International Journal of Applied Earth Observation and Geoinformation 118 (2023) 103245

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

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

How to quantify the travel ratio of urban public transport at a high spatial resolution? A novel computational framework with geospatial big data

Ganmin Yin [a,b], Zhou Huang [a,b,*], Liu Yang [a,b], Eran Ben-Elia [c], Liyan Xu [d], Bronte Scheuer [a,b], Yu Liu [a,b]

[a] *Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China*
[b] *Beijing Key Lab of Spatial Information Integration & Its Applications, Peking University, Beijing 100871, China*
[c] *Department of Geography and Environmental Development, Ben-Gurion University of the Negev, Israel*
[d] *College of Architecture and Landscape Architecture, Peking University, Beijing 100871, China*

ARTICLE INFO

Keywords:
Urban public transportation
Public transport travel ratio
High spatial resolution
Linear programming
Supply and demand
Geospatial big data

ABSTRACT

Improving the travel ratio of public transportation (PTR) is important for realizing low-carbon transportation and sustainable city development. However, limited by data resolution and model accuracy, existing research rarely involves the spatially refined calculation of PTR and the quantitative analysis of its influencing factors. In this study, based on multi-source geospatial big data, we propose a novel computational framework to solve the above problems. Specifically, we first design a linear programming-based three-step method, which realizes the calculation of PTR at 500-meter grid-pair scale for the first time; secondly, we develop a Beta-binomial model for regression analysis, which improves by more than 50% compared with traditional generalized linear models. The case of Wangjing area in Beijing shows that: the overall PTR in Wangjing is only 16%, which is much lower than the official expectation (45%), and less than 20% of origin–destination (OD) pairs meet the standard; among the influencing factors, the travel duration gap between public transportation and private cars, walking distance, number of transfers, and residential parking density have significant negative effects on PTR. Finally, this paper provides an implication of the proposed computational framework, i.e., the accurate detection of public transportation (PT) supply–demand imbalance areas, which proves its great potential in refined transportation optimization and sustainable urban planning.

1. Introduction

There is no doubt that public transportation plays a key role in urban transportation systems. In addition to meeting the travel needs of residents, a well-designed public transportation (PT) system can effectively reduce traffic congestion, promote energy conservation and emission reduction, and facilitate sustainable urban development. In most countries, PT is given top priority in urban transportation management (Currie et al., 2006; Haitao et al., 2019; Song et al., 2021).

Among the existing PT evaluation indicators, the public transportation travel ratio (PTR) is an important indicator to measure the status of PT supply and demand, and further guide the planning of transit networks. PTR refers to the proportion of PT trips to the total trips in a certain statistical area and period. According to the travel time range, mode and purpose, it can be classified as one-day PTR, motorized PTR, commuting PTR, etc. (Ling et al., 2014). For most cities, commuting trips account for the highest proportion of all trips and result in morning and evening peaks. Improving the PTR, especially in the commuting context, is significant for achieving low-carbon transportation and sustainable cities. Since traditional methods such as questionnaire surveys do not provide fine-scale and comprehensive travel data on residential trips, most existing studies on PTR have adopted a city-wide scale rather than a higher spatial resolution within cities, and thus cannot support the refined planning and optimization of the transit networks.

With the rise of technologies such as positioning, wireless communication, and mobile internet, as well as the popularization of location-aware mobile computing devices, an unprecedented amount of geospatial data with individual movement information has been generated and stored over the years. Geospatial big data forms a true record of urban resident's movements as it covers datasets such as smart card data, mobile signaling data, and mobile phone GPS location data (Bao

* Corresponding author at: Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China.
E-mail address: huangzhou@pku.edu.cn (Z. Huang).

https://doi.org/10.1016/j.jag.2023.103245
Received 2 December 2022; Received in revised form 3 February 2023; Accepted 23 February 2023
Available online 1 March 2023

et al., 2021; Huang et al., 2023; Wang et al., 2022). Moreover, it has significant advantages over traditional residential travel surveys in terms of wide coverage, low acquisition cost, and high spatiotemporal resolution. This explains why it has been widely used in the studies of human mobility, jobs-housing balance, and traffic flow prediction, etc. (Liu et al., 2015; Wang et al., 2022; Zhao et al., 2020), leading to new knowledge and discoveries. This is also good news for the extraction of the fine-scale PTR.

In this paper, we propose a computational framework for PTR calculation at a high spatial resolution based on multi-source geospatial big data. Taking the Wangjing area of Beijing as an example, we obtain PT commuting flow and total commuting flow based on smart card data and mobile phone location data. Since PT flow is between station pairs and total flow is between grid pairs, there is a problem of inconsistent research units. Therefore, we propose a three-step method based on linear programming to realize the allocation of the PT flow between station pairs to grid pairs (500 m). After the research units are unified, we calculate the ratio of PT flow to total flow, i.e., PTR. Based on this, a regression model called the Beta-binomial model is developed to analyze the influencing factors of PTR, which is more than 50% better than traditional models on all three goodness-of-fit indicators. Finally, this paper provides an implication of the proposed computational framework for the accurate detection of PT supply–demand imbalance areas, demonstrating its great potential in refined transportation optimization and sustainable urban planning.

The remainder of this paper is organized as follows. Section 2 reviews recent research works on public transportation travel ratio (PTR), travel mode selection, and PT supply and demand analysis. Section 3 introduces the study area and datasets. Section 4 describes the extraction process of PT commuting flow and further proposes a three-step (i.e., "route planning", "route matching", and "flow assignment") method for calculating the fine-scale PTR. Section 5 analyzes the results of PT commuting flow and PTR. Section 6 develops a regression model to quantitatively analyze the influencing factors of PTR and further proposes a strategy for discovering the areas with unbalanced PT supply and demand. Section 7 summarizes the research findings and suggests directions for future improvements.

2. Related works

2.1. Public transportation travel ratio

Public transportation travel ratio (PTR) is a recent research hotspot because it significantly reflects the effectiveness of PT systems. Shi and Ju (2015) explore the effects of PT facility, land use types, individual socioeconomic attributes on PTR, using transportation analysis zone (TAZ) as the research unit. Wen et al. (2016) choose three types of influencing factors, i.e., travel time, ride comfortability, per capita occupation of road area, to forecast PTR and study its time series changes in Nanjing. Zhang and Wang (2018) put forward a planning target for Shanghai's PTR in 2035. Xie (2018) analyzes the variation of PTR in seven international metropolises from 1999 to 2013, including Hong Kong, Tokyo, and Paris, etc. From the above, due to the lack of finely measured and widely covered travel data, most of the studies on PTR have been conducted at a city-wide scale, with a small proportion covering the TAZ scale. However, according to Benenson et al. (2017), in more than 70% of TAZs, the difference in PT accessibility between buildings is greater than normal, indicating internal PT supply heterogeneity is evident even at the TAZ scale. Therefore, it is necessary to extract PTR at a finer scale.

2.2. Travel mode selection

PTR is the result of a combination of many factors when urban residents are making travel mode choices. Many scholars have studied the factors influencing travel mode selection (Bresson et al., 2003; Ding, 2016; Outwater et al., 2011; Redman et al., 2013). In the commuting context, although commuters in different cities have different preferences for travel modes, the common influencing factors are transit infrastructure factors (e.g., public transportation efficiency and parking convenience) and socioeconomic factors (e.g., income level and the number of private cars). As for the regression model of travel mode selection, most of the existing studies adopt the logistic regression in generalized linear models, while a few adopt probit regression, with a difference between the two being the link function (Ben-Akiva and Bierlaire, 1999; Ding, 2016; Outwater et al., 2011). In this paper, the generalized linear model with the response variable of a binomial distribution is referred to as the "Binomial" regression model, which may exhibit an "over-dispersion" problem in use and lead to drawbacks in the fitting effect.

2.3. Public transportation supply and demand analysis

Depending on the research purposes, existing studies on PT supply and demand analysis have different focuses and generally adopts three perspectives: transit network, accessibility, and PTR. The PTR perspective can be referred to in Section 2.1.

The perspective of the transit networks focuses on measuring its supply capability. Chen et al. (2018) propose absolute and relative indicators of PT supply capability. The former is the number of bus stations weighted by the vehicle capacity and departure frequency in a research unit; and the latter is the result of dividing the absolute indicator by the number of residents, taking the potential demand into account. Both indicators are limited to the level of PT stations. Mishra et al. (2012) use the bus network as a graph and define the connectivity of bus stations, bus lines, and urban regions, which can measure the PT supply level at a single station, line, or region, but does not involve the interaction between two stations, lines, or regions. Similarly, Wang et al. (2020) propose a unified analysis method for multimodal PT networks that can predict local travel efficiency based on its topology characteristics, but does not consider demand factors. In general, the perspective of transit network focuses only on the supply capacity of individual regions, not on the inter-regional supply capacity, and does not fully consider travel demand.

As a key concept in understanding the relationship between transportation, land use, and human activities (de Alwis Pitts and So, 2017; Hansen, 1959; Mallick and Routray, 2001), accessibility is another perspective for analyzing the PT supply and demand. The simplest accessibility refers to the area or available opportunities (e.g., job, health care, and education) of the region that can be reached from a research unit by a certain travel mode within a certain period of time. On this basis, more complex accessibilities have been proposed in previous studies (Ben-Elia and Benenson, 2019; Benenson et al., 2017; Ferguson et al., 2013; Järv et al., 2018). Among them, Benenson et al. (2017) define the ratio of public transportation accessibility to private car accessibility as relative accessibility, but it focuses on the supply side of public transportation rather than the demand side. Hence, Ben-Elia and Benenson (2019) further consider the number of PT travelers to weight the difference between public transportation accessibility and private car accessibility to reflect the loss of accessibility for residents traveling by public transportation. However, accessibility can reflect the supply side of PT in a single study area, but fails to address interregional interactions, where accessibility varies widely across directions and distances from the same origin to different destinations.

3. Study area and datasets

3.1. Study area

This study focuses on commuters living or working in the Wangjing area within the Sixth Ring Road of Beijing, as shown in Fig. 1. As a comprehensive new area located in the northeast of central Beijing,

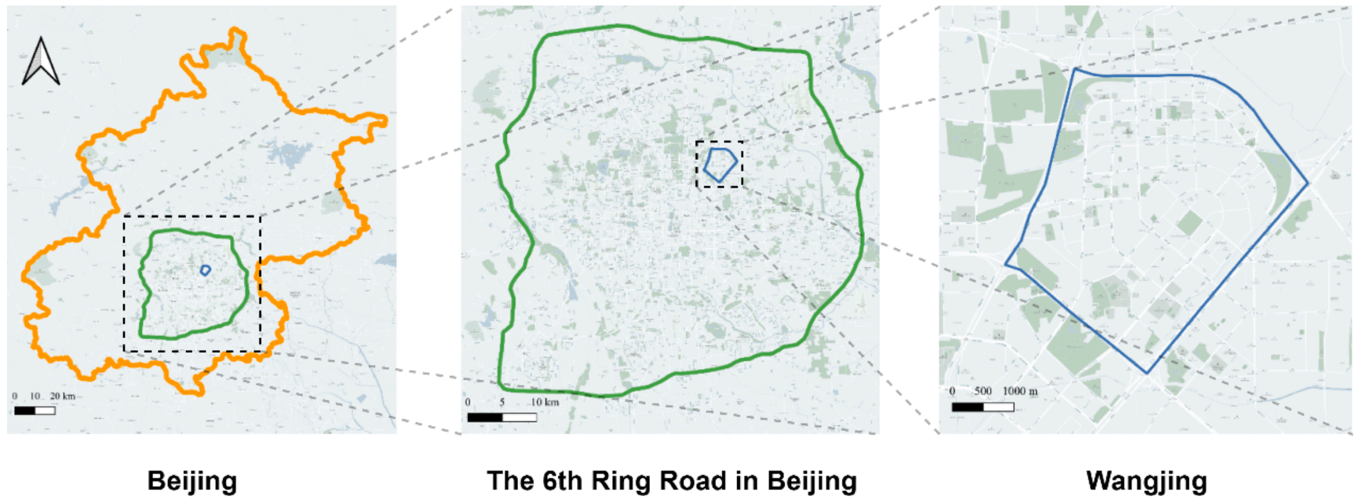

Fig. 1. Study area.

Wangjing is surrounded by four main roads: The Fourth Ring Road, the Fifth Ring Road, the Jingcheng Expressway, and the Capital Airport Expressway. It falls under the jurisdiction of Beijing's Chaoyang District, with an area of 17.8 square kilometers and a resident population of ~ 300,000 in 2010 (Bureau of Statistics of Chaoyang District, 2011). Wangjing has a dense concentration of companies with a large number of job opportunities, attracting many talents from other regions to work here. At the same time, it also has a large resident population, with many residents working outside the area. Therefore, Wangjing is considered an ideal case study area for commuting research.

3.2. Datasets

The datasets involved in this paper mainly consists of smart card data and total commuting flow data, provided by Amap, China's largest navigation e-map company (https://www.amap.com/).

The smart card data includes bus and subway travel records with card IDs, get on/off stations, etc., there are a total of 15 fields, as shown in Table 1, and each row represents a complete travel record. The raw smart card data covers a period of one month (June 2019) for the residents in Beijing. After counting, it includes 36.7 million records of 1.38 million people boarding or alighting from stations in Wangjing area.

Total commuting flow data is also available for June 2019, provided by Amap company. The data product is generated by machine learning based on GPS location data collected from many mobile apps such as Alipay, Tiktok, and Weibo. The data source covers more than 700 million users and has an accuracy rate of more than 90% for home and work locations compared to the ground truth of registered users, implying its high reliability (Yin et al., 2022). The field descriptions are shown in Table 2, including the anonymous user ID, longitude and latitude of home and workplace, and commuting time. According to statistics, the total number of commuters in Wangjing area is 406,394.

Table 1
Smart card data field information.

Field name	Type	Description
Card ID	String	Unique card id.
Traffic mode	String	"GJ" for bus, "DT" for subway.
Station name	String	The name of get on/off station.
Line name	String	The name of the line where the get on/off station is located.
Line direction	String	The direction of the line where the get on/off station is located: "0" or "1".
Station coordinates	Float	The longitude and latitude of get on/off station.
Swipe time	String	The timestamp to get on/off, like "YYYYmmddHHMMSS"

Table 2
Total commuting flow data field information.

Field name	Type	Description
User ID	String	Unique anonymous user id.
Home/workplace coordinates	Float	The longitude and latitude of home and workplace.
Commuting time	String	The commuting time interval, like "HHMM-HHMM", each interval is 15 min.

4. Methodology

This research consists of three main parts, the framework of which is shown in Fig. 2. The first part is the commuting flow extraction (Section 4.1), which focuses on the extraction of PT commuting flow based on smart card data, since the total commuting flow is the available data. The second part is the PT travel ratio calculation (Section 4.2), which consists of three steps: "route planning", "route matching" and "flow assignment". The third part is the influencing factors analysis (Section 6), which introduces the regression analysis process for PTR and the division strategy of PT supply and demand.

4.1. PT commuting flow extraction

PT commuting flow is extracted from smart card data through three steps: data preprocessing, transfer merging, and home and workplace inference.

4.1.1. Data preprocessing

Two abnormalities are excluded during the data preprocessing stage: (1) a total of about 5,000 records of early drop-offs; (2) excessive number of rides (some individuals have taken hundreds of rides, more than 10 per day).

Public transportation commuters are expected to take public transportation at a relatively high frequency. At the same time, they should exhibit relatively regular travel demand on weekdays. Therefore, those who take public transportation at least two weekdays per week are screened as potential commuters.

4.1.2. Transfer merging

To get complete trips, the transfers in a public transportation trip should be merged. Three principles should be followed for transfer merging: (1) interchange time < 30 min and interchange distance < 500

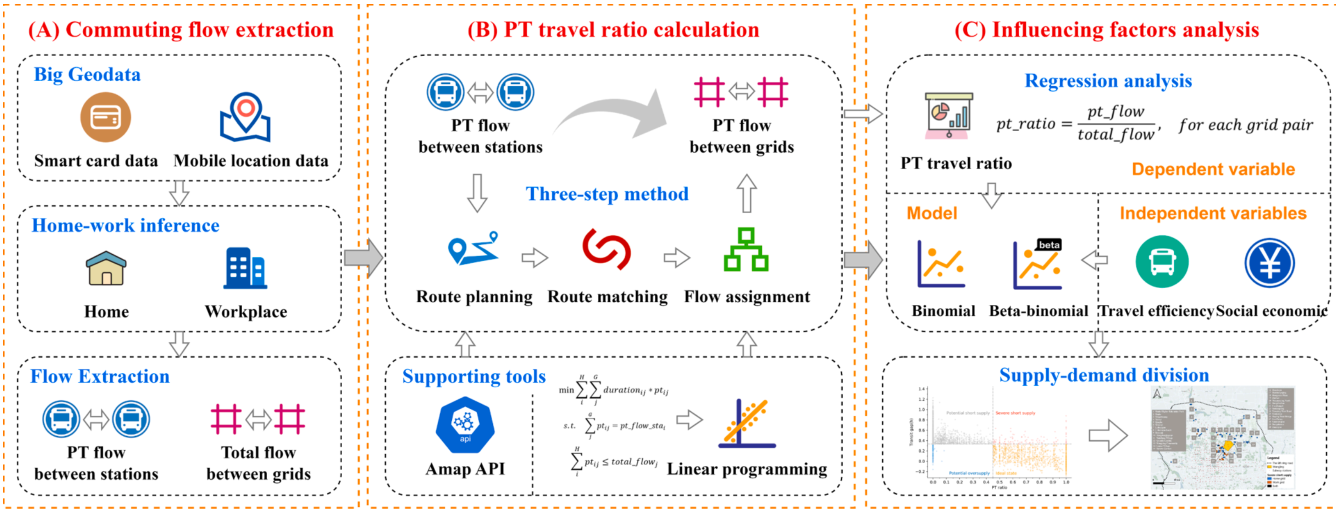

Fig. 2. The overall research framework: (A) Commuting flow extraction; (B) PT travel ratio calculation; (C) Influencing factors analysis.

m; (2) the interchange time < the total time of two consecutive rides; (3) If the distance from start to endpoint after the merge < 500 m, it should be marked as a round trip and the merge should be canceled.

4.1.3. Home and workplace inference

Since this research focused on commuting context, trips before 10:00 a.m. and after 16:00 p.m. are identified as potential commutes. For home extraction, the boarding point for the first trip before 10:00 a.m. and the drop-off point for the last trip after 16:00 p.m. are marked as candidates for home. For workplace extraction, the drop-off point of the last trip before 10:00 a.m. and the boarding point of the first trip after 16:00 p.m. are marked as candidates for workplace.

Based on the extracted potential candidates, the actual homes and workplaces can be further determined by the frequency of occurrence.

Algorithm 1 Calculate the PT Travel Ratio

```
Input: Total commuting grid pairs, GP; Public transport commuting routes, PT
Output: The PT travel ratio of all commuter grid pairs, PT_Ratio
function CalcPTR(GP, PT)
    // 1.Route Planning
    for each gp ∈ GP do
        let R be the fastest top 3 planning routes pf gp          ▷ By Amap API
        let r′ be the fastest planning route of gp
        for each r ∈ R do
            if r.dura − r′.dura > 15min then
                R ← R \ {r}
            end if
        end for
        gp.routes ← R
    end for

    // 2.Route Matching
    for each pt ∈ PT do
        for each gp ∈ GP do
            for each r ∈ gp.routes do
                if (pt.ori == r.ori) and (pt.des == r.des) then
                    pt.gp ← pt.gp ∪ {gp}
                end if
            end for
        end for
    end for

    // 3.Flow Assignment
    PT_Ratio ← ϕ
    for each pt ∈ PT do
        assign pt.flow to pt.gp                                   ▷ By Linear Programming
    end for
    for each gp ∈ GP do
        gp.pt_ratio ← gp.pt_flow / gp.total_flow
        PT_Ratio ← PT_Ratio ∪ {gp.pt_ratio}
    end for
    return PT_Ratio
end function
```

Fig. 3. The three-step calculation algorithm of PTR at a high spatial resolution.

Due to the regularity of commuting behavior, the home and workplace stations should occur at high frequency in pairs. Therefore, for each traveler, trips to and from the above candidates are screened to determine whether their occurrence frequency exceeds a certain threshold (4 in this study). If the threshold is exceeded, then the traveler is a commuter, the route is a commuting route, and the boarding and alighting points of the route are the corresponding actual home and work locations. Otherwise, the traveler is not a commuter.

4.2. PT travel ratio calculation

The PTR refers to the ratio of PT commuting flow in total commuting flow. To obtain PTR at a high spatial resolution (500 m grid), PT commuting flow between stations should be assigned to the home-work grid pairs. If the solution objective is given (e.g., the minimal total travel time or cost), the allocation of PT commuting flow to grid pairs can be realized through linear programming. On the one hand, linear programming is a simple but efficient optimization algorithm that has been used in many fields (Cormen et al., 2022; Gass, 2003; Schrijver, 1998), and on the other hand, the algorithm is also perfectly suited for the optimization task, i.e., the optimization objective and constraints are linear. The three-step calculation process of PTR at a high spatial resolution is shown in Algorithm 1 (Fig. 3).

4.2.1. Route planning

First, the grid pairs of total commuting flow are used as potential sources of home-workplaces of PT commuting flow, and route planning is performed for the aforementioned grid pairs to obtain their PT candidate routes. In this study, two PT routes with the same get on/off stations are considered the same regardless of the specific travel process in the trip. The Amap API (https://lbs.amap.com/) is adopted for route planning. As mentioned above, the study area is divided into 500 m × 500 m grids, and the home-workplace pair belonging to the same grid pair are considered to have the same PT routes as the grid center. Since there may be multiple PT routes between a home-workplace pair, the top three routes with the shortest travel time are first selected as candidates in route planning. Also, to exclude the excessively slow routes, those routes that are longer than the fastest route by more than 15 min are removed from the set of candidate routes, and the remaining ones are used as candidate routes between home and workplaces.

The total number of commuters is denoted as N, and the $GridPair = \{GP_1, GP_2, \cdots, GP_G\}$, where G is the total number of grid pairs. The j^{th} grid pair $GP_j = \{(gp_j^o,\ gp_j^d,\ total\_flow_j)|1 \leq j \leq G\}$, where gp_j^o, gp_j^d and $total\_flow_j$ respectively represent the origin, destination, and total flow between it, and $\sum_j^G total\_flow_j = N$. The planned routes between the grid pairs after route planning is denoted as $Route = \{r_1, r_2, \cdots, r_G\}$, where G is the total number of grid pairs. The planned routes of the j^{th} grid pair $r_j = \{r_{jk}(r_{jk}^o, r_{jk}^d)|1 \leq k \leq 3, 1 \leq j \leq G\}$, where r_{jk}^o and r_{jk}^d represent the get on/off stations of the planned routes, respectively. There are at most three routes in the set for each grid pair.

4.2.2. Route matching

Second, the planned routes are then matched with the PT commuting routes (extracted in Section 4.1) to get the potential source grid pairs of

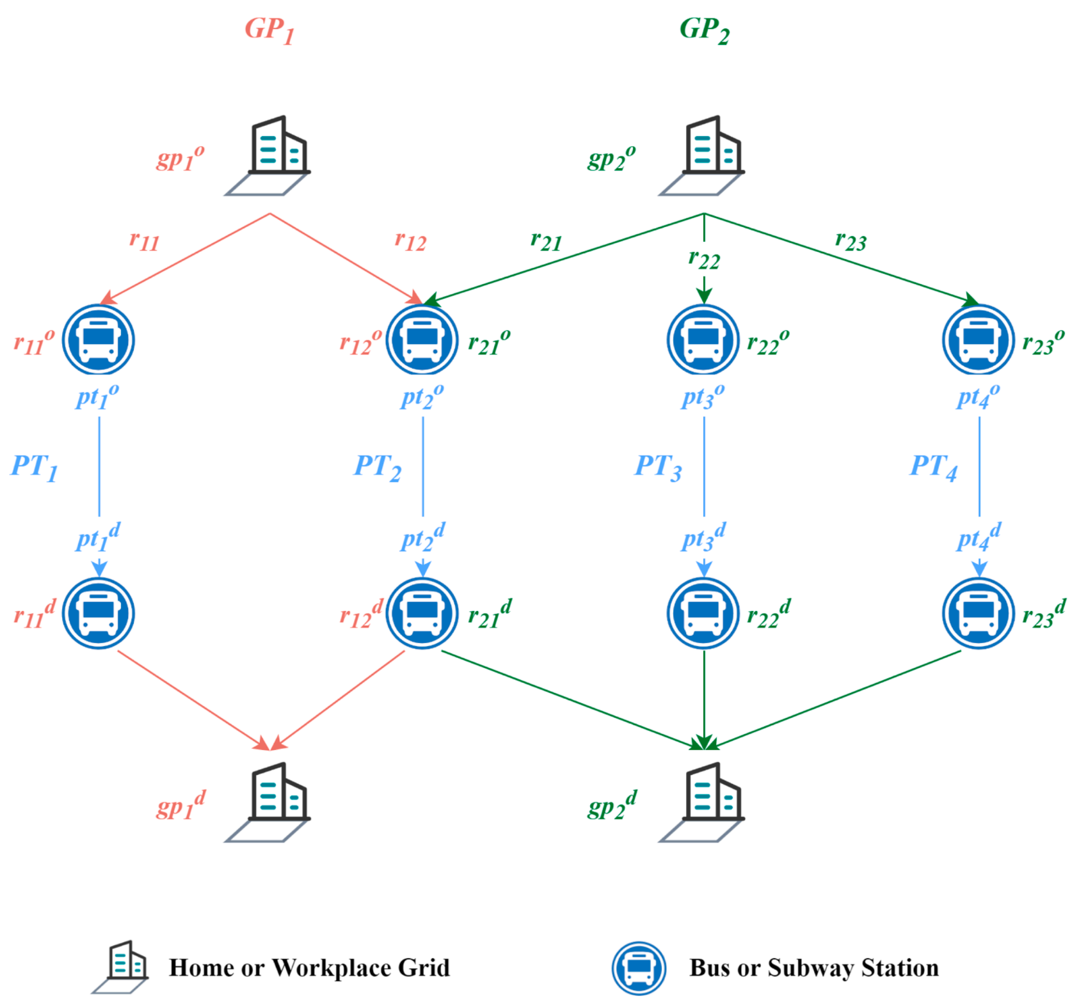

Fig. 4. Matching the planned routes with the PT commuting routes.

the PT commuting flow. The sum of PT commuting flow is denoted as M, and the PT routes $PublicTransport = \{PT_1, PT_2, \cdots, PT_H\}$, where H is the total number of PT routes. The i^{th} PT route $PT_i = \{(pt_i^o, pt_i^d, pt\_flow\_sta_i) | 1 \le i \le H\}$, where pt_i^o, pt_i^d, and $pt\_flow\_sta_i$ respectively represent the get-on station, get-off station, and PT flow on the route, and $\sum_i^H pt\_flow\_sta_i = M$. As stated above, since two PT routes with the same get on/off stations are considered the same, the matching of the planned routes to the PT routes only involves a comparison between (r_{jk}^o, r_{jk}^d) and (pt_i^o, pt_i^d).

Taking Fig. 4 as an example, the grid pair GP_1 has two planned routes: r_{11} and r_{12}, and the grid pair GP_2 has three planned routes: r_{21}, r_{22}, and r_{23}. Since $(r_{11}^o, r_{11}^d) = (pt_1^o, pt_1^d)$, $(r_{12}^o, r_{12}^d) = (pt_2^o, pt_2^d)$, $(r_{21}^o, r_{21}^d) = (pt_2^o, pt_2^d)$, $(r_{22}^o, r_{22}^d) = (pt_3^o, pt_3^d)$, and $(r_{23}^o, r_{23}^d) = (pt_4^o, pt_4^d)$, so r_{11} is matched with PT_1, r_{12} and r_{21} are both matched with PT_2, r_{22} is matched with PT_3, and r_{23} is matched with PT_4, respectively. Thus, the flow on PT_1 is from GP_1, the flow on PT_2 may come from both GP_1 and GP_2, and the flows on PT_3 and PT_4 are both from GP_2.

4.2.3. Flow assignment

Third, linear programming is then adopted to assign the PT commuting flow to the potential source grid pairs and calculate the PTR. The PT flow assigned by the i^{th} PT route to the j^{th} grid pair is denoted as pt_{ij}. In Fig. 4, $pt_{11} = pt_1$, $pt_{12} = 0$, $pt_{31} = 0$, $pt_{32} = pt_3$, $pt_{41} = 0$, and $pt_{42} = pt_4$. Since the flow on PT_2 may come from both GP_1 and GP_2, then $pt_{21} + pt_{22} = pt_2$, and the grid pairs should not be assigned with PT flow that exceeds their total flow (denoted as gp_j for the j^{th} grid pair), i.e., $\sum_i^4 pt_{ij} \le gp_j$, where $1 \le j \le 2$. The values of pt_{21} and pt_{22} are not unique under the above constraints. For the PT flow allocation, this study assumes that individuals always choose the fastest route to travel. This way, the allocation of the PT flow is transformed into a linear programming problem to minimize the overall travel time under the constraints of the total flow and PT flow between grid pairs, as shown by the following mathematical expressions:

$$\min \sum_i^H \sum_j^G duration_{ij} * pt_{ij} \tag{1}$$

$$s.t. \sum_j^G pt_{ij} = pt\_flow\_sta_i \tag{2}$$

$$\sum_i^H pt_{ij} \le total\_flow_j \tag{3}$$

$$duration_{ij} = \begin{cases} \text{The travel time of } r_{jk}, \text{ if } PT_i \text{ matches the planned route } r_{jk} \text{ of } GP_j \\ +\infty, \text{ if } PT_i \text{ does not match any planned routes of } GP_j \end{cases} \tag{4}$$

$$pt_{ij} \in [0, min(pt\_flow\_sta_i, \ total\_flow_j)], \ \text{and is an integer}. \tag{5}$$

Then, pt_{ij} is obtained through linear programming to obtain the PT flow $pt\_flow\_grid_j = \sum_i^H pt_{ij}$ allocated on the grid pair GP_j. So far, we have obtained PT flow and total flow between grid pairs, that is, the research units are unified. By definition, the PTR can be calculated as follows:

$$pt\_ratio_j = \frac{pt\_flow\_grid_j}{total\_flow_j} \tag{6}$$

5. Results

5.1. PT commuting flow

According to the extracted results, 70,415 commuters living or working in Wangjing choose public transportation, including 12,529 home-work station-pairs. In terms of the commuting distance (straight-line distance), 51% of commuters are in the [0, 10 km) range, 43% are in the [10 km, 20 km) range, and 6% are in the [20 km, 44 km) range, indicating that the commuting distance is mostly within 20 km. In addition, the longest distance is from outside Wangjing to the inside Wangjing (43 km). And 55 commuters have a commuting distance of 500 m or less, indicating that some commuters still prefer public transportation even for short-distance commutes.

In terms of the commuting time, due to the pre-screening of card swiping time, the on-duty and off-duty hours are respectively distributed between 0:00 a.m. − 10:00 a.m. and 16:00 p.m. − 24:00 p.m., as shown in Fig. 5 (b). It can be seen that the on-duty hours are more concentrated than the off-duty hours. The largest on-duty travel volume occurs between 8:00 a.m. and 9:00 a.m., accounting for 42%, while the largest off-duty travel volume occurs between 18:00 p.m. and 19:00 p.m., accounting for 31%, and the evening peak lasts for 4 hours between 17:00 p.m. and 21:00 p.m. As for the spatial distribution of the commuting routes, 73 hot commuting routes carry more than 100 passengers, of which 67 are subway routes and 6 are regular bus routes, as shown in Fig. 5 (c). This indicates that subway has become the most important travel mode for commuters in Wangjing area, especially lines 5, 8, 13, 14, and 15 carry a large number of commuters.

5.2. Travel ratio of public transportation (PTR)

After excluding the commuters assigned to the outside grids at the stations near the Wangjing border, the PT commuters living or working in Wangjing dropped from 70,415 to 65,217. Since the total commuting flow is 406,394, the overall PTR in Wangjing is 65,217/406,394 = 16.0%. In 2011, the Ministry of Transport of China states that the PTR (excluding walking) of the "transit city" demonstration cities with rail transit should reach at least 45% (Ministry of Transport of China, 2011), suggesting that Wangjing's PTR is still far below expectations. Moreover, we believe that the calculation results are reliable: on the one hand, it comes from the accuracy and coverage of the original data, with the validated accuracy of Amap positioning data exceeding 90% (Yin et al., 2022), and the smart card data covering a whole month of PT travel records of Beijing residents; on the other hand, the uncertainty of the results mainly comes from the linear programming in the flow assignment, although the minimum travel time as the optimization goal is not always realistic but still reasonable (Ahn and Rakha, 2008; Levinson and Zhu, 2013; Manley et al., 2015; Zhu et al., 2021), and its error is still generally controllable and does not affect the feasibility of the entire computational framework. Another uncertainty may arise from the so-called Braess's Paradox and Price of Anarchy (Roughgarden, 2005), namely the contradiction between individual optimality and global optimality, which may lead to problems with the assumption of linear programming that remains to be further explored. In 2020, the annual report released by the Beijing Transport Research Center states that the PTR in Beijing in 2019 is 31.8% (16.5% for subway and 15.3% for bus), but this value is for the whole of Beijing on the one hand, and for all trips on the other (Beijing Transport Research Center, 2020). Our study object is the PTR in commuting context in the Wangjing area, which may be the reason for the inconsistent results. In addition, the fine-scale PTR calculation results differ significantly from the city-scale results, indicating that the existing coarse-grained statistics may be overly optimistic in estimating urban transportation conditions, which also proves the importance of calculating PTR at high resolution.

For different travel directions, the number of PT commuters in Wangjing area from outside-to-inside, from inside-to-outside, and from inside-to-inside are 44,246, 17,219, and 3,752, respectively. In addition, the total number of commuters is 254,397, 110,766, and 41,231, with PTRs of 17.4%, 15.5%, and 9.1%, respectively. It can be seen that the outside-to-inside PTR is the highest, and the PTR from inside-to-inside is the lowest. After statistics, the average commuting distance from outside-to-inside, inside-to-outside and inside-to-inside is 11.3 km, 8.2 km and 1.8 km respectively, and we can tentatively conclude that the longer the commuting distance is, the larger the PTR is. Fig. 6 (b) shows the frequency distribution of the PTR results for all grid-pairs. Taking the

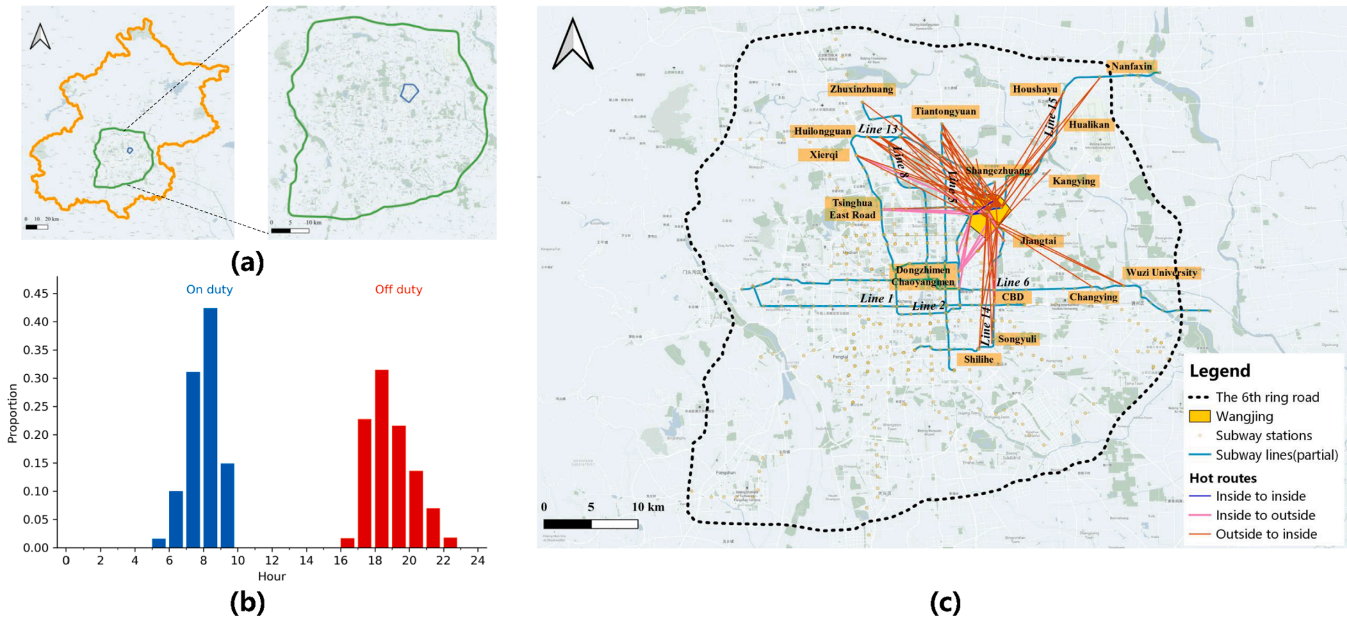

Fig. 5. Temporal and spatial distribution of the PT commuting flow: (a) study area, (b) temporal distribution, (c) hot commuting routes with flow intensity greater than 100.

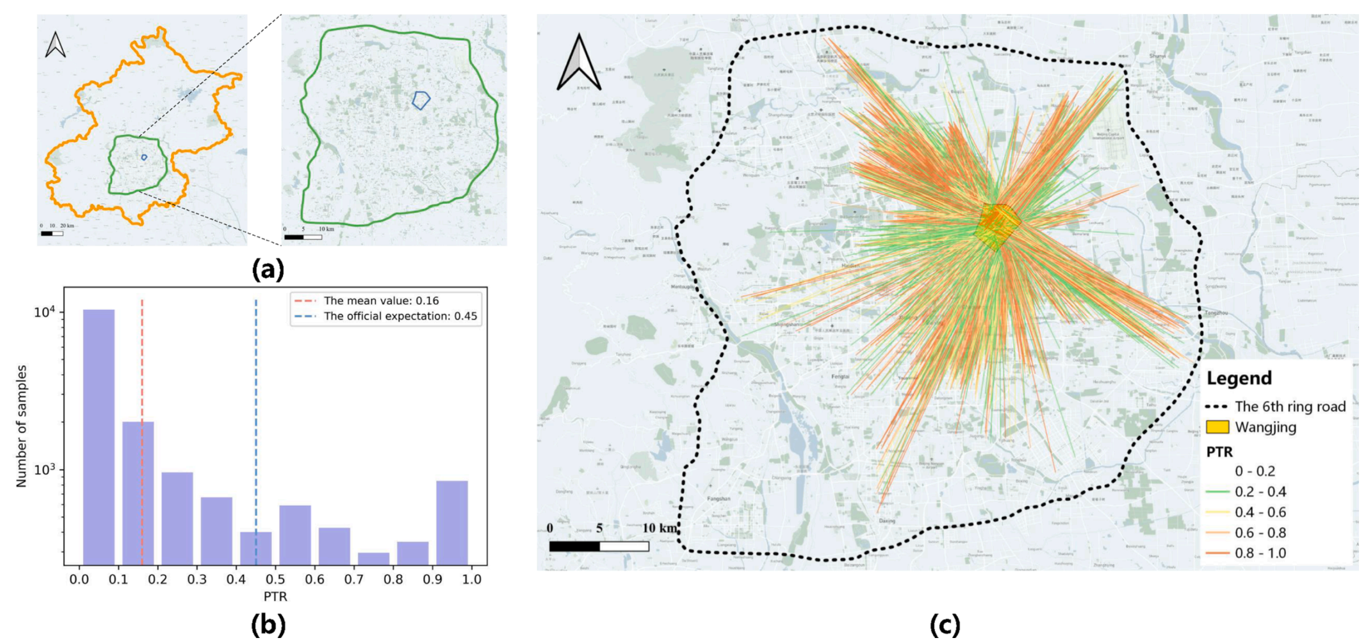

Fig. 6. The calculation results of PTR: (a) study area, (b) frequency histogram, (c) spatial distribution.

mean value (16%) and the official expectation (45%) as the benchmark, we can see that a large percentage of PTR (about 69.4%) is below the mean value, and less than 20% of the grid pairs meet the expectation, which indicates that more work needs to be done to improve the usage of public transportation in Wangjing area. Fig. 6 (c) shows the spatial distribution of PTR, and the OD pairs shows a relatively obvious directionality in general, with long-distance trips tending to have a larger PTR, which is roughly consistent with the previous conclusion.

6. Discussion

6.1. Influencing factors analysis of PTR

PTR is the ratio of PT flow to total flow between an OD pair, reflecting the willingness of urban residents to choose public transportation. Therefore, it is necessary to quantify the influencing factors of PTR, which on the one hand can help to enhance our understanding of human travel behavior, and on the other hand can guide transportation management and urban planning in a targeted manner.

For influencing factors, the following characteristics reflecting travel efficiency and socio-economic features are selected for regression analysis. These characteristics include: the travel time difference between public transportation and private cars (transit gap), walking distance, number of transfers, travel distance, parking density in home and workplace, and housing price, as shown in Table 3.

For the regression models, PTR can also be considered as the probability that a traveler chooses public transportation according to its definition, which depends on the influencing factors mentioned above.

Table 3
Influencing factors of PTR.

Influencing factors	Definition	Reflect
Transit gap	The travel time difference between PT and private cars.	The relative travel efficiency of PT.
Walking distance	The total walking distance when traveling by PT, including "first/last mile" and transfers.	Travel characteristics.
Number of transfers	The number of transfers when traveling by PT.	Travel characteristics.
Travel distance	The road network distance between home and workplace.	Travel characteristics.
Workplace parking density	The number of parking lots in the 1.5 km square buffer of the workplace.	Parking convenience.
Residential parking density	The number of parking lots in the 1.5 km square buffer of the home.	Private car ownership.
Housing price	The average housing price of the residential community.	Resident income level.

Assume that the total flow in a grid pair i is n_i, the travelers between it have the same probability of choosing public transportation, which is denoted as p_i. Therefore, p_i can be regarded as the PTR of the grid pair i. If the travelers are independent of each other, then the PT flow Y_i obeys binomial distribution: $Y_i \sim Binomial(n_i,\ p_i)$. Most studies on PTR have adopted the Binomial regression model, i.e., a generalized linear model with a response variable in a binomial distribution (McCullagh and Nelder, 2019; Shi and Ju, 2015; Wen et al., 2016; Xie, 2018). However, binomial models may suffer from excessive residuals, which are referred to as "over-dispersion". Over-dispersion increases fitting error, but is often seen in practical applications (McCullagh and Nelder, 2019). In this study, the reason may be that there is a strong correlation between travelers in the same OD grid-pairs, which undermines the independence assumption. Nested models have proven to be effective in mitigating the problem of over-dispersion. The parameter p_i is taken as a random variable, and obeys beta-distribution most frequently, thus constructing the Beta-binomial model (Chatfield and Goodhardt, 1970; Gange et al., 1996; McCullagh and Nelder, 2019; Williams, 1975). Details of the mathematical derivation, regression process and result comparison of the regression models are described in the Supplementary Material. Compared with M1 (the Binomial model), M3 (the Beta-binomial model) is reduced by 52.5%, 52.4% and 52.3% on all three indicators, which proves the necessity of considering the Beta prior distribution (Table S1 in Supplementary Material). The mathematical expression for the optimal model M3 is as follows:

$$log\frac{p_i}{1-p_i} = X^T\beta \Rightarrow p_i = \frac{1}{1+\exp\left(-X^T\beta\right)}$$
$$= \frac{1}{1+\exp\left(-\beta_0-\sum_{i=1,2,4,5}\beta_i x_i-\beta_3 x_3 x_4-\beta_6 x_1 x_4\right)} \quad (7)$$

Where x_1 is the transit gap, x_2 is the walking distance, x_3 is the number of transfers, x_4 is the travel distance, and x_5 is the residential parking density. Transit gap is measured in hours, walking distance and travel distance in kilometers, and the residential parking density in 1/2.25 square kilometers. And the regression coefficients are shown in Table 4.

As shown in Table 4, transit gap, walking distance, number of transfers, and residential parking density are negatively correlated with PTR, which is consistent with our common knowledge. The interaction term's coefficient between the travel distance and number of transfers is negative, indicating that with the increase in the travel distance, more transfers lead to a greater decline in PTR. In contrast, the interaction term's coefficient between the travel distance and transit gap is positive, indicating that with increase in travel distance, a smaller decline in PTR happens caused by the increase in the unit time of the transit gap. This is not surprising because people prefer public transportation if they are commuting long distances, as the proportion of the transit gap decreases as the total travel time increases. The relationship between the travel distance and PTR is influenced by the number of transfers and the transit gap. Assuming that the transit gap is 0, if the number of transfers is no more than 2, then the coefficient of the travel distance is positive. Otherwise, the coefficient is negative. Therefore, the travel distance essentially does not have a positive or negative correlation with PTR.

Table 4
The regression coefficients of explanatory variables.

Explanatory variable	Coefficient
Transit gap	−6.52
Walking distance	−1.57
Travel distance	0.09
Residential parking density	−0.001
Travel distance * number of transfers	−0.04
Travel distance * transit gap	0.13

6.2. Potential implication: Finding the areas with unbalanced PT supply and demand

Areas with imbalance between PT supply and demand can be determined on a combination of PTR and the transit gap. PTR represents the willingness of residents to choose public transportation between grid pairs. This study adopts a threshold value of 0.45. The transit gap reflects the relative travel efficiency of public transportation by indirectly considering the walking distance and number of transfers. In this study, 20 min is used as the cutoff for high and low travel efficiency, meaning that if the transit gap exceeds 20 min, the PT system is considered to be in an inefficient state, and vice versa.

First of all, as implied in Table 4, PTR is negatively correlated with the transit gap. At the same time, there are situations where PTR is high when the transit gap is large and PTR is low when the transit gap is small, and these are exactly the situations in the areas with imbalanced supply and demand. As shown in Fig. 7, the red dots indicate the large transit gap and high PTR, suggesting that commuters within these grid pairs significantly depend on public transportation, despite its inefficiency. High PTR and low travel efficiency imply a severe short supply. The yellow dots represent the ideal state, where both PTR and travel efficiency are high. The blue dots represent the high efficiency and low PTR, possibly due to the low willingness of commuters to take public transportation, suggesting a potential oversupply. The gray dots represent the low travel efficiency and PTR, where an increase in travel efficiency may increase PTR, implying a potential short supply. Overall, 115 grid pairs are in severe short supply, 968 are in ideal state, 117 are in potential oversupply, and 2,648 are in potential short supply.

As the key area for optimization, the areas in severe short supply are visualized for further study. To simplify, O(rigin) is preserved for grid pairs from outside-to-inside of Wangjing, and D(estination) is preserved for grid pairs from inside-to-outside, and the visualization results are shown in Fig. 8. The blue grids represent the O points from outside to inside Wangjing, i.e., the home grids, with 55 points. The orange grids represent the D points from inside to outside Wangjing, i.e., the workplace grids, with 16 points. The black grids represent the grids of commuters working in Wangjing and coming from Wangjing to work, including the grids where both of the above cases exist at the same time, with two grids.

Considering the causes of the transit gap and the flow intensity (Table 5), shuttle buses should be added, or the "subway + bus" lines should be customized to alleviate the supply shortage. Based on spatial proximity and directional consistency, some areas have been combined (the merged cells in Table 5): Shahe Higher Education Park - Shahe, Xierqi - Qinghe, Lishuiqiao - Beiyuan, Dongzhoujiayuan - Naizifang Village, Loutai Village - Sakura Garden, and Sanyuanqiao – Dongzhimen. The un-asterisked cells indicate that the PT flow in these areas is greater than 30. It is recommended to open direct shuttle buses to Wangjing in these areas. The cells whose IDs are marked with an asterisk

International Journal of Applied Earth Observation and Geoinformation 118 (2023) 103245

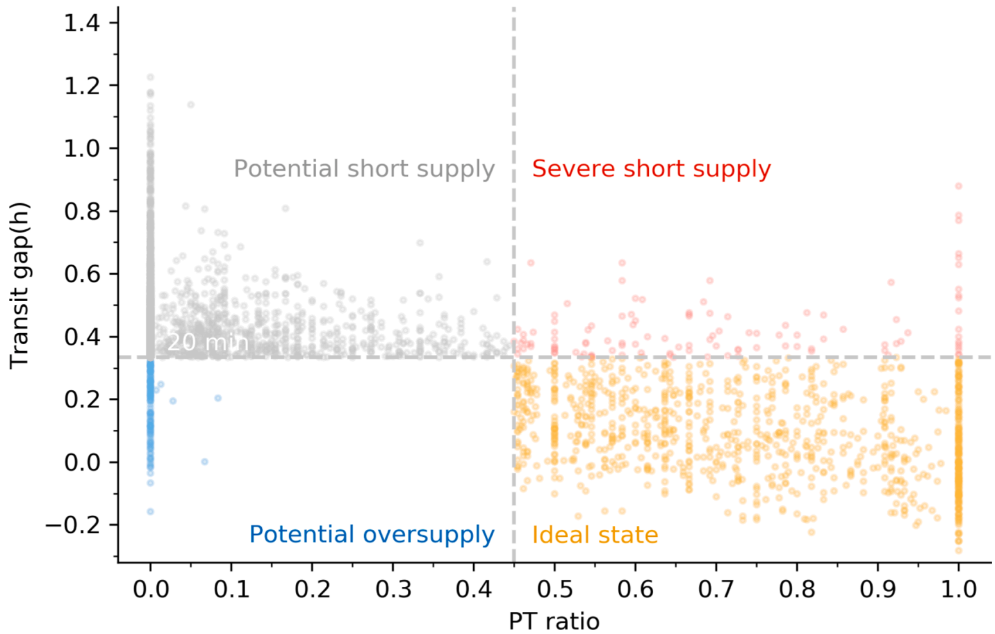

Fig. 7. The division result of public transportation supply and demand.

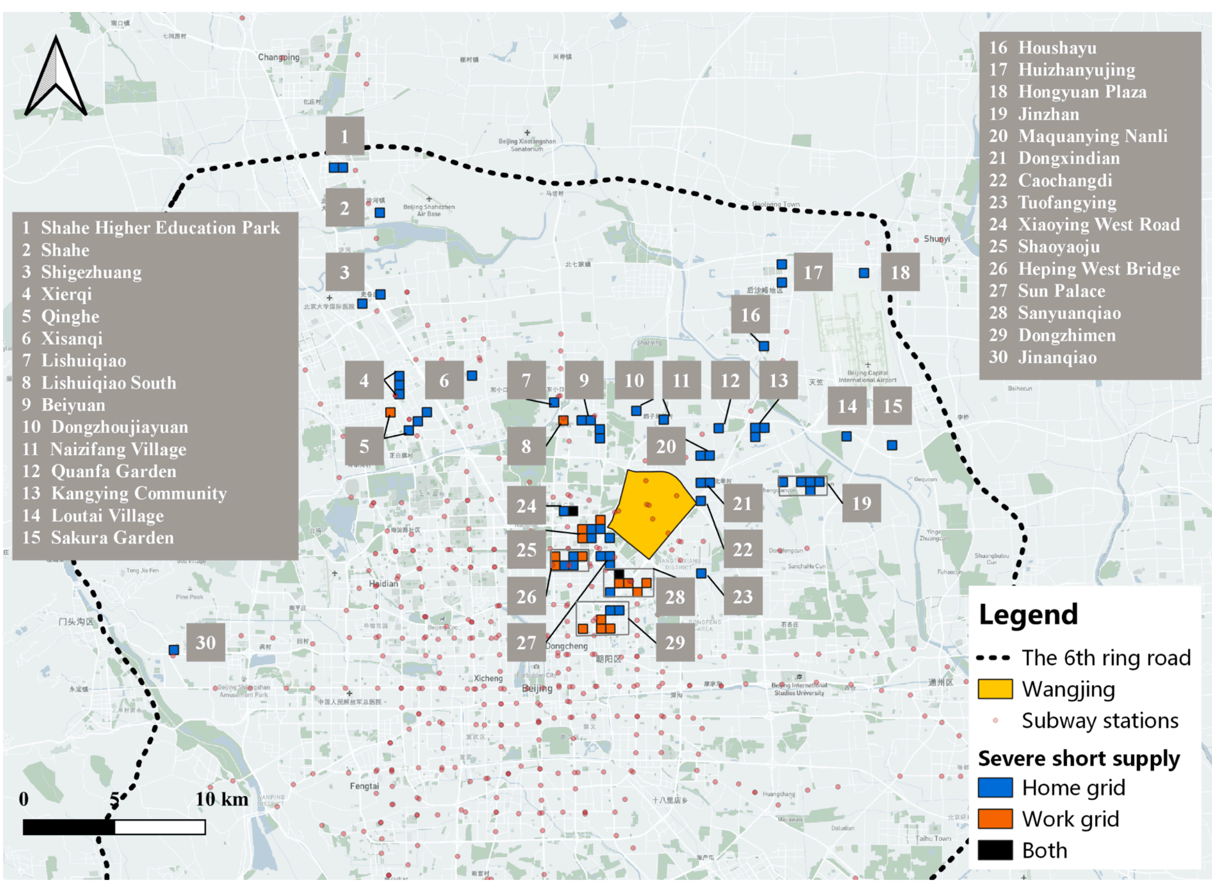

Fig. 8. Areas in severe short supply. The blue grids represent homes; the orange grids represent workplaces; the black grids represent both. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 5
Areas in severe short supply. The ID has no asterisk, indicating that the number of PT commuters exceeds 30, and direct shuttle buses to Wangjing can be added; the ID is marked with an asterisk, which means that the PT flow does not exceed 30, and buses can be added to connect to the subway stations or shuttle buses can be added in combination with the flow in the surrounding area.

ID	Region	PT flow/ Total flow	Sum	ID	Region	PT flow/ Total flow	Sum
			From outside to inside Wangjing				
1	Shahe Higher Education Park	61/73	68/85	16	Houshayu	76/105	76/105
2	Shahe	7/12		17	Huizhan -yujing	33/35	33/35
3	Shigezhuang	34/48	34/48	18*	Hongyuan Plaza	12/12	12/12
4	Xierqi	50/54	69/78	19	Jinzhan	116/178	116/178
5	Qinghe	19/24		20	Maquanying Nanli	42/48	42/48
6*	Xisanqi	13/13	13/13	21	Dongxindian	44/79	44/79
7	Lishuiqiao	6/13	66/100	22*	Caochangdi	16/27	16/27
9	Beiyuan	60/87		23*	Tuofangying	24/40	24/40
10	Dongzhoujia -yuan	20/29	36/60	24*	Xiaoying West Road	16/25	16/25
11	Naizifang Village	16/31		25	Shaoyaoju	69/110	69/110
12*	Quanfa Garden	19/27	19/27	26	Heping West Bridge	34/52	34/52
13	Kangying community	106/160	106/160	27	Sun Palace	48/75	48/75
14	Loutai Village	9/18	33/43	28	Sanyuanqiao	21/33	43/63
15	Sakura Garden	24/25		29	Dongzhimen	22/30	
				30*	Jinanqiao	5/11	5/11
			From inside to outside Wangjing				
5*	Qinghe	26/29	26/29	27	Heping West Bridge	58/103	58/103
8*	Lishuiqiao South	6/13	6/13	28	Sanyuanqiao	54/92	54/92
24*	Xiaoying West Road	6/11	6/11	29	Dongzhimen	150/219	150/219
25*	Shaoyaoju	27/50	27/50				

represent flow not exceeding 30. Buses can be added to connect to the subway stations or shuttle buses can be added in conjunction with the flow in surrounding areas.

7. Conclusion

In this paper, a novel computational framework is proposed for calculating PTR in high resolution and quantifying its influencing factors. First, PT commuting flow and total commuting flow can be obtained based on multisource geospatial big data. Then, we propose a linear programming-based three-step method to assign PT flow between stations to grids, which realizes the calculation of PTR between grid pairs (500 m) for the first time. Besides, we develop a Beta-binomial model to address the over-dispersion problem in existing studies, and quantitatively analyze the impact of various travel efficiency and socioeconomic characteristics on PTR. Finally, we provide a new strategy to evaluate the PT supply–demand status, and give some feasible suggestions to inspire the transit optimization in Wangjing.

Nevertheless, this study also has some limitations. For example, in calculating PTR, the PT flow is allocated through linear programming to seek the minimum total travel time, which may differ from the realistic travel behaviors of passengers. The population distribution of each grid cell also plays a role in the flow allocation, and should be considered. Travel questionnaires can be conducted in some areas to validate the calculation results of PTR and more cities can be used as case studies to test the robustness of the proposed approach. Also, additional influencing factors, such as the per capita income of residents and the degree of traffic congestion, can be included in the regression analysis for further exploration.

CRediT authorship contribution statement

Ganmin Yin: Conceptualization, Methodology, Software, Writing – original draft. **Zhou Huang:** Conceptualization, Methodology, Writing – review & editing. **Liu Yang:** Conceptualization, Methodology, Software. **Eran Ben-Elia:** Validation, Writing – review & editing. **Liyan Xu:** Validation, Visualization. **Bronte Scheuer:** Writing – review & editing. **Yu Liu:** Validation, Writing – review & editing.

Declaration of Competing Interest

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

Data availability

The authors do not have permission to share data.

Acknowledgements

We acknowledge the financial support from the National Natural Science Foundation of China (42271471, 42201454, 41830645), the International Research Center of Big Data for Sustainable Development Goals (CBAS2022GSP06), and China Postdoctoral Science Foundation (2022M710193). We also appreciate the detailed comments from the Editor and the anonymous reviewers.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jag.2023.103245.

References

Ahn, K., Rakha, H., 2008. The effects of route choice decisions on vehicle energy consumption and emissions. Transp. Res. Part D Transp. Environ. 13, 151–167. https://doi.org/10.1016/J.TRD.2008.01.005.
Beijing Transport Research Center, 2020. 2020 Beijing Transport Annual Report.
Ben-Akiva, M., Bierlaire, M., 1999. Discrete Choice Methods and their Applications to Short Term Travel Decisions. Springer, Boston, MA, Boston. https://doi.org/10.1007/978-1-4615-5203-1_2.
Bao, Y., Huang, Z., Li, L., Wang, Y., Liu, Y., 2021. A BiLSTM-CNN model for predicting users' next locations based on geotagged social media. Int. J. Geogr. Inf. Sci. 35, 639–660. https://doi.org/10.1080/13658816.2020.1808896.
Ben-Elia, E., Benenson, I., 2019. A spatially-explicit method for analyzing the equity of transit commuters' accessibility. Transp. Res. Part A Policy Pract. 120, 31–42. https://doi.org/10.1016/J.TRA.2018.11.017.
Benenson, I., Ben-Elia, E., Rofé, Y., Geyzersky, D., 2017. The benefits of a high-resolution analysis of transit accessibility. Int. J. Geogr. Inf. Sci. 31, 213–236. https://doi.org/10.1080/13658816.2016.1191637.

Bresson, G., Dargay, J., Madre, J.L., Pirotte, A., 2003. The main determinants of the demand for public transport: a comparative analysis of England and France using shrinkage estimators. Transp. Res. Part A Policy Pract. 37, 605–627. https://doi.org/10.1016/S0965-8564(03)00009-0.

Bureau of Statistics of Chaoyang District, 2011. Communiqué on Main Data of the Sixth National Census of Chaoyang District in 2010(in Chinese).

Chatfield, C., Goodhardt, G.J., 1970. The Beta-Binomial Model for Consumer Purchasing Behaviour. J. R. Stat. Soc. Ser. C (Applied. Stat. 19, 240–250. https://doi.org/10.2307/2346328.

Chen, Y., Bouferguene, A., Li, H.X., Liu, H., Shen, Y., Al-Hussein, M., 2018. Spatial gaps in urban public transport supply and demand from the perspective of sustainability. J. Clean. Prod. 195, 1237–1248. https://doi.org/10.1016/J.JCLEPRO.2018.06.021.

Cormen, T., Leiserson, C., Rivest, R., Stein, C., 2022. Introduction to algorithms.

Currie, G, Sarvi, A.M., Young, A.B., Sarvi, M., Young, B, Currie, Graham, Majid, A.E., Ae, S., Young, Bill, 2006. A new approach to evaluating on-road public transport priority projects: balancing the demand for limited road-space. Transp. 2006 344 34, 413–428. https://doi.org/10.1007/S11116-006-9107-3.

de Alwis Pitts, D.A., So, E., 2017. Enhanced change detection index for disaster response, recovery assessment and monitoring of accessibility and open spaces (camp sites). Int. J. Appl. Earth Obs. Geoinf. 57, 49–60. https://doi.org/10.1016/J.JAG.2016.12.004.

Ding, L., 2016. A study of travel behavior under transit priority strategies considering traffic state(in Chinese). Southeast University.

Ferguson, N.S., Lamb, K.E., Wang, Y., Ogilvie, D., Ellaway, A., 2013. Access to Recreational Physical Activities by Car and Bus: An Assessment of Socio-Spatial Inequalities in Mainland Scotland. PLoS One 8, e55638.

Gange, S.J., Muñoz, A., Sáez, M., Alonso, J., 1996. Use of the Beta-Binomial Distribution to Model the Effect of Policy Changes on Appropriateness of Hospital Stays. J. R. Stat. Soc. Ser. C (Applied. Stat. 45, 371–382. https://doi.org/10.2307/2986094.

Gass, S., 2003. Linear programming: methods and applications.

Haitao, H., Yang, K., Liang, H., Menendez, M., Guler, S.I., 2019. Providing public transport priority in the perimeter of urban networks: A bimodal strategy. Transp. Res. Part C Emerg. Technol. 107, 171–192. https://doi.org/10.1016/J.TRC.2019.08.004.

Hansen, W.G., 1959. How Accessibility Shapes Land Use. J. Am. Inst. Plann. 25, 73–76. https://doi.org/10.1080/01944365908978307.

Huang, Z., Yin, G., Peng, X., Zhou, X., Dong, Q., 2023. Quantifying the environmental characteristics influencing the attractiveness of commercial agglomerations with big geo-data. Environ. Plan. B Urban Anal. City Sci., 239980832311583 https://doi.org/10.1177/23998083231158370.

Järv, O., Tenkanen, H., Salonen, M., Ahas, R., Toivonen, T., 2018. Dynamic cities: Location-based accessibility modelling as a function of time. Appl. Geogr. 95, 101–110. https://doi.org/10.1016/J.APGEOG.2018.04.009.

Levinson, D., Zhu, S., 2013. A portfolio theory of route choice. Transp. Res. Part C Emerg. Technol. 35, 232–243. https://doi.org/10.1016/J.TRC.2013.03.001.

Ling, X., Yang, T., Shi, Q., 2014. Discussion on Public Transit Mode Share(in Chinese). Urban Transp. China 12, 26–33.

Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., Shi, L., 2015. Social Sensing: A New Approach to Understanding Our Socioeconomic Environments. Ann. Assoc. Am. Geogr. 105, 512–530. https://doi.org/10.1080/00045608.2015.1018773.

Mallick, R.K., Routray, J.K., 2001. Identification and accessibility analysis of rural service centers in Kendrapara District, Orissa, India: a GIS-based application. Int. J. Appl. Earth Obs. Geoinf. 3, 99–105. https://doi.org/10.1016/S0303-2434(01)85027-3.

Manley, E.J., Addison, J.D., Cheng, T., 2015. Shortest path or anchor-based route choice: a large-scale empirical analysis of minicab routing in London. J. Transp. Geogr. 43, 123–139. https://doi.org/10.1016/J.JTRANGEO.2015.01.006.

McCullagh, P., Nelder, J.A., 2019. Generalized Linear Models, 2nd Editio. ed. Routledge, New York. https://doi.org/10.1201/9780203753736.

Ministry of Transport of China, 2011. Notice on Issues Concerning the Development of the National Public Transport City Construction Demonstration Project (in Chinese).

Mishra, S., Welch, T.F., Jha, M.K., 2012. Performance indicators for public transit connectivity in multi-modal transportation networks. Transp. Res. Part A Policy Pract. 46, 1066–1085. https://doi.org/10.1016/J.TRA.2012.04.006.

Outwater, M.L., Spitz, G., Lobb, J., Campbell, M., Sana, B., Pendyala, R., Woodford, W., 2011. Characteristics of premium transit services that affect mode choice. Transportation (Amst). 38, 605–623. https://doi.org/10.1007/S11116-011-9334-0.

Redman, L., Friman, M., Gärling, T., Hartig, T., 2013. Quality attributes of public transport that attract car users: A research review. Transp. Policy 25, 119–127. https://doi.org/10.1016/J.TRANPOL.2012.11.005.

Roughgarden, T., 2005. Selfish routing and the price of anarchy.

Schrijver, A., 1998. Theory of linear and integer programming.

Shi, F., Ju, Y., 2015. Analysis on influencing factors of public transportation share: An empirical study of central Nanjing(in Chinese). City Plan. Rev. 76–84.

Song, Y., Wu, P., Hampson, K., Anumba, C., 2021. Assessing block-level sustainable transport infrastructure development using a spatial trade-off relation model. Int. J. Appl. Earth Obs. Geoinf. 105, 102585. https://doi.org/10.1016/J.JAG.2021.102585.

Wang, H., Huang, Z., Zhou, X., Yin, G., Bao, Y., Zhang, Y., 2022. DouFu: A double fusion joint learning method for driving trajectory representation. Knowledge-Based Syst 258. https://doi.org/10.1016/j.knosys.2022.110035.

Wang, Y., Zhu, D., Yin, G., Huang, Z., Liu, Y., 2020. A unified spatial multigraph analysis for public transport performance. Sci. Rep. 10 https://doi.org/10.1038/s41598-020-65175-x.

Wang, Y., Huang, Z., Yin, G., Li, H., Yang, L., Su, Y., Liu, Y., Shan, X., 2022. Applying Ollivier-Ricci curvature to indicate the mismatch of travel demand and supply in urban transit network. Int. J. Appl. Earth Obs. Geoinf. 106, 102666. https://doi.org/10.1016/J.JAG.2021.102666.

Wen, X., Zhang, Z., Yang, T., 2016. Research on public transportation share rate calculation model construction and empirical(in Chinese). J. Chongqing Jiaotong Univ. Sci. 35, 127–132.

Williams, D.A., 1975. 394: The Analysis of Binary Responses from Toxicological Experiments Involving Reproduction and Teratogenicity. Biometrics 31, 949. https://doi.org/10.2307/2529820.

Xie, C., 2018. Research on the travel effect of international metropolitan transportation infrastructure construction (in Chinese). Shanghai Academy of Social Sciences.

Yin, G., Huang, Z., Bao, Y., Wang, H., Li, L., Ma, X., Zhang, Y., 2022. ConvGCN-RF: A hybrid learning model for commuting flow prediction considering geographical semantics and neighborhood effects. Geoinformatica 1–21. https://doi.org/10.1007/s10707-022-00467-0.

Zhang, T., Wang, B., 2018. Research on public traffic share ratio of Shanghai transport in 2035 (in Chinese). Commun. Shipp. 5, 42–49.

Zhao, N., Cao, G., Zhang, W., Samson, E.L., Chen, Y., 2020. Remote sensing and social sensing for socioeconomic systems: A comparison study between nighttime lights and location-based social media at the 500 m spatial resolution. Int. J. Appl. Earth Obs. Geoinf. 87, 102058. https://doi.org/10.1016/J.JAG.2020.102058.

Zhu, Z., Mardan, A., Zhu, S., Yang, H., 2021. Capturing the interaction between travel time reliability and route choice behavior based on the generalized Bayesian traffic model. Transp. Res. Part B Methodol. 143, 48–64. https://doi.org/10.1016/J.TRB.2020.11.005.