Book-ahead ride-hailing trip and its determinants: findings from large-scale trip records in China

Wu Li [a, b], Shengchuan Zhao [a], Jingwen Ma [a], Otto Anker Nielsen [b], Yu Jiang [b, c]

[a] School of Transportation and Logistics, Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian 116024, China.

[b] DTU Transport, Department of Technology, Management, and Economics, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark.

[c] Lancaster University Management School (LUMS), Lancaster University, Lancaster, LA1 4YX, United Kingdom.

ABSTRACT: The cruising of ride-hailing vehicles generates negative externalities such as traffic congestion and vehicular emissions. These externalities can be mitigated by reducing cruising driving via operating book-ahead ride-hailing services, where the platform dispatches and routes drivers based on precise information on travelers' departure time and origin-destination (OD). However, the effects of factors influencing book-ahead ride-hailing trips have rarely been empirically examined with real data. Using six-month trip data from China, this study employs a gradient boosting decision tree (GBDT) method with hyperparameters optimized by the Bayesian optimization algorithm to examine the factors associated with book-ahead ride-hailing trips across OD pairs (hexagon cells-to-hexagon cells) at various spatial scales. The relative importance rankings generated from this study indicate that trip features, weather conditions, and accessibility to transportation hubs are significant determinants correlated with the usages of book-ahead ride-hailing. The partial dependence plots demonstrate the nonlinear threshold effects of these determinants on the hourly number of book-ahead ride-hailing per OD pair. Moreover, this study compares the differences in associations between peak and non-peak hours as well as weekdays and weekends. The disparity in the nonlinear threshold effects between weekdays and weekends is only observable during the evening peak period and not at other times. These findings provide valuable insights into developing practical strategies for promoting book-ahead ride-hailing services.

Keywords: Reserved transportation; Book-ahead trip; On-demand ride-hailing; Influence factors; Built environment; GBDT

Preprint

1. Introduction

The past decade has witnessed the rapid growth of the sharing economy and the boom in new mobility services such as ride-hailing and taxi-calling (Ravula, 2022; Rayle et al., 2016). Service providers like Uber, Lyft, and Didi Chuxing have capitalized on this trend, allowing passengers to use their smartphones to hail a car and a driver to meet their travel needs through their ride-hailing platforms (He, 2021; Wang and Yang, 2019). According to the 47th Statistical Report on Internet Development in China (China Internet Network Information Center, 2021), the number of ride-hailing app users in China reached 365 million by December 2020, which accounts for 36.9% of the total Internet users and represents an increase of 2.98 million since March 2020.

The rapid expansion of on-demand ride-hailing services in urban areas has raised concerns among authorities about their negative impacts on urban mobility. One of the most heated debates is whether ride-hailing services can cause external traffic congestion (Diao et al., 2021; Tirachini, 2020; Tirachini et al., 2020; Tirachini and Gomez-Lobo, 2020; Wei et al., 2022). According to Diao et al. (2021), ride-hailing services have increased traffic congestion by 0.9% in terms of the travel time index and by 4.5% in terms of congestion duration. The primary cause is the cruising of idle ride-hailing drivers on the road network in search of potential patronage, which occupies limited road space without contributing to serving travel demand (Gao et al., 2022; Wei et al., 2023). This puts authorities and transportation network companies (TNCs) in a difficult position. On the one hand, passengers expect short waiting times from their requested departure time. On the other hand, if ride-hailing platforms increase the number of drivers to cover more trip requests, a large number of cruising drivers searching for customers will inevitably lead to idle driving time and traffic congestion. Schaller (2018) pointed out that the number of ride-hailing vehicles in New York City increased by 59% from 2013 to 2017, resulting in an 81% increase in idle driving, a 15% decline in traffic speed, and a 36% increase in vehicle kilometers traveled.

To address concerns about the negative impacts of cruising ride-hailing vehicles on

Preprint

urban mobility, authorities, and transportation network companies have implemented various regulatory strategies, including capping ride-hailing fleet sizes, congestion surcharges, ride-splitting services, shared parking policies, and “schedule a ride” services that allow passengers to book trips in advance. Of these, “schedule a ride” is a promising approach to reduce cruising traffic in two ways. First, it provides ride-hailing platforms with precise information about the origin-destination (OD) and departure time of future trips, enabling the platform to dispatch and route drivers more efficiently and reduce idle driving. Second, it allows the platform to adjust vehicle fleet size and pre-assign ride-hailing vehicles to service areas based on submitted trip requests in advance. Compared to real-time ride-hailing services, which require immediate booking and trip departure, book-ahead ride-hailing offers a more sustainable urban mobility option.

Despite the prevalence of reservation scenarios in daily life, such as seat reservations for railway or airline travel and online bookings for medical services, only limited pilot programs concerning book-ahead ride-hailing are conducted in the real world. Currently, there is still a lack of overall understanding of ride-hailing trip reservations on a large scale with real data. While Yahia et al. (2021) examined the impacts and benefits of book-ahead rides on driver supply management for ride-sourcing platforms, there have been few empirical studies on book-ahead ride-hailing trips and their determinants. This knowledge gap is relevant to regulators and planners seeking to promote or limit the use of book-ahead ride-hailing services, as the benefits of such services will remain theoretical without a deeper understanding of their determinants. In response to this situation, our study aims to explore the relationship between book-ahead ride-hailing trips and their determinants, including trip characteristics, built environment attributes at origin and destination, and weather conditions. The contributions of our research are outlined below.

First, this study investigates a novel book-ahead ride-hailing service using large-scale trip records in China. Unlike real-time ride-hailing services, book-ahead ride-hailing allows passengers to submit their trip requests several hours or days in advance, allowing the platform to adjust the fleet size and pre-assign vehicles to service areas. Second, this study uses an interpretable machine learning approach to explore the relationships between book-ahead ride-hailing trips and their determinants, i.e., trip characteristics, built environment attributes, socio-economic characteristics at origin and destination, and weather conditions. To ensure model accuracy, a Bayesian optimization algorithm is employed to search for optimal hyperparameters, and the model is executed at three spatial scales. Third, this study uses relative importance analysis and partial dependence plots to identify the factors that contribute most to the usage of book-ahead ride-hailing and to visualize nonlinear associations. We conduct a heterogeneity analysis for different periods, including morning peak, evening peak, and off-peak hours, to compare the nonlinear effects of contributing factors. Finally, our research contributes to a comprehensive understanding of the association between book-ahead ride-hailing usage and its determinants. It identifies where and when it is most feasible to promote such services and shows how urban planning and traffic demand management could contribute to place-based strategies to support or intervene in this promising service.

The remainder of our paper is organized as follows. Section 2 reviews related studies. Section 3 presents our data profile and methodology. Section 4 discusses the relative importance of variables and partial dependence plots. Finally, Section 5 summarizes the key findings, offers corresponding policy implications, and sheds light on future research directions.

Preprint

2. Literature review

2.1. Reservation in transportation management

Reservation strategies have been identified as a complementary demand management approach within the context of traffic management (Lamotte et al., 2017). The most well-known application of reservation systems in traffic management is roadway reservation systems, which allow vehicles to reserve a spot on the freeway in advance, enabling them to use specific segments of the freeway within a predetermined period (Chen et al., 2022; Liu et al., 2015; Su and Park, 2015). Parking reservation schemes have also gained traction and have been proposed to address inadequate curbside parking spaces (Chen et al., 2015; Liu et al., 2014; Wang et al., 2022). In the context of trip reservations, De Feijter et al. (2004) utilized a simulation experiment to demonstrate the positive impact of trip booking on effective road capacity utilization and travel-time reliability. Ma et al. (2017) designed an autonomous vehicle-sharing and reservation system utilizing a linear programming approach. The proposed system enables travelers to book their trips in advance, and the system operator optimally arranges the autonomous vehicle pick-up and delivery before the requested time. Notably, Ouyang et al. (2021) proposed a modeling framework for many-to-many carpooling services with in-advance reservations in idealized settings. They further formulated an analytical model with a closed form to examine the effects of detours and waiting time restrictions on reservation-based carpooling services. Yahia et al. (2021) conducted a simulation experiment to analyze the impact of book-ahead rides on driver supply management for ride-sourcing platforms and found that an increase in book-ahead rides led to a reduction in the total number of drivers required. It's a sign that ride-hailing trip reservations are still lacking in critical aspects despite their attractiveness and promise.

2.2. Contributing factors associated with ride-hailing demand

Several studies have investigated the relationship between ride-hailing usage and its determinants. Table 1 outlines the differences between this study and the existing literature. It shows that four main types of variables were found to be associated with ride-hailing adoption: trip feature factors, socio-demographic factors, built environment attributes, and weather conditions. The roles of trip feature factors, including median or average travel duration, distance, and fare, have been extensively studied in ride-hailing usage analysis (Tu et al., 2021; Xu et al., 2021). Most of the existing literature on socio-demographic factors paid particular attention to aggregated population, employment, and income characteristics, such as population density, employment density, and medium median income in the census tract (Ghaffar et al., 2020; Marquet, 2020). The collection of these valuable data profiles is usually tied to the national or regional demographic census. One of the most well-known

Preprint

tools for assessing the built environment is Ewing and Cervero's (2010) 5Ds model containing density (such as population density and job density), diversity (such as land use mixture and jobs-housing balance), design (such as intersection density and street density), destination accessibility (such as job accessibility by transit or auto), and distance to transit (such as the nearest bus stop or subway station). The measurements of the built environment in current studies have stretched and expanded in specific research contexts (Bi et al., 2020; Ghaffar et al., 2020; Huang et al., 2021; Jin et al., 2022). In addition, several distinctive weather condition variables, including temperature, humidity, wind speed, and rainfall, have also been integrated into the modeling process (Liu et al., 2020; Shokoohyar et al., 2020). However, the most crucial role among the four main types of variables is not consistent. For instance, Tu et al. (2021) suggested that the collective influence of the built environment factors on the ride-splitting adoption rate is greater than that of demographic characteristics, whereas the findings from Xu et al. (2021) indicated that the socio-economic and demographic variables are the most important factors in predicting the ride-splitting adoption rate.

2.3. Related works on ride-hailing demand modeling

Since the survey conducted in San Francisco highlighted the significant benefits of ride-hailing services in reducing waiting times and providing fast point-to-point trips compared to conventional taxi services (Rayle et al., 2016), an increasing body of literature has explored the role of this emerging travel mode in urban mobility (Liu et al., 2022; Wang and Yang, 2019). Previous studies of ride-hailing demand modeling have paid particular attention to six main aspects: (1) identifying the spatial and temporal distribution characteristics of ride-hailing trips, such as ridership, duration, origin, and destination (He, 2021; Wang and Noland, 2021); (2) using various deep learning approaches, such as convolutional long short-term memory (Chen et al., 2022; Liu et al., 2023), and graph convolutional neural networks (Jin et al., 2020), to predict spatio-temporal online ride-hailing demand; (3) examining the factors that influence passengers' decisions to switch from traditional travel modes to ride-hailing, based on the discrete choice model (Azimi et al., 2020; Tarabay and Abou-Zeid, 2020); (4) inferring the relationship between customer satisfaction and loyalty to ride-hailing services using structural equation modeling (Nguyen-Phuoc et al., 2020); (5) optimizing matching and order dispatching algorithms between passengers and drivers to improve the efficiency and performance of ride-hailing services under network equilibrium (Bertsimas et al., 2019; Wei et al., 2022; Xu et al., 2021); and (6) investigating the impacts of contributed factors, such as built environment and socio-economic factors, on online ride-hailing trip demand to provide guidance for policy-makers and urban planners (Bi et al., 2020; Shokoohyar et al., 2020; Yu and Peng, 2019).

Concerning the sixth aspect of ride-hailing demand modeling, previous research has employed two modeling approaches to investigate the relationship between ride-hailing services and contributing factors: regression-based methods and interpretable machine learning methods. Common regression-based models used in prior studies include ordinary least squares (OLS) regression models (Brown, 2020; Ghaffar et al., 2020; Liu et al., 2021; Marquet, 2020; Shokoohyar et al., 2020), spatial regression models that incorporate spatial dependence (Dean and Kockelman, 2021; Huang et al., 2021; Lavieri et al., 2018), and geographically weighted regression models that integrate spatial heterogeneity (Bi et al., 2020; Yu and Peng, 2019). These regression-based models typically predefine a specific model form (such as linearity and logarithmic linearity) for the possible relationship between ride-hailing adoption and its determinants, resulting in the marginal effects of independent variables being considered constant or following a regular rule. Consequently, the most effective range and probable threshold of a contributing factor's effect on ride-hailing demand could be obscured (Ding et al., 2018).

In contrast to regression-based models, the interpretable machine learning approach overcomes the limitations of classical regression models requiring a predetermined functional form, thereby conveniently revealing nonlinear and threshold effects related to neighborhoods (Galster, 2018). The interpretable machine learning approach has become a powerful tool to examine the nonlinear effects and thresholds related to different travel

Preprint

modes' usage, such as driving distance (Ding et al., 2018), electric-bike ownership (Ding et al., 2019b), walking distance to transit (Tao et al., 2020a), active travel (Tao et al., 2020b), older adults' walking propensity (Yang et al., 2021), dockless bike-sharing usage (Wang et al., 2022), customized bus service use (Wang et al., 2023), and taxi charging station utilization (Cai et al., 2023). In the context of ride-hailing, two commonly used algorithms, gradient boosting decision trees (GBDT) (Jin et al., 2022; Tu et al., 2021) and random forest (RF) (Xu et al., 2021), have received significant attention. It can be predicted that evidence of nonlinear effects and thresholds is increasingly becoming prominent in ride-hailing mobility.

To sum up, the above-mentioned studies suggested that real-time ride-hailing services are increasingly important in urban mobility systems. However, the knowledge of book-ahead ride-hailing is limited. Based on the literature review, it can be concluded that empirical studies of book-ahead ride-hailing adoption and its determinants require further exploration. This study aims to bridge the gap by examining the nonlinear effects and thresholds of potential contributing factors on book-ahead ride-hailing utilization via an interpretable machine learning method. The study considers trip-level features, grid-level built environment and socio-economic characteristics, and city-level weather conditions in the modeling. Furthermore, the effect of these variables is compared between peak and non-peak hours as well as weekdays and weekends.

Preprint

Table 1. Summary of the influencing factors associated with ride-hailing usage using large-scale trip orders.

Reference	Ride-hailing type	Study area	Method	Dependent Variable	Key influencing factor				Temporal heterogeneity	
					TF	SD	BE	WC	Weekend	Peak time
Lavieri et al. (2018)	Real-time	Austin, Texas	SRM	Daily Number of trips	/	√	√	/	√	/
Yu and Peng (2019)	Real-time	Austin, Texas	GWR	Daily Number of trips	√	√	√	/	√	/
Bi et al. (2020)	Real-time	Chengdu, China	GWR	Ridership during a period	/	/	√	/	√	√
Shokoohyar et al. (2020)	Real-time	Philadelphia	RM	Travel time, fare	/	/	√	√	√	/
Marquet (2020)	Real-time	Chicago	RM	Daily Number of trips	/	√	√	/	√	/
Ghaffar et al. (2020)	Real-time	Chicago	RM	Daily Number of trips	/	√	√	√	√	/
Brown (2020)	Ride-splitting	Los Angeles	RM	Number of ride-sharing trips	/	√	√	/	/	/
Dean and Kockelman (2021)	Ride-splitting	Chicago	SRM	Proportion of ride-sharing trips	√	√	√	/	√	√
Xu et al. (2021)	Ride-splitting	Chicago	IML	Ride-sharing adoption rate	√	√	√	/	/	/
Huang et al. (2021)	Ride-splitting	Chengdu, China	SRM	Ride-sharing adoption rate	/	/	√	/	/	/
Liu et al. (2021)	Real-time	Haikou, China	RM	Ridership during a period	/	/	/	√	√	/
Tu et al. (2021)	Ride-splitting	Chengdu, China	IML	Ride-sharing adoption rate	√	√	√	/	/	/
Jin et al. (2022)	Real-time	Nanjing, China	IML	Hourly Number of trips	/	/	√	/	√	/
This study	Book-ahead	Haikou, China	IML	Hourly Number of trips	√	√	√	√	√	√

Note: (1) 'TF", "BE," "SD," and "WC" respectively represent the trip feature factors, built environment factors, socio-demographic factors, and weather condition factors in independent variables. (2) "RM": regression model; "SRM": spatial regression model; "GWR": geographically weighted regression model; "IML": interpretable machine learning method.

Preprint

3. Data profile and methodology

3.1. Study area

The data used in this study come from Haikou City (the capital city of Hainan province), China (see Fig. 1). Situated in the tropics and covering an area of 3,126 km^2, Haikou is a southern coastal city with abundant natural scenery. In 2020, the resident population was 2,873,400, an increase of 3.6% from the previous year. In the same year, its GDP reached 179.15 billion yuan. As one of the most important tourist destinations in China, Haikou's future urban transport planning is focused on emerging transport modes such as ride-hailing services.

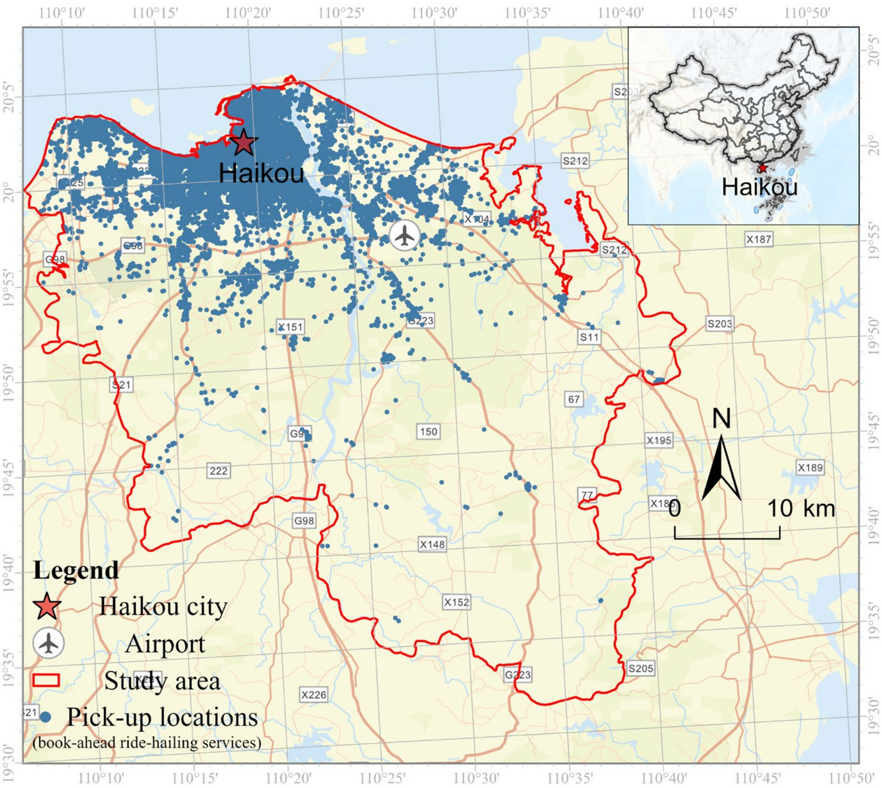

Fig. 1. Study area and distribution of pick-up locations of book-ahead ride-hailing.

3.2. Modeling unit segmentation

Fig. 1 shows that the pick-up locations of book-ahead ride-hailing trips are randomly distributed across the entire study area due to passengers' diverse travel preferences. To capture this varied trip demand, ride-hailing platforms in China often divide cities into hexagonal cells to predict general trip demand and driver supply and make operational decisions based on a specific time interval for each cell (Chen et al., 2022; Ke et al., 2019; Liu Y. et al., 2022; Liu K. et al., 2023). Therefore, we also divided our study area into hexagonal cells to aggregate book-ahead ride-hailing trips. This approach requires fewer cells to fully tessellate a given region than using triangular or square cells, and the distances between neighboring hexagons are equal (Asamer et al., 2016).

However, it is challenging to select a uniform size for hexagonal cells under different scenarios (Bi et al., 2020). Several studies suggested that 250 meters (Bi and Ye, 2021), 500 meters (Liao, 2021), and 660 meters (Ke et al., 2019) are all appropriate hexagonal side lengths for ride-hailing demand prediction and trip characteristic analysis. Fig. 2 shows a diagram of the hexagonal cell with different side lengths. To validate our results, we successively selected 250-meter, 500-meter, and 660-meter hexagons as modeling units to identify a finer spatial scale. This approach generated 14,254 250-meter hexagons, 3,670 500-meter hexagons, and 2,143 660-meter hexagons, respectively.

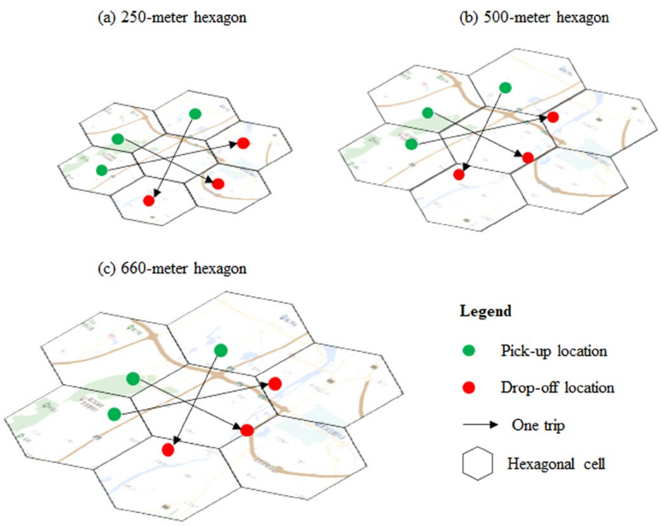

Fig. 2. Diagram of different hexagonal cells.

3.3. Variables and data profile

3.3.1. Dependent variable: hourly number of trips per OD pair

The Didi Chuxing GAIA Initiative[1] provided ride-hailing order data for all 14,160,162 trips taken in Haikou between May 1 and October 31, 2017. The data includes the order ID,

[1] https://gaia.didichuxing.com

Preprint

type, departure time, price estimate, the distance between origin and destination, pick-up location, and drop-off location for each trip. The data sample can be found in Appendix Table A1. It's worth noting that the ride-hailing order data includes two types of services: real-time ride-hailing and book-ahead ride-hailing. For our study, we focus solely on book-ahead ride-hailing trips. To generate the dependent variable, we first processed the book-ahead ride-hailing orders. We eliminated pick-up and drop-off locations outside our study area, removed duplicate trip records with the same order IDs, and deleted orders with null trip distance or fee records. Additionally, we filtered out trip orders with abnormal distances (less than 2 km or greater than 45 km) and fees (less than 9 RMB or greater than 300 RMB). After completing the data cleaning procedure, we kept 249,857 book-ahead ride-hailing trip records in the subsequent analysis of book-ahead ride-hailing demand at the hexagon level.

In this study, we use the hourly number of requests per OD pair (hexagon-to-hexagon) as the dependent variable, also referred to as hourly OD pair volumes. To calculate this variable, we first geocoded the pick-up/drop-off locations of trip orders using the latitude and longitude coordinates of the origin and destination. We then assigned the origin and destination of each book-ahead ride-hailing trip to the corresponding hexagon. Next, we aggregated the book-ahead ride-hailing trips that were picked up at the original hexagon i and dropped off at destination hexagon j during each hourly period. As discussed in Section 3.2, our study area is divided into hexagonal cells with different side lengths of 250 meters, 500 meters, and 660 meters. Thus, there are 201,176 OD pairs, 193,986 OD pairs, and 189,444 OD pairs for book-ahead ride-hailing trips under 250-meter hexagons, 500-meter hexagons, and 660-meter hexagons, respectively.

Fig. 3 shows the statistical distribution of hourly OD pair volumes (indicated by the green curve). On the horizontal axis, the frequency of OD pair volumes is represented, while on the vertical axis, the percentage of OD pairs over a certain volume frequency is represented. For instance, the first point (1, 100) in Fig. 3(a) indicates that 100% of OD pairs consist of one trip per hour. However, owning to OD pairs with one trip per hour may be occasional, so it is necessary to exclude them to ensure reliable subsequent analysis (Wang et al., 2022). To

Preprint

identify the "right" frequency of OD pair volumes, a knee-point detection method is applied (Liu et al., 2020). The subplots of Fig. 3 show that a blue line is drawn to connect the first and last points on the distribution curve (green) of hourly OD pair volumes. The distance of each point on the OD volume curve to the blue line is then calculated. Generally, the point with the maximum distance to the blue line is considered the knee point of the OD pair volume distribution. It can be observed that, under three different hexagonal side lengths, the knee point of the OD pair volume distribution is 3 (indicated by the red star). Therefore, only the OD pairs with over three trips per hour are further analyzed, which respectively account for around 4.30%, 5.35%, and 6.1% of all OD pairs under 250-meter, 500-meter, and 660-meter hexagons. Table 2 presents some information about hexagons, book-ahead ride-hailing trips, and OD pairs.

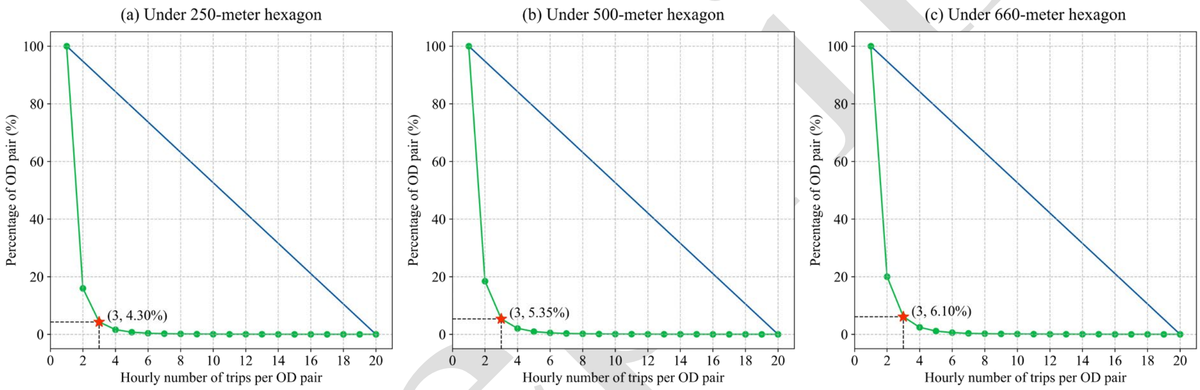

Fig. 3. The distribution of trips per grid-level OD pair.

Table 2. Statistical information on trips and OD pairs.

Hexagonal side length	**250 meters**	**500 meters**	**660 meters**
Number of hexagonal cells	14254	3670	2143
All trips of book-ahead ride-hailing	249,857	249,857	249,857
All grid-level OD pairs	201,176	193,986	189,444
Knee-point of hourly number of trips per OD pair	3	3	3
Dominant grid-level OD pairs	8,651	10,376	11,558
Percentage of dominant grid-level OD pairs	4.30%	5.35%	6.10%

3.3.2. Explanatory variables

Based on the results of previous studies (refer to Table 1), this study includes explanatory variables from five categories: trip-level features, grid-level points of interest (POIs), accessibility to facilities and socio-economic characteristics, and city-level weather conditions. As passengers' private information is unavailable, individual-level variables such as age, income, and gender are excluded.

Preprint

The indicators of trip features include the median trip distance and fee for each grid-level origin-destination (OD) pair, which is extracted from book-ahead ride-hailing order data. These indicators can help identify preferred travel distances and acceptable fees for book-ahead ride-hailing usage. It should be noted that the trip fee of the ride-hailing services is determined by trip length and travel time (Wei et al., 2021), but it can be influenced by trip demand. In turn, the trip fee also influences ride-hailing demand. In this study, we are mainly interested in the influence of trip fees on ride-hailing demand. The feedback effects of the trip demand on the trip fee are not considered in our model. Additionally, we also include a dummy variable that captures whether the trip was requested on a weekday or weekend.

Grid-level POIs, facilities accessibility, and socio-economic characteristics can provide insights into passengers' trip purposes. Valuable features related to grid-level POIs were derived from POI data obtained from the places application programming interface (API) of Gaode Map[1]. This data contains location information of various urban facilities, such as transportation facilities, workplaces, social services facilities, and others. POI data has been extensively used as a surrogate measure to estimate built environment characteristics (Bi et al., 2020; Yang et al., 2021). For this study, we extracted eight typical POIs (workplaces, residences, dining, recreation, education, shopping, medical, and accommodation) from the POI data. We used a POI-based entropy index to quantify land use (Huang et al., 2021; Li et al., 2020; Yue et al., 2017). The formula for the POI-based entropy index for a hexagonal cell is as follows:

$$E_{POI} = -\sum_{k=1}^{K} p_k \cdot \log_2 \left(p_k \right) , \tag{1}$$

where K is the number of POI types, and p_k denotes the percentage of the k^{th} POI type in a given hexagonal cell. A higher POI-based entropy index means a greater diversity of land use. To assess grid-level accessibility, we measured the minimum travel time from each hexagonal center to key locations (bus stop, coach station, railway station, airport, tourism spot, city center, county center, and town center) with the help of the route planning function of the Baidu Map API[2]. It is essential to point out that the minimum travel time for a given hexagonal center to

[1] https://lbs.amap.com/api
[2] https://lbsyun.baidu.com/products/direction

the nearest bus stop is walking time, but to other places, it is driving time. To measure grid-level socio-economic characteristics, we calculated the average GDP (Gross Domestic Product) of each hexagon based on disaggregated GDP data at a 1 km × 1 km grid area in 2017 (Zhao et al., 2017). It should be noted that the grid-level variables at both the origin hexagons and destination hexagons were incorporated into the models.

To examine the effects of weather conditions and changes on book-ahead ride-hailing usage, we considered five variables following Wang et al. (2022): AQI (air quality index), temperature, humidity, wind speed, and rainfall per hour. The data were obtained from the "Environment Cloud" API[1] of China and were merged with book-ahead ride-hailing trips based on the time index. Additional details about the explanatory variables are provided in Table 3.

Table 4 presents the descriptive statistics of these variables for the three types of hexagonal cells. The average hourly requests for book-ahead ride-hailing services per OD pair in the 250-meter, 500-meter, and 660-meter hexagons were close to 3.8 trips, indicating that the penetration of book-ahead ride-hailing services was low in Haikou City. As shown in Table 4, the average median travel distance and median travel fee were less than 12 km and 40 RMB, respectively. It can be observed from Table 4 that trip features and accessibility to key places were similar for the three cell sizes, but the mean values for POI features varied somewhat.

Preprint

[1] http://www.envicloud.cn/pages/product.html

Table 3. Explanatory variable definitions.

Explanatory variables	Description
Trip features	
Median travel distance	Median travel distance of all trips per grid-level OD pair (km)
Median travel fee	Median travel fee of all trips per grid-level OD pair (RMB)
Weekdays	Dummy variable, 0: weekends, 1: weekdays
POI features at origin hexagons and destination hexagons	
Workplace POIs	Number of workplace POIs in the hexagonal cell, including government, factory, industry, company, enterprise, etc.
Residence POIs	Number of residence POIs in the hexagonal cell, including apartments, houses, etc.
Dining POIs	Number of dining POIs in the hexagonal cell, including Chinese restaurants, Western restaurants, fast food outlets, snack bars, etc.
Shopping POIs	Number of shopping POIs in the hexagonal cell, including supermarkets, shopping malls, large department stores, etc.
Recreation POIs	Number of recreation POIs in the hexagonal cell, including scenic spots, stadiums, movie theaters, etc.
Schooling POIs	Number of schooling POIs in the hexagonal cell, including schools, colleges, universities, training institutions, etc.
Medical POIs	Number of medical POIs in the hexagonal cell, including hospitals, clinics, pharmacies, etc.
Accommodation POIs	Number of accommodation POIs in the hexagonal cell, including hotels, guesthouses, etc.
POI entropy index	An index to evaluate land use mixture based on the number of POI types
Accessibility features at origin hexagons and destination hexagons	
Accessibility to bus stop	Duration of walking from a hexagon center to the nearest bus stop (minutes)
Accessibility to coach station	Duration of driving from a hexagon center to the nearest coach station (minutes)
Accessibility to railway station	Duration of driving a hexagon center to the nearest railway station (minutes)
Accessibility to airport	Duration of driving from a hexagon center to the nearest international airport (minutes)
Accessibility to tourism spot	Duration of driving from a hexagon center to the nearest tourism spot (minutes)
Accessibility to city center	Duration of driving from a hexagon center to a municipal government office (minutes)
Accessibility to county center	Duration of driving from a hexagon center to the nearest county government office (minutes)
Accessibility to town center	Duration of driving from a hexagon center to the nearest government office or subdistrict office (minutes)
Socio-economic features at origin hexagons and destination hexagons	
Grid-level GDP	GDP value per hexagonal cell (RMB)
Weather condition	
AQI	City-level average air quality index per hour
Temperature	City-level average temperature per hour (℃)
Humidity	City-level average humidity per hour (%rh)
Wind speed	City-level average wind speed per hour (m/s)
Rainfall	City-level average rainfall per hour (mm)

Preprint

Table 4. Descriptive statistics.

Variables	250-meter hexagonal cell (8,651 OD pairs)				500-meter hexagonal cell (10,376 OD pairs)				500-meter hexagonal cell (11,558 OD pairs)			
Dependent variable	**Mean**	**Std.**	**Min**	**Max**	**Mean**	**Std.**	**Min**	**Max**	**Mean**	**Std.**	**Min**	**Max**
Hourly requests per OD pair	3.8	1.66	3	41	3.85	1.8	3	39	3.88	1.81	3	41
Explanatory variables												
Trip features												
Median travel distance (km)	11.13	6.76	2	40.08	11.96	7.03	2	40.08	12.63	7.24	2	40.4
Median travel fee (RMB)	35.99	26.76	9	230.4	38.18	26.78	9	230.4	40.04	26.82	9	230.4
Weekdays (dummy variable)	0.36	0.48	0	1	0.35	0.48	0	1	0.35	0.48	0	1
POI features at origin hexagons												
Workplace POIs (n)	24.64	44.45	0	382	89.72	124.73	0	753	161.65	210.75	0	1109
Residence POIs (n)	8.57	10.45	0	53	33.07	34.25	0	139	57.01	57.92	0	216
Dining POIs (n)	47.76	60.38	0	311	158.98	170.84	0	752	261.25	253.83	0	1027
Shopping POIs (n)	54	81.57	0	700	199.25	253.95	0	1711	353.02	411.03	0	2149
Recreation POIs (n)	5.19	8.42	0	75	17.69	21.27	0	121	30.41	37.23	0	194
Schooling POIs (n)	9.94	14.34	0	122	34.9	36.46	0	189	59.72	60.49	0	243
Medical POIs (n)	5.88	10.85	0	210	21.5	28.07	0	241	37.81	46.95	0	267
Accommodation POIs (n)	6.43	8.45	0	59	21.52	22.93	0	106	35.85	35.59	0	127
POI entropy index	2.82	0.7	0	3.81	3.06	0.55	0	3.94	3.13	0.5	0	3.83
POI features at destination hexagons												
Workplace POIs (n)	23.27	47.94	0	382	70.22	118.62	0	753	109.58	192.53	0	1109
Residence POIs (n)	6.82	10.08	0	53	24.92	33.1	0	139	37.98	53.27	0	216
Dining POIs (n)	36.83	57.55	0	311	108.88	161.25	0	752	173.28	244.82	0	1027
Shopping POIs (n)	49.34	88.43	0	700	149.62	250.43	0	1711	248.92	385.55	0	2149
Recreation POIs (n)	4.53	9.09	0	75	13	20.83	0	121	20.35	34.65	0	194
Schooling POIs (n)	8.09	15.77	0	122	23.38	35.43	0	189	37.37	57.14	0	243

Preprint

Medical POIs (n)	5.34	13.07	0	210	15.12	25.23	0	241	25.29	41.88	0	267
Accommodation POIs (n)	5.46	8.5	0	48	16.36	22.58	0	106	25.05	34.25	0	127
POI entropy index	2.35	1.08	0	3.81	2.04	1.56	0	3.94	1.95	1.59	0	3.83
Accessibility features at origin hexagons												
To bus stop (minutes)	8.88	16.93	0	431.97	23.19	31.97	0	430.05	46.4	71.13	0.05	688.98
To coach station (minutes)	13.53	6.01	0.12	50.08	12.9	5.85	0.33	50.05	12.34	5.89	2.1	52.23
To railway station (minutes)	13.5	7.57	0.08	52.72	17.63	8.42	1.42	52.38	16.71	7.92	1.82	54.27
To airport (minutes)	31.19	13.4	1.98	70.57	38.14	17.94	0.28	87.05	33.4	14.39	1.87	82.38
To tourism spot (minutes)	9.01	4.2	0.13	54.55	9.9	5.42	0.98	57.45	10.34	6.61	0.9	85.65
To city center (minutes)	32.16	10.91	2.23	81.02	31.77	11.02	2.23	81.73	33.4	11.05	4.55	98.87
To county center (minutes)	14.57	8.97	0.1	56.43	13.99	8.87	0.1	54.22	13.39	8.77	2.08	54.12
To town center (minutes)	9.57	5.82	0	41.95	9.88	5.68	0	42.03	11.27	7.15	0.33	43.93
Accessibility features at destination hexagons												
To bus stop (minutes)	5.6	10.75	0	290.32	14.37	23.18	0	430.05	44.79	60.09	0.05	525.18
To coach station (minutes)	9	5.33	0.12	48.75	8.33	4.96	0.33	49.75	16.01	9.13	2.1	52.03
To railway station (minutes)	8.2	5.62	0.08	56.17	10.38	8.07	1.42	55.6	18.05	9.38	1.82	59
To airport (minutes)	23.14	17.85	1.35	70.25	26.15	23.08	0.28	87.05	31.34	12.19	1.87	82.38
To tourism spot (minutes)	11.12	5	0.13	54.52	12.02	5.1	1	57.75	13.3	7.12	1.67	56.62
To city center (minutes)	32.97	9.54	2.23	81.68	33.79	9.91	2.23	81.73	37.64	10.61	4.55	79.15
To county center (minutes)	14.7	7.81	0.1	57.78	15.38	7.81	0.1	58.23	21.31	13.55	2.08	58.28
To town center (minutes)	13.15	8.72	0	41.95	14.61	8.18	0	39.07	13.71	6.4	0.33	38.43
Socio-economic features at origin hexagons												
Grid-level GDP (RMB)	1058	1634	0	5450	1150	1705	0	5450	1234	1706	0	5450
Socio-economic features at destination hexagons												
Grid-level GDP (RMB)	1057	1526	0	5450	1003	1490	0	5450	873	1414	0	5450
Weather condition												
AQI	26.9	15.74	7	145	26.81	15.68	7	145	26.68	15.52	7	145

Preprint

Temperature (℃)	27.08	2.7	18.4	37.6	27	2.65	18.4	37.6	26.95	2.59	18.4	37.6
Humidity (%rh)	84.05	10.7	40	98	84.43	10.56	40	98	84.76	10.43	40	98
Wind speed (m/s)	2.67	1.5	0	10.2	2.62	1.5	0	10.2	2.61	1.49	0	10.2
Rainfall (mm)	0.2	1.58	0	30.8	0.18	1.44	0	30.8	0.17	1.38	0	30.8

Preprint

3.4. Modeling approach

3.4.1. GBDT model

Following Ding et al. (2018) and Tao et al. (2020a), we used GBDT to estimate the relationship between book-ahead ride-hailing trips and their possible determinants. A GBDT model can make the objective relationships behind variables more flexible because there is no assumption about the model form (such as linear, logarithmic, or exponential). In addition, a high goodness of fit can be obtained from a small sample as long as the distribution of the output variable in the sample data is sufficiently comprehensive. Moreover, the relative importance and partial dependence plots provide tools to help interpret the model results.

As a prevalent ensemble learning method, GBDT is recognized as a combination of two algorithms: the regression tree and the boosting method (Friedman, 2001). For a dataset $D = \{(\boldsymbol{x}_1, y_1), (\boldsymbol{x}_2, y_2), \ldots, (\boldsymbol{x}_n, y_n)\}$ with N samples, assuming that for the n^{th} sample, $\boldsymbol{x}_n$ is a set of explanatory variables and y_n is the dependent variable, which is the hourly number of trips per OD pair, the key task of GBDT is adding a series of decision trees $f(\boldsymbol{x})$ to produce an approximation function $F(\boldsymbol{x})$ for mapping explanatory variables to the dependent variable (see Eq. (2)).

$$F(\boldsymbol{x}) = \sum_{m=1}^{M} f_m(\boldsymbol{x}) = \sum_{m=1}^{M} \beta_m h(\boldsymbol{x}; a_m), \tag{2}$$

where M represents the number of iterations, $h(\boldsymbol{x}; a_m)$ is a single decision tree, a_m is the average split location of each splitting variable in a single tree. β_m, the weight of the m^{th} decision tree, is estimated by minimizing a squared error loss function. Based on the gradient descent direction, the approximation function $F(\boldsymbol{x})$ is updated as shown in Equation (3):

$$\begin{cases} F_m(\boldsymbol{x}) = F_{m-1}(\boldsymbol{x}) + \beta_m h(\boldsymbol{x}; a_m) \\ (\beta_m, a_m) = \arg\min\limits_{\beta, a} \sum_{i}^{N} \left(y_i - \left(F_{m-1}(\boldsymbol{x}_i) + \beta h(\boldsymbol{x}_i; a) \right) \right)^2 \end{cases}. \tag{3}$$

A learning rate (shrinkage) parameter ξ is further utilized to scale the contribution of each decision tree to control the overfitting problem (Ding et al., 2019b; Friedman, 2001) as follows:

$$F_m(\boldsymbol{x}) = F_{m-1}(\boldsymbol{x}) + \xi \cdot \beta_m h(\boldsymbol{x}; a_m), \ \xi \in (0,1]. \tag{4}$$

3.4.2. Relative importance

The GBDT model produces relative importance indicators to examine and rank the influence of all explanatory variables on the dependent variable. In this study, $I_{x_p}^2$ is the relative importance of the p^{th} independent variable x_p on book-ahead ride-hailing trips, calculated via Equations (5) and (6).

$$I_{x_p}^2 = \frac{1}{M}\sum_{m=1}^{M} I_{x_p}^2\left(h_m\right), \tag{5}$$

$$I_{x_p}^2\left(h_m\right) = \sum_{j(h_m)=1}^{J(h_m)-1} d_j, \tag{6}$$

where d_j is the improvement of the squared error term by the j^{th} splitting, and J is the number of leaves on the m^{th} decision tree h_m. In addition, the sum of the relative importance of all exploratory variables is 100%.

3.4.3. Partial dependence plot

In addition, another advantage of GBDT is the ability to visualize the association between the exploratory and dependent variables by generating partial dependence plots (Friedman, 2001). These plots provide a graphical depiction of the marginal effect of an independent variable on an output variable while controlling for all other variables in the model (Ding et al., 2019b, 2018). Moreover, the partial dependence plots also show the effective range and threshold (Tu et al., 2021).

4. Results

4.1. Temporal patterns of book-ahead ride-hailing trips

The temporal pattern of book-ahead ride-hailing trips on weekdays and weekends is presented in Fig. 4. It can be observed that there is no big difference in the hourly distribution of book-ahead ride-hailing trips on weekdays and weekends. After 9:00, the number of requesting book-ahead ride-hailing services per hour on weekends is slightly higher than on weekdays. Moreover, A clear temporal pattern is witnessed, which can be approximately

Preprint

divided into a morning peak (4:00-8:00), an evening peak (16:00-20:00), and two off-peak valleys (8:00-16:00 and 20:00-4:00). During the morning peak, the average requests for book-ahead ride-hailing services per hour increase dramatically to its highest value of the day, about 160 trips per hour. Compared to the morning peak, the maximum number of requests for using this mode during the evening peak is lower, at about 110 trips per hour. During other off-peak hours, the hourly number of requesting book-ahead ride-hailing services is less than that of peak hours. The maximum number of requests during off-peak hours is only about 70 trips per hour.

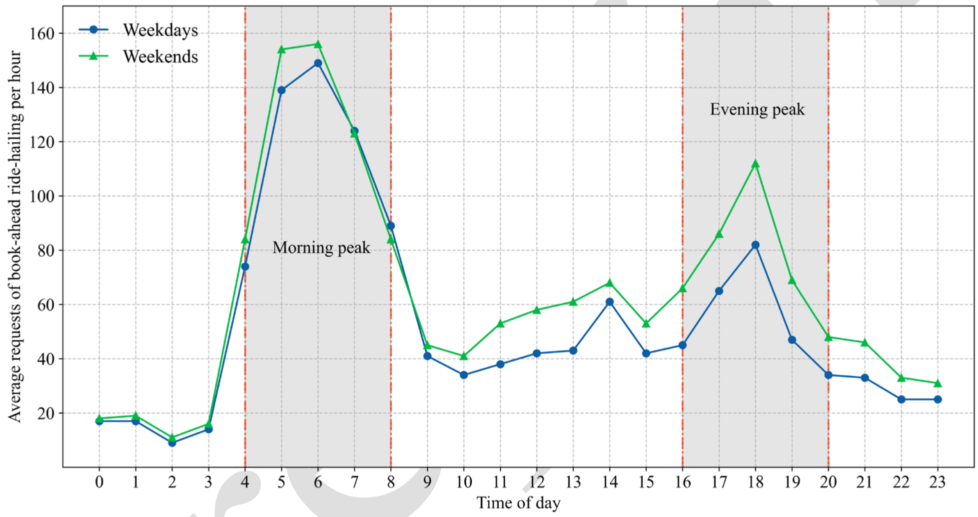

Fig. 4. Hourly distribution of book-ahead ride-hailing trips.

4.2. Comparison of model fitting results

The GBDT model was estimated using the “scikit-learn” package [1] in the Python environment. To optimize parameter settings and ensure the robustness of modeling results, we employed five-fold cross-validation. Besides, the sample set was randomly divided into the training subset (70%) and the testing subset (30%). Given the heterogeneous temporal usage pattern of book-ahead ride-hailing services (as illustrated in Fig. 4), we modeled three time periods: morning peak, evening peak, and off-peak hours for a comprehensive comparative analysis.

[1] https://scikit-learn.org/stable/modules/ensemble.html#gradient-tree-boosting

Before applying GBDT, it was necessary to find optimal model hyperparameters. Conventional approaches, such as grid search and random search, can assist in finding a suitable set of model hyperparameters, but they can be time-consuming and labor-intensive, particularly with a large-scale dataset (Joy et al., 2020; Yin and Li, 2022; Zhang et al., 2023). To address this issue, the Bayesian optimization (BO) algorithm was employed, as it has been proven to be a powerful tool for finding the most appropriate results with fewer iterations (Yin et al., 2022; Zhong et al., 2021). We applied a BO algorithm with 10 iterations to tune the optimal hyperparameter combination of shrinkage, the number of trees, and tree depth. The ranges of the three aforementioned hyperparameters were set as follows: the number of trees was set from 100 to 1,000; the maximum depth for each tree varied from 3 to 10, and the shrinkage parameter was set between 0.05 and 0.2. We then compared the performance of the root mean square error (RMSE) to determine the most suitable hyperparameter combination. This process was repeated for the three types of hexagonal cells.

Table 5. Optimal hyperparameters of GBDT models.

Hexagon	Period	Sample (OD pairs)	RMSE	Number of trees	Tree depth	Shrinkage
	Morning peak	4233	1.676	347	6	0.104
250-meter	Evening peak	2085	2.003	311	5	0.108
	Off-peak hours	2333	1.516	320	5	0.154
	Morning peak	5449	1.755	470	4	0.137
500-meter	Evening peak	2300	2.287	470	4	0.136
	Off-peak hours	2627	1.963	475	6	0.148
	Morning peak	6427	1.783	470	4	0.136
660-meter	Evening peak	2379	2.089	747	8	0.124
	Off-peak hours	2752	1.890	320	7	0.154

Table 5 shows the optimal values of the hyperparameter combination under different periods. It can be observed from Table 5 that among the three types of hexagons, the GBDT model in 250-meter hexagons optimized by the BO algorithm had the minimum RMSE, outperforming the other two types of hexagonal cells. The gap for RMSE among the three types of hexagons is closed, meaning that the hexagon size may not exert significant effects on model

Preprint

results. In 250-meter hexagonal cells, after conducting 10 tests, the shrinkage, number of trees, and tree depth for the morning peak model were respectively 0.104, 347, and 6. As for the evening peak model, the values were 0.108, 311, and 5, and for the off-peak model, they were 0.154, 320, and 5. Given the better performance of GBDT in 250-meter hexagonal grids, subsequent analysis on variable importance and partial dependence plot was conducted based on these hyperparameter combinations.

4.3. Relative importance of independent variables

Table 6 displays the relative contribution of each independent variable to the book-ahead ride-hailing trips for three models (morning peak model, evening peak model, and off-peak hours mode). The higher an independent variable's relative importance value, the greater its contribution to estimating the hourly number of book-ahead ride-hailing trips per OD pair. At the end of the table, we report the model performance via the RMSE and R-squared. All R-squared values are around 0.93, proving that the GBDT models have a high accuracy. Collectively, we find trip features valuables (average RI = 5.97%) have the most prominent factor contributing to book-ahead ride-hailing usage, followed by weather condition variables (average RI = 4.79%), accessibility features variables both in origin hexagons (average RI = 3.21%) and destination hexagons (average RI = 1.53%). The POI features variables (average RI = 1.06% (origin hexagons), 0.97% (destination hexagons)), and grid-level socio-economic features (average RI = 1.20% (origin hexagons), 0.79% (destination hexagons)) do not present a very significant association with book-ahead ride-hailing usage.

In three trip features valuables, median travel distance (morning peak RI = 10.96%, evening peak RI = 7.48%, off-peak hours RI = 6.04%) and median travel fee (morning peak RI = 9.00%, evening peak RI = 8.39%, off-peak hours RI = 7.38%) are the most critical factor that contributes to book-ahead ride-hailing usage during all periods. The dummy variable (weekdays) only shows high relative importance during the evening peak (RI = 2.85%). The partial dependence plots in the next section will report the marginal effects of these variables on book-ahead ride-hailing usage.

Preprint

Interestingly, four variables associated with weather conditions exhibit high relative importance on hourly book-ahead ride-hailing trips per OD pair, including AQI (average RI = 7.58%), temperature (average RI = 4.63%), humidity (average RI = 4.91%), and wind speed (average RI = 5.83%). During the morning peak, wind speed (RI = 8.06%) and temperature (RI = 6.15%) are the most prominent factors that affect book-ahead ride-hailing demand. During evening peak and off-peak hours, AQI (evening peak RI = 9.18%, off-peak hours RI = 8.04%) becomes the most critical factor affecting the hourly amount of book-ahead ride-hailing trips per OD pair. It is not unreasonable to speculate that unfriendly weather, such as high temperatures, wind, and humidity, could influence book-ahead ride-hailing usage. For instance, when passengers notice that weather conditions could be worse, they may give up the planned travel modes and turn to book a reliable ride-hailing vehicle in advance.

In the accessibility variables, the accessibility between original hexagons (average RI = 6.47%) or destination hexagons (average RI = 7.58%) and the airport contribute most to predicting book-ahead ride-hailing trips per OD pair. Accessibility to the railway station and accessibility to the bus stop are also major factors. Additionally, accessibility to the city center or county center is found to be related to book-ahead ride-hailing services. The above findings are generally in agreement with expectations. First, this indicates that book-ahead ride-hailing services are possible to be used to connect with mainly transportation hubs. For instance, we found that a large number of book-ahead ride-hailing trips often occur around the airport (Xu et al., 2021) and the railway terminal (Li et al., 2019). This is probably because these trip requests are subject to flight and high-speed railway schedules. When passengers are informed to arrive at these transportation hubs very early in the morning or well past midnight, they may be attracted to book-ahead ride-hailing services because taxi and public transit services are scarce at that time. Second, some book-ahead ride-hailing trips start or end around bus stops, which might result from inconvenient first- or last-mile accessibility between bus stops and the origin or destination. If travelers take more time from the origin to the pick-up bus stop or from the drop-off bus stop to the destination, they may choose the book-ahead ride-hailing services instead of walking.

Preprint

For accessibility to the airport, the RI during off-peak hours (origin RI = 16.74%, destination RI = 3.41%) is higher than that during morning peak (origin RI = 1.97%, destination RI = 0.59%) and evening peak (origin RI = 2.83%, destination RI = 2.44%). It is concluded that book-ahead ride-hailing services are preferred to be used by passengers during off-peak hours than peak hours, which is not a surprise. On the one hand, during off-peak hours, especially in the early morning or well past midnight, other modes, such as airport buses and taxis, may not have enough service frequency and convergence. This means that passengers will have to wait a long time to exit the airport to board a bus or taxi. Conversely, the long waiting time could be avoided if they choose book-ahead ride-hailing services. On the other hand, passengers do not prefer to use book-ahead ride-hailing services during peak hours because passengers could worry about road congestion during the peak period, causing ride-hailing services to be unreliable and incurring high travel costs. This finding shows that book-ahead ride-hailing services may become a supplement for airport buses or taxis during off-peak periods.

Preprint

Table 6. Relative importance of all independent variables in 250-meter hexagon cells

Variables	Morning peak	Evening peak	Off-peak hours
	RI (%)	RI (%)	RI (%)
Trip features	*6. 95*	*6. 24*	*4. 74*
Median travel distance	10.96	7.48	6.04
Median travel fee	9.00	8.39	7.38
Weekdays	0.90	2.85	0.79
POI features (origin)	*1. 36*	*1. 06*	*0. 75*
Workplace POIs	1.35	0.59	0.91
Residence POIs	1.00	0.55	0.41
Dining POIs	1.35	1.17	0.62
Shopping POIs	1.80	1.19	0.67
Recreation POIs	0.57	0.33	0.26
Schooling POIs	1.45	1.06	0.72
Medical POIs	1.29	3.23	0.34
Accommodation POIs	1.06	0.20	1.16
POI entropy index	2.39	1.18	1.67
POI features (destination)	*0. 73*	*1. 08*	*1. 09*
Workplace POIs	0.48	0.69	0.67
Residence POIs	0.43	0.88	0.26
Dining POIs	0.96	0.85	1.09
Shopping POIs	1.01	1.09	1.59
Recreation POIs	0.78	1.18	0.92
Schooling POIs	0.80	1.05	0.78
Medical POIs	0.43	0.45	0.66
Accommodation POIs	0.73	1.45	0.67
POI entropy index	0.98	2.12	3.20
Accessibility features (origin)	*2. 98*	*2. 31*	*4. 34*
To bus stop	3.74	1.83	2.50
To coach station	2.13	1.26	2.47
To railway station	2.13	1.27	1.88
To airport	1.97	2.83	16.74
To tourism spot	2.01	3.39	2.21
To city center	3.76	3.99	3.77
To county center	6.12	1.92	2.55
To town center	1.98	2.03	2.58
Accessibility features (destination)	*1. 13*	*1. 87*	*1. 58*
To bus stop	1.13	2.50	1.73
To coach station	1.10	2.26	1.77
To railway station	1.09	3.16	1.73
To airport	0.59	2.44	3.41
To tourism spot	1.73	1.94	1.11

Preprint

To city center	1.85	0.68	0.91
To county center	0.75	0.72	0.63
To town center	0.79	1.28	1.38
Socio-economic features (origin)	*1.11*	*1.68*	*0.83*
Grid-level GDP	1.11	1.68	0.83
Socio-economic features (destination)	*0.77*	*0.80*	*0.79*
Grid-level GDP	0.77	0.80	0.79
Weather condition	*5.11*	*5.22*	*4.04*
AQI	5.51	9.18	8.04
Temperature	6.15	4.34	3.40
Humidity	5.64	5.57	3.51
Wind speed	8.06	4.59	4.85
Rainfall	0.18	2.41	0.43
Measurement of fit			
Samples (OD pairs)	4,233	2,085	2,333
Root mean square error (RMSE)	0.407	0.384	0.427
R-squared	0.934	0.956	0.929

Note: RI represents relative importance. Italicized numbers in the table represent average relative importance.

Preprint

4.4. Associations among key independent variables

Upon examining the relative importance of all independent variables, it is crucial to depict the marginal effects of significant contributing factors on book-ahead ride-hailing demand by the partial dependence plot. Identifying threshold effects and effective ranges of key variables from partial dependence plots can provide valuable guidance for planning, operating, and managing ride-hailing platforms. As highlighted in Table 6, certain contributing variables, such as trip features, weather conditions, and accessibility variables, displayed high relative importance concerning book-ahead ride-hailing adoption. Further analysis in this section will focus primarily on these variables.

To interpret the partial dependence plots of the GBDT model, we infer the direction of effects by observing how the hourly number of book-ahead ride-hailing trips per OD pair changes with increases in one of the aforementioned variables. We also examine whether salient nonlinear effects and thresholds exist in the partial dependence plots. To prevent fake nonlinear effects resulting from sparse samples, we include the sample distribution of the independent variables at the bottom of each plot.

Moreover, we pay special attention to comparing the differences in partial dependence plots between weekdays and weekends in various periods. Through this comparison, we can identify any unique patterns in using book-ahead ride-hailing services during weekdays and weekends. This information can help ride-hailing platforms tailor their services to meet the needs of different users during specific times, leading to increased user satisfaction and overall platform success.

4.4.1. Trip features

As expected, we find complex nonlinear effects of two trip feature variables: median travel distance and median travel fee on hourly trips of book-ahead ride-hailing per OD pair. Here, special attention is paid to the impacts of median travel distance. The results of the median travel fee variable are omitted because the effects are similar. Fig. 5 illustrates how book-ahead ride-hailing trips per hour per OD pair respond to changes in travel distance between weekdays and weekends. The effects of travel distance on book-ahead ride-hailing usage do not have a

Preprint

huge difference between weekdays and weekends during the morning peak (see Fig. 5 a). In this period, there is an overall V-shaped relationship between book-ahead ride-hailing trips and travel distance. As the travel distance increases, the hourly number of book-ahead ride-hailing trips per OD pair first decreases, then increases, and finally stabilizes. An inflection point can be found at around 15 km. On the contrary, during other periods, travel distance presents an overall positive effect on book-ahead ride-hailing usage (see Fig. 5 b and c). Longer travel distance leads to higher usage, but it shows a wave-like increased pattern. The positive effect of travel distance is more significant over 10 km.

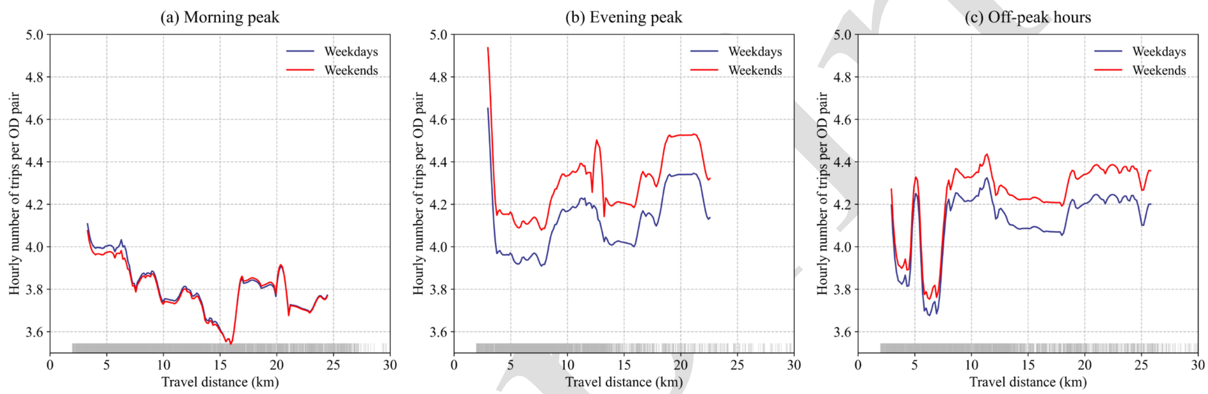

Fig. 5. Partial dependence plot for median travel distance.

The above findings are meaningful for designing book-ahead ride-hailing services. We can infer that passengers are more likely to use book-ahead ride-hailing services to accomplish a short-distance trip during the morning peak from Fig.5 a. This is plausible because most trips during the morning peak are for commuting purposes, and their destinations are usually around workplaces. For trips correlated with these activities, choosing public transit like bus or subway is more reliable due to lower external congestion effects. However, inconvenient first-/last-mile accessibility to bus stops or metro stations will reduce the willingness to use public transit because travelers usually get tired of long walking from the origin to bus stops or metro stations. Thus, book-ahead ride-hailing may play a role in short-distance boarding services for the first/last mile during the morning peak. During the evening peak and off-peak hours, book-ahead ride-hailing services may be used for special activities, recreation, dating, gathering, etc., after finishing one day of work. Passengers might plan their subsequent activities several hours

Preprint

ahead of schedule. It is convenient and time-saving for travelers to achieve these trips via book-ahead ride-hailing services because of less waiting time.

4.4.2. Weather conditions

The study examined the nonlinear effects of weather conditions on the adoption of book-ahead ride-hailing services. Specifically, we analyzed the impact of temperature and wind speed on the hourly trips of book-ahead ride-hailing per OD pair during different periods between weekdays and weekends.

Fig. 6 presents the partial dependence plot of the temperature variable during various periods. The plot indicates that the temperature is also V-shaped (see Fig. 6 a and b) associated with the hourly trips of book-ahead ride-hailing per OD pair during the morning and evening peak periods, while their relationship is negative during off-peak hours. During the morning and evening peaks, as the temperature rises from 20 °C to 27.5 °C, the hourly trips of book-ahead ride-hailing per OD pair drop stepwise. After 27.5 °C, the trip volume per OD pair increases significantly from around 3.5 trips to 4.5 trips per OD pair. In contrast, during off-peak hours, the number of book-ahead ride-hailing trips gradually decreases as the temperature rises from 20 °C to 37.5 °C.

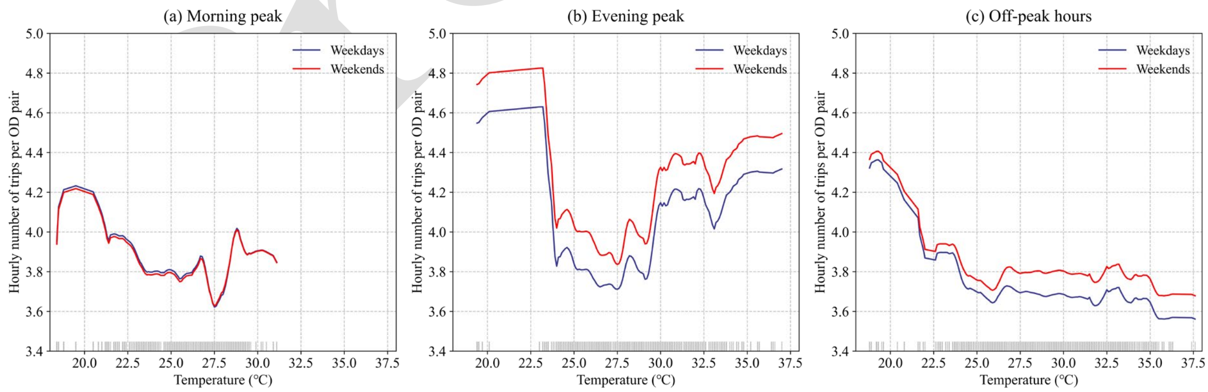

Fig. 6. Partial dependence plot for temperature.

Similarly, Fig. 7 shows that wind speed is positively associated with the book-ahead ride-hailing adoption during the morning and evening peaks, while the relationship between book-ahead ride-hailing trips and wind speed is negative during off-peak hours. During morning and evening peaks, hourly trips of book-ahead ride-hailing per OD pair do not significantly change when the wind speed is below 4 m/s (see Fig. 7 a and b). However, after an increase in wind speed from 4 m/s to 10 m/s, the hourly book-ahead ride-hailing trips per OD pair steadily grew to a high level (around hourly 4.4 trips per OD pair). In contrast, during off-peak hours, the hourly number of book-ahead ride-hailing per OD pair continuously declines as the wind speed increases (see Fig. 7 c).

The explanation for this result is similar to the temperature variable. In turbulent windy conditions, book-ahead ride-hailing services attract a large number of commuting trips, thus producing high book-ahead ride-hailing usage during morning and evening peaks. During off-peak hours, book-ahead ride-hailing services may experience high usage from passengers taking non-commuting trips. These services can provide personalized services for them and reduce their costs. However, if in turbulent windy conditions, non-commuting trips would be dropped, thus exerting a low book-ahead ride-hailing demand in this situation.

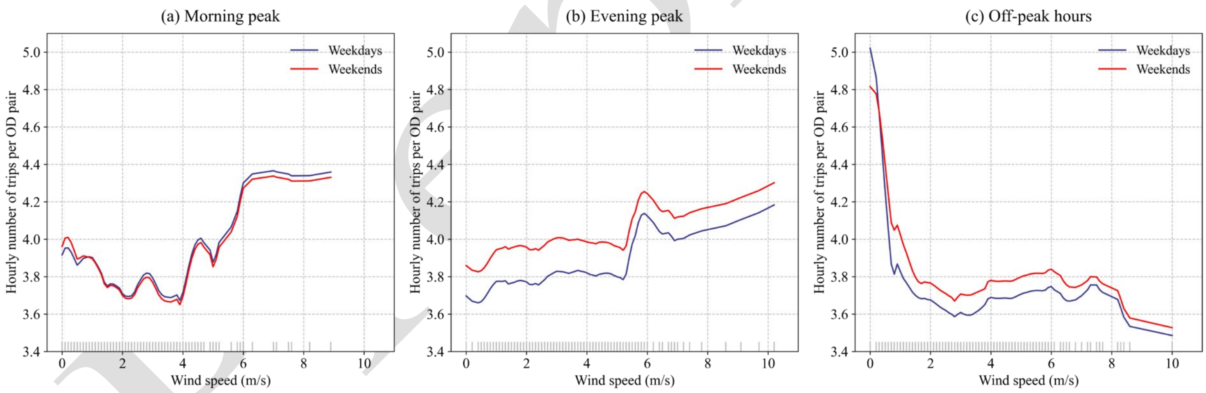

Fig. 7. Partial dependence plot for wind speed.

4.4.3. Accessibility to transportation hubs

We investigated the nonlinear associations between hourly trips of book-ahead ride-hailing services and accessibility variables to key urban places, focusing on how accessibility connecting to transportation hubs impacts book-ahead ride-hailing usage. The results of this analysis will support book-ahead ride-hailing services to be a first- or last-mile feeder mode reaching transportation hubs. Furthermore, we found that the nonlinearity pattern between accessibility variables and hourly trip volume of book-ahead ride-hailing at trip starts and trip

Preprint

ends is quite similar. Thus, we only present the results at the trip's origin.

Fig. 8 illustrates the influence of accessibility to the airport on book-ahead ride-hailing trips during all periods between weekdays and weekends. As anticipated, overall accessibility to the airport is positively associated with the hourly number of book-ahead ride-hailing trips per OD pair (see Fig. 8 a, b, and c). As the minimum travel time from the hexagonal center to the airport increases, the accessibility to the airport becomes poorer, and the hourly number of book-ahead ride-hailing trips decreases stepwise. Specifically, when the travel time from the hexagons to the airport ranges from 0 to 20 minutes, the hourly trip volume of book-ahead ride-hailing per OD pair decreases from around 4.8 trips to 3.6 trips. When accessibility to the airport exceeds 20 minutes, the hourly number of trips per OD pair tends to remain stable, indicating that the accessibility variable to the airport has neglectable effects on book-ahead ride-hailing usage when it exceeds its threshold. The notable threshold of 20 minutes suggests the feasibility of book-ahead ride-hailing serving as the pick-up service for passengers arriving at the airport in Haikou City, China. This result would help ride-hailing platforms to identify new book-ahead ride-hailing trip-generating areas and offer an alternative to shuttle buses from the airport.

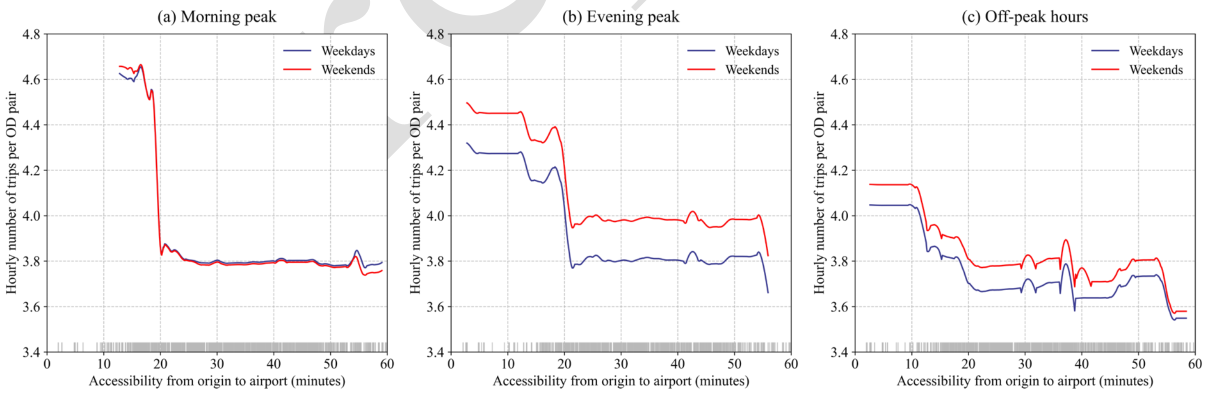

Fig. 8. Partial dependence plot for accessibility to the airport.

We also observed a positive association between the accessibility to the railway station and book-ahead ride-hailing demand across different periods (see Fig. 9 a, b, and c), similar to the accessibility to the airport (see Fig. 8 a, b, and c). The positive effects of accessibility to the railway station on book-ahead ride-hailing trip volume are only effective with certain thresholds. For instance, as the travel time from the original hexagon to the station increases across three periods, the hourly number of book-ahead ride-hailing trips decreases step by step. When the travel time from the original hexagon to the station is more than 10 minutes (threshold), the partial dependence plots for accessibility to the railway station start to remain constant. When the travel time between the original hexagon and the station ranges from 0 to 10 minutes, the hourly number of book-ahead ride-hailing trips per OD pair drops significantly from 4.2 to 3.7 trips. These findings suggest that a buffer area within 10-minute driving accessibility is practical for the integrated use of book-ahead ride-hailing services and railways in our study area. Interestingly, after about 20 minutes, during the morning peak (see Fig.9 a), the hourly number of book-head ride-hailing trips per OD pair has an inverted V-shaped increase for unknown reasons.

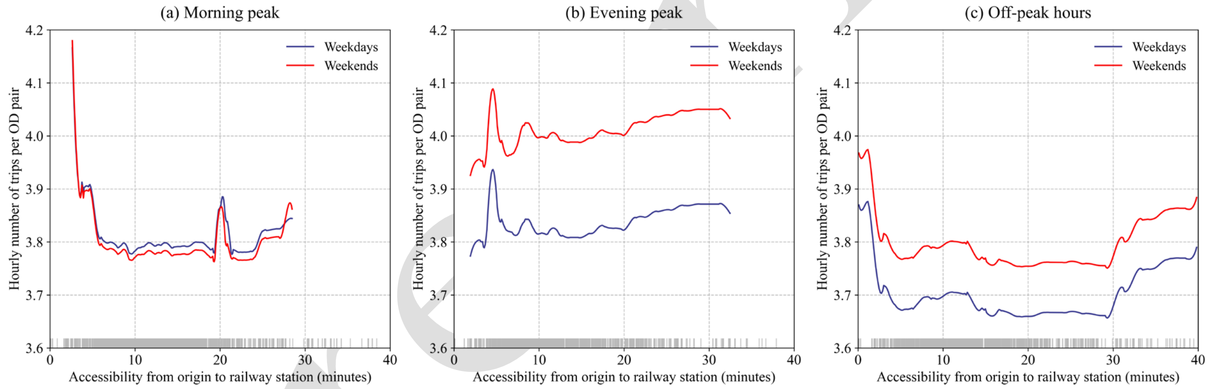

Fig. 9. Partial dependence plot for accessibility to the railway station.

Compared with the accessibility to the airport and railway station, our analysis reveals that the accessibility to the nearest bus stop exhibits distinct effects on book-ahead ride-hailing demand during peak and off-peak hours (see Fig.10 a, b, and c), with larger marginal effects observed on weekends than on weekdays during the evening peak and off-peak hours. Specifically, during morning and evening peak hours, the curves depicting the hourly number of book-ahead ride-hailing trips per OD pair grow gradually within the interval of zero and 20 minutes, after which the curves sharply increase, then drop, and eventually reach a steady state before the accessibility to the nearest bus stop reaches 30 minutes. In contrast, during off-peak hours, a significant surge in ride-hailing trips is observed once the accessibility to the nearest bus stop exceeds 30 minutes. These findings suggest that when the origin is located near a bus stop, the volume of book-ahead ride-hailing trips in the vicinity of bus stops tends to be lower. This could be attributed to the possibility that travelers may opt for alternative bus routes or rely on private cars. Additionally, travelers are also more likely to choose walking or biking to reach transit stations directly. Conversely, when the distance between the origin and the bus stop is too far, travelers may switch to other modes, such as book-ahead ride-hailing or taxis, to connect with the bus. Compared to taxi services, book-ahead ride-hailing is more attractive as it can reduce passengers' waiting time between requesting a trip and boarding a vehicle. Moreover, our analysis reveals that the acceptable feeding time from the origin to the nearest bus stop is longer during morning and evening peak hours (20 minutes) than during off-peak hours (30 minutes). This observation can be attributed to the fact that necessary trips are more common during peak hours, and travelers may be in a hurry to reach their destinations. In contrast, during off-peak hours, these trips are optional, and travelers can arrive at the bus stop more relaxedly.

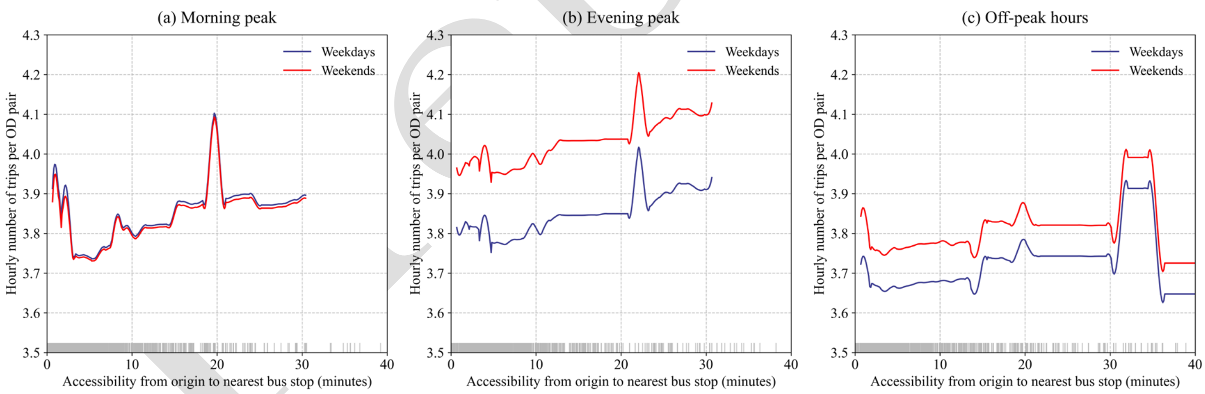

Fig. 10. Partial dependence plot for accessibility to the nearest bus stop

5. Policy discussion and conclusions

5.1. Practice insights

The empirical findings of this study provide not only operational strategies for TNCs to extend ride-hailing services but also practical measures to formulate intermodal friendly land-use policies to encourage book-ahead ride-hailing services in the future.

Preprint

We recommend that promoting book-ahead ride-hailing services should begin with a focus on several pilot programs and then gradually expand to attract passengers and encourage continuous usage. Specifically, the target audiences for pilot programs could be long-distance commuters without access to a private car, as they have shown heavy adoption of book-ahead ride-hailing during peak hours from our analysis. To serve these audiences, TNCs could launch an on-demand commuting services package based on book-ahead ride-hailing that connects their residence and workplace. In addition, given the emergence of book-ahead ride-hailing demand during uncomfortable weather conditions such as high temperatures and turbulent wind, it is suggested that book-ahead ride-hailing services could be recommended to citizens in advance, coupled with local future weather conditions. The ride-hailing platforms could optimize scheduling and improve the efficiency of ride-hailing vehicle dispatch with passengers' book-ahead trip information received ahead of time. Moreover, preferential pricing could incentivize travelers to remain loyal to these services.

From a multimodal intercity travel perspective, our study has revealed a substantial increase in the demand for book-ahead ride-hailing services within the vicinity of airports and railway stations. This finding provides empirical evidence in support of advancing modal transfer between different modes, such as using book-ahead ride-hailing as a feeder mode to the airport/railway station. To alleviate potential traffic congestion around transportation hubs, transportation planners could set up dedicated pick-up/drop-off zones for book-ahead ride-hailing services at these locations. Besides, airport and railway agencies can share real-time information about intermodal travel plans with TNCs to enable them to allocate ride-hailing services efficiently. By promoting the integration between book-ahead ride-hailing and airport/rail transit, we not only provide a convenient first or last connection for customers but also curtail the number of empty ride-hailing vehicles on the road, contributing to the reduction in vehicle miles and congestion.

Finally, our study has identified significant threshold effects of accessibility to the nearest bus stop on book-ahead ride-hailing adoption, which evidences the importance of promoting coordinated development between public transit and ride-hailing services and sheds light on

Preprint

setting the operation constraints. In particular, integrating book-ahead ride-hailing with public transit can indeed be an effective strategy with satisfactory accessibility. Our findings indicate that the maximum tolerable threshold for peak-hour usage is a 20-minute walking time, while for off-peak hours, it is around 30 minutes. We suggest that the buffer area around the bus stop between a 20- and 30-minute walking time is likely to attract the highest transit-integrated demand. Accordingly, TNCs may allocate ride-hailing services in advance to this specific zone on weekdays and weekends. Moreover, public transit agencies could explore collaboration with TNCs to increase their commuting ridership during peak hours. If TNCs offer first- and last-mile feeder services for suburban commuters without private cars or those located in a low transit demand area, public transit agencies can cultivate a continuous usage habit of book-ahead ride-hailing services to connect with bus services. Afterward, they can further optimize bus route layouts with book-ahead ride-hailing services instead of solely relying on feeding public transit routes.

5.2. Conclusions

Cruising ride-hailing vehicles on urban streets results in additional empty vehicle miles and fuel costs, detrimental to urban mobility (Gao et al., 2022; Wei et al., 2023). Book-ahead ride-hailing services enabling travelers to pre-book rides before their trips can potentially reduce cruising traffic and waiting times for passengers, thereby improving urban mobility (Yahia et al., 2021). Effective management strategies and policy interventions for such services require understanding the relationship between book-ahead ride-hailing demand and its determinants. This study used a six-month book-ahead ride-hailing trip dataset from Haikou City, China, and incorporated grid-level points of interest (POIs), accessibility to facilities, socio-economic characteristics, and city-level weather conditions to investigate potential associations between book-ahead ride-hailing usage and its determinants at three spatial scales. A GBDT model was used, and relative importance indicators and partial dependence plots were adopted to identify important determinants and their nonlinear threshold effects. Additionally, the study examined the difference in relative importance and nonlinear associations between

Preprint

peak and non-peak hours and weekdays and weekends.

The key findings of this study are as follows. Firstly, the GBDT model performs better at 250-meter hexagons than at 500-meter and 660-meter hexagons. Secondly, trip feature variables, such as travel distance and fee, contribute an average of over 5.79% to the predictive power of the GBDT model. Thirdly, weather condition variables, such as temperature and wind speed, and accessibility to transportation hubs, such as airports, railway stations, and bus stops, are significant determinants that consistently influence book-ahead ride-hailing trips across all periods. Fourthly, the partial dependence plot of the GBDT model shows a nonlinear association and threshold effects between the volume of book-ahead ride-hailing trips per origin-destination (OD) pair and the aforementioned determinants. Finally, the difference in nonlinear effects and thresholds between weekdays and weekends is only evident during the evening peak period rather than at other periods.

5.3. Future work

This study has some limitations that suggest directions for future research. Firstly, the findings of this study highlight the relevance of place-based policies to improve the operation and management of book-ahead ride-hailing services. However, the context-dependent nature of the built environment (Tao et al., 2020a, 2020b) and regional spatial heterogeneity (Wang et al., 2022) means that effective ranges and thresholds may not be directly transferrable to other cities. Therefore, city-specific policies should be formulated based on empirical studies that account for the local context and related factors.

Secondly, this study uses three hexagonal side lengths to test the reliability of the model results. The requests for book-ahead ride-hailing services are aggregated at an hourly time interval of grid partitions. However, the modifiable areal unit problem (MAUP) (Fotheringham and Wong, 1991) and the modifiable temporal unit problem (MTUP) (Cheng and Adepeju, 2014) may still exist. Therefore, other hexagonal side lengths and time intervals should be investigated in future research (Chen et al., 2022; Liu et al., 2023).

Thirdly, due to the difficulty in obtaining personal private data (Wang et al., 2022; Xu et

Preprint

al., 2021), disaggregated variables concerning travelers' socio-demographic attributes have not been considered in this study. Incorporating these variables would enhance the reliability and credibility of the results. However, ride-hailing companies will only have access to such data, if they conduct additional surveys on the users.

Fourthly, this study focuses on the impact of given trip fees on the ride-hailing demand, and the feedback effects of the trip demand on the trip fee are not explicitly considered. In future studies, developing an overall optimization framework that could optimize the fee with given criteria like maximizing profit or socio-economic benefit would be interesting. In the optimization framework, this study would be a sub-component of a fixed-point problem with price setting at the upper level and demand modeling at the lower level.

In addition, more advanced explainable methods, like the ALE (accumulated local effects) plot (Wang et al., 2023) and SHAP (SHapley Additive exPlanations) method (Cai et al., 2023), could be compared with the partial dependence plot to interpret the results of machine learning models comprehensively. Finally, a comparative study of the differences in nonlinear associations between real-time and book-ahead trips is warranted.

Preprint

Appendix.

Table A1. Book-ahead ride-hailing trip data sample (May 1, 2017).

Order ID	Order Type	Origin Longitude	Origins Latitude	Destination Longitude	Destination Latitude	Departure Time	Distance (m)	Price (RMB)	Time (min)
17592364194458	reserve	110.3196	20.0165	110.4634	19.9369	2017-05-01 04:10:00	23656	79	34
17592340856943	reserve	110.3508	19.9899	110.4621	19.9389	2017-05-01 05:00:00	16517	46	21
17592365824167	reserve	110.4385	19.9653	110.4634	19.9369	2017-05-01 05:10:00	6258	15	9
17592366330089	reserve	110.3517	19.9901	110.4634	19.9369	2017-05-01 05:40:00	17140	47	22
17592357000948	reserve	110.2841	19.9969	110.4621	19.9389	2017-05-01 05:50:00	24587	72	34
17592361539842	reserve	110.3701	20.0585	110.2566	20.0185	2017-05-01 07:00:00	13972	39	22
17592366368380	reserve	110.3747	19.9954	110.3433	19.9841	2017-05-01 07:10:00	5564	15	7
17592353538714	reserve	110.3218	19.9663	110.2900	20.0125	2017-05-01 09:40:00	9911	50	18
17592370335494	reserve	110.4658	19.9405	110.2600	20.0111	2017-05-01 12:10:00	28408	81	40
17592366439288	reserve	110.3092	19.9946	110.3306	19.9952	2017-05-01 13:00:00	2576	9	5
17592372305548	reserve	110.343	19.9712	110.3216	19.9119	2017-05-01 13:20:00	11030	50	31
17592348603375	reserve	110.3433	19.9841	110.3039	20.0224	2017-05-01 13:40:00	9022	23	27
17592378836388	reserve	110.2965	20.0138	110.3433	19.9841	2017-05-01 18:00:00	10301	27	16
17592381792376	reserve	110.3543	19.9934	110.3239	20.0336	2017-05-01 18:50:00	7597	21	30

Preprint

References

Asamer, J., Reinthaler, M., Ruthmair, M., Straub, M., Puchinger, J., 2016. Optimizing charging station locations for urban taxi providers. Transp. Res. Part Policy Pract. 85, 233–246. https://doi.org/10/gfpnqp

Azimi, G., Rahimi, A., Asgari, H., Jin, X., 2020. Role of Attitudes in Transit and Auto Users' Mode Choice of Ridesourcing. Transp. Res. Rec. J. Transp. Res. Board 2674, 1–16. https://doi.org/10/gm8kh3

Bertsimas, D., Jaillet, P., Martin, S., 2019. Online Vehicle Routing: The Edge of Optimization in Large-Scale Applications. Oper. Res. 67, 143–162. https://doi.org/10/gfthdr

Bi, H., Ye, Z., Wang, C., Chen, E., Li, Y., Shao, X., 2020. How Built Environment Impacts Online Car-Hailing Ridership. Transp. Res. Rec. J. Transp. Res. Board 2674, 745–760. https://doi.org/10/gpb3xc

Brown, A.E., 2020. Who and where rideshares? Rideshare travel and use in Los Angeles. Transp. Res. Part Policy Pract. 136, 120–134. https://doi.org/10/gnr5vn

Cai, H., Wu, F., Cheng, Z., Li, B., Wang, J., 2023. A large-scale empirical study on impacting factors of taxi charging station utilization. Transp. Res. Part Transp. Environ. 118, 103687. https://doi.org/10.1016/j.trd.2023.103687

Chen, E., Ye, Z., 2021. Identifying the nonlinear relationship between free-floating bike sharing usage and built environment. J. Clean. Prod. 280, 124281. https://doi.org/10/ghgpzh

Chen, Z., Liu, K., Wang, J., Yamamoto, T., 2022. H-ConvLSTM-based bagging learning approach for ride-hailing demand prediction considering imbalance problems and sparse uncertainty. Transp. Res. Part C Emerg. Technol. 140, 103709. https://doi.org/10.1016/j.trc.2022.103709

Chen, Z., Yin, Y., He, F., Lin, J.L., 2015. Parking Reservation for Managing Downtown Curbside Parking. Transp. Res. Rec. J. Transp. Res. Board 2498, 12–18. https://doi.org/10.3141/2498-02

Cheng Qixiu, Chen Yinghao, Liu Zhiyuan, 2022. A bi-level programming model for the optimal lane reservation problem. Expert Syst. Appl. 189, 116147. https://doi.org/10.1016/j.eswa.2021.116147

Cheng, T., Adepeju, M., 2014. Modifiable Temporal Unit Problem (MTUP) and Its Effect on Space-Time Cluster Detection. PLoS ONE 9, e100465. https://doi.org/10.1371/journal.pone.0100465

China Internet Network Information Center, 2021. The 47th Statistical Report on Internet Development (in Chinese). [Online] Available at: http://www.cac.gov.cn/2021-

Preprint

02/03/c_1613923423079314.htm [Accessed 15 April 2021]

Dean, M.D., Kockelman, K.M., 2021. Spatial variation in shared ride-hail trip demand and factors contributing to sharing: Lessons from Chicago. J. Transp. Geogr. 91, 102944. https://doi.org/10.1016/j.jtrangeo.2020.102944

De Feijter, R., Evers, J.J.M., Lodewijks, G., 2004. Improving Travel-Time Reliability by the Use of Trip Booking. IEEE Trans. Intell. Transp. Syst. 5, 288–292. https://doi.org/10.1109/TITS.2004.837817

Diao, M., Kong, H., Zhao, J., 2021. Impacts of transportation network companies on urban mobility. Nat. Sustain. 4, 494–500. https://doi.org/10.1038/s41893-020-00678-z

Ding, C., Cao, X., Dong, M., Zhang, Y., Yang, J., 2019a. Non-linear relationships between built environment characteristics and electric-bike ownership in Zhongshan, China. Transp. Res. Part Transp. Environ. 75, 286–296. https://doi.org/10/ghtmb7

Ding, C., Cao, X. (Jason), Næss, P., 2018. Applying gradient boosting decision trees to examine non-linear effects of the built environment on driving distance in Oslo. Transp. Res. Part Policy Pract. 110, 107–117. https://doi.org/10.1016/j.tra.2018.02.009

Ding, C., Cao, X., Liu, C., 2019b. How does the station-area built environment influence Metrorail ridership? Using gradient boosting decision trees to identify non-linear thresholds. J. Transp. Geogr. 77, 70–78. https://doi.org/10/ghtmb6

Ewing, R., Cervero, R., 2010. Travel and the Built Environment. J. Am. Plann. Assoc. 76, 265–294. https://doi.org/10/bkffn3

Fotheringham, A.S., Wong, D.W.S., 1991. The Modifiable Areal Unit Problem in Multivariate Statistical Analysis. Environ. Plan. Econ. Space 23, 1025–1044. https://doi.org/10.1068/a231025

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Stat. 29. https://doi.org/10/fbgj35

Galster, G.C., 2018. Nonlinear and Threshold Effects Related to Neighborhood: Implications for Planning and Policy. J. Plan. Lit. 33, 492–508. https://doi.org/10.1177/0885412218793693

Gao, J., Li, S., Yang, H., 2022. Shared parking for ride-sourcing platforms to reduce cruising traffic. Transp. Res. Part C Emerg. Technol. 137, 103562. https://doi.org/10/gpfrgr

Ghaffar, A., Mitra, S., Hyland, M., 2020. Modeling determinants of ridesourcing usage: A census tract-level analysis of Chicago. Transp. Res. Part C Emerg. Technol. 119, 102769. https://doi.org/10.1016/j.trc.2020.102769

He, Z., 2021. Portraying ride-hailing mobility using multi-day trip order data: A case study of

Preprint

Beijing, China. Transp. Res. Part Policy Pract. 146, 152–169. https://doi.org/10.1016/j.tra.2021.01.017

Huang, G., Qiao, S., Yeh, A.G.-O., 2021. Spatiotemporally heterogeneous willingness to ridesplitting and its relationship with the built environment: A case study in Chengdu, China. Transp. Res. Part C Emerg. Technol. 133, 103425. https://doi.org/10.1016/j.trc.2021.103425

Jin, G., Cui, Y., Zeng, L., Tang, H., Feng, Y., Huang, J., 2020. Urban ride-hailing demand prediction with multiple spatio-temporal information fusion network. Transp. Res. Part C Emerg. Technol. 117, 102665. https://doi.org/10.1016/j.trc.2020.102665

Jin, T., Cheng, L., Zhang, X., Cao, J., Qian, X., Witlox, F., 2022. Nonlinear effects of the built environment on metro-integrated ridesourcing usage. Transp. Res. Part Transp. Environ. 110, 103426. https://doi.org/10.1016/j.trd.2022.103426

Joy, T.T., Rana, S., Gupta, S., Venkatesh, S., 2020. Fast hyperparameter tuning using Bayesian optimization with directional derivatives. Knowl.-Based Syst. 205, 106247. https://doi.org/10.1016/j.knosys.2020.106247

Ke, J., Yang, H., Zheng, H., Chen, X., Jia, Y., Gong, P., Ye, J., 2019. Hexagon-Based Convolutional Neural Network for Supply-Demand Forecasting of Ride-Sourcing Services. IEEE Trans. Intell. Transp. Syst. 20, 4160–4173. https://doi.org/10/gpb3xm

Lamotte Raphaël, de Palma André, Geroliminis Nikolas, 2017. On the use of reservation-based autonomous vehicles for demand management. Transp. Res. Part B Methodol. 99, 205–227. https://doi.org/10.1016/j.trb.2017.01.003

Lavieri, P.S., Dias, F.F., Juri, N.R., Kuhr, J., Bhat, C.R., 2018. A Model of Ridesourcing Demand Generation and Distribution. Transp. Res. Rec. J. Transp. Res. Board 2672, 31–40. https://doi.org/10.1177/0361198118756628

Li, A., Zhao, P., Huang, Y., Gao, K., Axhausen, K.W., 2020. An empirical analysis of dockless bike-sharing utilization and its explanatory factors: Case study from Shanghai, China. J. Transp. Geogr. 88, 102828.

Li, W., Pu, Z., Li, Y., (Jeff) Ban, X., 2019. Characterization of ridesplitting based on observed data: A case study of Chengdu, China. Transp. Res. Part C Emerg. Technol. 100, 330–353. https://doi.org/10.1016/j.trc.2019.01.030

Liao, Y., 2021. Ride-sourcing compared to its public-transit alternative using big trip data. J. Transp. Geogr. 95, 103135. https://doi.org/10.1016/j.jtrangeo.2021.103135

Liu, K., Chen, Z., Yamamoto, T., Tuo, L., 2023. Exploring the Impact of Spatiotemporal Granularity on the Demand Prediction of Dynamic Ride-Hailing. IEEE Trans. Intell.

Preprint

Transp. Syst. 24, 104–114. https://doi.org/10.1109/TITS.2022.3216016

Liu, K., Qiu, P., Gao, S., Lu, F., Jiang, J., Yin, L., 2020. Investigating urban metro stations as cognitive places in cities using points of interest. Cities 97, 102561. https://doi.org/10/gjqzdw

Liu, S., Jiang, H., Chen, Z., 2021. Quantifying the impact of weather on ride-hailing ridership: Evidence from Haikou, China. Travel Behav. Soc. 24, 257–269. https://doi.org/10.1016/j.tbs.2021.04.002

Liu, W., Yang, H., Yin, Y., 2015. Efficiency of a highway use reservation system for morning commute. Transp. Res. Part C Emerg. Technol. 56, 293–308. https://doi.org/10.1016/j.trc.2015.04.015

Liu Wei, Yang Hai, Yin Yafeng, 2014. Expirable parking reservations for managing morning commute with parking space constraints. Transp. Res. Part C Emerg. Technol. 44, 185–201. https://doi.org/10.1016/j.trc.2014.04.002

Liu, Y., Jia, R., Ye, J., Qu, X., 2022. How machine learning informs ride-hailing services: A survey. Commun. Transp. Res. 2, 100075. https://doi.org/10.1016/j.commtr.2022.100075

Ma, J., Li, X., Zhou, F., Hao, W., 2017. Designing optimal autonomous vehicle sharing and reservation systems: A linear programming approach. Transp. Res. Part C Emerg. Technol. 84, 124–141. https://doi.org/10.1016/j.trc.2017.08.022

Marquet, O., 2020. Spatial distribution of ride-hailing trip demand and its association with walkability and neighborhood characteristics. Cities 106, 102926. https://doi.org/10/ghtvfm

Nguyen-Phuoc, D.Q., Su, D.N., Tran, P.T.K., Le, D.-T.T., Johnson, L.W., 2020. Factors influencing customer's loyalty towards ride-hailing taxi services – A case study of Vietnam. Transp. Res. Part Policy Pract. 134, 96–112. https://doi.org/10.1016/j.tra.2020.02.008

NYCTLC, 2019. New York state's congestion surcharge. https://www1.nyc.gov/site/tlc/about/congestion-surcharge.page.

Ouyang, Y., Yang, H., Daganzo, C.F., 2021. Performance of reservation-based carpooling services under detour and waiting time restrictions. Transp. Res. Part B Methodol. 150, 370–385. https://doi.org/10.1016/j.trb.2021.06.007

Pierog, K., 2019. Chicago approves traffic congestion tax on ride-hailing services. https://www.reuters.com/article/us-chicago-ridehailing-tax/chicago-approvestraffic-congestion- tax- on- ride- hailing- services- idUSKBN1Y02BV.

Ravula, P., 2022. Monetary and hassle savings as strategic variables in the ride-sharing market.

Preprint

Res. Transp. Econ. 94, 101184. https://doi.org/10.1016/j.retrec.2022.101184

Rayle, L., Dai, D., Chan, N., Cervero, R., Shaheen, S., 2016. Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. Transp. Policy 45, 168–178. https://doi.org/10/ggxjzz

Schaller, B., 2018. The new automobility: Lyft, Uber and the future of American cities. Technical report, Schaller Consulting, Brooklyn, NY.

Shokoohyar, S., Sobhani, Ahmad, Sobhani, Anae, 2020. Impacts of trip characteristics and weather condition on ride-sourcing network: Evidence from Uber and Lyft. Res. Transp. Econ. 80, 100820. https://doi.org/10.1016/j.retrec.2020.100820

Su Peng, Park Byungkyu Brian, 2015. Auction-based highway reservation system an agent-based simulation study. Transp. Res. Part C Emerg. Technol. 60, 211–226. https://doi.org/10.1016/j.trc.2015.07.018

Tao, T., Wang, J., Cao, X., 2020a. Exploring the non-linear associations between spatial attributes and walking distance to transit. J. Transp. Geogr. 82, 102560. https://doi.org/10/ghtmb9

Tao, T., Wu, X., Cao, J., Fan, Y., Das, K., Ramaswami, A., 2020b. Exploring the Nonlinear Relationship between the Built Environment and Active Travel in the Twin Cities. J. Plan. Educ. Res. 0739456X2091576. https://doi.org/10/ghp5zg

Tarabay, R., Abou-Zeid, M., 2020. Modeling the choice to switch from traditional modes to ridesourcing services for social/recreational trips in Lebanon. Transportation 47, 1733–1763. https://doi.org/10/gm8kbj

Tirachini, A., 2020. Ride-hailing, travel behaviour and sustainable mobility: an international review. Transportation 47, 2011–2047. https://doi.org/10.1007/s11116-019-10070-2

Tirachini, A., Chaniotakis, E., Abouelela, M., Antoniou, C., 2020. The sustainability of shared mobility: Can a platform for shared rides reduce motorized traffic in cities? Transp. Res. Part C Emerg. Technol. 117, 102707. https://doi.org/10.1016/j.trc.2020.102707

Tirachini, A., Gomez-Lobo, A., 2020. Does ride-hailing increase or decrease vehicle kilometers traveled (VKT)? A simulation approach for Santiago de Chile. Int. J. Sustain. Transp. 14, 187–204. https://doi.org/10.1080/15568318.2018.1539146

Tu, M., Li, W., Orfila, O., Li, Y., Gruyer, D., 2021. Exploring nonlinear effects of the built environment on ridesplitting: Evidence from Chengdu. Transp. Res. Part Transp. Environ. 93, 102776. https://doi.org/10/gk7b3m

Wang, H., Yang, H., 2019. Ridesourcing systems: A framework and review. Transp. Res. Part B Methodol. 129, 122–155. https://doi.org/10.1016/j.trb.2019.07.009

Preprint

Wang, J., (Jason) Cao, X., Liu, K., Wang, D., 2023. Exploring the nonlinear effects of built environment characteristics on customized bus service. Transp. Res. Part Transp. Environ. 114, 103523. https://doi.org/10.1016/j.trd.2022.103523
Wang, S., Li, Z., Xie, N., 2022. A reservation and allocation model for shared-parking addressing the uncertainty in drivers' arrival/departure time. Transp. Res. Part C Emerg. Technol. 135, 103484. https://doi.org/10/gnwffw
Wang, S., Noland, R.B., 2021. Variation in ride-hailing trips in Chengdu, China. Transp. Res. Part Transp. Environ. 90, 102596. https://doi.org/10/gpb3xr
Wang, Y., Zhan, Z., Mi, Y., Sobhani, A., Zhou, H., 2022. Nonlinear effects of factors on dockless bike-sharing usage considering grid-based spatiotemporal heterogeneity. Transp. Res. Part Transp. Environ. 104, 103194. https://doi.org/10/gpdnnr
Wei, B., Du, B., Saberi, M., Waller, S.T., Liu, W., 2023. On ride-sourcing services of electric vehicles considering cruising for charging and parking. Transp. Res. Part Transp. Environ. 118, 103716. https://doi.org/10.1016/j.trd.2023.103716
Wei, K., Vaze, V., Jacquillat, A., 2022. Transit Planning Optimization Under Ride-Hailing Competition and Traffic Congestion. Transp. Sci. 56, 725–749. https://doi.org/10.1287/trsc.2021.1068
Wei, S., Feng, S., Ke, J., Yang, H., 2022. Calibration and validation of matching functions for ride-sourcing markets. Commun. Transp. Res. 2, 100058. https://doi.org/10.1016/j.commtr.2022.100058
Xu, Y., Yan, X., Liu, X., Zhao, X., 2021. Identifying key factors associated with ridesplitting adoption rate and modeling their nonlinear relationships. Transp. Res. Part Policy Pract. 144, 170–188. https://doi.org/10.1016/j.tra.2020.12.005
Xu, Z., Chen, Z., Yin, Y., Ye, J., 2021. Equilibrium Analysis of Urban Traffic Networks with Ride-Sourcing Services. Transp. Sci. 55, 1260–1279. https://doi.org/10.1287/trsc.2021.1078
Yahia, C.N., de Veciana, G., Boyles, S.D., Rahal, J.A., Stecklein, M., 2021. Book-ahead & supply management for ridesourcing platforms. Transp. Res. Part C Emerg. Technol. 130, 103266. https://doi.org/10.1016/j.trc.2021.103266
Yang, J., Cao, J., Zhou, Y., 2021. Elaborating non-linear associations and synergies of subway access and land uses with urban vitality in Shenzhen. Transp. Res. Part Policy Pract. 144, 74–88. https://doi.org/10/gjt52g
Yang, L., Ao, Y., Ke, J., Lu, Y., Liang, Y., 2021. To walk or not to walk? Examining non-linear effects of streetscape greenery on walking propensity of older adults. J. Transp. Geogr. 94,
Preprint

103099. https://doi.org/10.1016/j.jtrangeo.2021.103099

Yin, J., Li, N., 2022. Ensemble learning models with a Bayesian optimization algorithm for mineral prospectivity mapping. Ore Geol. Rev. 145, 104916. https://doi.org/10.1016/j.oregeorev.2022.104916

Yin, R., Liu, Z., Zheng, N., 2022. A Simulation-Based Model for Continuous Network Design Problem Using Bayesian Optimization. IEEE Trans. Intell. Transp. Syst. 23, 20352–20367. https://doi.org/10.1109/TITS.2022.3176918

Yu, H., Peng, Z.-R., 2019. Exploring the spatial variation of ridesourcing demand and its relationship to built environment and socioeconomic factors with the geographically weighted Poisson regression. J. Transp. Geogr. 75, 147–163. https://doi.org/10/gnvj9f

Yue, Y., Zhuang, Y., Yeh, A.G.O., Xie, J.-Y., Ma, C.-L., Li, Q.-Q., 2017. Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. Int. J. Geogr. Inf. Sci. 31, 658–675. https://doi.org/10/gdrbsd

Zgheib, N., Abou-Zeid, M., Kaysi, I., 2020. Modeling demand for ridesourcing as feeder for high capacity mass transit systems with an application to the planned Beirut BRT. Transp. Res. Part Policy Pract. 138, 70–91. https://doi.org/10.1016/j.tra.2020.05.019

Zhang, Y., Lin, R., Zhang, H., Peng, Y., 2023. Vibration prediction and analysis of strip rolling mill based on XGBoost and Bayesian optimization. Complex Intell. Syst. 9, 133–145. https://doi.org/10.1007/s40747-022-00795-6

Zhao, N., Liu, Y., Cao, G., Samson, E.L., Zhang, J., 2017. Forecasting China’s GDP at the pixel level using nighttime lights time series and population images. GIScience Remote Sens. 54, 407–425. https://doi.org/10.1080/15481603.2016.1276705

Zhong, S., Gong, Y., Zhou, Z., Cheng, R., Xiao, F., 2021. Active learning for multi-objective optimal road congestion pricing considering negative land use effect. Transp. Res. Part C Emerg. Technol. 125, 103002. https://doi.org/10.1016/j.trc.2021.103002

Preprint