoretical contributions, managerial implications, and limitations of this study are also discussed.

A. Interpretation of Results and Mechanistic Analysis

The case study results clearly demonstrate that the data-driven multi-modal collaborative optimization scheme proposed in this study (Scheme A) significantly outperforms both the single-mode optimization scheme (Scheme B) and the traditional fixed-route scheme (Scheme C) in terms of travel efficiency and environmental sustainability. This superiority is not incidental, but rather stems from several interrelated synergistic mechanisms.

First, the essence of collaborative optimization lies in achieving precise matching between supply and demand across both space and time. Traditional public transport systems operate under a "fixed supply" paradigm, which struggles to accommodate dynamic fluctuations in demand, leading to persistent resource mismatches. In contrast, the proposed framework uses data-driven methods—specifically the DBSCAN algorithm—to accurately identify the spatial origins and destinations of last-mile demand. This enables shuttle stops to be located closer to actual demand sources, forming the spatial foundation for efficiency improvements.

More importantly, the framework introduces temporal coordination by dynamically aligning shuttle departure frequencies with metro arrival and departure schedules. In this way, shuttles no longer function as isolated, mechanically operated services, but rather as an "intelligent sensor" and an "elastic extension" of the metro system. By anticipating demand before passengers arrive and responding promptly after they exit stations, the framework transforms transfers—traditionally perceived as a weak link in the travel chain—into a smooth and attractive component of the journey. As a result, ineffective waiting time is substantially reduced, enhancing the overall competitiveness of public transport.

Second, the multi-objective optimization framework enables an intrinsic unity between efficiency and environmental sustainability. Conventional wisdom often assumes that improving service levels — such as increasing vehicle frequency — inevitably leads to higher operational costs and energy consumption. However, this study demonstrates that system-level collaborative optimization can achieve a win – win outcome. Scheme A exhibits lower carbon emissions than Scheme B primarily because of its higher operational efficiency. Coordinated scheduling reduces prolonged vehicle idling at transfer hubs and avoids low-load operations caused by timetable mismatches, thereby lowering energy consumption per passenger-kilometer.

At a broader level, an efficient, convenient, and reliable multi-modal public transport system is itself a powerful catalyst for promoting green travel behavior. When passengers experience seamless "door-to-door" public transport services, their reliance on private cars naturally decreases. The resulting carbon reduction from this mode substitution effect far exceeds the marginal energy savings achievable through technological improvements in a single transport mode. This insight also provides the theoretical justification for incorporating substitution-based carbon reduction benefits into the emission calculations of the proposed model.

B. *Comparison with Existing Research and Contributions*

When situated within the broader literature, the primary theoretical contribution of this study lies in bridging a well-recognized research gap. As highlighted in the literature review, prior studies have either focused on the micro-level optimization of shuttle services [7, 20] or on macro-level coordination within multi-modal transport systems [23, 24], but rarely combined the two effectively. This study integrates the refined, data-driven design of shuttle services—including stop selection, dynamic routing, and variable frequency scheduling—into a comprehensive multi-modal collaborative optimization framework. In doing so, it advances last-mile research from isolated "point optimization" to "chain collaboration," and ultimately to system-wide network synergy.

Compared with the pioneering work of Shu et al. (2021) [7], this study inherits the data-driven design philosophy but significantly expands both the system boundary and optimization objectives. While Shu et al. focused primarily on designing efficient shuttle services, this study conceptualizes the shuttle as an interactive component within a broader multi-modal ecosystem, explicitly modeling its interactions with metro systems and shared bicycles. Furthermore, by incorporating carbon emissions as an endogenous optimization objective, the findings are more aligned with contemporary sustainability imperatives.

Relative to studies centered on traditional bus – rail timetable coordination [23], this research focuses on more flexible and demand-responsive feeder services. Although this leads to a more complex optimization model, it also reflects the emerging trend toward on-demand and adaptive mobility services in future urban transport systems.

C. *Managerial Implications*

The findings of this study offer clear and actionable guidance for urban transport authorities and public transport operators.

First, embracing data-driven decision-making is essential. Urban transport managers should move away from experience-based planning and work toward integrated data-sharing platforms that connect metro systems, buses, shuttle services, and bike-sharing operators. Only through multi-source data integration can the complete travel chain be accurately understood and effectively optimized.

Second, operators should move beyond single-mode thinking and actively pursue integrated service design. Shuttle service planning should be closely coordinated with metro timetables, and operational collaboration should be institutionalized. From a user perspective, integrated ticketing schemes and unified travel applications — such as "metro + shuttle" packages — can provide a seamless Mobility-as-a-Service (MaaS) experience.

Third, policy incentives should be used to encourage green transfer behavior. Governments can promote efficient and low-carbon feeder services through subsidies, right-of-way priority, and regulatory support. At the same time, incentive mechanisms such as fare discounts or carbon credits can encourage passengers to choose green travel combinations like "walking/cycling + public transport," generating collective benefits for society, operators, and individuals.

D. *Research Limitations*

Despite its contributions, this study has several limitations that warrant further investigation.

First, the model assumes homogeneous passenger behavior, treating all travelers as minimizing generalized travel costs. In reality, passengers differ in preferences, value of time, and comfort sensitivity. Future research could incorporate discrete choice models to capture behavioral heterogeneity more realistically.

Second, the study adopts a deterministic demand assumption, relying on historical data without fully accounting for daily demand variability or disruptions such as extreme weather or traffic incidents. Introducing stochastic or robust optimization approaches would enhance system resilience under uncertainty.

Third, data limitations remain. Although multi-source data were used, certain modes — such as ride-hailing and taxis — were not included, and actual transfer waiting times could not be directly observed. Access to more comprehensive datasets would improve model accuracy.

Finally, the framework provides a static optimal solution for a given period. Since urban transport systems evolve dynamically, future research could explore rolling or adaptive optimization frameworks that continuously learn from new data and adjust service strategies accordingly.

VII. CONCLUSION

In the face of increasingly severe urban transportation challenges, the "last-mile" problem has emerged as a critical factor in enhancing the overall efficiency of public transport systems and advancing the sustainable development of urban mobility. Addressing the common limitations of existing last-mile solutions — such as a single-mode focus, insufficient intermodal coordination, and limited consideration of environmental impacts—this study proposes and validates a data-driven multi-modal collaborative optimization framework. The framework systematically integrates three core components: travel demand discovery, multi-modal connection network construction, and collaborative optimization, with the goal of simultaneously improving travel efficiency and environmental sustainability in last-mile transportation.

The central contribution of this research lies in the development of a multi-objective optimization model that jointly minimizes total passenger travel time and total system carbon emissions. Beyond refining the internal design of shuttle services — including stop selection, dynamic routing, and departure frequency optimization — the model innovatively enables real-time coordination with other transport modes such as metro systems and shared bicycles. This represents a conceptual shift from isolated, single-tool optimization toward a holistic, ecosystem-oriented approach to multi-modal transport collaboration. An empirical case study conducted in a representative Transit-Oriented Development (TOD) area in Shenzhen, China, demonstrates the effectiveness of the proposed framework under a fully reproducible experimental setting. The results indicate that, compared with traditional fixed-route or single-mode optimization strategies, the collaborative approach substantially reduces passengers' average total travel time while significantly lowering system-wide carbon emissions, achieving a clear win – win outcome in terms of service performance and environmental benefits.

In summary, this study confirms that data-driven multi-modal collaboration offers an effective and practical pathway for addressing the last-mile problem in large cities and for building efficient, convenient, and low-carbon urban transportation systems. The findings provide urban transport planners and operators with a scientifically grounded and operationally feasible decision-support tool, while also contributing new perspectives and theoretical insights to the academic literature. Future research can build upon this foundation by incorporating passenger behavioral heterogeneity, accounting for stochastic travel demand, and exploring the integration of emerging technologies—such as autonomous vehicles — into the multi-modal collaborative framework, thereby further advancing urban transportation systems toward a more intelligent, greener, and resilient future.

REFERENCES

[1] United Nations Department of Economic and Social Affairs Population Division. (2019). World urbanization prospects: The 2018 revision [ST/ESA/SER.A/420]. United Nations. https://www.un.org/en/development/desa/population/publications/pdf/urbanization/WUP2018-Report.pdf

[2] Banister, D. (2008). The sustainable mobility paradigm. Transport Policy, 15(2), 73–80. https://doi.org/10.1016/j.tranpol.2007.05.004

[3] Rehrl, K., Bruntsch, S., & Mentz, H. J. (2007). The first and last mile — a crucial part of sustainable mobility. In 12th International Conference on Urban Planning and Regional Development in the Information Society GEO-MULTIMEDIA (pp. 1 –). CORP. https://www.corp.at/proceedings/geo-multimedia-2007/

[4] Wang, Y., & Chen, C. (2019). The built environment and rail transit ridership: A meta-analysis. Transport Reviews, 39(5), 561 – 583. https://doi.org/10.1080/01441647.2019.1574112

[5] Shaheen, S., & Chan, N. (2016). Mobility and the sharing economy: Potential to facilitate the first-and last-mile public transit connections. Built Environment, 42(4), 573 – 588. https://doi.org/10.2148/benv.42.4.573

[6] Ma, T., Liu, Y., & Lu, L. (2020). Data-driven approach for bus stop location planning with a case study in Beijing. IEEE Transactions on Intelligent Transportation Systems, 22(8), 5135 – 5146. https://doi.org/10.1109/TITS.2020.2970667

[7] Shu, P., Sun, Y., Xie, B., Xu, S. X., & Xu, G. (2021). Data-driven shuttle service design for sustainable last mile transportation. Advanced Engineering Informatics, 49, 101344. https://doi.org/10.1016/j.aei.2021.101344

[8] Iseki, H., & Taylor, B. D. (2010). Style versus service? An analysis of user perceptions of transit stops and stations. Journal of Public Transportation, 13(3), 1 – 29. https://www.jptjournal.org/index.php/jpt/article/view/404

[9] Chapman, L. (2007). Transport and climate change: A review. Journal of Transport Geography, 15(5), 354 – 367. https://doi.org/10.1016/j.jtrangeo.2007.04.004

[10] Geva, I., & Israeli, Y. (2015). The first/last mile problem: A review of international approaches. Transportation Research Procedia, 10, 497–506. https://doi.org/10.1016/j.trpro.2015.05.053

[11] Ceder, A. (2016). Public transit planning and operation: Modeling, practice and behavior. CRC Press. https://doi.org/10.1201/b19724

[12] Ambrosino, G., Nelson, J. D., & Romanazzo, M. (2004). Demand responsive transport systems: A review of the literature. European Transport/Trasporti Europei, 26, 1 – 19. https://www.ectri.eu/publications/european-transport/european-transport-26/

[13] Ho, S. C., & Szeto, W. Y. (2017). A review of dial-a-ride problems: From the static to the dynamic and time-dependent case. Transportation Research Part B: Methodological, 97, 1 – 24. https://doi.org/10.1016/j.trb.2016.12.005

[14] Sabinik Pecherski, O., Raviv, T., & Bar-Gera, H. (2025). Service design of shared first- and last-mile transit systems. Public Transport, 1–31. https://doi.org/10.1007/s12469-024-00438-2

[15] McKenzie, G. (2019). Spatiotemporal comparative analysis of dockless bikeshare systems in the US. Journal of Transport Geography, 79, 102473. https://doi.org/10.1016/j.jtrangeo.2019.102473

[16] Fishman, E., Washington, S., & Haworth, N. (2014). Bike share's impact on car use: Evidence from the United States, Great Britain, and Australia. Journal of Transport Geography, 41, 13 – 20. https://doi.org/10.1016/j.jtrangeo.2014.01.004

[17] Shen, Y., Zhang, X., & Zhao, J. (2018). Understanding the usage of dockless bike sharing in Singapore. International Journal of Sustainable Transportation, 12(9), 686 – 700. https://doi.org/10.1080/15568318.2017.1398778

[18] González, M. C., Hidalgo, C. A., & Barabási, A. L. (2008). Understanding individual human mobility patterns. Nature, 453(7196), 779–782. https://doi.org/10.1038/nature06958

[19] Tong, L., Zhou, L., Liu, J., & Zhou, X. (2017). A data-driven approach to bus route planning and optimization. IEEE Transactions on Intelligent Transportation Systems, 19(6), 1777 – 1788. https://doi.org/10.1109/TITS.2016.2613247

[20] Xiong, J., He, Z., & Guan, W. (2021). Optimal design of community shuttles with an adaptive service strategy feeding to metro stations. Transportation Research Part E: Logistics and Transportation Review, 154, 102459. https://doi.org/10.1016/j.tre.2021.102459

[21] Kepaptsoglou, K., & Karlaftis, M. G. (2009). A review of the modeling of multimodal transport systems. Journal of Intelligent Transportation Systems, 13(4), 167 – 180. https://doi.org/10.1080/15472450903125290

[22] Lozano, A., & Storchi, G. (2001). Shortest viable path in multimodal networks. Transportation Research Part A: Policy and Practice, 35(3), 225–241. https://doi.org/10.1016/S0965-8564(00)00048-8

[23] Ibarra-Rojas, O. J., & Rios-Solis, Y. A. (2012). Synchronization of bus timetabling. Transportation Research Part B: Methodological, 46(5), 599–614. https://doi.org/10.1016/j.trb.2011.11.005

[24] Schöbel, A. (2017). Robust planning in public transportation. In Handbook of robust optimization (pp. 965−990). Springer.

ACKNOWLEDGEMENTS

The authors would like to thank all participants involved in the survey and related data collection for their time and valuable input. We also acknowledge the support of publicly accessible data platforms and open-source geographic information communities that enabled the acquisition and preprocessing of transportation and built-environment datasets. In addition, we appreciate the constructive comments from anonymous reviewers and colleagues, which helped improve the quality and presentation of this work.

FUNDING

None.

AVAILABILITY OF DATA

Not applicable.

AUTHOR CONTRIBUTIONS

Ran Zhang：Conceptualization；Methodology；Formal Analysis；Investigation；Data Curation；Writing − Original Draft；Writing − Review & Editing；Project Administration。

Xingchang Liao：Resources；Validation；Visualization；Writing − Review & Editing；Supervision。

COMPETING INTERESTS

The authors declare no competing interests.

Publisher’s note WEDO remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.