MYU Tokyo

S & M 3899

Transit-oriented Development and Traffic Congestion in Beijing: An Empirical Analysis Based on Geo Big Data

Yongxuan Liu[1] and Cong Liao[2*]

[1]Beijing Institute of Surveying and Mapping, No. 15 Yangfangdian Road, Haidian District, Beijing 100045, China
[2]CAUPD Beijing Planning & Design Consultants LTD., No. 5 Chegongzhuang West Road, Beijing 100044, China

(Received July 29, 2024; accepted January 20, 2025)

Keywords: TOD performance, travel behavior, node-place model, multilevel model, Beijing

Transit-oriented development (TOD) has become a popular planning strategy for many large cities moving toward sustainable urban spatial structure and transport systems. However, the challenges that need to be overcome in the development process of TOD have not been extensively investigated in previous studies owing to insufficient data on TOD areas, which would reduce the potential benefits of TOD. In this study, we examine the relationship between traffic congestion and TOD performance, and measure the relative importance of TOD components on congestion in station areas of Beijing during weekday morning peak hours. Using the node-place model and a clustering method, we will use geo-tagged big data, such as taxi trajectory data and metro card swiping data originating from multiple sensors, to assess TOD performance and traffic congestion in Beijing subway station areas. An analytical framework is proposed to understand their relationships, and a quantitative analysis is conducted using a multilevel regression model. The results indicate that 94.58% of metro station areas in Beijing face light or higher congestion in morning peak hours. The performance of the 'transport' component needs to be improved compared with the 'development' and 'interrelation' components of TOD. The 'transit', 'development', and 'interrelation' components of TOD are significantly associated with Beijing's morning peak hours. It is difficult to reduce congestion simply by hoping for a traffic modal shift to public transit use because of the lack of public transportation availability. Optimizing the structure of travel demand in TOD areas and creating a pedestrian friendly environment were shown to be related to congestion reduction. The results of this study provide insights for developing targeted strategies to reduce congestion in TOD areas and promote better TOD performance in station areas.

1. Introduction

Since its rediscovery in the 1990s, transit-oriented development (TOD) has been not only an important concept in academics, but also a strategic practice toward smart growth in urban and transportation planning.(1) The concept of TOD advocates a high-density development pattern, land use diversity, and pedestrian- and bicycle-friendly environments around public transit

*Corresponding author: e-mail: liaocong233@pku.edu.cn
https://doi.org/10.18494/SAM5255

ISSN 0914-4935

nodes as well as integration of land use and transport systems.[2] To better guide the development by TOD, many studies have been conducted to analyze the effects of TOD.[3,4]

Many groups have investigated various positive impacts of TOD. At the city level, TOD helps shape a polycentric urban spatial structure and mitigate urban sprawl.[5] The TOD-informed urban spatial structure will influence individuals' travel behavior, such as travel demand and modal split.[6,7] People living in TOD areas with convenient public transit and pedestrian-friendly environments tend to walk and take public transportation more and drive less.[8] The TOD-informed urban spatial structure also shows that time-competitive public transit and land use management within TOD areas have spatial accessibility benefits.[9,10] Moreover, greenhouse gas emissions and economic costs for TOD users can be reduced owing to few vehicle miles traveled (VMT).[11] At the station level, TOD influences residents attitude or propensity to live in certain types of TOD neighborhoods (i.e., residential self-selection).[12] TOD neighborhoods with different built environments have been found to have a dominant influence on travel mode choice. The development of diversified economic activities in TOD areas also leads to higher relative property valuations.[13] In addition, a more pedestrian-friendly environment will promote more physical activities and healthier lifestyles.[14]

While previous studies have revealed some benefits of implementing TOD, research studies on the challenges that need to be overcome in the development process of TOD are lacking.[15] On the one hand, there is insufficient data on TOD areas to quantify those challenges. TOD areas usually encompass only a few hundred meters around a station.[16,17] This makes it difficult for traditional statistics to support TOD-related empirical analysis. On the other hand, it is natural to target well-developed TOD areas that have typical TOD characteristics for study.[11] For those areas that are developing or have the potential to be designated as TOD areas, the literature provides limited clues to understanding their characteristics and effects. The disintegration of transportation systems and land development in these areas makes research on TOD evaluation and policy analysis more challenging.[18] However, such regions need more precise policy guidance for TOD development. The potential benefits of TOD will be undermined by any lack of attention to the challenges associated with TOD in these areas.

Identifying these TOD-related challenges is important to further improve the performance of TOD areas. In several studies on TOD-related challenges, traffic congestion, which has become one of the most serious problems in urban development worldwide, is a topic of discussion. Congestion can lead to several negative impacts, such as high economic cost, time loss, pollution, and health risks.[19] Although in the long run, in well-designed TOD areas, congestion can be reduced by restricting the use of cars and decreasing VMT, the reality of disruptive congestion in developing TOD areas necessitates a more detailed analysis of traffic congestion. A recent study of traffic congestion related to TOD was carried out by Zhu *et al.*[20] Their findings suggest that public transit ridership increases in TOD areas, but local congestion worsens with total delays increasing by 15.6% in the simulated scenario because the mixed land use in the TOD area generates more trips than trips switched to non-automobile modes. The risk of congestion in TOD areas can arise for several reasons. First, mixed land uses will result in population clustering and increased traffic around stations. Population clustering around station areas without a simultaneous modal shift from driving to walking, cycling or public transit will increase the risk of encountering traffic congestion.[21,22] Second, car use remains high in TOD areas because the public transit network is not fully saturated and covers only a limited area.(23) Third, residents' attitudes toward TOD and preferences for using public transit influence the choice of the traffic mode.(24)

To the best of our knowledge, there are few studies on analyzing the relationship between TOD and traffic congestion conditions in real-world scenarios at the city level. A quantitative study of the relationship between TOD performance and traffic congestion is valuable for goal-oriented TOD improvements. In this study, we attempt to investigate the relationship between TOD performance and traffic congestion. The emergence of geotagged big data originating from multiple sensors provides granular data previously unattainable in TOD areas through traditional datasets. Geotagged big data, including taxi trajectory data and subway smart card records will be used to identify traffic congestion during weekday morning peak hours in TOD areas. TOD performance will be evaluated on the basis of the node-place model. The relationship between them will be tested using a multilevel regression model (MLM) in which the effects of different TOD types are controlled. We will use Beijing as the case study area.

The structure of this study is as follows. Section 2 is an explanation of materials and methods. In Sect. 3, we describe the results and analysis. The discussion is presented in Sect. 4. In Sect. 5, we give the conclusions and policy implications.

2. Materials and Methods

2.1 Study area

In this study, we chose the subway station areas in Beijing as the case study area. As the capital of China, Beijing is one of the most important metropolises in the country with 21.73 million inhabitants and 12.2 million employees in 2016 (Beijing Municipal Statistics Bureau, 2017). As early as 1969, the construction of Beijing's first subway was begun. Since the beginning of the 21st century, the Beijing municipal government has placed a high priority on public transport priority policies. By the end of 2015, the total length of Beijing's subway system was 666 km. In 2016, Beijing had 20 subway lines and 277 subway stations. Beijing also implements good policies to promote TOD. In the Beijing Urban Master Plan (2016–2035), the adoption of TOD strategies to promote the optimization of the transport system and urban spatial structure is stipulated. Therefore, identifying traffic congestion in Beijing's TOD areas can help further improve the TOD performance of station areas.

Since the 1990s, Beijing's TOD has undergone significant progress. Early explorations began with individual transit stations, with the Sihui station being the first area for the collaborative development of station construction and property projects. After the 2008 Beijing Olympics, the city intensified its TOD efforts, proposing the creation of micro-centers around transit stations, establishing mechanisms for comprehensive land use around these stations, and designating areas for integrated use around newly built stations when conditions allow. Station development has been categorized into station and facility planning, utilization of residual space, and upper-level development. In 2020, Beijing approved the "Directory of Transit Micro-Centers in Beijing (First Batch)". Transit micro-centers are defined as spaces that fully integrate and interact with transit stations, have high accessibility, a high degree of land intensification, diverse urban functions, and a strong sense of place and identity. On the basis of the actual conditions of various districts, 71 micro-centers were selected, covering 14 districts.

Over the past 30 years, Beijing has made significant efforts to develop public transportation, and by 2021, the green travel rate in the central urban area reached 74%. Among these, the proportion of trips taken by subway has increased substantially, reaching 21.6% in 2020, gradually replacing bus travel as the preferred mode of public transportation. With the implementation of car quota control policies, the proportion of trips made in private cars has stabilized at around 35%. According to the 2017 Beijing Transport Development Annual report published by the Beijing Transport Institute (http://www.bjtrc.org.cn/), the six most urbanized districts in Beijing had a modal share of public transport (bus and subway) of more than 49% in 2016. The motor vehicle fleet in Beijing is 5.7 million. Of these, 4.5 million are cars. Figure 1 shows an overview of the study area.

2.2 Analytical framework

An analytical framework is proposed to establish the relationship between TOD performance and traffic congestion. In this study, we analyze the specific components and performance of TOD using the node-place model proposed by Bertolini.[(1)] This model integrates land use and transport into a unified analytical framework, considering public transit stations as nodes in a transportation network and their surroundings as development places. The model breaks down the node value and place value into indicators that represent the performance of nodes and places, reflecting them onto x and y coordinates. It categorizes existing stations into the following five types. (1) Balanced: transportation and land use are mutually supportive and sustainable. (2) Pressure: located in the upper part of the balanced zone; the values of both node and place have reached their limits, potentially leading to competition that hinders mutual

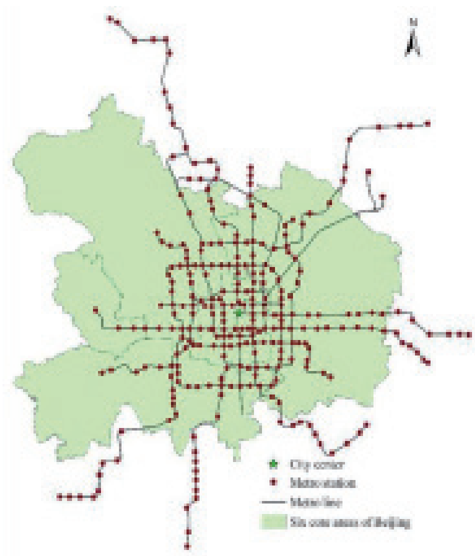

Fig. 1. (Color online) Overview of the study area.

development. (3) Subordinate: located at the lower part of the balanced zone; low levels of transportation supply and land use, with significant potential for development. (4) Imbalanced Node: located in the upper half of the diagonal; transportation performance is far superior to land use. (5) Imbalanced Place: located in the lower half of the diagonal; land is overdeveloped, exceeding the service level that the current transportation capacity can provide. Using this model, a large number of studies on assessing the TOD performance of station areas in different contexts, such as Switzerland,(16) Australia(25) and Portugal,(26) have been carried out. Recent theoretical extensions of the node-place model emphasize the importance of the interrelationship component (Oriented) between transit (T) and development (D),(27) as the lack of morphological and functional links between transit and station area development will result in transit adjacent development (TAD).

To comprehensively investigate the relationship between TOD performance and traffic congestion conditions, the analytical framework shown in Fig. 2 is adopted to illustrate the relationship between TOD and congestion using an extended form of the node-place model.

First, there is an increased demand for travel in TOD areas owing to development with high density and land use diversity.(9) These demands are related to work commuters and travel needs associated with daily activities and services.(8) In the context of the jobs–housing imbalance, scholars have found a significant amount of long-distance commuting.(28) These travel demands lead to an increase in traffic volume and add the risk of encountering traffic congestion.

Second, the availability of public transport helps to reduce the congestion. Increased level of public transit service will reduce reliance on automobiles.(29) A modal shift from driving to

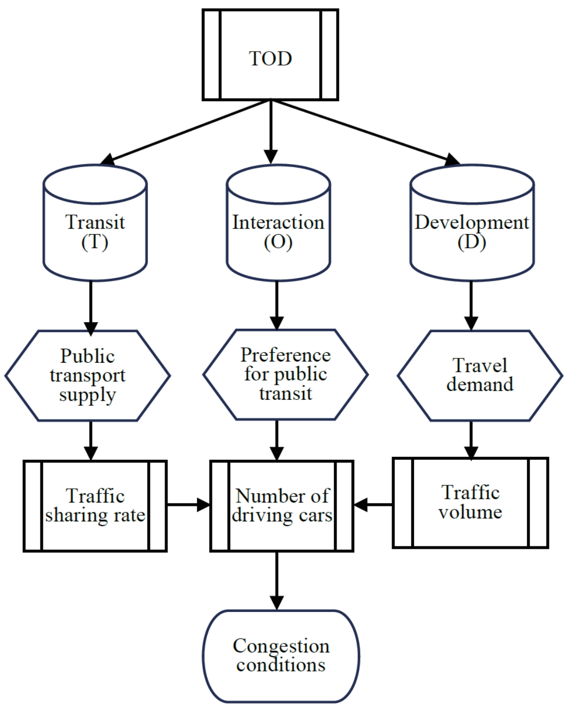

Fig. 2. Analytical framework of association between TOD performance and congestion.

taking public transit will reduce the traffic share of automobiles, which can reduce congestion in TOD areas. Conversely, inadequate public transport availability will lead to increased car use and higher congestion risk.(20) In addition, poor metro experiences and peak hour subway restrictions can cause people to shift to cars.

Third, the interrelationship between node and place will influence the individuals' preferences for using public transit and nonmotorized modes of travel. For instance, a modal shift from driving to walking or cycling in a pedestrian-friendly environment will reduce the risk of experiencing traffic congestion in TOD areas.(30) Convenient connectivity between stations and residences will also encourage people to shift to public transportation. The increased time cost of driving a car on poor roads in TOD areas would also affect people's preference for using public transportation.

In short, we hypothesize that congestion conditions are influenced by travel demand, public transit sharing rates, and individual preferences for public transit. We also hypothesize that these factors are sensitive to the TOD performance.

2.3 Datasets

Four types of geotagged big data were collected and applied, as follows: (1) weekday records of smart metro card use within one week (2016.06.20–2016.06.24), obtained from Beijing Bus Group; (2) resident and employed population data in 2017, inferred from China Unicom's mobile phone data; (3) point-of-interest data for Beijing in 2016, obtained from Amap (https://www.amap.com). These three types of data were used in the TOD evaluation. (4) Taxi GPS trajectory data on weekdays during one week (2016.06.20–2016.06.24) were used to analyze traffic congestion conditions in the TOD areas.

Before evaluating the performance of TOD for station areas, it is important to determine the size of the buffer zone delineating the site area. In previous studies, the buffer size varied from 400 to 1500 m in Euclidean distance around the station in different contexts.(31) In this study, referring to Lyu *et al.*(27) for Beijing's TOD evaluation, we used a 700 m Euclidean distance as the buffer size to define the TOD area.

The data format and preprocessing are presented. Table 1 shows the format of the preprocessed taxi trajectory data. There are over 300 million trajectory records in the entire study area and over 20 million records in the TOD area during the study period. The average sampling interval of cabs is about 30 s. We process these geotagged trajectory data by parallel programming in Python. The trajectory data within the TOD area were first extracted. Among these station areas, the extraction of trajectory records was unsuccessful in two station areas. To reduce data bias, we excluded them along with other invalid records in the preprocessing process. To reduce the effect of stopping on the average speed, the sampling data were removed and obtained when the engine switch flag was off.

Table 1
Format of taxi trajectory data.

Plate number of vehicles	Update time	Latitude of upload position	Longitude of upload position	Speed (km/h)
33756	20160620075425	40.38	116.49	18

The data formats of the smart card data and demographic data are shown in Tables A1 and A2 in the Appendix. The Unicom Smart Footprint Platform records relevant data from millions of mobile users and performs necessary cleaning and preprocessing on this data. From the monthly stay duration and frequency of stay, it identifies users' residences, workplaces, and primary activity locations. After preprocessing the data to remove irrelevant attributes and invalid records, a sample of each smart card record contains six attributes. Each record in the demographic data includes the coordinates of the place of residence, the coordinates of the place of employment, and the number of residents and employed persons.

To eliminate the inconsistencies between the data, we normalized the data by min–max normalization. All positively and inversely correlated indicators were normalized to 0 to 1 using different formulas. Similar treatment was applied to the congestion-related variables described below.

2.4 Methodology

First, we use taxi trajectory data in Beijing to capture traffic congestion during morning peak hours in the TOD areas of Beijing. Second, we identify and measure specific TOD components. Indicators related to transit (T), development (D), and their interrelation (O) in TOD areas are used as TOD components to assess TOD performance through principal component analysis (PCA) and the k-means clustering method. Finally, we apply MLM to test the proposed framework.

2.4.1 Identifying traffic congestion in station areas

There is no established definition and evaluation criteria for identifying traffic congestion. Traffic flow speed is considered to be an intuitive factor in identifying congestion. The difference between the average speed and the threshold speed (or free flow speed) on a road network is often used to describe congestion conditions.(32) The average speed in the road network can be converted into actual travel time. Some composite indices can also be used to describe congestion, such as the traffic congestion index(33,34) and the traffic performance index.(35) Some studies have revealed a conversion relationship between these indices and the average speed.(36)

To determine congestion conditions, taxi GPS trajectory data have been widely used in recent years to obtain traffic flow speeds.(37) Compared with fixed sensors installed on certain road sections, taxi GPS trajectory data have a wider coverage and are more easily accessible. In this study, following previous research on congestion identification,(38) we identify congestion based on the average vehicle speed and threshold speed in Beijing. The average vehicle speed on the road network within the TOD areas is first calculated. Then, the congestion levels of selected roads within the TOD area are classified in accordance with the Beijing traffic congestion level classification standard.

According to data released by the Beijing Transportation Committee, traffic congestion in Beijing is primarily concentrated during two peak periods: morning and evening. The morning peak usually occurs between 7:00 and 9:00 a.m., which is the busy period for commuting to work and school. Compared with the evening peak, travel destinations during the morning peak are more clearly defined. To analyze the relationship between TOD and congestion, we have chosen the morning peak as the study period. According to some existing Beijing-based case studies, 7:00 a.m.–9:00 a.m. is considered to be the morning peak hours based on empirical or some data-based analysis results,(39) although peak-hour commuting includes departure before and after peak hours.(40) Therefore, we also designated 7:00 a.m.–9:00 a.m. as the morning peak hours. The congestion calculation procedure for TOD areas is as follows.

(1) Extract the GPS trajectory records of all TOD areas from 7:00 a.m.–9:00 a.m.
(2) Calculate the total path length and travel time of each vehicle in each TOD area by vehicle license plate number.
(3) Calculate the average travel speed of each vehicle on all paths in all TOD areas.
(4) Take the average speed of all paths on weekdays as the average speed of the road network within the TOD area.

According to the definition of congestion given by the Beijing Municipal Bureau of Quality and Technology Supervision (2011), the types of traffic state can be classified as smooth, basic smooth, light, moderate, and severe congestion. Table 2 shows different congestion levels corresponding to different average speed thresholds. Considering the development characteristics of TOD areas, we classify the congestion using the criteria of branch roads.

2.4.2 Evaluation of specific components and performance of TOD areas

Referring to the indicators used in previous studies on TOD evaluation, we constructed a system of indicators for TOD components. The specific descriptions of the indicators and measurement are presented in Table A3 in the Appendix.

The expected impact of the TOD component on congestion is assumed. The development characteristics (D component) are considered to be related to travel demand in the station area. TOD advocates high-density and diverse development in the station areas. High-density development is a sign of a high TOD level and therefore potentially high travel demand.(41) Greater travel demand will increase the potential risk of congestion. We chose total employment (D1), total resident population (D2), and area of buildings around station areas (D3) to represent high-density development in TOD areas(42) and expected negative correlations between D1, D3, and the average speed in the station area. Because of the diversity of development and public transit in the station areas, people can access a variety of services in the TOD area and reduce the number of long-distance trips. Therefore, we expect a positive correlation between D2 and the average speed of the station area. The number of educational/health/cultural establishments

Table 2
Classification of congestion levels for different road types.

Road level	Smooth (km/h)	Basic smooth (km/h)	Light congestion (km/h)	Moderate congestion (km/h)	Severe congestion (km/h)
Expressway	$V > 65$	$50 < V \leq 65$	$35 < V \leq 50$	$20 < V \leq 35$	$V \leq 20$
Arterial road	$V > 40$	$30 < V \leq 40$	$20 < V \leq 30$	$15 < V \leq 20$	$V \leq 15$
Branch road	$V > 35$	$25 < V \leq 35$	$15 < V \leq 25$	$10 < V \leq 15$	$V \leq 10$

Note: V represents the average speed on the road.

(D4) and the number of retail/hotel and catering establishments (D5) show the relationship between different economic activities and congestion. A greater diversity of activities in the station areas will increase traffic demand and potentially increase vehicle use. We expect a negative correlation between D4, D5, and the average speed of the station area.

The traffic characteristics (T component) are mainly the functional and morphological characteristics of the subway system. A possible relationship between traffic characteristics and station area congestion can be expected, as greater use of public transport leads to a decrease in the use of motor vehicles, thereby increasing the mode share of public transportation.(22) We included the total number of subway passengers from 7:00 a.m.–9:00 a.m. (T1) and the density of bus stops around the station area (T2) to represent the capacity of public transportation(17) and expected a positive relationship between T1, T2, and the average speed in the station area. We used the number of directions served by the subway (T3), the departure intervals of the subway (T4), the number of entrances and exits (T5), and average distance between subway stations (T6) to determine the effects of user-friendliness of the public transit system on congestion. We expect a negative correlation between T3–T6 and the average speed in TOD areas.

The interrelationship characteristics (O component) emphasize the preference for using the subway. A pedestrian-friendly environment of the station area contributes to the willingness to walk from the station to its surrounding destinations. An increase in the number of people willing to use the subway will correspondingly reduce the use of motor vehicles.(8) We combined the average distance of a walkway (O1), road density (O2), and number of signals (O3), and reflected the pedestrian environment(40) and expect a positive correlation between O1, O2, O3, and the average speed in the station area. The average distances from the station to the workplace (O4) and from the station to the residence (O5) indicate the convenience of using the subway. If the distance from a station to surrounding destinations is too far, the likelihood of using the motor vehicle mode would increase. We expect a negative correlation between O4, O5, and the average speed in the station area.

To evaluate the TOD performance of the station areas, we clustered the station areas on the basis of the constructed indicators using the *k*-means clustering method. After clustering, each type of cluster possesses some similarity in certain TOD elements. The *k*-means method is a commonly used clustering approach. The goal of this method is to maintain homogeneity within the station areas and heterogeneity between the station areas. The basic computational process starts by randomly selecting *k* elements as *k* clustering centers. Then the similarity of the remaining samples to the *k* centers is calculated separately. These elements are grouped into the clustering centers with the greatest similarity. The clustering criterion can be defined as

$$SE = \sum_{k=1}^{K} \sum_{t=1}^{n} \left(p - m_k \right)^2 , \tag{1}$$

where p denotes the sample involved in clustering; k denotes the index of clustering centers; m_k denotes the mean of the kth clustering center, and t denotes the index of n attributes. To reduce the complexity of the data, PCA is first used to generate a new set of uncorrelated low-dimensional variables before clustering.

The determination of the number of clusters is very important for the interpretation of the clustering results. There are dozens of metrics and algorithms that can be used to determine the number of clustering centers, such as the sum of square errors. To facilitate further analysis and interpretation, the number of clustering centers should be as cognitive as possible. In this study, the number of clusters is determined using the silhouette coefficient and interpretability.

2.4.3 Linking specific components of TOD to station area congestion

To test whether the proposed analytical framework can explain the traffic congestion during the Beijing morning peak hours, we first run an OLS model with specific components of the TOD as explanatory variables and congestion as the dependent variable. The model can be simplified as

$$Y_i = \beta_0 + \beta_1 X_i + C_i + \epsilon_i , \tag{2}$$

where Y_i refers to the average speed of the TOD areas, X_i is the independent variable associated with a specific TOD segment, C_i is a vector of control variables, and ϵ_i is the error term of TOD area *i*.

Station areas in the same TOD cluster are similar in some characteristics. Ignoring the differences between different TOD clusters can lead to an underestimation of the standard error of the regression coefficients. To control the effects of different TOD clusters, we use MLM. MLM is suitable for datasets with nested structures.(43) In MLM, the residual components are recognized at each level, which will overcome the limitation of independent samples in the same group. The equation of MLM is

$$Y_{ij} = \beta_{0j} + \beta_1 Transit_{ij} + \beta_2 Oriented_{ij} + \beta_3 Development_{ij} + C_{ij} + e_{ij} , \tag{3}$$

$$\beta_{0j} = \beta_0 + \mu_{0j} , \tag{4}$$

$$\mu_{0j} \sim \mathrm{N}\left(0, \sigma^2{}_{\mu 0}\right), \tag{5}$$

$$e_{ij} \sim \mathrm{N}\left(0, \sigma^2{}_{e}\right), \tag{6}$$

where Y_{ij} represents the average speed of station *i* in the *j*th cluster, $Transit_{ij}$ $Oriented_{ij}$, and $Development_{ij}$ represent the specific TOD component of station *i* in the *j*th cluster, and β_1, β_2, and β_3 are coefficients. C_{ij} is a vector of control variables. β_{0j} is the intercept. e_{ij} and μ_{0j} are the error terms at levels *i* and *j*, respectively, and usually follow a normal distribution. $\sigma^2{}_{\mu 0}$ and $\sigma^2{}_e$ are the variances at levels *i* and *j*, respectively.

To control the effect of other non-TOD factors on traffic congestion, we include some control variables in the model. C1, C2, and C3 are used to represent road accessibility and connectivity. Higher-rated roads typically have more lanes, greater roadway capacity, and less road disruption and are less likely to be congested. We expect a positive correlation between C1–C3 and the average speed in the station areas.

3. Results and Analysis

3.1 Output results of traffic congestion in the study area

Table 3 shows the traffic status statistics of each station area in Beijing during the morning peak period. The traffic congestion conditions during the morning peak in Beijing are very clear. More than 94.58% of the station areas are lightly congested or worse. The total average speed around all stations is 15.23 km/h, which is close to the threshold of moderate congestion. Among them, none of the station areas are in a smooth state, 13 station areas are in a basic smooth state, 113 station areas are lightly congested, 113 station areas are moderately congested, and 36 station areas are severely congested. The average speed of the road network in the station area ranged from a minimum of 6.46 km/h to a maximum of 30.18 km/h. Figure 3 shows the statistics of traffic states for station areas during morning rush hours.

To analyze the spatial distribution of station areas with different congestion states, the nearest neighbor index was calculated. The ratio of observed distance to a random expected distance (NNRatio index) indicates the degree of dispersion clustering. If the ratio is greater than 1, the distribution tends to be dispersed. If the ratio is less than 1, the distribution tends to be clustered.

Table 4 summarizes the results of average nearest neighbor analysis of station areas with different congestion states. The Z score indicates whether the identified pattern is statistically significant. The bottom row in Table 4 indicates that the overall distribution pattern of station areas in the study area is clustered, with the NNRatio of 0.73. The distribution pattern of station areas with severe congestion is clustered, with the NNRatio of 0.73. It is conceivable that the distribution of station areas with moderate congestion may be more clustered than those with severe and light congestion. Compared with the distribution characteristics of station areas with severe congestion, the clustered characteristics of station areas with light congestion are less obvious. Unlike station areas with congestion, the distribution of station areas in the basic smooth state is dispersed with an NNRatio of 1.47.

3.2 TOD performance in the study area

Table 3
Statistics of traffic states for station areas during morning rush hours.

Types of congestion	Threshold	Average speed (km/h)	Number of stations	Proportion (%)
Smooth	$V > 25$		0	0
Basic smooth	$25 < V \leq 35$	27.50	13	4.70
Light	$15 < V \leq 25$	18.69	113	40.79
Moderate	$10 < V \leq 15$	12.38	113	40.79
Severe	$V \leq 10$	8.88	36	13.00
No records		—	2	0.72

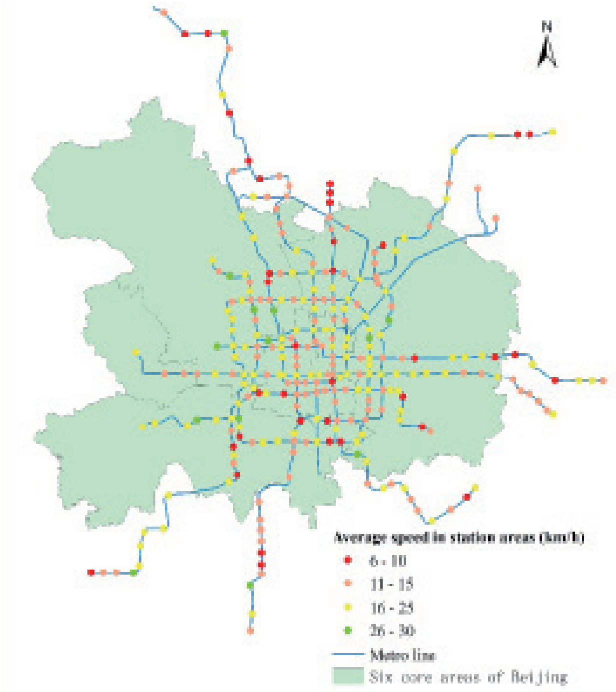

Fig. 3. (Color online) Spatial distribution of congestion states of station areas.

Table 4
Average nearest neighbor analysis results of station areas in different congestion states.

Types of congestion	Number of station areas	Expected distance (m)	Observed distance (m)	NNRatio	*Z* score
Basic smooth	13	4861.56	7122.50	1.47	3.21[a]
Light	113	2349.91	1964.91	0.83	−3.33[b]
Moderate	113	2608.63	1810.04	0.69	−6.22[b]
Severe	36	4164.23	3032.13	0.73	−3.12[a]
All	275	1679.31	1228.06	0.73	−8.52[a]

a: Significant at 0.01.
b: Significant at 0.001.

In this section, we describe the overall characteristics of TOD performance. The synthetic node, place, and their interrelationship indices are constructed from equally weighted averages of the corresponding indicators.(16) The results of the synthetic indices show that Beijing's overall TOD performance needs further optimization, with mean values of 0.30, 48, and 0.58 for the node, place, and their interrelationship indices, respectively. The mean index values for place and interrelation are higher than that of the node index. This suggests that the capacity of public transportation may have difficulty keeping up with the traffic volume attracted by high-density development in TOD areas.

Figure 4 shows the spatial distribution of the different types of station areas by k-means clustering. The results of PCA showed that the first five principal components could explain 68.36% of the total information in this study. On the basis of the silhouette coefficient and interpretability, all station areas were classified into five categories.

Table 5 shows the average values of the three TOD components in the different clusters (e.g.,

Table 5). Among the five clusters, the station areas in Cluster 1 (labeled ‘TOD areas in the airport’) are all located near the airport terminal. Their TOD index performance is poor because of specific service objectives. The station areas in Cluster 2 (labeled ‘Less developed TOD areas in outer suburbs’) are primarily located in the outer suburbs and near the end of the subway lines. They have the lowest mean index values for the O and D characteristics and the second lowest value for the T characteristic. The station areas in Cluster 3 (labeled ‘Less developed TOD areas in suburbs’) are mainly located in suburban areas and along subway lines leading to suburban areas. They have lower mean index values for the T, O, and D characteristics and better performances than those of outer suburban TOD areas. The station areas in Cluster 4 (labeled ‘Development-oriented TOD areas in core areas’) are located primarily in core areas. They have average values for the T and O characteristics and higher average values for the D characteristic.

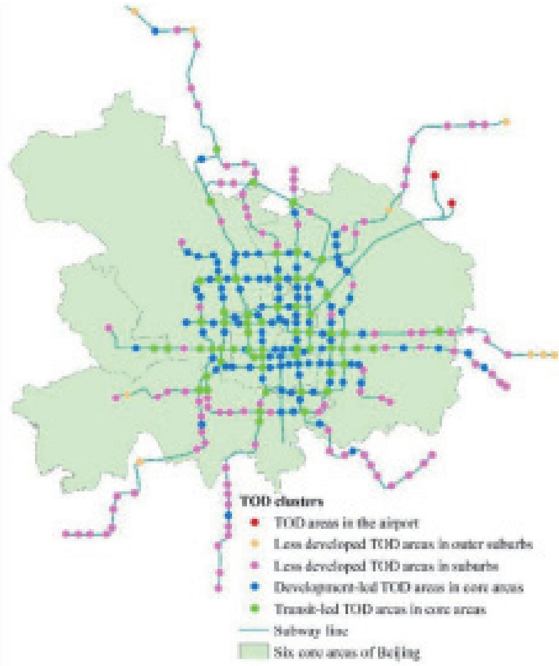

Fig. 4. (Color online) Distribution of station areas in different clusters.

Table 5
Average values of three TOD components for different clustering types.

Clusters	T characteristic	O characteristic	D characteristic
Cluster 1: TOD areas in the airport ($N = 2$)	0.13	0.32	0.24
Cluster 2: Less developed TOD areas in outer suburbs ($N = 9$)	0.20	0.30	0.12
Cluster 3: Less developed TOD areas in suburbs ($N = 99$)	0.26	0.49	0.38
Cluster 4: Development-oriented TOD areas in core areas ($N = 106$)	0.32	0.64	0.57
Cluster 5: Transit-oriented TOD areas in core areas ($N = 59$)	0.44	0.66	0.58
Average value	0.30	0.58	0.48

The station areas in Cluster 5 (labeled 'Transit-oriented TOD areas in core areas') are primarily located in the core areas, and they are more likely to occur where public transit lines intersect. They have the highest mean index values for T, O, and D characteristics.

3.3 Relationship between traffic congestion and TOD performance

Prior to the regression analysis, the collinearity between the independent variables was tested. The variance inflation factors (VIF) for all variables are less than 5, which ensured the reliability of the subsequent regression analysis.

Table 6 shows the results obtained with the OLS and multilevel models. The R-square for OLS reached 0.461. The regression results show that the analytical framework developed from the TOD performance can explain the average speed on the road network in station areas. Although the overall trend of the regression results of the two models is consistent, there are differences in the regression coefficients and statistical significance. To assess the effectiveness of the multilevel model, the likelihood ratio test was introduced in this study to compare the values of −2*loglikelihood in the proposed model and the base model, with no statistically significant differences between the different TOD clusters under the null hypothesis. The difference in −2*loglikelihood was compared with the chi-squared distribution. The results show

Table 6
Results obtained using different regression models.

Variables	OLS modeling	Multi-level modeling
Intercept	1.237***	1.236***
Total number of subway riders from 7:00 a.m.–9:00 a.m. (T1)	−0.143**	−0.111***
Density of bus stops around the station area (T2)	−0.171***	−0.132***
Number of directions served by subway (T3)	−0.031	−0.017
Departure intervals of subway (T4)	0.035	0.019
Number of entrances and exits (T5)	0.018	0.022
Average distance between subway stations (T6)	−0.024	−0.028
Average distance of walkway (O1)	0.016**	0.062**
Road density (O2)	0.075	0.047
Number of signals (O3)	−0.114*	−0.071**
Average distance from station to workplace (O4)	0.022	0.035
Average distance from station to residence (O5)	0.156**	0.149***
Total employment population (D1)	−0.233***	−0.152***
Total resident population (D2)	0.314***	0.240***
Building area around the station areas (D3)	−0.175**	−0.195***
Number of educational/health/cultural establishments (D4)	−0.045	−0.038
Number of retail/hotel and catering establishments (D5)	−0.227**	−0.134***
Number of entrances and exits for motorway (C1)	0.264***	0.307***
Is there a nearby highway (C2)	0.365***	0.051***
Is there a nearby urban expressway (C3)	0.143***	0.110***
Number of samples at level 1	275	275
Classes at level 2		5
R^2	0.461	
Likelihood ratio test	−308.206	−472.55***
Variance partition coefficient		9.09 %

Note: *, **, and *** suggest statistically significant at the 5, 1, and 0.1% levels, respectively.

that the p-value is extremely small, indicating that the null hypothesis should be rejected. Thus, within the station areas, there is a difference in the mean velocity between clusters. We used the variance partition coefficient (VPC), which represents the proportion of total residual variation due to intergroup differences, to measure the similarity between station areas within the same cluster. The VPC indicated that the similarity of station areas was as high as 9.09%. These results suggest that the multilevel model is better than the OLS model in the study area.

Transit characteristics are closely related to the average speed in the TOD areas. T1 and T2, which represent the traffic capacity, have a negative relationship with the traffic speed in the morning peak hours. Although some subway stations can accommodate more passengers, the expected modal shift from driving to taking public transit did not occur. Places with a large number of subway users also have a large travel demand. Car use in these TOD areas remains high because of a mismatch between public transportation capacity and travel demand. Inadequate public transport capacity stimulates the use of motorized travel modes and increases traffic congestion. T3, T4, and T5 represent the morphological characteristics of subway stations and subway lines and have little to do with improving morning peak congestion.

The interrelationship characteristics between nodes and places are positively related to the average speed in TOD areas. The average distance of the walkway (O1) is positively correlated with average speed. This suggests that a friendly walking environment will promote the use of nonmotorized modes and thus reduce the use of automobiles. The number of traffic signals (O3) has a negative correlation with the average speed. The more signals there are, the slower the speed of vehicles. In some places, signals are used to facilitate pedestrians to cross the road. The results of this study show that the installation of signals also has a side effect on average speed, which increases the risk of traffic congestion. The average distance from the station to the residence (O5) has a positive relationship with the average speed. It is important to note that the inverse correlation of O5 is rescaled by min–max normalization. The further the distance from the subway station to the residence, the more people tend to use a car.

The development characteristics in the TOD components are correlated with the average speed of the TOD areas. Interestingly, the regression results are reversed for the total employment (D1) and total resident (D2) populations. This suggests that people are more likely to use their cars for employment-related trips during the morning peak hours. For places with a large residential population, people tend to use their cars less. The building area around station areas (D3) is negatively related to the average speed in the TOD areas. A large building area indicates a large concentration of population, which can lead to increased traffic volume and congestion risk around such station areas. Different activities in the station areas have different effects on average speed. The number of retail/hotel and catering establishments (D5) has a negative correlation with average speed. These facilities are typically located in areas with large populations, which will in turn attract travel demand.

Among the control variables, C1, C2, and C3 have a positive relationship with average speed. This indicates that better road conditions reduce congestion. A road of higher class faces less pedestrian disturbance. This suggests that the separation of pedestrians and vehicles on roads may be an effective way to reduce congestion.

4. Discussion

The purpose of this study is to reveal the congestion conditions in TOD areas and their relationship with specific components of TOD, using Beijing as an example. We now discuss the main findings of the empirical analysis.

The analysis shows that congestion in TOD areas is evident during the morning peak in Beijing. Previous research has shown that traffic congestion is more severe in the urban center than in the surrounding areas in Beijing.(18) This study has revealed that the spatial distribution of station areas with different congestion conditions is more complex. There are also many severely congested TOD areas in the suburbs owing to the high demand for work trips in the fringe areas.

Compared with the travel demand brought about by high-density development in TOD areas, the capacity of the subway system in Beijing is under great pressure during the morning peak hours. The capacity and demand of public transportation are in an unbalanced state. Because of the high travel demand, the capacities of most of the subway lines are already saturated during the morning peak hours. Some stations must restrict the flow of people during the morning peak hours to maintain proper operation of the subway. The poor experience of public transportation during the morning peak hours further forces some people to turn to cars. Over the last decade, Beijing has implemented a number of traffic restriction control policies, such as a car lottery and license plate number restrictions. These measures have reduced car use to some extent from the supply side. However, owing to the inadequate public transit availability, there is limited potential to reduce congestion during the morning peak hours through a modal shift from driving to taking public transit. Identifying areas with high levels of TOD but inadequate public transit availability can help with better transportation planning.

The congestion challenge in TOD areas can be answered by the development component of TOD areas. The imbalance between employment and housing has led to a large increase in travel demand in Beijing.(44) This suggests that optimizing the development structure within TOD areas is necessary. The results of this study show that the increase in the residential population in the station area is associated with an increase in average speed. On the one hand, people living in a TOD area can travel to other places via public transport. On the other hand, the diversity of activities in the TOD area also enables people to access services within the TOD area.(2) Therefore, an appropriate increase in the residential population in TOD areas will help alleviate congestion.

The interrelationship between nodes and places is important not only in revealing the degree of interconnectedness between transport and land use, but also in understanding the people's preference for public transportation. Previous analyses of travel behavior in TOD areas have shown that people prefer to walk or take public transit and drive less in pedestrian-friendly environments in TOD areas.(45,46) This was reflected in the results of our analysis. Therefore, the interrelationship between land use and public transportation nodes should be emphasized in TOD planning.

Some further research is needed. There is a feedback loop between transportation and land use.(47) The construction of public transportation facilities can promote population clustering. In turn, the population attracted by land development helps to further improve the use of public transport. Therefore, the long-term effects of TOD on traffic congestion need to be documented and analyzed. Many non-TOD factors can also affect traffic congestion, such as people's attitudes toward public transportation, so a comparison of TOD- and non-TOD-related factors is needed. More data should be used in congestion analysis to reduce the bias in detecting congestion. In addition, travel peaks on weekends are different from weekdays. Comparative analysis between weekdays and weekends can be considered to identify different travel characteristics.

5. Conclusions

This study revealed the congestion in TOD areas and their relationships with TOD performance during morning peak hours using Beijing subway station areas as an example. We proposed an analytical framework based on an extended form of the node and place model to understand the relationship between TOD performance and congestion. The framework is quantitatively examined through a multilevel regression model. Geotagged big data originating from multiple sensors, such as taxi trajectory data and metro smart card data are used to assess congestion and TOD performance.

The results show that TOD areas in Beijing face significant congestion during the morning peak hours, with an average speed of only 15.23 km/h. The numbers of station areas with severe, moderate, and light congestion are 36, 113, and 113, respectively, accounting for more than 94.58% of all station areas. The distribution of moderately congested station areas is more clustered compared with those of light and severely congested station areas. Through cluster analysis, the TOD performance of station areas can be classified into five different clusters, i.e., TOD areas in the airport, less developed TOD areas in the outer suburbs, less developed TOD areas in the suburbs, development-led TOD areas in core areas, and transit-led TOD areas in core areas. Each cluster has different characteristics in terms of TOD components. Beijing's public transport characteristics require improvement compared with the development and interrelationship characteristics of TOD areas.

The results of multilevel regression illustrate that specific components of TOD are closely related to traffic congestion in TOD areas. For the transit component, the increase in public transport ridership does not reduce congestion because of the saturated public transport capacity during the morning peak hours. For the development component, a high density of development in a TOD area leads to population clustering, which increases the risk of congestion. Places with large residential populations are less likely to experience congestion than places with large employment populations. Regarding the oriented component, better walkability and smaller average distances from stations to residences help reduce congestion. In addition, better road conditions also reduce congestion.

Our findings have several implications. In terms of the transit component in TOD, the capacity of public transportation cannot meet the high volume of travel demand in the morning peak hours in Beijing's TOD areas. An approach to reducing congestion by simply encouraging a modal shift to public transit without improving its capacity may have limited effectiveness. In those areas with high levels of TOD but insufficient public transportation, there is still a need to increase the capacity of public transportation through transport planning. In terms of the development component of TOD, increasing living functions can help reduce congestion. TOD planning needs to consider not only attracting the employment population but also promoting the matching of employment and housing within TOD areas. Optimizing the functional structure, such as increasing the proportion of affordable housing in TOD areas, can be an important way to adjust traffic demand. In terms of the interrelationship between transportation and development, establishing a pedestrian-friendly environment and connectivity within TOD areas can help increase people's willingness to use public transportation. In addition, the diversion of pedestrians and vehicles also helps to reduce congestion.

References

1 L. Bertolini: Planning Pract. Res. **14** (1999) 199. https://doi.org/10.1080/02697459915724
2 R. Cervero: Urban Plann. Overseas **357** (2005) 10. https://doi.org/10.1186/s13104-015-1325-3
3 S. Lee, C. Yi, and S.-P. Hong: Cities **35** (2013) 69. https://doi.org/10.1016/j.cities.2013.06.010
4 Y. Huang, D. Parker, and L. Minaker: J. Transp. Geogr. **90** (2021) 1. https://doi.org/10.1016/j.jtrangeo.2020.102940
5 K. A. Ratner and A. R. Goetz: Cities **30** (2013) 31. https://doi.org/10.1016/j.cities.2012.08.007
6 K. Park, R. Ewing, B. C. Scheer, and S. S. Ara Khan: Transp. Res. Rec. **2672** (2018) 31. https://doi.org/10.1177/0361198118774159
7 Y. Choi and S. Guhathakurta: J. Transp. Land Use **13** (2020) 343. https://doi.org/10.5198/JTLU.2020.1530
8 J. Zhao, W. Deng, Y. Song, and Y. Zhu: Cities **35** (2013) 114. https://doi.org/10.1016/j.cities.2013.07.002
9 E. Papa and L. Bertolini: J. Transp. Geogr. **47** (2015) 70. https://doi.org/10.1016/j.jtrangeo.2015.07.003
10 G. Lyu, L. Bertolini, and K. Pfeffer: Int. J. Sustainable Transp. **14** (2020) 533. https://doi.org/10.1080/15568318.2019.1578841
11 A. Nasri and L. Zhang: Transp. Policy **32** (2014) 172. https://doi.org/10.3141/2413-05
12 P. L. Mokhtarian and X. Cao: Transp. Res. Part B Methodol. **42** (2008) 204. https://doi.org/10.1016/j.trb.2007.07.006
13 A. I. Kay, R. B. Noland, and S. DiPetrillo: J. Transp. Geogr. **39** (2014) 131. https://doi.org/10.1016/j.jtrangeo.2014.06.017
14 L. D. Frank, M. A. Andresen, and T. L. Schmid: Am. J. Preventive Med. **27** (2004) 87. https://doi.org/10.1016/j.amepre.2004.04.011
15 A. Ibraeva, G. H. de Almeida Correia, C. Silva, and A. P. Antunes: Transp. Res. Part A Policy Pract. **132** (2020) 110. https://doi.org/10.1016/j.tra.2019.10.018
16 D. E. Reusser, P. Loukopoulos, M. Stauffacher, and R. W. Scholz: J. Transp. Geogr. **16** (2008) 191. https://doi.org/10.1016/j.jtrangeo.2007.05.004
17 S. Zemp, M. Stauffacher, D. J. Lang, and R. W. Scholz: J. Transp. Geogr. **19** (2011) 670. https://doi.org/10.1016/j.jtrangeo.2010.08.008
18 P. Zhao, H. Yang, L. Kong, Y. Liu, and D. Liu: Transp. Res. Part A Policy Pract. **116** (2018) 290. https://doi.org/10.1016/j.tra.2018.06.017
19 M. Sweet: J. Planning Lit. **26** (2011) 391. https://doi.org/10.1177/0885412211409754
20 Z. Zhu, M. Lee, Y. Pan, H. Yang, and L. Zhang: Transp. Plann. Technol. **41** (2018) 816. https://doi.org/10.1080/03081060.2018.1526882
21 A. Loukaitou-Sideris: J. Urban Design **15** (2010) 49. https://doi.org/10.1080/13574800903429399
22 R. Tamakloe, J. Hong, and J. Tak: Cities **108** (2021) 1. https://doi.org/10.1016/j.cities.2020.102988
23 A. Faghri and M. Venigalla: Transp. Res. Rec. **2397** (2013) 72. https://doi.org/10.3141/2397-09
24 M. Matsuyuki, N. Aizu, F. Nakamura, and K. Leeruttanawisut: Case Stud. Transp. Policy **8** (2020) 1341. https://doi.org/10.1016/j.cstp.2020.09.005
25 M. Kamruzzaman, D. Baker, S. Washington, and G. Turrell: J. Transp. Geogr. **34** (2014) 54. https://doi.org/10.1016/j.jtrangeo.2013.11.002
26 D. S. Vale: J. Transp. Geogr. **45** (2015) 70. https://doi.org/10.1016/j.jtrangeo.2015.04.009
27 G. Lyu, L. Bertolini, and K. Pfeffer: J. Transp. Geogr. **55** (2016) 40. https://doi.org/10.1016/j.jtrangeo.2016.07.002
28 J. Zhou and Y. Long: Transp. Res. Rec. **2418** (2014) 1. https://doi.org/10.3141/2418-01
29 J. McIntosh, R. Trubka, J. Kenworthy, and P. Newman: Transp. Res. Part D Transp. Environ. **33** (2014) 95. https://doi.org/10.1016/j.trd.2014.08.013

30 G. Sun, J. Zacharias, B. Ma, and N. M. Oreskovic: Cities **56** (2016) 91. https://doi.org/10.1016/j.cities.2016.03.001
31 Y. J. Singh, A. Lukman, J. Flacke, M. Zuidgeest, and M. Van Maarseveen: Transp. Policy **56** (2017) 96. https://doi.org/10.1016/j.tranpol.2017.03.013
32 J. Qin, G. Mei, and L. Xiao: Sustainability **13** (2020) 266. https://doi.org/10.3390/su13010266
33 J. Yang, S. Chen, P. Qin, F. Lu, and A. A. Liu: J. Environ. Econ. Manage. **88** (2018) 114. https://doi.org/10.1016/j.jeem.2017.09.007
34 P. Zhao and H. Hu: Cities **92** (2019) 164. https://doi.org/10.1016/j.cities.2019.03.022
35 Z. Kan, L. Tang, M.-P. Kwan, C. Ren, D. Liu, and Q. Li: Comput. Environ. Urban Syst. **74** (2019) 229. https://doi.org/10.1016/j.compenvurbsys.2018.11.007
36 K. Wu, Y. Chen, J. Ma, S. Bai, and X. Tang: Transp. Res. Part D Transp. Environ. **51** (2017) 203. https://doi.org/10.1016/j.trd.2016.06.005
37 Y. Shi, D. Wang, J. Tang, M. Deng, H. Liu, and B. Liu: Int. J. Geogr. Inf. Sci. **35** (2021) 1449. https://doi.org/10.1080/13658816.2021.1905820
38 S. Sun, J. Chen, and J. Sun: Int. J. Distrib. Sens. Netw. **15** (2019) 1550147719847440. https://doi.org/10.1177/1550147719847440
39 Y. Zhang, E. Yao, R. Zhang, and H. Xu: J. Transp. Geogr. **76** (2019) 191. https://doi.org/10.1016/j.jtrangeo.2019.04.002
40 E. Ben-Elia and D. Ettema: Transp. Res. Part F Psychol. Behav. **14** (2011) 354. https://doi.org/10.1016/j.trf.2011.04.003
41 Y. J. Singh, P. Fard, M. Zuidgeest, M. Brussel, and M. van Maarseveen: J. Transp. Geogr. **35** (2014) 130. https://doi.org/10.1016/j.jtrangeo.2014.01.014
42 C. D. Higgins and P. S. Kanaroglou: J. Transp. Geogr. **52** (2016) 61. https://doi.org/10.1016/j.jtrangeo.2016.02.012
43 I. H. Langford, G. Bentham, and A. L. McDonald: Stat. Med. **17** (1998) 41. https://doi.org/10.1002/(SICI)1097-0258(19980115)17:1
44 Y. Long and J.-C. Thill: Comput. Environ. Urban Syst. **53** (2015) 19. https://doi.org/10.1016/j.compenvurbsys.2015.02.005
45 B. G. Nichols and K. M. Kockelman: Energy Policy **68** (2014) 232. https://doi.org/10.1016/j.enpol.2013.12.062
46 M.-J. Jun, K. Choi, J.-E. Jeong, K.-H. Kwon, and H.-J. Kim: J. Transp. Geogr. **48** (2015) 30. https://doi.org/10.1016/j.jtrangeo.2015.08.002
47 L. Bertolini: Eur. Spatial Res. Policy **12** (2005)77. https://doi.org/10.2968/066004008

About the Authors

Yongxuan Liu received his master's degree from China University of Mining and Technology (Beijing) in 2014. From 2014 to 2024, he worked at Beijing Institute of Surveying and Mapping. He obtained the title of Senior Engineer in 2021. His research interests are in geographic information system. (liuyongxuan123@126.com)

Cong Liao received his B.S. degree from Chuzhou University, China, in 2011, and his M.S. and Ph.D. degrees from the China University of Mining & Technology, and Beijing Normal University, China, in 2014 and 2018, respectively. From 2018 to 2022, he was a postdoctoral student at Peking University, China. Since 2022, he has been an engineer at CAUPD Beijing Planning & Design Consultants LTD. His research interests are in urban planning and geospatial analysis. (liaocong233@pku.edu.cn)

Appendix

Table A1
Typical sample record of metro smart card.

Card number	Swipe time	Line number	Embarking metro station	Line number	Disembarking metro station
19151591	20160620081432	1	12	2	18

Table A2
Data format of population data.

Number	Longitude of residence grid	Latitude of residence grid	Longitude of employment grid	Latitude of employment grid	Number of employees	Number of residents
1	117.17	40.15	116.84	40.17	1	2

Table A3
Description of TOD variables.

Variables	Measurement method	Expected direction
Total number of subway riders from 7:00 a.m.–9:00 a.m. (T1)	Count the number of swiping records of metro stations in metro smart card data	Positive
Density of bus stops around the station area (T2)	Count the number of bus stops within a station area in POI data of Beijing	Positive
Number of directions served by metro (T3)	Get the transfer status of each metro station from Beijing Metro official website (http://www.mtr.bj.cn/article/line)	Positive
Departure intervals of metro (T4)	Get the data from Beijing transport development annual report (https://www.bjtrc.org.cn/List/index/cid/7.html)	Positive
Number of entrances and exits (T5)	Get the number of entrances and exits from POI data of Beijing	Positive
Average distance between subway stations (T6)	Determine average distance from each service direction to the current and the next metro station (http://www.mtr.bj.cn/service/line/distable/line-1.html)	Positive
Average distance of walkway (O1)	Calculate the total length of the walkway within station area referring to the open street map of Beijing	Positive
Road density (O2)	Calculate the ratio of total length of street network within TOD buffer to that within the area of TOD referring to the open street map of Beijing	Positive
Number of traffic signals (O3)	Count the number of signal lamps within station area referring to the open street map of Beijing	Negative
Average distance from station to workplace (O4)	Calculate the average distance between station and all entities of economic activities within TOD areas	Negative
Average distance from station to residence (O5)	Calculate the average distance between station and housing entities within TOD areas	Negative
Total employment population (D1)	Count the number of employed persons within TOD area from China Unicom's mobile phone data	Negative
Total resident population (D2)	Count the number of residents within TOD area from China Unicom's mobile phone data	Positive
Building area around the station areas (D3)	Calculate the area of buildings around station areas using AOI data in Amap. (https://www.amap.com/)	Negative
Number of educational/health/cultural establishments (D4)	Count the number of facilities in station area in POI data of Beijing	Negative
Number of retail/hotel and catering establishments (D5)	Count the number of facilities in station area in POI data of Beijing	Negative
Number of entrances and exits for motorway (C1)	Count the number of entrances and exits for motorway in station areas referring to open street map of Beijing	Positive
Is there a nearby highway (C2)	Determine the value referring to open street map of Beijing	Positive
Is there a nearby urban expressway (C3)	Determine the value referring to open street map of Beijing	Positive