




ORIGINAL PAPER

Open Access



How do taxi usage patterns vary and why? A dynamic spatiotemporal analysis in Beijing

Jiaoe Wang^{1,2}, Fangye Du³, Jie Huang^{1,2*}  and Yu Liu⁴

Abstract

Existing studies lack attention to taxi usage dynamics, considering its trip proportion over other travel modes and its influencing factors at fine spatiotemporal resolutions. To fill these gaps, we propose a method for examining taxi usage in a grid of 1 km × 1 km cells per hour during a one-day cycle in Beijing. This method measures the differences between taxi trips from taxi trajectory data and mobile signaling data in the same week in January 2017. To explain the spatiotemporal variation in taxi usage, multiple linear models were used to investigate taxi usage dynamics with alternative transport modes, socioeconomic factors, and built environments. In summary, this study proposes to develop an indicator to measure taxi usage using multiple data sources. We confirm that taxi usage dynamics exist in both temporal and spatial dimensions. In addition, the effects of taxi usage factors vary over each hour in a one-day cycle. These findings are useful for urban planning and transport management, in which the dynamic interactions between taxi demand and distribution of facilities should be included.

Keywords Taxi usage, Built environment, Temporal dynamic, Mobile signaling data, Taxi trajectory data

1 Introduction

Taxi transport is an indispensable mode of transport in large cities. It complements other public transport modes in terms of flexible door-to-door services and 24/7 operations. The significance of the taxi industry is also reflected in its fleet size and the number of passengers served (Qian & Ukkusuri, 2015). By the end of 2020, for example, approximately 72,000 taxicabs were generating over 331 million trips in Beijing. However, balancing passenger demand and taxicab supply is a critical

challenge for the taxi industry. Oversupply might lead to unwanted vacant trips and introduce negative externalities such as additional congestion and emissions (Çetin & Yasin Eryigit, 2011). In contrast, an inadequate supply may result in a lengthy wait time for passengers (Schaller, 2005). Understanding the spatial and temporal dynamics of taxi usage patterns is significant when the popularity of on-demand transport services increases and smart transportation is available (Wang & Ross, 2018).

In previous studies, two streams of investigation into usage patterns by transport mode have evolved. One stream compares the mode of choice among different transport modes, such as public transit, private cars, taxis, bicycles, and walking, using questionnaire survey data (Buehler, 2011; Du et al., 2021). The other stream compares usage patterns and ridership across space (Kang & Qin, 2016; Zhong et al., 2015). With the support of trip records, such as smart card data and taxi trajectory data, the spatial distribution of ridership has been examined along with the influencing factors. Both between-modal and within-modal comparisons are important to understand usage patterns of transport

*Correspondence:

Jie Huang
huangjie@igsnr.ac.cn

¹ Key Laboratory of Regional Sustainable Development Modeling, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

² College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China

³ School of Public Affairs and Administration, University of Electronic Science and Technology of China, Chengdu 611731, Sichuan, China

⁴ Institute of Remote Sensing and Geographical Information Systems, School of Earth and Space Sciences, Peking University, Beijing 100871, China



modes as well as examine ridership spatial distribution because both consider the trade-off with other travel modes when residents take taxis and indicate the irreplaceability of taxis. In addition, the temporal dynamics of taxi usage on a micro scale, such as hour by hour, is also a critical attribute because the spatial patterns of taxi usage vary by hour. However, the current literature has not sufficiently examined the usage pattern with both between-modal and within-modal comparisons from fine spatial and temporal perspectives. This might be due to the lack of multiple sources of data at a fine-grained scale, which is a useful methodology.

Recent studies have shown that sampling biases among multi-source data can be used to investigate taxi usage at a finer spatial and temporal resolution (Wang, Huang, & Du, 2020). Therefore, the research objectives of this study could be summarized as follows: (1) we propose an indicator to measure the taxi usage index using GPS-traced taxi trips and mobile signaling data for the same week in January 2017; (2) using multi-source data in Beijing, we measure the dynamics of the taxi usage index by hour in a daily cycle in $1\text{ km} \times 1\text{ km}$ grid cells and explore the influencing factors associated with taxi usage from a spatiotemporal perspective.

2 Literature review

2.1 Taxi usage and mobility patterns

There are two streams of research on taxi usage. First is between-modal or within-modal comparison of taxi ridership among different transport modes, such as public transit, private cars, bicycles, and walking (Böcker et al., 2017; Buehler, 2011; Du et al., 2021). Under this stream, mode-choice prediction has received considerable research attention because it is an important component in travel demand estimation, transport planning, and urban planning (Yang et al., 2019). Traditionally, information on choice of travel mode is mainly collected from questionnaires. For example, Schwanen and Mokhtarian (2005) measured the travel mode choice of populations in the San Francisco Bay Area by asking 8000 randomly selected responders about “the frequency of travel by a certain mode and trip purpose.” Du et al. (2021) collected data on travel mode choices to seek healthcare by conducting a health-seeking behavior survey of 915 patients at nine top-tier hospitals in Beijing between August 1st and 17th, 2019. These studies reflected the static travel mode choice at the individual level, further using the questionnaire data.

The second stream examines the spatial dynamics of taxi usage patterns or taxi ridership. Qian and Ukkusuri (2015) explored the spatial variation of taxi ridership using large-scale New York City taxi data and associated taxi ridership with various socio-demographic and

built-environment variables. Similar to taxi ridership, public transit ridership has also been studied by comparison over space. Bachir et al. (2019) proposed a method to infer dynamic origin-destination flows by transport mode using mobile phone data and observed different mobility patterns, by driving and by rail, between Paris and its suburbs. In addition, Liu et al. (2012) investigated the variations in both pick-ups and drop-offs using a taxi trajectory dataset collected in Shanghai and reported the underlying location difference, which reflects land usage patterns across the study area. In the era of big data, the usage of a transport mode by location can be evaluated using trip records from taxis, buses, and subways (Huang et al., 2019; Huang et al., 2021; Liu et al., 2021; Zhu et al., 2022).

Moreover, mobile phone data are widely used to investigate mobility patterns of the general population, and some achievements in this area have been made to date (Calabrese et al., 2013; Wesolowski et al., 2014). For example, Ratti et al. (2006) mapped the dynamics of urban activity in the metropolitan area of Milan, Italy using the Erlang values of cell phone stations. Candia et al. (2008) investigated calling activity patterns at the individual level and showed that the interevent time of consecutive calls is heavy-tailed. Yuan et al. (2012) provide a deeper understanding of how mobile phone usage correlates with individual travel behavior by exploring the correlation between mobile phone call frequencies and three indicators of travel behavior: radius, eccentricity, and entropy. Compared with transport records, such as taxi trajectory and smart card data, as well as mobile phone data can cover a wider group of people, representing the mobility pattern of the general population.

2.2 Factors associated with taxi usage

Factors associated with mode choice can be classified into three types: alternative transportation (e.g., private car, taxi, public transit); socioeconomic characteristics (e.g., income, age, employment); and built environment factors (e.g., road density, land use density, land use diversity) (Farinloye et al., 2019; Gan et al., 2020; Wang et al., 2022). First, public transit ridership and mode choice are closely related to the accessibility of alternative transportation (Du et al., 2021). For example, in a case study in Beijing, Wang et al. (2022) reported that the temporal and spatial dynamics of bus ridership are closely related to alternative transportation. Socioeconomic characteristics have been evaluated based on population, employment, and average household income, which usually have a positive correlation with public transit/taxi ridership (He et al., 2019; Zhao et al., 2013; Zhu et al., 2018). Among these factors, population is more closely related to the amount of travel demand, and employment can

significantly influence taxi usage patterns. In addition, the average household income reflects affordability and affects taxi ridership. As average household income is related to privacy, house prices in the corresponding space are often used as proxies to reflect economic status (Freeman, 2018; Gan et al., 2020). In terms of built environment factors, previous studies have primarily focused on three dimensions: density, diversity, and design (Cervero & Kockelman, 1997). In addition, road and land use densities, land use diversity, and land-use types, such as residential, commercial, and industrial areas, are all useful variables in explaining taxi/public transit ridership and mode choice (Cao et al., 2009). As land-use type is difficult to quantify directly in most cities, the density of facilities, such as retailers, schools, banks, hospitals, hotels, and restaurants, is widely used as a proxy for land-use type (He et al., 2019).

In summary, previous studies have mainly focused on factors associated with the spatial distribution of taxi ridership. Little attention has been paid to the factors associated with temporal dynamics. In this study, taxi usage patterns were investigated through a comparison over space and across different transportation modes, and influencing factors were examined in a daily cycle, by hour. In addition, factors associated with taxi ridership are included in the reviewed literature.

2.3 Methods to explore factors related to taxi usage

Regarding the research methods, global models, including multiple linear regression with ordinary least squares (Chow et al., 2006), Poisson regression models (Kuby et al., 2004), Aggregate logit models (Taylor & Fink, 2013) and Tobit regression models (Chiou et al., 2015), have been widely used to explore the determinants of public transit and taxi ridership. As the dependent variable is a continuous variable and obeys a normal distribution after the transformation, multiple linear models estimated by ordinary least squares were employed to investigate the factors associated with taxi usage patterns.

3 Research design

3.1 Study area and data

The study area, comprising six central municipal districts of Beijing, is a major urban area in which most taxi trips are generated (Wang, Huang, & Du, 2020). To investigate taxi usage patterns at a fine spatial scale, we generated 1 km × 1 km grid cells that were compatible with the resolution of the data sources (Fig. 1). These datasets regarding taxi trip data, mobile signaling data, points of interest (POIs), and other built-environment factors were then collected in the same grid cells. Overall, approximately 1500 grid cells were used in the analysis.

3.1.1 Taxi trip data

Taxi is a widely used travel mode in urban settings. With the popularity of mobile phones, taxis can be booked online and ride-hailing has become common. The emergence of ride-hailing has solved the taxi shortage problem to some extent. As ride-hailing cannot be stopped and is more likely to be associated with crime, taxis are still more popular than ride-hailing. Therefore, this research collected the trajectories of taxis, which contain information on each taxi's location, time, status (vacant or occupied), and speed every 10 seconds. We can identify the pick-up and drop-off points of each trip according to these continuous records to obtain the origins and destinations of the taxi trips. With a random sampling process that ensures representative spatial coverage and avoids confidentiality issues (Yao et al., 2019), we collected data for approximately 8.8 million taxi trips taken by 17,984 taxis during weekdays between January 9th and January 13th, 2017. It is worth noting that taxi trips on weekdays are only included in this study because residents' trip purposes on weekdays and weekends vary considerably, and trips on weekdays are more likely to be associated with necessary activities. According to the origin and destination (i.e., OD) locations of taxi trips, the average numbers of departure and arrival trips in each cell were counted by hour. As the temporal dynamics during weekdays are different from those during weekends, this research analyzes taxi-based activity during weekdays. As shown in Fig. 2(a), on average, taxi trips occurred mainly between 07:00 and 21:00.

3.1.2 Mobile signaling data (MSD)

The original MSD recorded an anonymous and unique ID, time, and location for each mobile user. To address privacy concerns, mobile user locations were tracked into grid cells. The origin and destination grid cells of trips can be tracked when mobile users generate movements between different grid cells between two adjacent time intervals. Then we collected about 64.7 million OD pairs between January 9th and January 13th, 2017 (five weekdays) from China Unicom, one of the three largest telecommunication corporations. We counted the number of trips originating from each cell as well as the number of trips arriving at each cell so that each grid cell had two MSD attributes, namely, origins and destinations by hour. It is worth noting that OD pairs identified from MSD only indicate the movement of mobile users, but the mode of transport for each movement is unknown. In other words, these OD pairs represent trips for all transport modes. The general distribution of the origins by hour is presented in Fig. 2(b). The MSD origins are mainly concentrated between 08:00 and 18:00, with one

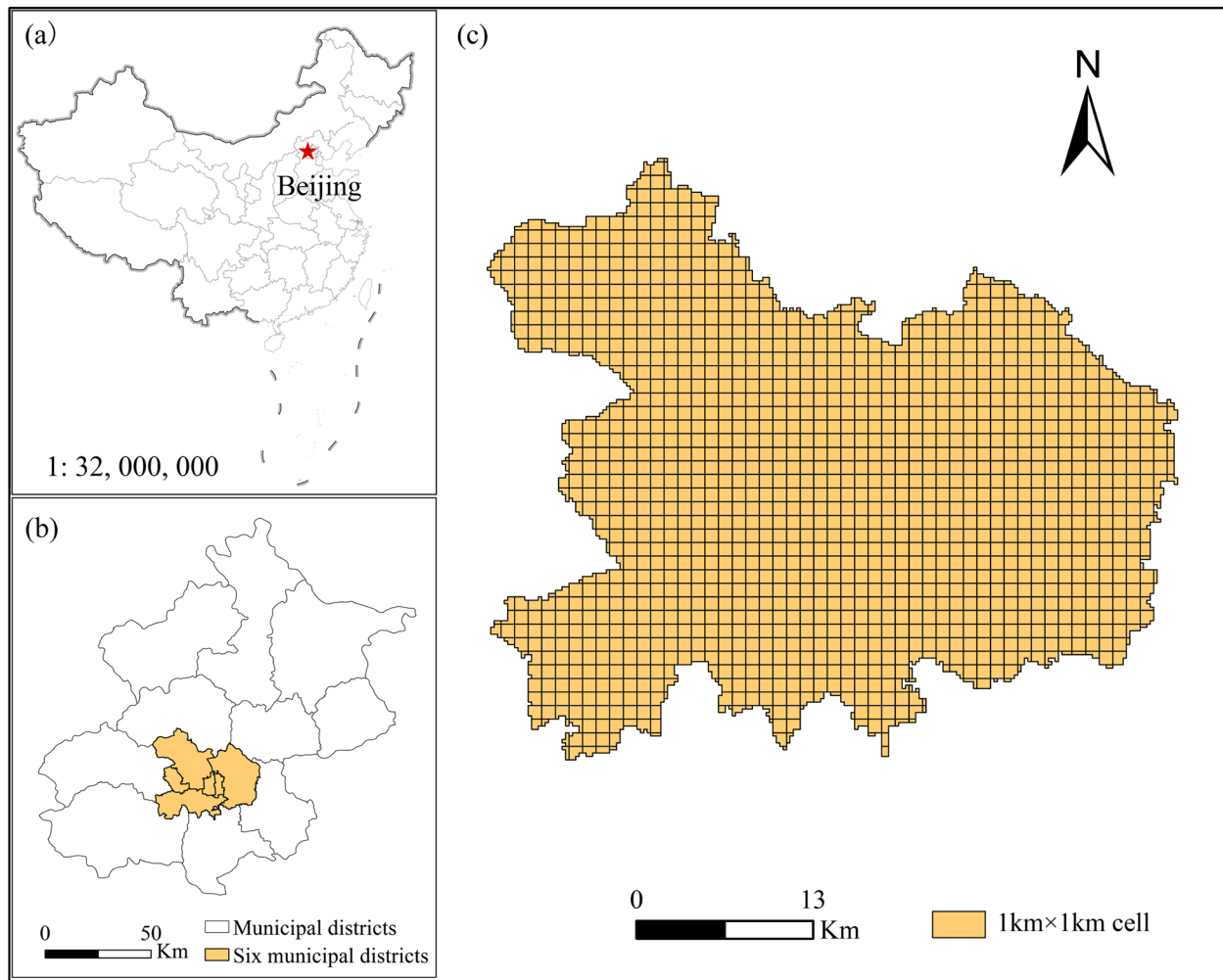


Fig. 1 **a** Location of Beijing in China; **b** Location of the study area in Beijing; **c** Study area (the six central municipal districts of Beijing) and unit (1 km x 1 km grid cell)

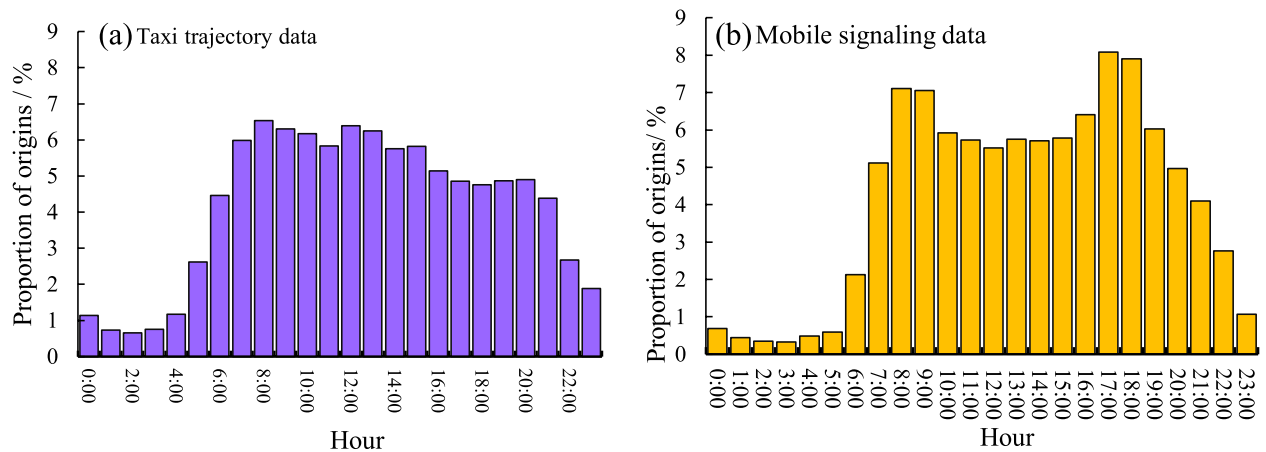


Fig. 2 Distribution of travelers by hour between January 9th and January 13th, 2017 from taxi trajectory data (a) and mobile signaling data (b)

morning and one evening peak identified. The difference between Fig. 2(a) and (b) indicates that quantitative gaps exist between taxi trips and MSD (trips by all modes) every hour; therefore, their spatial differences must be examined.

3.1.3 Variables selection

According to the reviews on factors associated with ridership in Section 2.2, alternative transportation, socioeconomic factors, and built environment variables were selected in this study. First, taxi usage is influenced by the distribution of alternative modes of transportation, such as buses and subways. Here, bus service coverage and subway service coverage in each grid are used to represent alternative transportation.

Second, taxi usage is related to socioeconomic factors, such as affordability, population density, and employment. People's affordability matters for taxi usage, as the monetary cost of traveling by a taxi is usually higher than that of other modes. In the absence of data on socioeconomic status, such as income, house prices are used to represent residence affordability. In this study, house price data were collected from the Soufang website (<https://bj.fang.com/>, accessed in October 2018), and the average house price for each grid cell was estimated based on the prices and locations of residences. Population refers to the population density, measured by the population scale in each grid, which was collected from the Beijing population census (2010), the latest census data obtained by the authors. Employment refers to employment density, measured by the employment scale in each grid, which was collected from the Beijing Economic Census (2014).

This research includes built environment factors related to land use, road areas, and facilities supporting daily life. Taxi usage is also associated with road network accessibility. Therefore, the road area in each grid was selected to represent the extent of the road network. The road network in Beijing, acquired from the OpenStreetMap website (<http://www.openstreetmap.org/>, accessed in October 2018), was used to calculate the length of roads in each grid cell. Land use diversity and density were calculated as the category and number of POIs in each grid, respectively. Recent studies (Gong et al., 2016) have incorporated POI data, indicating that urban dwellers visit with dataset attributes including latitude, longitude, name, and facility category. The POI categories include restaurants, recreation (e.g., bars, KTVs, parks, and gyms), enterprises, and residences. In this study, bus service coverage, subway service coverage, enterprises, residences, recreation facilities, and health facilities were counted in each 1 km × 1 km grid cell using Baidu Maps

(2018). The specific variable descriptions of the grid cells are listed in Table 1.

3.2 Methods

3.2.1 Taxi usage index

We propose a taxi usage index to quantify the concentration of taxi-based activities in a spatial unit (i.e., grid cell in this study). The taxi usage index measures the usage of taxis in two dimensions. First, it presents taxi usage in a grid cell compared with other grid cells, which presents a comparison in a spatial dimension. We defined a taxi trip ratio (TR) to calculate the number of taxis departing from a grid cell (N_{te}) over the total number of taxis departing from all grid cells studied (N_{tw}), indicating the relative degree of concentration. Second, the taxi usage index assesses taxi usage compared with other travel modes, as OD pairs measured by the MSD can present human movement in a general manner. In such a case, the taxi usage index measures usage in the dimension of travel mode. For all travel modes, the trip ratio (AR) is given by the number of trips departing from the grid cell (N_e) over the total number of trips departing from all the grid cells studied (N_w). AR was designed to reveal the distribution of trip departures using a grid cell. The taxi usage index is given by the following equation.

$$\text{Taxi usage index} = TR/AR \quad (1)$$

where $TR = N_{te}/N_{tw}$ and $AR = N_e/N_w$. A higher taxi usage index implies a higher concentration of taxi-based activities when compared to other grid cells and other travel modes, and vice versa.

3.2.2 Multiple linear regression

Multiple linear regression is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal is to model the linear relationship between the independent and dependent variables. Here, we fit a multiple regression with a step-wise selection method to model the relationship between the taxi travel preference index and built environment factors:

$$I_c = \beta_0 + \sum \beta_i * X_i \quad (2)$$

where β_i is the partial regression coefficient. X_i is shown in Table 1.

Before modeling, multicollinearity issues were examined by calculating the variance inflation factor between independent variables. We found that population, employment, land use density, land use diversity, public facilities, living facilities, government, sports and leisure, transportation hub, and financial area variables had

Table 1 Description of influence factors

Variables	Description	Source
<i>Alternative transportation</i>		
Bus service coverage ^a	Area of a 400-m buffer zone for bus stations in each grid	Baidu Map (2018)
Subway service coverage	Area of an 800-m buffer zone for subway stations in each grid	Baidu Map (2018)
<i>Socio-economic factors</i>		
House price	Average listing price of each residential quarter in each grid	Soufang Website (2018)
Population	Population density, measuring the population scales in each grid	Beijing population census (2010)
Employment	Employment density, measuring the employment scale in each grid	Beijing economic census (2014)
<i>Built environment factors</i>		
Land use diversity	Types of land use (POI category) in each grid	Baidu Map (2018)
Land use density	Number of POIs in each grid	Baidu Map (2018)
Road area	Area of road in each grid	Baidu Map (2018)
Catering facility	Number of catering facilities, such as restaurants and snack bars, in each grid	Baidu Map (2018)
Tourist spot	Number of tourist spots, such as park and interest, in each grid	Baidu Map (2018)
Public facility	Number of public toilets in each grid	Baidu Map (2018)
Enterprise	Number of enterprises in each grid	Baidu Map (2018)
Shopping spot	Number of shopping spots, such as shopping malls, supermarkets, and stores, in each grid	Baidu Map (2018)
School	Number of school (i.e., universities, colleges, etc.) in each grid	Baidu Map (2018)
Residence	Number of residential areas in each grid	Baidu Map (2018)
Living facility	Number of living facilities (i.e., barber shop)	Baidu Map (2018)
Recreation facility	Number of recreation facilities (i.e., KTV, cinema)	Baidu Map (2018)
Healthcare facility	Number of hospitals	Baidu Map (2018)
Government	Number of government agency	Baidu Map (2018)
Sport and leisure	Number of stadiums and parks	Baidu Map (2018)
Transportation hub	Number of transportation hubs (i.e., train station, coach station)	Baidu Map (2018)
Financial spot	Number of banks	Baidu Map (2018)

^a Note that the subway and most bus routes do not operate from 12:00 pm to 5:00 am in Beijing; thus, the number of bus and subway stops are omitted in the multiscale geographically weighted regression models from 12:00 pm to 5:00 am

multicollinearity issues with other variables, and were thus removed.

4 Results

In this section, we first explore the spatial pattern and spatiotemporal dynamics of the taxi usage index by calculating and visualizing its spatial pattern as 1 km × 1 km grid cells over 24 hours in a day. The number of trips from taxi trip data and mobile signaling data for each hour was calculated as the average from January 9th to January 13th, 2017. Then, an overall model and 24 multiple linear models for 24 hours were constructed and compared to examine the factors associated with the taxi usage index. As the smallest number of taxi trips at 2:00 am is approximately 40,000 trips, there are adequate trips and data records in the 24 models.

4.1 Spatial and temporal dynamics of taxi usage patterns

The spatial patterns of the taxi usage index by hour are shown in Fig. 3. Overall, the taxi usage index in the core areas was always higher than that in the periphery. This

can be explained as follows: first, taxi services are convenient in core areas with more daily travel and higher taxi demand. Second, travelers in core areas are advantaged in terms of socioeconomic status, and are more likely to afford expensive travel costs (Huang et al., 2018).

Temporal dynamics can be observed in the spatial patterns of the taxi usage index, and three periods in a day (24 hours) can be delineated. The first period was from 11:00 pm to 5:00 am the following day. During this period, *Sanlitun, Tuanjiehu, Beixiaguan, Shaungjing, Niujie, Hepingli, and Zhanlanlu subdistricts* showed a higher taxi usage index, which might be related to overtime journeys and recreational activities. The second period lasted from 5:00 am to 8:00 am. Grid cells with a high taxi usage index were relatively concentrated. They are mainly located in *Wangjing, Taiyanggong, Hepingjie, Hepingli, Anzhen, Desheng, Xiangheyuan, Zuojiazhuang, Sanlituan, Maizidian, and Tuanjiehu subdistricts*. The third period was from 8:00 am to 2:00 pm. The grid cells with high taxi usage indices were mainly concentrated in *Zuojiazhuang, Sanlitun, Dongzhimen,*

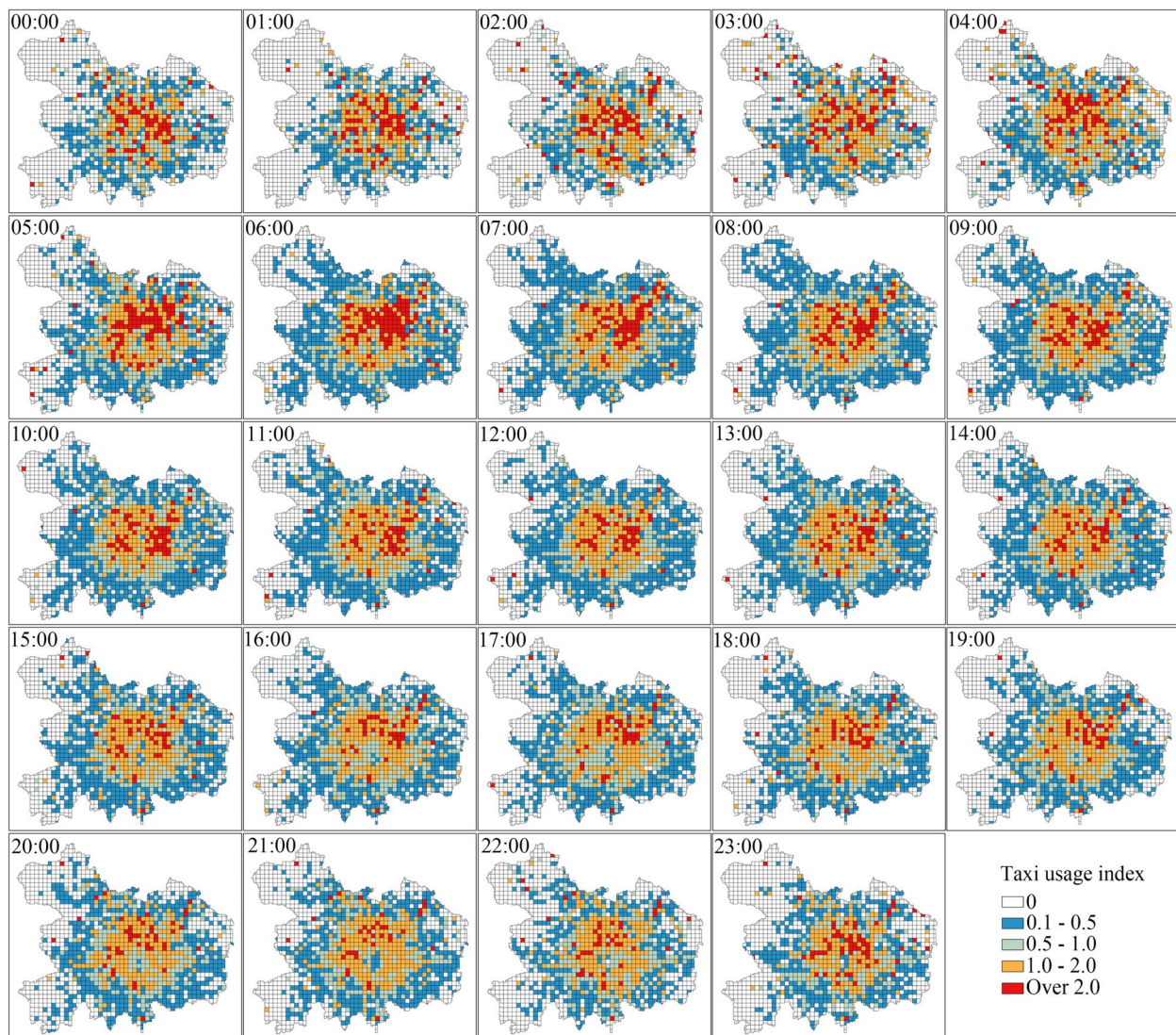


Fig. 3 Spatial patterns of the taxi usage index by hour

Maizidian, Tuanjiehu, Hujialou, Jianwai, Jianguomen, and Chaowai subdistricts. During the rest of the time, grid cells with a high taxi usage index were relatively dispersed.

Rapid urbanization has led to a mixed distribution of facilities, making it difficult to distinguish among those that affect taxi usage. Several multiple linear regressions were employed to investigate the factors associated with the temporal dynamics of taxi usage.

4.2 Multiple linear regression results

Overall, independent variables performed well in explaining the taxi usage index, as the pseudo- R^2 values of the overall model and 24 models were all over

0.4. Thereby, we summarized the results and discussed the regression coefficients of the factors associated with taxi usage (Tables 2 and 3).

4.2.1 Multiple linear results for the overall taxi usage patterns

The results indicated that house price, road density, bus service coverage, subway service coverage, residence, and healthcare facilities are significantly associated with taxi usage patterns. Among these, there are positive correlations between housing prices and taxi usage patterns. This might be because taxis are more expensive than other transport modes. Travelers located in areas with high housing prices are usually economically advantaged (Freeman, 2018); thus, they are more likely to afford and

Table 2 Results of Multiple Linear Regression on the origin side

Variable	Coefficient
House price	0.27**
Road density	0.64**
Bus service coverage	-0.18**
Subway service coverage	-0.17**
Residence	0.07**
Enterprise	0.06**
School	0.01
Healthcare facility	0.05**
Shopping spot	0.02
Catering facility	0.07
Recreation facility	0.02
Tourist spot	0.04
Intercept	0.11

** $p < 0.05$

choose taxis. Road density has a positive effect on taxi usage. This is because road density increases with road connectivity (Mo et al., 2018) and taxis are more likely to operate on roads with high connectivity. Bus and subway service coverage have a negative association with taxi usage patterns. Owing to the three travel modes in urban transportation, public transit and taxis can provide alternatives for passengers and share passenger flow among each other (Wang et al., 2022). We also find that enterprises and residences are positively associated with taxi usage patterns. An explanation for this finding is that areas with more enterprises and residences often have a more concentrated population, and thus may induce more travel demand overall (Liu et al., 2012). Furthermore, healthcare facilities are positively correlated with taxi usage patterns, which might also be related to the large travel demand for activities in these facilities. However, schools, as well as shopping, catering and recreational facilities, and tourism spots had no significant correlation with taxi usage patterns. These findings imply that taxi usage may vary depending on travel purposes.

Table 3 Coefficients of multiple linear regression in different hours

Variables	0:00	1:00	2:00	3:00	4:00	5:00	6:00	7:00	8:00	9:00	10:00	11:00
House price	0.29**	0.13**	0.14**	0.15**	0.14**	0.12**	0.16**	0.19**	0.26**	0.29**	0.23**	0.19**
Road density	0.13**	0.18**	0.19**	0.08	0.11**	0.10	0.07	0.11**	0.08**	0.18**	0.16**	0.24**
Bus service coverage	-	-	-	-	-	-	-0.18**	-0.12**	-0.15**	-0.19**	-0.03**	-0.13**
Subway service coverage	-	-	-	-	-	-	-0.15**	-0.12**	-0.15**	-0.12**	-0.11**	-0.10**
Residence	0.36**	0.35**	0.41**	0.36**	0.42**	0.39**	0.39**	0.46**	0.38**	0.35**	0.06**	0.04
Enterprise	0.07**	0.09**	0.19**	0.01**	0.29**	0.48	0.01	0.07	0.07**	0.17**	0.15**	0.09**
School	0.02	0.06	0.28	0.26	0.43	0.15	0.30	0.17	0.12	0.57	0.10	0.11
Healthcare facility	0.07	0.05	0.15	0.22	0.12	0.10	0.14	0.96	0.06	0.09**	0.09**	0.15**
Shopping spot	0.11	0.09	0.46	0.70	0.59	0.12	0.45	0.76	0.13	0.09	0.02	0.05
Catering facility	0.05	0.05	0.07	0.13	0.23	0.06	0.08	0.15	0.09	0.01**	0.11	0.15
Recreation facility	0.03	0.02	0.08	0.27	0.56	0.83	0.06	0.19	0.03	0.06	0.24	0.49
Tourist spot	0.04	0.01	0.02	0.14	0.12	0.13	0.29	0.84	0.03	0.06**	0.15**	0.16**
Intercept	0.13	-0.10	0.09	0.18	0.06	0.07	0.08	0.12	0.14	0.04	0.07	0.15
Variables	12:00	13:00	14:00	15:00	16:00	17:00	18:00	19:00	20:00	21:00	22:00	23:00
House price	0.43**	0.26**	0.18**	0.24**	0.28**	0.27**	0.34**	0.29**	0.28**	0.25**	0.23**	0.27**
Road density	0.37**	0.29**	0.28**	0.13**	0.25**	0.24**	0.18**	0.22**	0.18**	0.24**	0.24**	0.23**
Bus service coverage	-0.19**	-0.15**	-0.13**	-0.11**	-0.13**	-0.18**	-0.16**	-0.14**	-0.11**	-0.15**	-0.11**	-0.19**
Subway service coverage	-0.11**	-0.10**	-0.08	-0.19**	-0.08	-0.13**	-0.16**	-0.13**	-0.15**	-0.19**	-0.11**	-0.09**
Residence	0.04	0.05**	0.06**	0.56**	0.05**	0.07**	0.38**	0.44**	0.36**	0.37**	0.35**	0.35**
Enterprise	0.10**	0.01**	0.12**	0.13**	0.13**	0.18**	0.18**	0.06**	0.04**	0.31**	0.19**	0.02**
School	0.11	0.09	0.14	0.17	0.08	0.13	0.08	0.28	0.09	0.55	0.07	0.01
Healthcare facility	0.16**	0.13**	0.12**	0.09**	0.14**	0.15**	0.37**	0.29**	0.24	0.75	0.29	0.51
Shopping spot	0.12	0.07	0.45	0.27	0.38	0.63	0.10	0.51**	0.11**	0.15**	0.18	0.24
Catering facility	0.13	0.13	0.17	0.13	0.14	0.09	0.21**	0.35**	0.42**	0.46**	0.23	0.14
Recreation facility	0.36	0.50	0.27	0.14	0.18	0.14	0.15	0.17**	0.19**	0.17**	0.19**	0.11**
Tourist spot	0.12**	0.11**	0.14**	0.16**	0.18**	0.16**	0.15**	0.17**	0.15	0.17	0.16	0.12
Intercept	0.09	0.10	0.12	0.09	0.09	0.11	0.09	0.12	0.14	0.19	0.21	0.16

** $p < 0.05$

Among the indicators significantly associated with the taxi usage index, the correlation coefficient of road density is the largest, followed by house price, bus service coverage, and subway service coverage. This might be because taxis tend to cruise where roads are dense, thus presenting a larger taxi usage index. Since taxis are an expensive mode of transport, the economic status of individuals in different spaces also contributes to taxi ridership. Whereas, taxi usage is negatively correlated with the density of bus and subway service coverage. In addition, there is a weaker correlation between the density of other facilities and taxi usage index. Among the facilities studied, the correlation between residence and the taxi usage index was the largest, followed by enterprises and healthcare facilities.

4.2.2 Temporal dynamics of regression coefficients

In general, the effects of bus service coverage, subway service coverage, enterprises, healthcare facilities, shopping spots, catering facilities, recreational facilities, and tourist spot variables showed significant temporal dynamics. Enterprises are positively associated with taxi usage patterns from 8:00 am to 5:00 am the next day, induced by business activities from enterprises to other facilities. From 9:00 am to 7:00 pm, the larger the number of healthcare facilities, the higher the taxi usage index. An explanation might be that taxis are like private cars, which are more convenient and comfortable than other transport modes, such as public transit and bicycles. Hence, for patients with physical discomfort, taxi is a preferential transport mode. Furthermore, taxis are also a better choice than private cars because most well-known hospitals in Beijing are located in the inner city where parking spaces are lacking (Wang, Du, et al., 2020). However, healthcare facilities had no significant correlation with the taxi usage index during their off-work hours (from 6:00 pm to 8:00 am the next day).

The recreational facility variable is positively related to the taxi usage index from 7:00 pm to 12:00 pm, whereas it is not significantly related to the taxi usage index during the remaining hours. This is reasonable because populations are more likely to visit recreational facilities during off-work hours on weekdays. The phenomenon in which many recreational trips in Beijing are serviced by taxis can be explained by the following reasons. First, recreational facilities are usually located in areas with a high development density and provide limited parking spaces. Second, activities in recreation facilities (e.g., bars) usually involve alcohol consumption, which increases taxi demand owing to drunk-driving restrictions in China. There was a positive correlation between tourist spots and taxi usage patterns from 9:00 am to 8:00 pm, which is related to the large number of tourists during this time.

However, the correlation between house prices, road density, residence, school, and taxi usage patterns does not have obvious temporal dynamics. Among these, house prices and road density are statistically significant at almost all hours. It can be reasoned that the economic status of individuals will always influence whether they choose taxis. There was no significant association between school and taxi usage patterns over any course of time. In conclusion, taxis are not the main mode of transportation for parents to take their children to and from school.

As we conducted the multiple linear analysis for 24 hours in a one-day cycle, the coefficients of independent variables at each hour were compared vertically. Thus, we can find the factors that have the largest correlation with the taxi usage index. The results indicated that the factors most associated with the taxi usage index varied in different periods. From 10:00 am to 6:00 pm, the correlation coefficients between house price, road density, and taxi usage index were the largest. From 6:00 pm to 10:00 am the next day, the correlations between residence and the taxi usage index are the largest, which is probably because it is when residents are at home and thus depart from home to other locations.

5 Conclusions and discussion

Comparisons among different transport modes over space offer significant dimensions for understanding taxi usage patterns, namely travel mode choice and transport ridership. However, the existing literature focuses only on one of these. Regarding related factors, existing studies have focused on those associated with transport ridership and travel mode choice, leaving the factors associated with the temporal dynamics of taxi usage patterns unexplored. To fill these gaps, we proposed a new indicator to measure taxi usage in a grid of 1 km × 1 km cells in Beijing, utilizing multi-source data harvested in the same week. Moreover, we examined the spatiotemporal dynamics of the taxi usage index in a 24-hour cycle and explained the driving factors behind these dynamics with alternative transport, socioeconomic factors, and built environment factors. Overall, this research offers an example of utilizing big data in the study of travel behavior with a novel research framework. The calculation method for the taxi usage index can also be extended to other cities and cases.

Some policy implications can be generated from this study. First, taxi ridership should not be solely considered when allocating taxi services, and taxi usage interacting with other travel modes should be considered in transportation management. Second, the temporal dynamics of the factors influencing taxi usage should be included in taxi ridership sourcing. Factors such as residence,

enterprise, healthcare facility, shopping spot, catering facility, recreation facility, and tourist spot had different associations with taxi usage patterns in different time, while others such as house price, road density, residence, and school did not. These findings imply that we should consider the influence of the built environment differently; specifically, taxi cruises around residences all day, while more taxi services are advised to concentrate on enterprises, healthcare facilities, and tourism spots in the daytime. In addition, more taxi services should be in areas with recreation facilities, catering facilities, and shopping spots in the evening.

However, one limitation was the POI data, which could not reflect the scale of each point. Hence, investigation of the density and areas of the facilities was limited. In addition, the Beijing census data were collected in 2010, which has a 7-year lag with taxi trip data and mobile signaling data. However, it is the most recent data that we could collect, as the census in China is carried out every 10 years. In future work, more evidence on the spatiotemporal dynamics of the taxi usage index and its related factors will be explored and compared to summarize and extract general findings. In addition, taxi usage often presents significantly different patterns on weekdays and weekends. This study mainly focuses on the temporal dynamics of taxi usage, and the disparity in taxi usage between weekdays and weekends will be further explored.

Acknowledgements

We thanks to Smart Steps Digital Technology CO. offering mobile signal data for this study.

Code availability

Not applicable.

Authors' contributions

Jiaoe Wang: Conceptualization, Writing – original draft, Writing – review & editing. Fangye Du: Data collection, Formal analysis, Visualization, Writing – original draft. Jie Huang: Conceptualization, Writing – original draft, Writing – review & editing. Yu Liu: Resources, Writing – review & editing. The author(s) read and approved the final manuscript.

Funding

This research was financially supported by the National Natural Science Foundation of China (Grants No. 42121001 and 42225106), and Youth Innovation Promotion Association of Chinese Academy of Sciences (Grant No. 2021049).

Availability of data and materials

Mobile signaling data is offered Smart Steps Digital Technology CO. Taxi trips included in this paper are confidential. Housing price data are obtained from Fang.com. Built environment factors are calculation from OpenStreetMap and Baidu map.

Declarations

Competing interests

We confirm no conflicts of interest.

Received: 11 November 2022 Revised: 12 February 2023 Accepted: 26

February 2023

Published online: 06 March 2023

References

- Bachir, D., Khodabandelou, G., Gauthier, V., Yacoubic, E. M., & Puchinger, J. (2019). Inferring dynamic origin-destination flows by transport mode using mobile phone data. *Transportation Research Part C: Emerging Technologies*, 101, 254–275.
- Böcker, L., Amen, V. P., & Helbich, M. (2017). Elderly travel frequencies and transport mode choices in greater Rotterdam, the Netherlands. *Transportation*, 44, 831–852.
- Buehler, R. (2011). Determinants of transport mode choice: A comparison of Germany and the USA. *Journal of Transport Geography*, 19(4), 644–657.
- Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr., J., & Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, 26, 301–313.
- Candia, J., Gonzalez, C. M., Wang, P., Schoenharl, T., Madey, G., & Barabasi, A. (2008). Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, 41, 224015.
- Cao, X., Mokhtarian, P., & Handy, S. (2009). The relationship between the built environment and nonwork travel: A case study of northern California. *Transportation Research Part A: Policy and Practice*, 43(5), 548–559.
- Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 2, 199–219.
- Çetin, T., & Yasin Eryigit, K. (2011). Estimating the effects of entry regulation in the Istanbul taxicab market. *Transportation Research Part A: Policy and Practice*, 45(6), 476e484.
- Chiou, Y., Jou, R., & Yang, C. (2015). Factors affecting public transportation usage rate: Geographically weighted regression. *Transportation Research Part A: Policy and Practice*, 78, 161–177.
- Chow, L. F., Zhao, F., Liu, X., Li, M. T., & Ubaka, I. (2006). Transit ridership model based on geographically weighted regression. *Transportation Research Record Journal of the Transportation Research Board*, 1972(1), 105–114.
- Du, F., Mao, L., & Wang, J. (2021). Determinants of travel mode choice for seeking healthcare: A comparison between elderly and non-elderly patients. *Journal of Transport Geography*, 92, 103023.
- Farinloye, T., Mogaji, E., Aririguzoh, S., & Kieu, A. T. (2019). Qualitatively exploring the effect of change in the residential environment on travel behaviour. *Travel Behaviour and Society*, 17, 26–35.
- Freeman, S. (2018). *Relationships of structural characteristics and homeowner socioeconomics with urban vegetation and bird communities*. Texas Tech University.
- Gan, Z., Yang, M., Feng, T., & Timmermans, J. P. H. (2020). Examining the relationship between built environment and metro ridership at station-to-station level. *Transportation Research Part D: Transport and Environment*, 82, 102332.
- Gong, L., Liu, X., Wu, L., & Liu, Y. (2016). Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, 43(2), 103–114.
- He, Y., Zhao, Y., & Tsui, K. L. (2019). Exploring influencing factors on transit ridership from a local perspective. *Smart and Resilient Transportation*, 1, 2–16.
- Huang, J., Hu, Y., Wang, J., & Li, X. (2021). A tale of two cities: Jobs-housing balance and urban spatial structures from the perspective of transit commuters. *Environment and Planning B: Urban Analytics and City Science*, 48(6), 1543–1557.
- Huang, J., Levinson, D., Wang, J., & Jin, H. (2019). Job-worker spatial dynamics in Beijing: Insights from smart card data. *Cities*, 86, 83–93.
- Huang, J., Levinson, D., Wang, J. E., Zhou, J. P., & Wang, Z. J. (2018). Tracking job and housing dynamics with smartcard data. *Proceedings of the National Academy of Sciences*, 115(50), 12710–12715.
- Kang, C., & Qin, K. (2016). Understanding operation behaviors of taxicabs in cities by matrix factorization. *Computers, Environment and Urban Systems*, 60, 79–88.
- Kuby, M., Barranda, A., & Upchurch, C. (2004). Factors influencing light-rail station boardings in the United States. *Transportation Research Part A: Policy and Practice*, 38(3), 247.

- Liu, Y., Singleton, A., Arribas-bel, D., & Chen, M. (2021). Identifying and understanding road-constrained areas of interest (AOIs) through spatiotemporal taxi GPS data: A case study in new York City. *Computers, Environment and Urban Systems*, 86, 101592.
- Liu, Y., Wang, F., Xiao, Y., & Gao, S. (2012). Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning*, 106, 73–87.
- Mo, B., Shen, Y., & Zhao, J. (2018). Impact of built environment on first- and last-mile travel mode choice. *Transportation Research Record: Journal of the Transportation Research Board*, 2672(6), 40–51.
- Qian, X. W., & Ukkusuri, V. S. (2015). Spatial variation of the urban taxi ridership using GPS data. *Applied Geography*, 59, 31–42.
- Ratti, C., Pulselli, R. M., Williams, S., & Frenchman, D. (2006). Mobile landscapes: Using location data from cell phones for urban analysis. *Environment and Planning B: Urban Analytics and City Science*, 33(5), 727–748.
- Schaller, B. (2005). A regression model of the number of taxicabs in US cities. *Journal of Public Transportation*, 8(5), 63.
- Schwaben, T., & Mokhtarian, P. L. (2005). What affects commute mode choice: Neighborhood physical structure or preferences toward neighborhoods? *Journal of Transport Geography*, 13(1), 83–99.
- Taylor, B. D., & Fink, C. N. Y. (2013). Explaining transit ridership: What has the evidence shown? *Transportation Letters*, 5(1), 15–26.
- Wang, F., & Ross, L. C. (2018). Machine learning travel mode choices: Comparing the performance of an extreme gradient boosting model with a multinomial logit model. *Transportation Research Record: Journal of the Transportation Research Board*, 2672(47), 35–45.
- Wang, J., Du, F., Huang, J., & Liu, Y. (2020). Access to hospitals: Potential vs. observed. *Cities*, 100, 102671. <https://doi.org/10.1016/j.cities.2020.102671>
- Wang, J., Huang, J., & Du, F. (2020). Estimating spatial patterns of commute mode preference in Beijing. *Regional Studies, Regional Science*, 7(1), 382–386.
- Wang, J., Li, Y., Jiao, J., Jin, H., & Du, F. (2022). Bus ridership and its determinants in Beijing: A spatial econometric perspective. *Transportation*. <https://www.s11116-021-10248-7>.
- Wesolowski, A., Stresman, G., Eagle, N., Stevenson, J., Owaga, C., Marube, E., Bousema, T., Drakeley, C., Cox, J., & Buckee, O. C. (2014). Quantifying travel behavior for infectious disease research: A comparison of data from surveys and mobile phones. *Scientific Reports*, 4, 5678.
- Yang, H., Dijst, M., Feng, J., & Ettema, D. (2019). Mode choice in access and egress stages of high-speed railway travelers in China. *Journal of Transport and Land Use*, 12(1), 701–721.
- Yao, X., Wu, L., Zhu, D., Gao, Y., & Liu, Y. (2019). Visualizing spatial interaction characteristics with direction-based pattern maps. *Journal of Visualization*, 22(3), 555–569.
- Yuan, Y., Raubal, M., & Liu, Y. (2012). Correlating mobile phone usage and travel behavior – A case study of Harbin, China. *Computers, Environment and Urban Systems*, 36(2), 118–130.
- Zhao, J., Deng, W., Song, Y., & Zhu, Y. (2013). What influences metro station ridership in China? Insights from Nanjing. *Cities*, 35, 114–124.
- Zhong, C., Manley, E., Arisona, M. S., Batty, M., & Gerhard, S. (2015). Measuring variability of mobility patterns from multiday smart-card data. *Journal of Computational Science*, 9, 125–130.
- Zhu, Y., Chen, F., Wang, Z., & Deng, J. (2018). Spatio-temporal analysis of rail station ridership determinants in the built environment. *Transportation*, 46(6), 2269–2289.
- Zhu P., Huang J, Wang J, Liu Y., Li J., Wang M., & Qiang W. (2022). Understanding taxi ridership with spatial spillover effects and temporal dynamics. *Cities*, 125(2), 103637.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.