Environmental impacts of ridesplitting considering modal substitution and associations with built environment

Citation for the original published paper (version of record):
Zhang, Z., Gao, K., He, H. et al (2024). Environmental impacts of ridesplitting considering modal substitution and associations with
built environment. Transportation Research Part D: Transport and Environment, 130.
http://dx.doi.org/10.1016/j.trd.2024.104160

(article starts on next page)

Transportation Research Part D 130 (2024) 104160

Contents lists available at ScienceDirect

Transportation Research Part D

journal homepage: www.elsevier.com/locate/trd

Environmental impacts of ridesplitting considering modal substitution and associations with built environment

Zhe Zhang [a,b], Kun Gao [b,*], Hong-Di He [a,*], Shaohua Cui [b], Liyang Hu [c], Qing Yu [d], Zhong-Ren Peng [e,f]

[a] Center for ITS and UAV Applications Research, State Key Laboratory of Ocean Engineering, School of Ocean and Civil Engineering, Shanghai Jiao Tong University, Shanghai, CN-200240, China
[b] Department of Architecture and Civil Engineering, Chalmers University of Technology, Goteburg, SE-412 96, Sweden
[c] School of Transportation, Southeast University, Nanjing, Jiangsu, CN-211189, China
[d] School of Urban Planning & Design, Peking University, Shenzhen, CN-518055, China
[e] International Center for Adaptation Planning and Design, College of Design, Construction and Planning, University of Florida, Gainesville, FL 32611-5706, USA
[f] Healthy Building Research Center, Ajman University, Ajman, United Arab Emirates

ARTICLE INFO

Keywords:
Ridesplitting
Modal substitution
Carbon emission
Built environment
Shared mobility

ABSTRACT

Ridesplitting promises to reduce emissions. Previous studies mainly compared ridesplitting with single rides like taxis, leaving its impact on other modes unclear. This study quantifies the reduction or increase in CO_2 emissions due to ridesplitting from a modal substitution perspective and explore the nonlinear impacts of built environment factors on emission reductions. Considering different urban contexts, we choose Chengdu and Xi'an as representative examples for analysis. The results indicate that most ridesplitting trips lead to an increase in CO_2 emissions when compared to other modes. In Chengdu, a mere 8.92% of ridesplitting trips result in emission reduction, whereas in Xi'an, this figure stands at 4.68%. Emission reduction is predominantly linked to taxi substitution. Moreover, many built environment factors exhibit positive relationship with the increase in emission resulting from ridesplitting, and present nonlinear and threshold effects. The findings offer a framework to assess ridesplitting's environmental benefits, aiding urban planners and policymakers.

1. Introduction

On-demand mobility services such as transportation network companies (also known as ride-sourcing or ride-hailing), matching drivers with passengers via smartphone applications, have become one of the most popular sharing services thanks to their flexibility, convenience and efficiency (Tingstad Jacobsen et al., 2023). In recent years, some argue that the benefits of ride-sourcing services may be outweighed by its harm due to shifting users from more sustainable transportation modes, undermining public transportation and increasing vehicle ownership, vehicle-miles-traveled(VMT), which leads to aggravated traffic congestion and environmental pollution (Taiebat et al., 2022; Tirachini, 2020; Cui et al., 2024). Ride-hailing leads to more than 38% modal shift from public transit in Chengdu, China from a perspective of affordability (Qiao and Yeh, 2023). Referring to comprehensive investigation including Lyft, UberX, LyftLine and UberPool, ride-hailing leads to approximately 83.5% more VMT than would have been driven had ride-hailing not existed (Henao and Marshall, 2019). Compared to 22% in a counterfactual scenario without transportation network companies,

* Corresponding authors.
E-mail addresses: gkun@chalmers.se (K. Gao), hongdihe@sjtu.edu.cn (H.-D. He).

https://doi.org/10.1016/j.trd.2024.104160
Received 7 December 2023; Received in revised form 6 February 2024; Accepted 11 March 2024
Available online 16 March 2024

weekday vehicle hours of delay increased by 62% between 2010 and 2016 (Erhardt et al., 2019). The announcement of the ride-hailing service initially resulted in a decrease in PM2.5 pollution in the month preceding the service launch and during the early months following its introduction. However, in the long term, the increased number of overall trips has contributed to a rise in vehicle pollution (Barnes et al., 2020).

Ridesplitting, as a specific type of ride-hailing service, is expected as a remedy for the above-mentioned downsides by enabling different riders to share a vehicle while traveling along similar routes (Shaheen et al., 2016b). For existing studies on ridesplitting, the impacts of ridesplitting on vehicle kilometers traveled (Chen et al., 2021), vehicle use and purchase willingness (Zheng et al., 2019), cost and time saving (Tu et al., 2019), adoption rate (Xu et al., 2021), travel behavior (Chen et al., 2017) have been investigated. However, only a few studies quantified environmental benefits of ridesplitting based on field data. Li et al. (2021) found that ridesplitting significantly reduced emissions of CO_2, CO, NO_x, and HC by 28.7%, 32.5%, 27.7%, and 31.2%, respectively, based on ride-sourcing dataset in Chengdu, China. 22% of vehicle hours traveled could be reduced due to ridesplitting (Li et al., 2019). Therefore, ridesplitting could generate potential benefits in alleviating traffic congestion and reducing traffic pollution compared to ride-sourcing without sharing. These studies provide reference for understanding the potential of emission reduction from ridesplitting. However, most of these studies evaluate environmental benefits by comparing ridesplitting with single rides (i.e., regular non-ridesplitting service). The role of ridesplitting in the entire transportation system has not been well understood. Apart from single rides, ridesplitting may attract passengers from more sustainable transportation modes, such as bus and metro transit. The various advantages of ridesplitting services, such as affordable fares and high priority, may lead to a shift of public transit passengers towards ridesplitting option (Zhu et al., 2020), resulting in additional emissions due to ridesplitting. In Hangzhou of China, the implementation of ridesplitting reduced road vehicle usage by 52,731 per day compared to single rides, including taxi, Express, and DiDi Private Car. When considering the modal shift from public transportation to ridesplitting, the actual decrease in road vehicle usage was only 3,051 per day (Zheng et al., 2019). Therefore, it is crucial to consider modal shifts among different transport modes when examining the environmental benefits of ridesplitting. This is essential for gaining a precise understanding of how ridesplitting contributes to sustainable transportation systems.

The modal shift of ridesplitting trips refer to the fact that a traveler would use other transport modes (e.g., car or public transit) for the same trips if ridesplitting did not exist. In the realm of ridesplitting, and more broadly, within the framework of ride-hailing services, users' transportation mode choice behavior have been investigated based on different data sources (Shaheen et al., 2016a; Chen et al., 2017; Zheng et al., 2019; Wang et al., 2019; Shi et al., 2021; Rafiq and McNally, 2023; Wali, 2023). The methods generally employed in existing studies are questionnaire-based surveys or data-driven analysis. Questionnaire-based surveys assess substitution rates using survey questions such as "How would you complete this trip when ridesplitting is not available?" (Wang et al., 2019; Zheng et al., 2019; Wali, 2023). In terms of data-driven analysis, machine learning techniques are utilized to examine the correlation between users' willingness to use various ride-hailing services, including ridesplitting, and factors such as demographics and the built environment (Chen et al., 2017; Huang et al., 2021). This approach typically involves data that cover the different modes of ride-hailing used for each trip. However, a limitation of this approach is that, due to the lack of comprehensive survey questionnaire data, it primarily analyzes the factors at an aggregated level and tends to overlook modes beyond ride-hailing, such as public transit. In contrast to relying solely on data-driven analysis, our study merges questionnaire-based surveys with data-driven techniques. This hybrid approach allows for a more thorough examination of the substitution effects of ridesplitting on other transport modes, leading to a more precise and detailed estimation of its impact on carbon emission reduction for each trips. By integrating direct user feedback with empirical data, this study provides a more comprehensive understanding of transportation mode choices, thereby enhancing the accuracy and depth of our findings.

Moreover, exploring the effects of various built environment factors on shared mobility has become a focal point in urban transportation planning and management. It has been reported that built environment factors have a significant influence on ride-sourcing demand (Yu and Peng, 2019) and other travel mode choices (Gao et al., 2023b, 2021a). Especially, several built environment factors, such as distance to the city center, land use diversity and road density, pointedly affect the adoption of ridesplitting (Tu et al., 2021). These will consequently impact the emission reduction from ridesplitting. The understanding of relations between emission reduction from ridesplitting and built environment factors can help reveal why and how to expand the environmental benefits of ridesplitting in different urban contexts.

To address the aforementioned gaps, this study uses a trip-level approach to infer which transport mode is substituted by ridesplitting. The travel mode choice modeling is employed to consider empirical factors significantly influencing the transport modes substituted by ridesplitting (such as travel time and costs of available options at the origin and destination locations), which helps prevent biases caused by distinctions in different spatiotemporal contexts. Based on modal substitution of ridesplitting to other transport modes (i.e., metro, bus and taxi), quantified emission reduction from each ridesplitting trip is estimated. Simultaneously, the spatial heterogeneity in emission reduction from ridesplitting is further investigated by exploring nonlinear relations between built environment factors and reduced emission from ridesplitting. The potential contributors to spatial heterogeneity of emission reduction from ridesplitting are unraveled. The findings could be utilized by urban managers and transportation network companies to expand the environmental benefits of ridesplitting.

The remaining sections are structured as follows. Section 2 describes empirical data used in this study. Section 3 elaborates on the technical details of methods, followed by the results and discussions in Section 4. The concluding remarks are summarized in the last section.

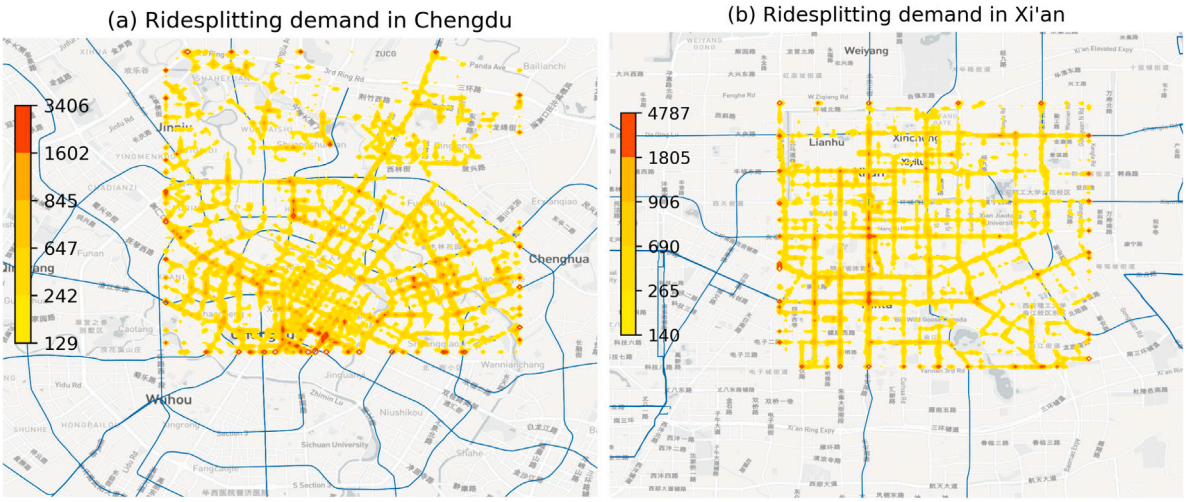

Fig. 1. Spatial distribution of ridesplitting trip origins.

2. Datasets

This study utilizes ride-sourcing transactional data acquired from Didi Chuxing through the GAIA Open Data Initiative (https://gaia.didichuxing.com). The dataset encompasses critical records such as order IDs, diver IDs, and trajectory specifics, including longitude and latitude information of vehicles every 3 s for every ride-sourcing trip. This comprehensive dataset spans a period from October 1, 2018, to November 30, 2018, encompassing both Chengdu and Xi'an. Chengdu, situated as a major metropolis in the southwest, and Xi'an, recognized as the largest city in the northwest, represent two of China's most significant ride-sourcing markets. However, considerable demographic distinctions exist between these cities, encompassing factors such as population, income distribution, and cultural dynamics. For instance, while Chengdu registered a population of 21.268 million in 2022, Xi'an's population stood at 12.996 million, approximately half of Chengdu's. Moreover, the differences in built environments and urban contexts further contribute to the distinctiveness between these two cities. These disparities, encompassing structural, demographic, and cultural facets, likely lead to divergent usage patterns in ridesplitting services observed across Chengdu and Xi'an.

Recognizing a distinct weekly periodicity, we specifically select a standard seven-day week (from November 3–9, 2018) for our analysis. The dataset, on average, comprises over 220,000 daily trip records in Chengdu, serviced by more than 40,000 drivers. In contrast, the average daily count in Xi'an includes around 140,000 trip records facilitated by over 20,000 drivers. The dataset encompasses both single rides and shared rides, with the latter denoting scenarios where riders jointly share their trip, thus dividing the incurred costs among them (Zhang et al., 2023). Shared rides are also referred to as 'ridesplitting trips' within this investigation. In order to identify ridesplitting trips, we adopt a criterion established by Li et al. (2019), which emphasizes temporal overlap between two rides serviced by the same driver. In this study, the ride-sourcing dataset is segmented according to driver ID, creating multiple sub-datasets where each encompasses trips conducted by a same driver. For each sub-dataset, we conduct a comparison of the departure and arrival times across all trip pairs. When there is a temporal overlap between any pair of trips, it indicates the presence of ridesplitting trips. Using this process, we obtain ridesplitting trips that account for approximately 7% of all raw ride-sourcing orders. However, it is essential to note a potential scenario where drivers accept new orders before completing the previous one, resulting in temporal overlap between two single rides without actual shared ridership. We exclude ridesplitting trips featuring temporal overlaps of less than 3 min, which are regarded as single rides. After this, the ridesplitting rides account for about 3% compared to raw data. Additionally, our focus in this research lies specifically on ridesplitting trips involving two riders sharing the same vehicle. We remove the ridesplitting trips involving more than two riders. We also filter out outliers, specifically ridesplitting trips with durations under 3 min or exceeding 1 h, which are considered implausible. The potentially duplicated records are eliminated as well as. After cleaning and selecting, the dataset utilized in this study comprises 65,622 ridesplitting trips in Chengdu and 49,926 ridesplitting trips in Xi'an.

This study spans an area of 8.5×8.5 km^2 in Chengdu, encompassing famous business clusters like Tianfu Square and Chunxi Road. The study area in Xi'an spans 8.5×9.0 km^2 and includes renowned business clusters such as Xi'an Tower and Xiaozhai. Fig. 1 Illustrates the spatial distributions of ridesplitting trips in Chengdu (Fig. 1(a)) and Xi'an (Fig. 1(b)). In Chengdu, a noticeable disparity in origins of ridesplitting trips is observed between the southern and northern parts of the study area. Additionally, there is a discernible surge in ridesplitting demand around business clusters like Xi'an Tower and Xiaozhai, surpassing the demand observed in other areas.

3. Methods

This section delineates the methodologies employed to assess CO_2 reduction through ridesplitting and investigate the nonlinear effects of the built environment on this reduction. Initially, we deduce the transport mode substituted by ridesplitting for each trip, operating under the assumption that passengers would have opted for the inferred substituted mode in the absence of ridesplitting. Subsequently, a comparative analysis is conducted between the CO_2 emissions generated from ridesplitting trips and those associated with the transport mode replaced by ridesplitting. The disparity in CO_2 emissions is then deemed as the emission reduction resultant from ridesplitting. Furthermore, we integrate CO_2 reduction and built environment into each analysis zone (AZ) and explore the nonlinear relationship between them using interpretable machine learning techniques.

3.1. Trip-level inference approach

This study adopts a trip-level approach to ascertain the transport mode substituted by ridesplitting, following methodologies outlined in previous works (Gao et al., 2023a, 2021a). The primary objective of this method is to determine the alternative transportation mode that a user would have opted for in the absence of ridesplitting services. To estimate the substituted mode for each ridesplitting trip, we employ a three-step process. Initially, we extract the specific detailed information such as the trip's origin, destination and departure time from each ridesplitting trip. Subsequently, we use the Application Programming Interface (API) provided by the Baidu Map Open Platform to obtain alternative transport modes and their associated level-of-service attributes for each trip. This step involves estimating crucial factors like travel time and cost, taking into account the accessibility of various transport modes, realistic network and traffic conditions. Finally, we apply a travel mode choice model to estimate the transport mode that would have been used in place of ridesplitting for each trip. This model operates under the hypothetical scenario of non-existence of ridesplitting services. We input the level-of-service factors of alternative transportation modes into this model to estimate the preferred modes prior to the emergence of ridesplitting. This approach allows us to gauge the plausible substituted transport mode for ridesplitting trips.

Unlike the conventional use of average substitution rates across all trips, this trip-level approach aims to infer the substituted transport mode for each specific ridesplitting trip, resulting in a more precise estimation. On the one hand, the selection of a particular transport mode for a specific trip is contingent upon the accessibility of transportation options, which tend to vary due to spatial disparities in the origins and destinations of distinct trips. On the other hand, travelers take into account diverse factors, such as travel time, cost, and comfort, when choosing their preferred mode of transport among the available options. Even when trips share similar distances, travel durations, and purposes, there exists notable variability in the likelihood of selecting a specific transport mode. Consequently, inferring the substituted transport mode by ridesplitting at the trip level proves indispensable for practical and analytical purposes.

3.1.1. Extracting alternative transport modes

The Baidu Maps API serves as an expansive toolkit specifically tailored for location-based applications and solutions. Its functionality encompasses robust route planning recommendations that leverage real-world urban context, integrating variables such as road networks, transit routes, timetables, and traffic conditions. For our study, the origin, destination, and departure time details of each recorded ridesplitting trip are employed as input parameters for the API. Due to the API's limitation in handling past departure times, we execute a retrieval process for corresponding route plans concerning all available transport modes. This is initiated when the recorded day of the week and hour of the day coincide with those registered for each ridesplitting trip. For instance, in the case of a ridesplitting trip recorded at 8:00:00 on November 5, 2018 (Monday), the departure time inputted to the API aligns with 8:00:00 of the following Monday following the API query moment. This meticulous approach is employed to ensure logical and precise alignment of the data with the respective trips.

This study encompasses bus, metro, taxi, and walking as potential transport modes that ridesplitting may substitute. Notably, the analysis excludes the consideration of ridesplitting as a substitute for private car usage. This limitation stems from the inherent constraints of our dataset, which does not provide information on whether a traveler has access to a private vehicle. While assumptions could be made about the likelihood of having access to private cars based on ownership statistics, such presumptions may introduce significant biases. Moreover, the travel patterns linked with private car usage, particularly in estimating parking durations, are laden with uncertainties (Xue et al., 2024). These factors are crucial as they profoundly affect the economic costs and consequently influence transportation mode choices. The complexity introduced by these uncertain parking durations escalates the potential for analytical biases, thereby precluding the precise assessment of the impacts related to private car usage. Therefore, this study restricts its focus to publicly accessible transportation options, namely bus, metro, taxi, and walking. It is imperative to note that the distinction between traditional taxis and solo rides in ride-hailing or e-hailing services is not explicitly delineated in this study. Rather, the reference to the "taxi" transport mode encompasses all solo rides across diverse forms. Furthermore, the outputs generated by the API offer route planning details for available transport options, including corresponding attributes such as travel time, monetary cost, and trip distance for each specific trip. The route planning for public transit often entails transfers between bus and metro, which implies that the coupled structure between the two modes can be well considered by API (Gao et al., 2021b). We categorize the routes involving transfers as the mode taking longer travel time.

3.1.2. Estimating transport mode substitution

We adopt Multinomial Logit (MNL) model to estimate the transport mode substituted by a ridesplitting trip, which is widely used in travel choice behavior, including travel mode choice, travel route choice and etc. The travel mode choice modeling in this study is based on two assumptions.

Assumption 1. We assume that travelers choose the transport mode based on the most highest subjective utility, comprehensively considering both time and economic costs. Although other factors such as in-vehicle crowding can also be considered, we use time and cost on account of the available data extracted from map API.

Assumption 2. We assume that the travel mode choice model in Shanghai is adapted for Chengdu and Xi'an. Although the survey dataset was collected from Shanghai, a city differing from Chengdu in our study, both cities are significant metropolitan areas in China, representing regions that are highly developed economically and culturally (Wang and Zhou, 2017). This similarity minimizes the potential bias due to geographical disparity. We assume that the bias resulted by this disparity is considered to be minimal.

For rider j, the utility of selecting transport mode i is determined as

$$U_{ij} = \begin{cases} \sum_z \beta_{ijz} x_{iz} + \tau_i + \varepsilon_{ij}, & i\epsilon\{taxi, walking\} \\ \sum_z \beta_{ijz} x_{iz} + \tau_i + \varepsilon_{ij} + \sigma, & i\epsilon\{metro, bus\} \end{cases} \tag{1}$$

where x_{iz} denotes the zth level-of-service factor of transport mode i, referring to travel time, monetary cost or walking distance in this study. β_{ijz} is the corresponding coefficient of x_{iz}. τ_i is a constant coefficient representing rider j's unobserved preference to transport mode i. ε is the random error term to model the stochastic. In terms of bus and metro, an error component between bus and metro σ is added in the model to consider the nested structure between the two public transit options (Gao et al., 2021b). Here we take buses as an example. When compared to taxi (denoted by k), the probability of choosing the bus as transport mode (denoted by i) is

$$\begin{aligned} Pr(y_{ij} = 1) &= Pr\left(U_{ij} >_{kj}\right) \\ &= Pr\left(\sum_z \beta_{ijz} + \tau_i + \varepsilon_{ij} + \sigma > \sum_z \beta_{kjz} + \tau_k + \varepsilon_{kj} + \sigma\right) \\ &= Pr\left(\sum_z \beta_{ijz} + \tau_i + \sigma - (\sum_z \beta_{kjz} + \tau_k) > \varepsilon_{kj} - \varepsilon_{ij}\right) \end{aligned} \tag{2}$$

$y_{ij} = 1$ indicates that traveler j choose option i for the trip. ε_{ij} is assumed to be identical and independent Gumbel distributions (Ben-Akiva and Lerman, 1985), so that we can get

$$Pr(y_{ij} = 1) = \frac{exp(\theta_i U_{ij}^s)}{\sum_z exp(\theta_i U_{ij}^s)} \tag{3}$$

where

$$U_{ij}^s = \begin{cases} \sum_z \beta_{ijz} + \tau_i, & i\epsilon\{taxi, walking\} \\ \sum_z \beta_{ijz} + \tau_i + \sigma, & i\epsilon\{metro, bus\} \end{cases} \tag{4}$$

θ_i is the index to denote if transport mode i is available or not for a specific trip. U_{ij}^s denotes simplified utility function. Table 1 illustrates the parameters of β_{ijz}, τ_i and σ, which were calibrated by Gao et al. (2023a). The transport mode i with the highest probability is determined as the substituted option, namely, the transport mode that the user would choose if there was no ridesplitting.

3.2. The impacts of ridesplitting on CO_2 emission

This study defines the discrepancy in CO_2 emissions between ridesplitting and the corresponding substituted transport mode for identical trips as the CO_2 emission reduction attributed to ridesplitting trips. Our methodology involves distinct calculations of emissions for both the ridesplitting and substituted modes concerning the same trip. Subsequently, a comparative analysis is conducted to ascertain the escalation or reduction in CO_2 emissions. This comparison is made under the assumption that passengers would select the substituted mode in the absence of ridesplitting services. It should be noted that the "emission" in the following contents refers to CO_2 emission.

3.2.1. CO_2 Emission from a ridesplitting trip

Fig. 2 shows that an example of a ridesplitting trip consists of overlapping segments where two different passengers are in the same vehicle (i.e., sharing the vehicle) and solo segments where only one passenger is in the vehicle. Emission from overlapping segments is equally divided into the trip emission of each passenger. For each ridesplitting trip, the emission from the trip of one passenger is presented by

$$E_{ridesplitting} = E_{solo} + \frac{E_{overlapping}}{2}, \tag{5}$$

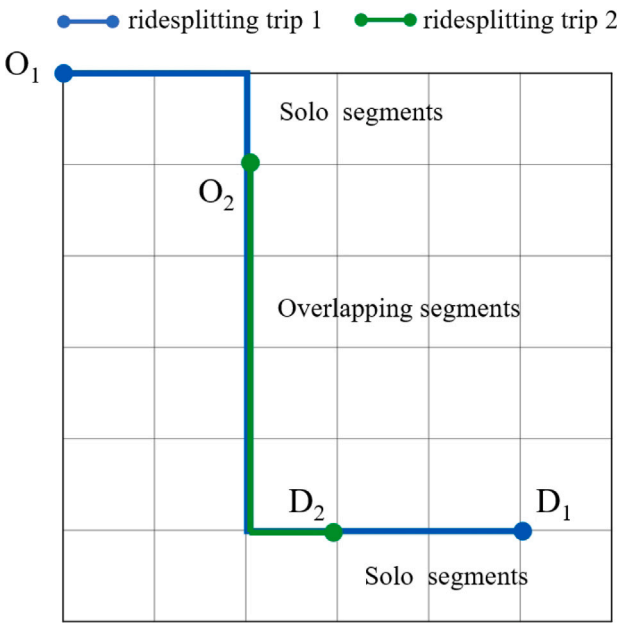

Fig. 2. Overlapping and solo segments in a ridesplitting trip.

Table 1
Parameters in MNL model (Gao et al., 2023a).

Parameters	Values
β_{cost}	−0.0355
$\beta(traveltime_{bus})$	−0.0307
$\beta_{(traveltime_{metro})}$	−0.0279
$\beta_{(traveltime_{taxi})}$	−0.017
$\beta_{(traveltime_{walking})}$	−0.0614
σ	1.39
τ_{bus}	0
τ_{metro}	1.38
τ_{taxi}	−1.29
$\tau_{walking}$	−0.3

where E_{solo} and $E_{overlapping}$ are the emissions from solo segments and overlapping segments, respectively.

We estimate E_{solo} and $E_{overlapping}$ using COPERT model. COPERT model, initially developed in accordance with European emission standards, has gained widespread application in estimating vehicular emissions across different countries. Recently, it has been adapted to calculating emissions from ride-sourcing vehicles in Chinese contexts (Yan et al., 2020; Li et al., 2021; Zhang et al., 2023; Ding et al., 2024; Li et al., 2024). This adaptation is especially pertinent considering that the majority of ride-sourcing vehicles in China comply with the China IV emission standard, which is equivalent to the European 4 standard (Ding et al., 2024). Therefore, in our research, we employ the COPERT model in accordance with the European 4 standard, ensuring a relevant and accurate assessment of emissions. Given that ridesplitting trips consist of multiple short-distance segments, we initially compute the emissions for each segment using the COPERT model. The cumulative emissions from all segments are then considered to represent the total emissions for a single trip. The emission of the segment j is calculated by

$$E_j = \sum EF_{j,j+1} \times d_{j,j+1}, \tag{6}$$

where $d_{j,j+1}$ (unit: km) denotes the distance between GPS points j and j+1. $EF_{j,j+1}$ (unit: g/km) is the emission factor and determined by

$$EF_{j,j+1} = \frac{\alpha \times v_{j,j+1}^2 + \beta \times v_{j,j+1} + \gamma + \frac{\delta}{v_{j,j+1}}}{\varepsilon \times v_{j,j+1}^2 + \zeta \times v_{j,j+1} + \eta}, \tag{7}$$

where $v_{j,j+1}$ denotes the average speed between GPS points j and $j+1$. The parameters $\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta$ are obtained based on the emission standards of vehicles and the type of pollutants. For CO_2 emissions from ride-sourcing vehicles, the specific parameters are as follows: $\alpha = 3.32 \times e^{-1}, \beta = -17.6, \gamma = 1450, \delta = 1.76 \times e^{-11}, \varepsilon = 8.01 \times e^{-4}, \zeta = 9.13 \times e^{-2}, \eta = 3.51$, referring to Li et al. (2021). Our dataset for this study includes GPS trajectories of the vehicle from each ridesplitting trip. By leveraging the timestamps

and coordinates from these trajectories, we can obtain the distance and average speed between two GPS points, of which the time difference is 3 s (Cui et al., 2023). Thus, the CO_2 emissions between two GPS points can be estimated using the COPERT model. The CO_2 emission of a trip can be acquired by aggregating the CO_2 emissions of segments of the trip directly.

3.2.2. CO_2 Emission from other transport modes

Previous studies has explored emission factors using diverse models such as COPERT, MOVES, and HBEFA, utilizing varied activity data like driving records and traffic conditions to compute emission factors for distinct vehicle types. These factors provide a quantitative representation of the volume of pollutants discharged per unit distance (g/km), time (g/s), or fuel consumption (g/kg) (Smit et al., 2010; He et al., 2024). While these models are proficient in estimating emissions from different vehicles, their adequacy for estimating public transport emissions remains uncertain. Estimations of bus and metros emissions stem from amalgamated results from multiple passengers, contrasting with individual passenger emissions. Consequently, this study adjusts public transport emissions accounting for vehicle occupancy (e.g., the number of passengers). Several studies have developed emission factors for buses and metros, taking into account the number of passengers in the context of Chinese urban contexts (Liu et al., 2023; Wang et al., 2015; Yu et al., 2020). Specifically, this study employs emission factors denoted per passenger per kilometer (passenger-kilometer, PKM). Notably, the CO_2 emission factors for bus and metro are 28.60 g/PKM and 48.85 g/PKM, respectively (Wang et al., 2015).

Regarding taxis, this study exclusively accounts for taxis transporting a single passenger as the transport mode substituted by ridesplitting. Thus, PKM is also appropriate for calculating CO_2 emission from taxi. The CO_2 emission factors for taxi taking one passenger is 165.98 g/PKM (Wang et al., 2015). The total CO_2 emission of a trip is determined as

$$E_i = \sum EF_i \times d_{i,m} \tag{8}$$

where $d_{i,m}$ denotes the distance traveled by passenger i using a particular transport mode m. EF_m denotes CO_2 emission factor for transport mode m. If a trip is a combination of bus and metro, the total emissions are calculated by the sum of CO_2 emissions generated by the sub-trips of using the bus and the metro.

Finally, the CO_2 emission reduction resulting from a ridesplitting trip refers to the difference between emission from the ridesplitting trip and the emissions of using the inferred mode for the same trip.

$$ED = E_{substituted} - E_{ridesplitting}, \tag{9}$$

where ED can be either positive or negative. When $ED > 0$ indicates that ridesplitting reduces CO_2 emissions in contrast to the inferred substituted mode. Whereas $ED < 0$, a ridesplitting trip increases CO_2 emissions compared to the scenario without ridesplitting and using the inferred transport mode for the same trip.

3.3. The potential effects of built environment on emission reduction

The results of substitutions to other transport modes are crucial in calculating emission reductions due to ridesplitting. We investigate the effects of transport-related built environment factors considering two aspects. First, the modal substitution of a ridesplitting trip to other transport modes is associated with built environment factors. The transport mode substituted by ridesplitting for each trip is influenced, either directly or indirectly, by various urban built environment factors, such as the accessibility of public transport stations and road density. Secondly, at aggregated level, the total number of ridesplitting trips is influenced by the built environment, with geographical variations leading to diverse travel demands. The demand of using ridesplitting trips in an area is related to the aggregated emission effects of ridesplitting.

Utilizing the available dataset and the methodologies outlined earlier, we embark on quantifying the CO_2 emission reduction attributed to individual ridesplitting trips, followed by aggregating these outcomes across diverse urban contexts. We divide the study area into grids with 500 × 500 m, and each grid is an AZ. The aggregated CO_2 emission reductions from trips are allocated to respective grids by leveraging trip origins, typically reflective of demand patterns. Making the best of the results, we further unravel the influencing factors on spatial variation in CO_2 emission reduction due to ridesplitting in different urban contexts. To unravel these influences, we assess built environment factors pivotal in shaping travel mode preferences and travel demand, consequently impacting the resultant CO_2 emission reduction through ridesplitting. Employing machine learning techniques, we model and interpret the nonlinear impacts of these factors on CO_2 emission reduction stemming from ridesplitting.

3.3.1. Built environment factors

The investigated built environment factors encompass density, diversity, design, destination accessibility, and distance to transit, commonly known as the 5-D built environment factors (Cervero et al., 2009; Ewing and Cervero, 2010; Gao et al., 2021b). All these factors are computed for each AZs, as illustrated in Table 2.

(1) Density. Population density is one of the factors considered. The population dataset is sourced from the Resource and Environment Science Data Center of the Chinese Academy of Sciences (https://www.resdc.cn/DOI/). It constitutes a raster dataset presenting population density with a resolution of 1 km2, integrating diverse elements such as land use types, nighttime light intensity, residential density, and other pertinent demographic factors.

(2) Diversity. Diversity factors encompass POI (Point of Interest) mixture and land use ratios, comprising residential land use ratio, public service land use ratio, commercial land use ratio, industrial land use ratio, park and square land use ratio. These factors

Table 2
Built environment factors used in this study.

5-D	factors	data sources	computation methods
Density	population density	Resource and Environment Science Data Center of the Chinese Academy of Science	provided by dataset
Diversity	POI mixture land use ratios	POI dataset from Amap	POI entropy TF–IDF
Design	road densities	OpenStreetMap	length of road per km^2
Destination accessibility	work facility density leisure facility density education facility density park and square facility density	POI dataset from Amap	number of POIs per km^2
Distance to transit	ARBus ARMetro ARParking	POI dataset from Amap	The proportional overlap between transit stations' influence areas and respective AZs

Table 3
Reclassified land use categories and corresponding POIs in Amap.

Land use categories	Code	Big-category in Amap	Mid-category in Amap
Residential land use	1	Commercial House	Residential Area
Public services land use	2	Governmental Organization & Social Group	all mid-category (for example, Governmental organization, Social Group, etc.)
	3	Medical Service	all mid-category (for example, hospital, pharmacy, etc.)
	4	Science/Culture & Education Service	all mid-category (for example, museum, school, etc.)
Commercial land use	5	Commercial House	Industrial Park, Building, Commercial House Related
	6	Daily Life Service	all mid-category (for example, ticket office, laundry, etc.)
	7	Finance & Insurance Service	all mid-category (for example, bank, insurance company, etc.)
	8	Food & Beverages	all mid-category (for example, Chinese food restaurant, coffee house, etc.)
	9	Shopping	all mid-category (for example, supermarket, clothing store, etc.)
	10	Accommodation Service	all mid-category (for example, hotel, hostel, etc.)
	11	Sports & Recreation	all mid-category (for example, sports stadium, recreation center, etc.)
	12	Enterprises	Company, Enterprises, Famous Enterprise
Industrial land use	13	Enterprises	Factory
Transport service land use	14	Transportation Service	all mid-category (for example, railway station, parking lot, etc.)
Park and square land use	15	Tourist Attraction	Park & Square

are derived from the POI dataset obtained from Amap (https://lbs.amap.com/), comprising 23 big-categories and over 200 mid-categories that delineate POI types. We utilize two approaches to compute these factors. Firstly, we employ selected big-categories to assess POI mixture using entropy within each AZ. Concurrently, we reclassify POIs to determine land use ratios. Following the guidelines of the Code for Classification of Urban Land Use and Planning Standards of Development Land (GB 50137-2011), we reclassify these POIs into six distinct land use categories: residential land use, public services land use, commercial land use, industrial land use, park and square land use ratio, and transport service land use, representing the primary functions of an urban

area (Chen et al., 2022). The relationship between land use categories and the respective POI types from the Amap dataset is outlined in Table 3. Subsequently, land use ratios are computed employing the Term Frequency–Inverse Document Frequency (TF–IDF) methodology, as detailed in Gao et al. (2021b).

(3) Design. The design factors encompass various road densities, specifically the density of expressways, major arterials, minor arterials, and branches. These datasets are sourced from OpenStreetMap, which comprehensively catalogs four primary urban road types in China, namely expressways, major arterials, minor arterials, and branches (Yan et al., 2019).

(4) Destination accessibility. Destination accessibility is evaluated through several metrics including work facility density, leisure facility density, education facility density, and park and square facility density. Work and leisure facility densities are quantified by aggregating the number of corresponding POIs identified by codes within the set and as outlined in Table 3. Education facility density is calculated based on the 'school' mid-category POIs obtained from Amap, while park and square facility density relies on the 'park and square' mid-category POIs from Amap.

(5) Distance to transit. To assess transit accessibility comprehensively, we consider indicators reflecting the coverage and accessibility of transportation facilities. These metrics, specifically ARBus, ARMetro, and ARParking, are designed to provide insights into transport convenience within each Analysis Zone (AZ). While previous studies commonly utilize the count of transit stations within AZs, a drawback lies in the oversight of transit station influence areas. To mitigate this, we introduce ARBus and ARMetro, adapted from Li et al. (2020), which evaluate the proportional overlap between transit stations' influence areas and respective AZs. The influence area for each transit station is defined as a buffer zone with a radius determined as the mean of the distances between adjacent transport stations (Li et al., 2020). For this study, a radius of x=83 meters and x=438 meters is adopted for bus and metro stations in Chengdu, while x=75 and x=603 are used for bus and metro stations in Xi'an, respectively. Expanding on this concept, we introduce ARParking, representing the proportion of overlap between parking lot influence areas and corresponding AZs. The influence area for parking lots is determined by a radius of x=37 and x=45 for Chengdu and Xi'an, respectively, based on field dataset measurements.

3.3.2. Interpretation of nonlinear effects

We initiate our analysis by employing machine learning (ML) algorithms to establish connections between CO_2 reduction and the built environment. In contrast to traditional methods like Least Squares Regression and Geographically Weighted Regression, ML algorithms offer an advantage in unveiling intricate nonlinear effects (Gao et al., 2023a; Liu et al., 2022). Among these algorithms, XGBoost stands out as a highly efficient and versatile model, renowned for its robustness in handling diverse data types and tasks. Its widespread utilization in real-world applications owes to its superior performance and ability to estimate feature importance accurately. In this study, we leverage XGBoost to elucidate the nonlinear impacts of built environment factors on CO_2 emission reduction resulting from ridesplitting. The performance of the model is evaluated through metrics such as the coefficient of determination (R^2) and adjusted R^2.

Next, we delve into interpreting the nonlinear effects of the built environment using ML interpretation techniques. A critical step in this investigation involves assessing the significance of each feature to understand their relationship with the dependent variable. In XGBoost regression, feature importance serves as a common method for evaluating the significance of features. It relies on the concept of Gain, measuring the average improvement in model loss attributed to individual features during the tree-building process. However, considering a more fundamental perspective, feature importance can be viewed as the reduction in model performance when a feature is permuted. Consequently, permutation-based importance offers a more fundamental evaluation of feature importance. This method quantitatively measures the impact of permuting a feature by examining the reduction in R_2. Yet, permutation-based importance faces challenges in handling correlated features, which is a common limitation in many importance interpretation methods. To mitigate this, we conduct an examination for potential multicollinearity among all built environment factors to prevent any biases. Specifically, we filter influencing factors to ensure that the correlation coefficient between all the built environment factors remains below 0.7. Additionally, explanatory variables with a variance inflation factor (VIF) greater than 5 are omitted from the analysis to maintain the accuracy of results.

Permutation importance offers insights into the average impact on CO_2 emission reduction. Furthermore, we delve deeper into understanding the specific nonlinear relationships between crucial built environment factors and emission reductions derived from ridesplitting. To quantitatively illustrate the nonlinear effect of each factor, we employ partial dependence analysis (PDA) (Friedman, 2001). PDA serves to both visualize and quantify the relationship between an independent variable and the model's predictions. Its primary aim is to reveal the marginal effects of features on the dependent variable within ML estimators. By generating partial dependence plots, we can visually comprehend how the model's predicted outcomes vary as the feature of interest changes, while other features remain constant. This method facilitates an investigation into how predictions are partially dependent on various feature values.

4. Results and discussion

This section provides an exposition of the outcomes concerning the substitution rates of ridesplitting services to various transport modes, the resultant CO_2 emission reduction, and the associations of these outcomes with built environment factors. Simultaneously, comprehensive discussions and implications are furnished alongside the results for a more nuanced understanding.

Fig. 3. Substitution rates of ridesplitting to other modes by distance in Chengdu (a) and Xi'an (b).

4.1. The substitutions of ridesplitting to different transport modes

Utilizing the trip-level inference approach, we estimate the specific transport mode substituted by each ridesplitting trip. In Chengdu, the substitution rates of ridesplitting to walking, bus, metro, and taxi are 5.20%, 41.62%, 22.68%, and 30.50%, respectively. In Xi'an, the corresponding rates are 3.70%, 60.10%, 15.67%, and 20.54%, respectively. Generally, walking exhibits the lowest probability of being substituted, while buses emerge as the most frequently replaced mode, particularly evident in Chengdu, where the substitution rate to buses could exceed 60%. Comparatively, in Xi'an, the substitution rates to buses are higher than in Chengdu, while the substitution rates to metro and taxis in Xi'an are lower than in Chengdu. This observed divergence can potentially be attributed to variations in infrastructure, demographic profiles, and other contextual factors. Our analysis specifically explores these discrepancies from the viewpoints of metro accessibility and the city's average income levels. The higher rates to buses and lower rates to metro in Xi'an could be linked to the relatively lower accessibility of metro services, especially when compared with Chengdu. Consequently, passengers in Xi'an, when considering public transportation options, tend to favor buses, which are more readily available. Furthermore, the subdued preference for taxis in Xi'an might be associated with the city's lower average income relative to Chengdu. It is a reasonable conclusion that passengers with more modest incomes are less likely to opt for the higher-cost taxi services.

Given the established correlation between travel mode choice and trip distance (Gao et al., 2021a), we explore variations in the substitution rates of ridesplitting trips concerning different transport modes with respect to trip distances. Fig. 3 delineates these substitution rates within 500-meter intervals. Notably, for short trips of less than 500 m, the overwhelming majority of ridesplitting trips (over 96%) are in lieu of walking. As trip distance increases, the likelihood of substituting bus, metro, and taxi escalates. However, beyond a certain threshold trip distance, a discernible decline is observed in the substitution rates to buses and taxis. Conversely, the substitution rates to the metro consistently rise with increasing trip distance. Notably, differences exist between Chengdu and Xi'an in this regard. Substitution rates to the metro in Chengdu surpass those in Xi'an, with the probability of ridesplitting trips substituting the metro exceeding 50% for distances beyond 12.5 km in Chengdu. In contrast, in Xi'an, substitution rates to the metro remain below 50% across all trip distances. In summary, ridesplitting is more inclined to replace the metro for longer distances in comparison to other modes, while it exhibits a higher likelihood to replace buses and taxis for mid-range distance trips (e.g., 2–10 km).

Fig. 4 illustrates the spatial patterns of substitution rates of ridesplitting to other transport modes, namely metro, bus, taxi, and walking. The intensity of the darker blue shading within a grid signifies higher probability of ridesplitting trips replacing a specific transport mode in that particular grid. Substantial spatial variations are discernible across different AZs. Fig. 4 (a1) and (b1) elucidate the spatial variation of substitution rates to metro in Chengdu and Xi'an, respectively. The red nodes and black lines in both figures denote metro stations and routes, respectively. Notably, these rates exhibit discernible variations along metro lines. Proximity to a metro station correlates with a higher probability of ridesplitting trips replacing the metro. Fig. 4 (a2) additionally showcases spatial patterns for substitution rates to buses in Chengdu, where variations are observable across spatial locations. The red nodes denote bus stations, and grids with a higher density of bus stations correspond to a larger probability of ridesplitting trips replacing buses. However, a distinctive contrast in substitution rates to buses between the two cities can be observed. In Xi'an (b2), substitution rates to buses are evenly distributed across the AZs. This divergence can be attributed to distinct road network layouts, with Chengdu adopting a ring road network and Xi'an featuring a conventional grid network. Consequently, the spatial distribution of bus stops in Xi'an is uniform, while in Chengdu, it is not. The substitution rates to the two cities exhibit different spatial variations, but it is evident in both two cities that substitution rates to buses are associated with bus stop density and road network.

Figs. 4 (a3) and 4(b3) depict the spatial variability of substitution rates to taxis. An intriguing observation is that areas with lower metro accessibility tend to exhibit higher substitution rates to taxis, which is reasonable as more users tend to use taxis in places without convenient access to metro stations if there was no ridesplitting service. Finally, Fig. 4 (a4) and (b4) reveal that walking consistently registers the lowest substitution rates across all grids.

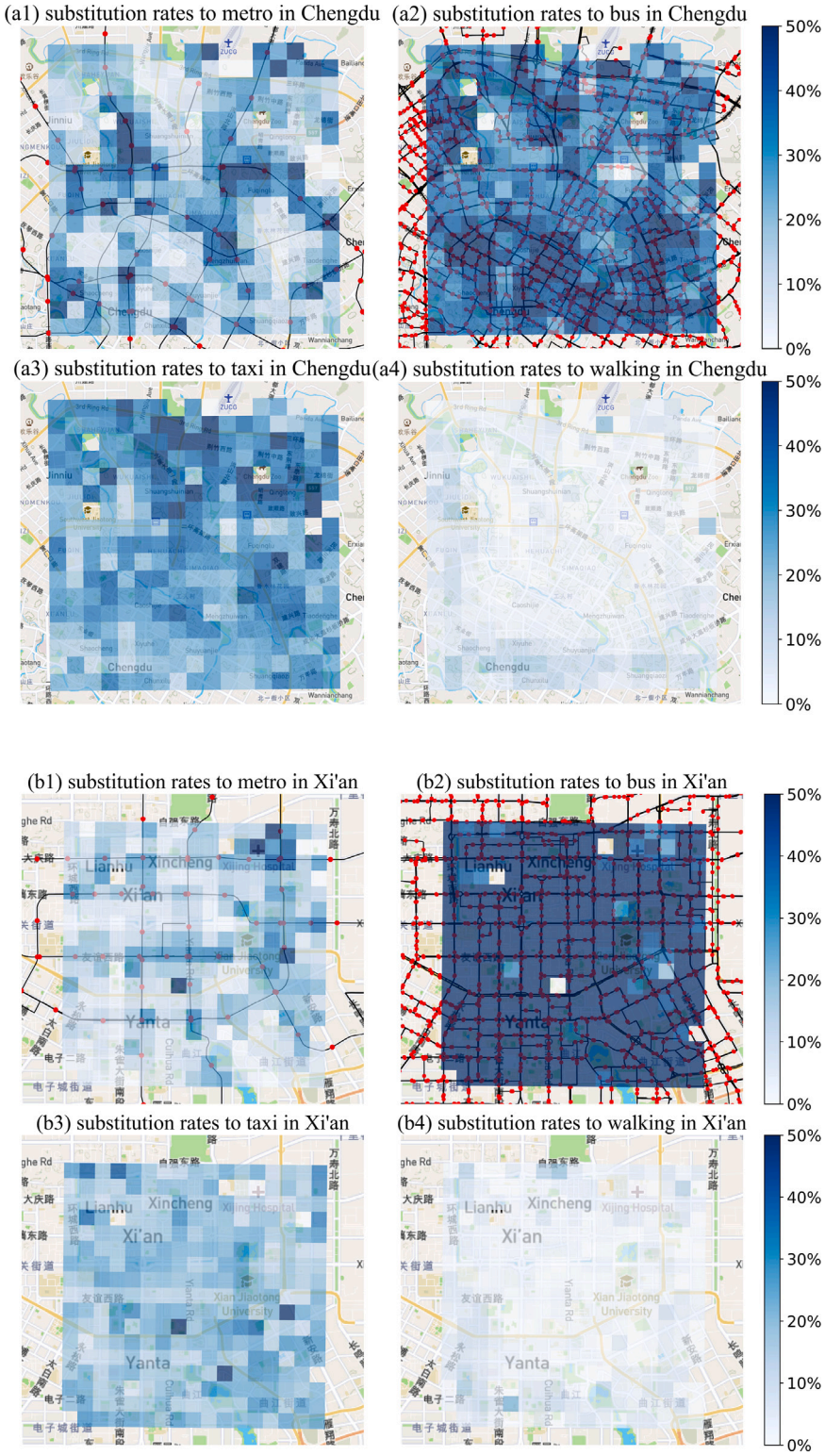

Fig. 4. Substitution rates of ridesplitting to other modes in two cities. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

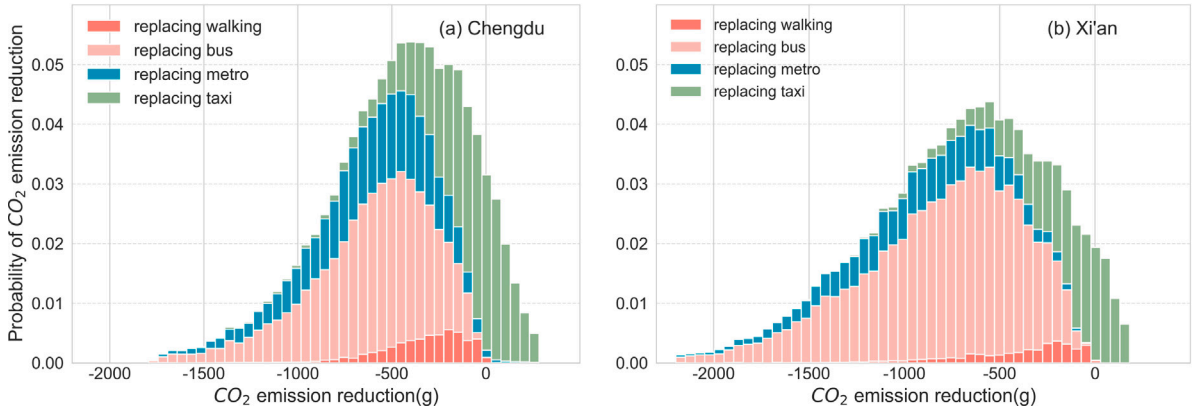

Fig. 5. Emission reduction from ridesplitting trips due to replacing other transport modes.

4.2. The impacts of ridesplitting on CO_2 emission

We conduct a comparative analysis of CO_2 emissions emanating from substituted transportation modes and ridesplitting for identical trips to gauge the emission reduction potential associated with ridesplitting. Fig. 5 elucidates the probability distribution of emission reduction resultant from ridesplitting. The X-axis delineates positive values signifying reduced emissions (in grams) due to ridesplitting, while negative values denote increased emissions attributed to ridesplitting compared to replaced transport modes. Concurrently, the Y-axis represents the probability of emission reduction arising from ridesplitting trips supplanting alternative transportation modes. Within each interval, different color bars denote the likelihood of emission reduction consequent to ridesplitting trips displacing specific transport modes. It is noteworthy that a predominant proportion of ridesplitting trips culminate in an increase of CO_2 emissions. In the Chengdu study area, ridesplitting trips effecting CO_2 emission reduction constitute merely 8.92% of the total trips, whereas in the Xi'an study area, 4.68% of ridesplitting trips yield emission reduction. The preponderance of emission reduction instances emanates from the substitution of ridesplitting to taxis. Quantitatively, among ridesplitting trips in Chengdu resulting in emission reduction, 97.85% substitute taxi, and 2.11% and 0.04% replace metro and bus, respectively. In Xi'an, among ridesplitting trips yielding emission reduction, 99.95% and 0.05% entail taxi and metro substitution, respectively. This underscores that the observed emission reduction in ridesplitting trips primarily stems from the displacement of taxis.

Fig. 6 illustrates the impacts of ridesplitting on CO_2 emissions at both the aggregated grid level and the average trip level in two cities. At both levels, the maximum values of CO_2 emission reduction in analysis zones (AZs) are negative in both cities. This indicates that ridesplitting leads to an overall increase in CO_2 emissions in aggregated AZs, despite some ridesplitting trips indeed producing emission reduction by replacing taxis. The larger the absolute value, the greater the increase in CO_2 emissions. More importantly, noticeable differences are observed in different AZs. Specifically, at the aggregated grid level, AZs closer to the city center in Chengdu experience a more substantial increase in CO_2 emissions due to ridesplitting. In the Xi'an study area, AZs on the western side show a greater increase in CO_2 emissions. AZs with a higher increase in CO_2 emissions due to ridesplitting at the aggregated level may stem from the locations of business clusters, which significantly affect the demand for ridesplitting. The spatial distributions of business clusters align with the AZs with large emission increases in both cities. This implies that a larger demand for ridesplitting trips leads to a higher increase in emissions at the aggregated grid level compared to scenarios without ridesplitting. Conversely, at the average trip level, the spatial patterns of increased CO_2 emissions per trip display a different and even opposing trend compared to the aggregated level results. AZs with higher aggregated increases in CO_2 emissions tend to have lower average increases in CO_2 emissions per trip, which should be ascribed to the higher probability of ridesplitting replacing taxis. AZs with lower average increases in CO_2 emissions per trip are mainly distributed in suburban or rural–urban areas with less convenient public transit service than business areas, which may be the reason for a higher probability of ridesplitting substituting taxis.

4.3. The effects of built environment factors on CO_2 emission reduction

We develop and train discrete XGBoost regression models with the objective of examining the associations between built environment factors and the CO_2 emission impacts of ridesplitting at the aggregated zone level in two distinct urban contexts. The decision to employ distinct models for each city is based on the discernment of substantially distinct outcomes concerning the influences of built environment factors. Both independent and dependent variables have been amalgamated into analysis zones (AZs), where the dependent variable is the reduction in CO_2 emissions (g). Five-fold cross-validation is applied to ascertain optimal hyper-parameters. In Chengdu, the R_2 value for the optimal model is 0.78, signifying commendable model performance and generalization capacity. The adjusted R_2 for the regression model in Chengdu is 0.72, denoting that the selected variables explicate 72% of the variability in R_2 emission reduction across diverse AZs in Chengdu. Conversely, in Xi'an, the R_2 value for the optimal model is 0.50, accompanied by an adjusted R_2 of 0.42. These metrics suggest that the identified variables account for 42% of the variance in CO_2

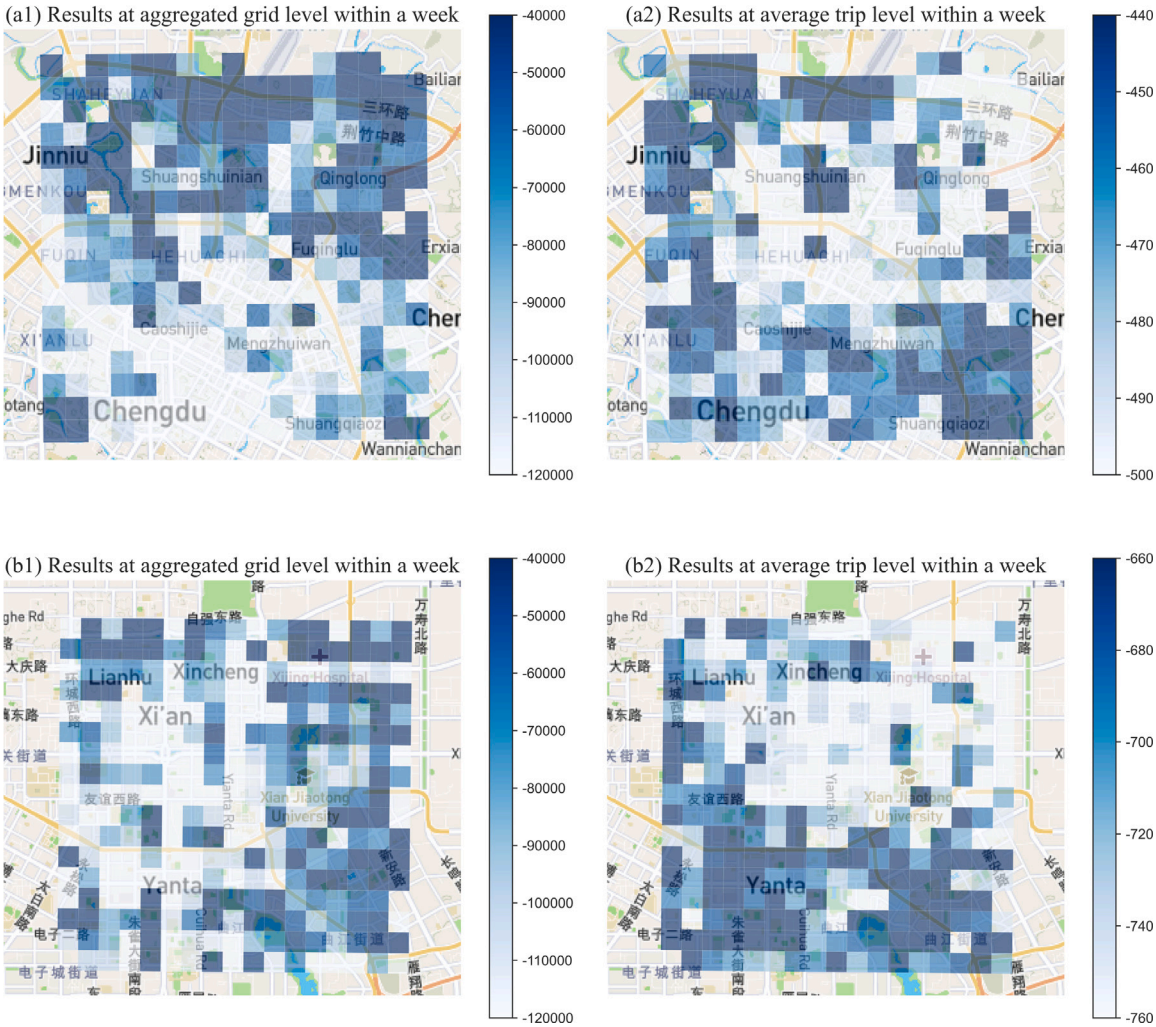

Fig. 6. The impacts of ridesplitting on CO_2 emission (g) in Chengdu (a1–a2) and Xi'an (b1–b2).

emission reduction across discrete AZs in Xi'an. The disparities in model performance between the two cities underscore the unique and context-specific nature of the relationships between built environment factors and the CO_2 emission impacts of ridesplitting.

Fig. 7 illustrates the permutation importance of built environment factors in the cities of Chengdu and Xi'an. The depicted figures highlight discernible distinctions in the built environment factors associated with CO_2 emission reduction between the two cities. In Fig. 7(a), it becomes evident that variables pertaining to destination accessibility and distance to transit exert the most pronounced impact on CO_2 emission reduction resulting from ridesplitting in Chengdu. Subsequently, factors related to land use ratios, population density, and road design follow in significance. Notably, the pivotal factors encompass leisure facility density, ARParking, work facility density and ARBus. These built environment factors contribute to improvements in R_2 of 0.22, 0.13, 0.08 and 0.06, respectively. Contrastingly, Fig. 7(b) elucidates that major arterials density, ARParking and leisure facility density are the preeminent factors relating to CO_2 emission reduction resulting from ridesplitting in Xi'an. These factors contribute to an R_2 improvement exceeding 0.05. Other factors including population density, ARBus and expressways density also show pretty large relation with the dependent variable, but with a lower contribution to the R_2 (between 0.02–0.05). The divergent importance rankings of built environment factors underscore the city-specific nuances in the determinants of CO_2 emission reduction resulting from ridesplitting.

Moreover, we make the best of partial dependence analysis to interpret the relationship of a factor with the dependent variable. The following partial dependence analysis focuses on the built environment factors with importance score of more than 0.02, namely, factors contributing to the R_2 improvement of more than 0.02. Fig. 8 and Fig. 9 illustrate the nonlinear effects of built environment factors on CO_2 emission reduction in Chengdu and Xi'an, respectively. Dependent variable is defined as CO_2 emission reduction (g), so the positive effects of built environment means that it links to higher CO_2 emission reduction resulting from ridesplitting.

As illustrated in Fig. 9, both leisure facility density and ARParking manifest initial negative association with CO_2 emission reduction when they fall below specific threshold values. Subsequently, these effects stabilize upon surpassing the threshold values.

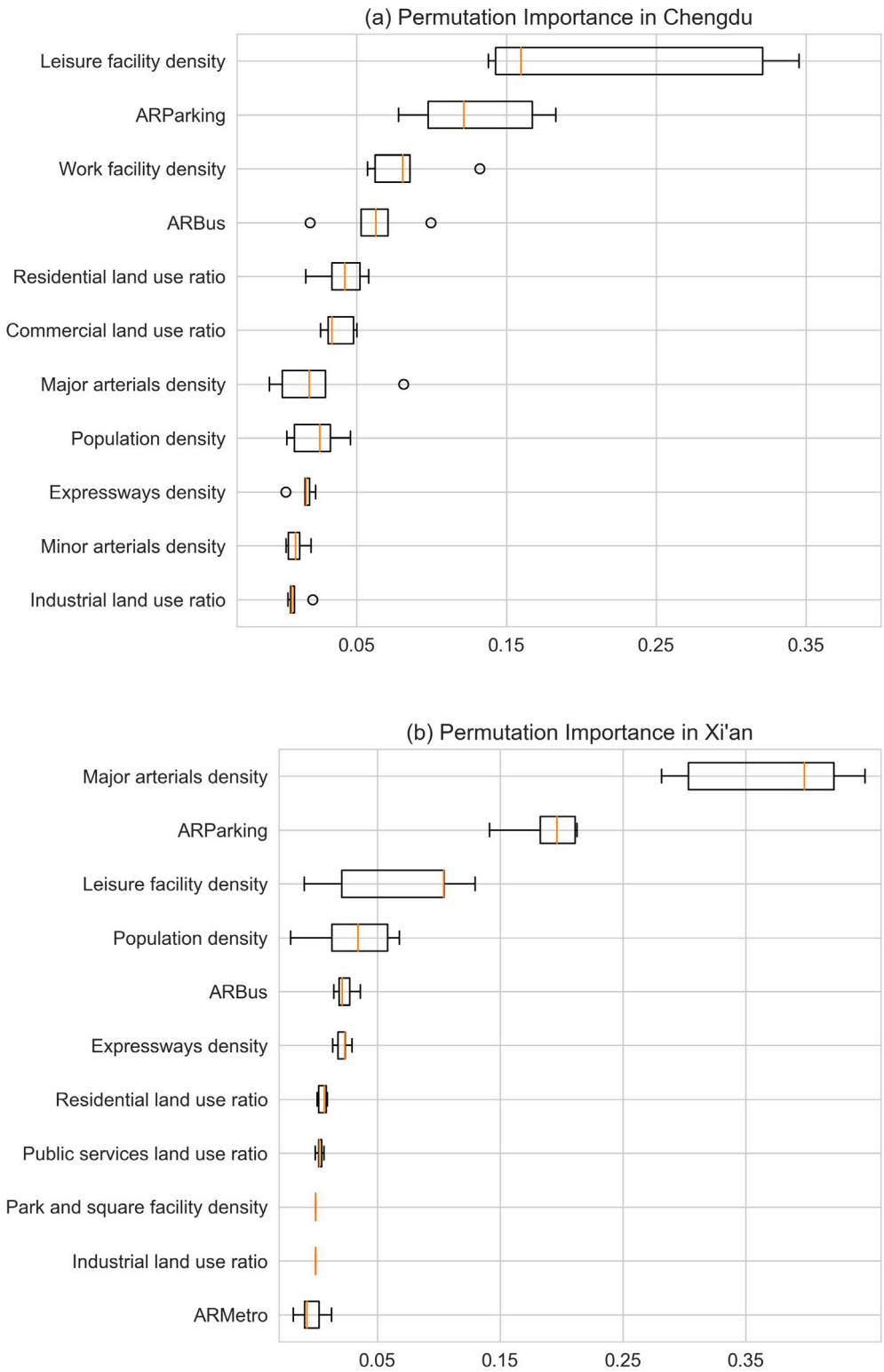

Fig. 7. Permutation importance of built environment factors.

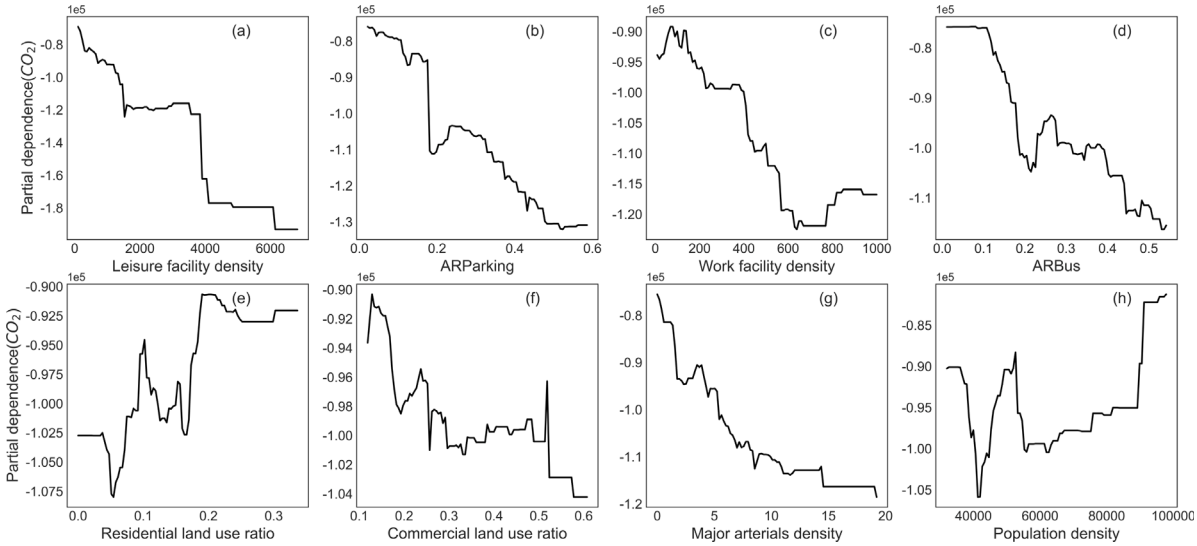

Fig. 8. Nonlinear effects of built environment factors in Chengdu.

Precisely, the identified threshold values for leisure facility density and ARParking are 6148 and 0.48, respectively. Throughout this progression, as the independent variables advance towards their respective threshold values, CO_2 emission reduction exhibits fluctuations, but the overarching trend remains characterized by negative relations. Consequently, our analysis primarily underscores this predominant trend. Work facility density initially demonstrates a negative impact on CO_2 emission reduction when it is below 770. However, when work facility density surpasses this threshold, CO_2 emission reduction experiences a discernible increase. Overall, ARBus, commercial land use ratio, and major arterials density link negatively to CO_2 emission reduction. This observation may be attributed to the fact that these factors denote district attractiveness and transportation convenience, signifying a heightened demand for ridesplitting. As depicted in the analysis presented in Fig. 6, where most ridesplitting trips result in increased CO_2 emissions, a higher demand for ridesplitting trips correlates with a more substantial increase in emissions stemming from ridesplitting (i.e. less emission reductions). Moreover, with an increase in residential land use ratio, emission reduction undergoes pronounced fluctuations before gradually ascending and ultimately stabilizing. A comparable trend pattern is observed in population density.

Fig. 9 reveals associations of six most pounced built environment factors with the CO_2 emission reduction resulting from ridesplitting in Xi'an. Several independent variables, including major materials density, ARParking, leisure facility density, ARBus and expressways density, manifest negative impacts on emission reduction within certain ranges. Outside these ranges, these variables do no influence emission reduction. Specifically, the ranges for major materials density, ARParking, leisure facility density, ARBus and expressways density are [2.22, 10.70], [0.27, 0.48], [752, 1868], [0.08, 0.25] and [2.66, 44.05], respectively. This observation may can also be explained by the fact that a higher demand for ridesplitting trips indicates a greater increase in emissions resulting from ridesplitting. As for population density, an increase within the range of 60317 and 84504 initially does not affect much before a threshold but leads to a decline in CO_2 emission reduction when exceeding the threshold. A commonly found pattern from the results is that a lot of factors present nonlinear and threshold effects on the reduced CO_2 emission.

Due to diverse urban context, demographics and other potential factors, various built environment factors have nonlinear effects on the increased emission resulting from ridesplitting in both cities. The threshold effects varies in different cities. The purpose of identifying these thresholds is to understand the critical points at which different built environment factors begin to significantly impact CO_2 emission reduction and how the effects on emission reduction differ below or above these thresholds. The identification of threshold values for various factors holds practical implications that can inform strategies for ridesplitting services. For example, in areas that are far from the urban center and marked by low leisure facility density, limited ARParking availability, sparse work facility density, and fewer ARBus options in Chengdu, ridesplitting emerges as a promising avenue for achieving more pronounced reductions in emissions. The combination of low leisure facility density, minimal ARParking, limited work facility density, and scarce ARBus services typically characterizes underdeveloped regions where public transit options are insufficient or less convenient. In these contexts, ridesplitting becomes a crucial component of daily travel, filling the gap left by less accessible public transport. This suggests that ridesplitting can be strategically employed to offer flexible mobility options in areas with specific environmental considerations. It also underscores the potential of ridesplitting to be strategically leveraged in areas with unique environmental and infrastructural challenges. By offering a flexible mobility solution, ridesplitting can significantly contribute to lowering overall carbon emissions, especially in parts of the city where other sustainable transport options might be limited. Consequently, incorporating ridesplitting into the broader urban transport strategy could be instrumental in addressing the specific mobility needs of these areas, while simultaneously advancing the city's environmental sustainability goals.

Urban planners and policymakers can utilize the insights from our study to design more sustainable transportation systems and incentivize ridesplitting, particularly in areas where it can significantly reduce emissions. For instance, policymaker could

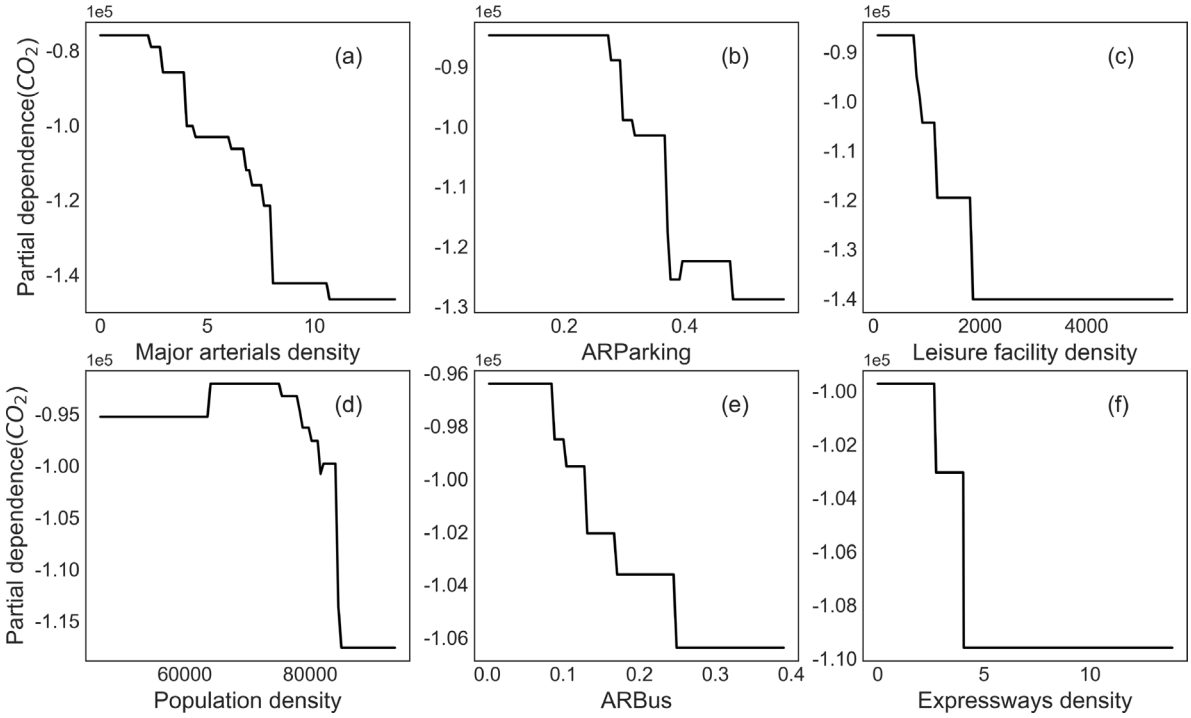

Fig. 9. Nonlinear effects of built environment factors in Xi'an.

implement dynamic incentive schemes for ridesplitting based on threshold values. In areas where essential amenities and public transport are below the threshold levels, enhanced incentives for ridesplitting could be offered as a compensatory measure for the limited sustainable transit options. Furthermore, policymaker could also adjust parking policies based on the ARParking threshold. In areas characterized by low ARParking availability, measures like escalating parking fees for single-occupancy vehicles could be introduced, thereby incentivizing ridesplitting and supporting environmental objectives. Given the variations in urban planning and characteristics across different cities, policymakers could formulate tailored policies based on local conditions. Recognizing the nuanced impact of built environment factors and understanding the specific contexts in which ridesplitting operates can enhance the environmental benefits and contribute to sustainable urban mobility solutions.

5. Conclusions

This study undertakes an assessment of the potential CO_2 reduction attributable to ridesplitting services and delves into the nonlinear impacts of the built environment on variation in emission based on field big data, especially considering mode substitution for different transport modes using multivariate data resources. The main findings can be summarized as follows.

In Chengdu, the substitution rates of ridesplitting trips to walking, bus, metro, and taxi are 5.20%, 41.62%, 22.68%, and 30.50%, respectively. In Xi'an, the substitution rates to walking, bus, metro, and taxi are 3.70%, 60.10%, 15.67% and 20.54%, respectively. Trip distance exerts discernible influence on substitution rates, the proportions of substituting bus and taxi demonstrating an ascent followed by a descent, but the substitution rates to the metro consistently increase with the increase in trip distance.

The majority of ridesplitting trips do not yield a reduction in CO_2 emissions compared to substituted transport modes. In Chengdu, a mere 8.92% of ridesplitting trips result in emission reduction, whereas in Xi'an, this figure stands at 4.68%. Emission reduction is predominantly linked to taxi substitution. Additionally, ridesplitting leads to an overall increase in CO_2 emissions in aggregated analysis zones, despite some ridesplitting trips indeed producing emission reduction by replacing taxis.

Additionally, there are nonlinear and threshold associations between different built environment factors and CO_2 emission reductions resulting from ridesplitting. Factors such as ARParking play significant roles in CO_2 emissions reduction resulting from ridesplitting in both cities. However, specific threshold values for these factors differ between the studied cities. Therefore, tailoring ridesplitting strategies in operation and matching in different urban contexts, taking into account these threshold values, could be helpful for optimizing environmental benefits while ensuring efficiency of the service.

Ridesplitting does not necessarily lead to a decrease in CO_2 emission. However, emissions are just one aspect of evaluating shared mobility system. Ridesplitting can be applied in certain scenarios to simultaneously produce environmental benefits and offer flexible mobility options. This study is rooted in Chengdu and Xi'an, but the methodological framework and insights derived can serve as a scholarly guide and reference for analogous analyses in diverse urban settings, contingent upon the availability of

pertinent data. The findings in this study provide references for the design and refinement of ridesplitting services within sustainable transportation frameworks. There are several limitations need to be listed for further extending the present study. Due to the data limitations, we did not consider all possible replaced transport modes by ridesplitting, which could be compensated by other data sources in the future. Moreover, we assume that the travel mode choice model in Shanghai can be adapted for Chengdu and Xi'an, considering that they are all large metropolises. Ideally, using survey data from the same city when constructing the choice model would yield more precise outcomes. The calibrated travel choice models could be improved by more comprehensive behavioral data in Chengdu and Xi'an to enhance accuracy. Furthermore, our investigation delves into the role of certain built environment factors on CO_2 emission reductions. Yet, due to constraints in our data, we acknowledge that our analysis might not encompass all potential variables. Specifically, factors extending beyond the built environment such as demographic attributes including income and age, which could have direct or indirect impacts on emission reductions, have not been thoroughly examined. Additionally, the public transit station data used in this study corresponds to their configuration as of 2023 due to data limitation. The study area encompasses almost the busiest districts of Chengdu and Xi'an, where public transit accessibility was already well-established in 2018. We assume that new stations introduced between 2019 and 2023 did not significantly impact our study area. This could be calibrated as long as the relevant data becomes available. These may be promising research directions for future work on the basis of the outcomes in this study.

CRediT authorship contribution statement

Zhe Zhang: Visualization, Methodology, Formal analysis, Data curation, Conceptualization. **Kun Gao:** Writing – review & editing, Validation, Supervision, Resources, Methodology, Funding acquisition, Conceptualization. **Hong-Di He:** Writing – review & editing, Validation, Supervision, Conceptualization. **Shaohua Cui:** Writing – review & editing. **Liyang Hu:** Writing – review & editing. **Qing Yu:** Conceptualization. **Zhong-Ren Peng:** Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This study was partially funded by the National Natural Science Foundation of China (No. 12072195) and Shanghai Foundation for Development of Science and Technology, China (No. 23692104000). The study is also supported by Swedish Energy Agency and JPI Urban Europe (e-FAST, P2022-00414 and e-MATS, P2023-00029) and Area of Advance Transport at Chalmers University of Technology. The authors would like to acknowledge DiDi Chuxing GAIA Initiative for providing the observed ride-sourcing data (https://gaia.didichuxing.com). Z. Zhang acknowledges a scholarship from the China Scholarship Council (CSC) (202206230104).

Appendix

See Tables 1–3.

References

Barnes, S.J., Guo, Y., Borgo, R., 2020. Sharing the air: Transient impacts of ride-hailing introduction on pollution in China. Transp. Res. D 86, 102434.
Cervero, R., Sarmiento, O.L., Jacoby, E., Gomez, L.F., Neiman, A., 2009. Influences of built environments on walking and cycling: lessons from bogotá. Int. J. Sustain. Transp. 3 (4), 203–226.
Chen, L., Lu, Y., Ye, Y., Xiao, Y., Yang, L., 2022. Examining the association between the built environment and pedestrian volume using street view images. Cities 127, 103734.
Chen, X.M., Zahiri, M., Zhang, S., 2017. Understanding ridesplitting behavior of on-demand ride services: An ensemble learning approach. Transp. Res. C 76, 51–70.
Chen, X., Zheng, H., Wang, Z., Chen, X., 2021. Exploring impacts of on-demand ridesplitting on mobility via real-world ridesourcing data and questionnaires. Transportation 48, 1541–1561.
Cui, S., Xue, Y., Gao, K., Lv, M., Yu, B., 2023. Adaptive collision-free trajectory tracking control for string stable bidirectional platoons. IEEE Transactions on Intelligent Transportation Systems 24 (11), 12141–12153.
Cui, S., Xue, Y., Gao, K., Wang, K., Yu, B., Qu, X., 2024. Delay-throughput tradeoffs for signalized networks with finite queue capacity. Transp. Res. B 180, 102876.
Ding, F., Chen, S., Zhao, Z., 2024. Incorporating walking into ride-hailing: The potential benefits of flexible pick-up and drop-off. Transp. Res. D 127, 104064.
Erhardt, G.D., Roy, S., Cooper, D., Sana, B., Chen, M., Castiglione, J., 2019. Do transportation network companies decrease or increase congestion? Sci. Adv. 5 (5), eaau2670.
Ewing, R., Cervero, R., 2010. Travel and the built environment, journal of the American planning association. J. Am. Planning Assoc..
Gao, K., Li, A., Liu, Y., Gil, J., Bie, Y., 2023a. Unraveling the mode substitution of dockless bike-sharing systems and its determinants: A trip level data-driven interpretation. Sustainable Cities Soc. 98, 104820.

Gao, K., Yang, Y., Gil, J., Qu, X., 2023b. Data-driven interpretation on interactive and nonlinear effects of the correlated built environment on shared mobility. J. Transp. Geogr. 110, 103604.
Gao, K., Yang, Y., Li, A., Li, J., Yu, B., 2021a. Quantifying economic benefits from free-floating bike-sharing systems: a trip-level inference approach and city-scale analysis. Transp. Res. A 144, 89–103.
Gao, K., Yang, Y., Li, A., Qu, X., 2021b. Spatial heterogeneity in distance decay of using bike sharing: An empirical large-scale analysis in shanghai. Transp. Res. D 94, 102814.
He, H., Lu, D., Zhao, H., Peng, Z., 2024. Characterizing CO2 and NOx emission of vehicles crossing toll stations in highway. Transp. Res. D 126, 104024.
Henao, A., Marshall, W.E., 2019. The impact of ride-hailing on vehicle miles traveled. Transportation 46 (6), 2173–2194.
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. C 133, 103425.
Li, W., Pu, Z., Li, Y., Ban, X.J., 2019. Characterization of ridesplitting based on observed data: A case study of Chengdu, China. Transp. Res. C 100, 330–353.
Li, W., Pu, Z., Li, Y., Tu, M., 2021. How does ridesplitting reduce emissions from ridesourcing? A spatiotemporal analysis in chengdu, China. Transp. Res. D 95, 102885.
Li, W., Wang, L., Pu, Z., Cheng, L., Yang, L., 2024. What determines the real-world CO2 emission reductions of ridesplitting trips? Travel Behav. Soc. 35, 100734.
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.
Liu, Y., Jia, R., Ye, J., Qu, X., 2022. How machine learning informs ride-hailing services: a survey. Communications in Transportation Research 2, 100075.
Liu, J., Li, J., Chen, Y., Lian, S., Zeng, J., Geng, M., Zheng, S., Dong, Y., He, Y., Huang, P., et al., 2023. Multi-scale urban passenger transportation CO2 emission calculation platform for smart mobility management. Appl. Energy 331, 120407.
Qiao, S., Yeh, A.G.-O., 2023. Is ride-hailing competing or complementing public transport? A perspective from affordability. Transp. Res. D 114, 103533.
Rafiq, R., McNally, M.G., 2023. An exploratory analysis of alternative travel behaviors of ride-hailing users. Transportation 50 (2), 571–605.
Shaheen, S.A., Chan, N.D., Gaynor, T., 2016a. Casual carpooling in the san Francisco Bay Area: Understanding user characteristics, behaviors, and motivations. Transp. Policy 51, 165–173.
Shaheen, S., Cohen, A., Zohdy, I., et al., 2016b. Shared mobility: current practices and guiding principles. Technical Report, United States. Federal Highway Administration.
Shi, K., Shao, R., De Vos, J., Cheng, L., Witlox, F., 2021. The influence of ride-hailing on travel frequency and mode choice. Transp. Res. D 101, 103125.
Smit, R., Ntziachristos, L., Boulter, P., 2010. Validation of road vehicle and traffic emission models–a review and meta-analysis. Atmosph. Environ. 44 (25), 2943–2953.
Taiebat, M., Amini, E., Xu, M., 2022. Sharing behavior in ride-hailing trips: A machine learning inference approach. Transp. Res. D 103, 103166.
Tingstad Jacobsen, S.E., Lindman, A., Kulcsár, B., 2023. A predictive chance constraint rebalancing approach to mobility-on-demand services. Communications in Transportation Research 3, 100097.
Tirachini, A., 2020. Ride-hailing, travel behaviour and sustainable mobility: an international review. Transportation 47 (4), 2011–2047.
Tu, M., Li, Y., Li, W., Tu, M., Orfila, O., Gruyer, D., 2019. Improving ridesplitting services using optimization procedures on a shareability network: A case study of chengdu. Technol. Forecast. Soc. Change 149, 119733.
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. D 93, 102776.
Wali, B., 2023. What drives the joint demand for ride-hailing and carsharing services? understanding consumers' behaviors, attitudes, & concerns. Transp. Res. C 157, 104373.
Wang, Z., Chen, X., Chen, X.M., 2019. Ridesplitting is shaping young people's travel behavior: Evidence from comparative survey via ride-sourcing platform. Transp. Res. D 75, 57–71.
Wang, Z., Chen, F., Fujiyama, T., 2015. Carbon emission from urban passenger transportation in Beijing. Transp. Res. D 41, 217–227.
Wang, D., Zhou, M., 2017. The built environment and travel behavior in urban China: A literature review. Transp. Res. D 52, 574–585.
Xu, Y., Yan, X., Liu, X., Zhao, X., 2021. Identifying key factors associated with ridesplitting adoption rate and modeling their nonlinear relationships. Transp. Res. A 144, 170–188.
Xue, Y., Wang, C., Ding, C., Yu, B., Cui, S., 2024. Observer-based event-triggered adaptive platooning control for autonomous vehicles with motion uncertainties. Transp. Res. C 159, 104462.
Yan, Y., Li, Y., Tong, P., 2019. Development of road functional classification in China: An overview and critical remarks. In: Intelligent Interactive Multimedia Systems and Services: Proceedings of 2018 Conference 11. Springer, pp. 347–355.
Yan, L., Luo, X., Zhu, R., Santi, P., Wang, H., Wang, D., Zhang, S., Ratti, C., 2020. Quantifying and analyzing traffic emission reductions from ridesharing: A case study of shanghai. Transp. Res. D 89, 102629.
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.
Yu, W., Wang, T., Xiao, Y., Chen, J., Yan, X., 2020. A carbon emission measurement method for individual travel based on transportation big data: the case of nanjing metro. Int. J. Environ. Res. Public Health 17 (16), 5957.
Zhang, Z., Gao, K., He, H.-D., Yang, J.-M., Jia, R., Peng, Z.-R., 2023. How do travel characteristics of ridesplitting affect its benefits in emission reduction? evidence from chengdu. Transp. Res. D 123, 103912.
Zheng, H., Chen, X., Chen, X.M., 2019. How does on-demand ridesplitting influence vehicle use and purchase willingness? A case study in Hangzhou, China. IEEE Intell. Transp. Syst. Mag. 11 (3), 143–157.
Zhu, Z., Qin, X., Ke, J., Zheng, Z., Yang, H., 2020. Analysis of multi-modal commute behavior with feeding and competing ridesplitting services. Transp. Res. A 132, 713–727.