



Early birds, night owls, and tireless/recurring itinerants: An exploratory analysis of extreme transit behaviors in Beijing, China

Ying Long^a, Xingjian Liu^b, Jiangping Zhou^{c,*}, Yanwei Chai^d

^a Hang Lung Center for Real Estate, Tsinghua University, China; School of Architecture, Tsinghua University, China

^b The University of Hong Kong, China

^c University of Queensland, Australia

^d Peking University, China

ARTICLE INFO

Article history:

Received 23 February 2016

Received in revised form 21 April 2016

Accepted 20 August 2016

Available online xxx

Keywords:

Extreme transit behavior

Public transit

Smart card data (SCD)

Travel survey

Big data

ABSTRACT

This paper seeks to understand extreme public transit riders in Beijing using both traditional household surveys and emerging new data sources such as Smart Card Data (SCD). We focus on four types of extreme transit behaviors: public transit riders who (1) travel significantly earlier than average riders ('early birds'); (2) ride in unusual late hours ('night owls'); (3) commute in excessively long distance ('tireless itinerants'); and (4) make significantly more trips per day ('recurring itinerants'). SCD are used to identify the spatiotemporal patterns of these four extreme transit behaviors. In addition, household surveys are employed to supplement the socioeconomic background and tentatively profile extreme travelers. While the research findings are useful to guide urban governance and planning in Beijing, our methodology and procedures can be extended to understand travel patterns elsewhere.

© 2016 Published by Elsevier Ltd.

1. Introduction

Extreme conditions often capture our attention and point to important underlying mechanism; we have learned a great deal about our cities by examining the extremes, such as the emergence and dynamics of the most dominant city of a nation, the most depressed city of a region, as well as the most popular gateway city among immigrants. In the past decade or so, against the backdrops of the global financial crisis, increased numbers of the unemployed, self-employed and part-time workers, the rise of telecommuters as well as the relocation of low-paying jobs, extreme commuters have received increasing academic and public attention in recent years. As extreme commuting accounts for an increased portion of daily residential trips, recent analysis starts to look into travelers making unusually long, early, late, and/or frequent trips, which have discretely explored or described by Barr, Fraszczyk, and Mulley (2010); Gregor (2013); Jones (2012); Landsman (2013); Marion and Horner (2007); Moss and Qing (2012); Rapino and Fields (2013); U.S. Census (2005).

As scholarly work on extreme travelers has been largely developed based on North American and European cities, we are interested in extending the framework to understand extreme trips in China. We will examine extreme travel behaviors in public transit, as Chinese cities have historically relied on public transit, and the Chinese government has sought public transit as a major remedy for congestions, pollution, and other issues caused by the rising car ownership.

* Corresponding author.

Email addresses: (Y. Long); xliu6@hku.hk (X. Liu); jp.zhou@uq.edu.au (J. Zhou); chyw@pku.edu.cn (Y. Chai)

To date, extreme traveler analyses have mostly used traditional data such as travel diaries and household surveys. More recently, emerging big data sources such as transit smart card data have been utilized to investigate the phenomenon. Transit smart card data record rich information about individual trips (e.g., origin, destination, and time length) and thus could be a useful supplementary data source for understanding travel behaviors. For example, since the 1990s, the use of smart cards has become prevalent in a large number of Chinese cities, partly owing to the development of the Internet and the advancement of mobile communication technologies (Blythe, 2004). Furthermore, Intelligent Transportation Systems (ITS) that incorporate smartcard-automated fare systems had been in place in over 100 Chinese cities as of 2007 (Zhou & Long, 2014). The combination of these conventional and emerging data sources could offer new opportunities to deepen our understandings of cities and related routine activities such as traveling and commuting (Batty, 2012; Liu et al., 2015).

While most literature on extreme travelers focuses on excessively long trips, we extend the definition of extreme travelers into four types with the context of Chinese society. The other three types of extreme travelers are public transit riders (1) who make significantly more trips ('recurring itinerants'), (2) travel significantly earlier than average riders (the 'early birds') during weekdays, and (3) ride in unusual late hours (the 'night owls') during weekdays. More specifically, we seek to identify these extreme travelers in Beijing, characterize their spatiotemporal trajectories, profile their socioeconomic backgrounds, and propose necessary policy implications on the phenomenon.

Extreme transit behavior is becoming more prevalent in China and moving to the center of government agendas (Long & Thill, 2015). Without a solid understanding of extreme transit patterns, government programs could easily result in misinformed interventions and policy failures. Therefore, this study seeks to explore the spatiotemporal trajectories of extreme travelers based on refined definitions of 'extreme transit behavior', using Beijing as a case study. Our analysis would leverage the power of emerging public transit data and seek to answer: (1) Are there large number of extreme travelers? (2) Where do these extreme travelers live and work? (3) What are their socioeconomic characters? In addition, our analysis will also tentatively address concerns about the causes of extreme transit behavior.

The remainder of this article is organized in five main sections. First, we relate our analysis to the ongoing debate on the causes and consequences of extreme transit behavior. Second, we detail our data sources, which include both smart card data and conventional household travel surveys. Third, we provide working definitions of extreme transit behavior and describe our empirical framework, which is a modification of the methodology detailed in Long and Thill (2015). Fourth, we summarize key findings based on our exploratory analyses. We conclude with a discussion of empirical contributions and avenues for future research.

2. Relevant literature

Passenger trips have long been of interest to transportation planners and modelers. In existing literature, they have been primarily classified according to trip purpose, time of day, day of week, mode, person type, frequency, activity duration, and route choice (Meyer & Miller, 2001). Emerging big data such as smart card data have enabled scholars to examine more types or aspects of passenger trips, over more time horizons and in large sample sizes, as compared to traditional data (Bagchi and White, 2005). Multi-day data of transit riders, for instance, were once difficult to collect if we rely on traditional methods such as surveys or interviews to collect data. But smart card data make it possible. In addition, smart card data can greatly facilitate our studies of activity space, locations and departure time of about 80% of all transit riders (Chu & Chapleau, 2010).

Extra information provided by smart card data enables us to better plan and manage our transit services (Frumin & Zhao, 2012). Utilizing those data, for instance, we can now identify and visualize over 80% of the transit riders' route choices (Tao, Corcoran, Mateo-Babiano, & Rohde, 2014a, Tao, Rohde, & Corcoran, 2014b). Such data also capture a great deal about revealed preference and could thus be useful for evidence-based planning and/or travel demand modeling (Janosikova, Slavik, & Kohani, 2014; Tao et al., 2014a).

Most smart card data, however, are not designed to capture all information about passenger trips, especially socioeconomic information about trip makers and their trip purposes (Pelletier, Morency, & Trépanier, 2011). Extra work is needed to connect smart card data to more information about trip makers. A small but an increased number of scholars have done so. By combing smart card data and socioeconomic information at a fine-grained spatial level, Mohamed, Côme, Baro, and Oukhellou (2014) show how we can detect which social groups (e.g. workers and students) travel, when/why they travel and where they travel to and from. Kusakabe and Asakura (2014) propose and implement a data fusion approach to inferring trip purposes of transit riders. In the same vein, Lee and Hickman (2014) develop heuristic rules and learning algorithms to infer trip purposes of smart card users. Refereed articles similar to the above cited ones, however, as a whole are still small in quantity. They nevertheless have already shown great potential of smart card data. Alone, they may only pro-

vide limited information about transit riders. When they are combined with other small data such as household surveys, land use and census, the combined data can allow us to understand transit riders better in theory for every time horizon and to make more informed city-planning and public-policy decisions based on the improved understandings (e.g., see Batty, 2013a).

Despite the progresses mentioned above, little has been done on extreme travelers based on smart card data. Most of the existing work on extreme travelers is by reporters, based on discrete evidences and/or personal stories. Landsman (2013), for instance, identifies several "extreme commuters" in New York, USA, who have a work trip taking 90 min or more each way and cites other sources to show that those commuters are not just a few. Similarly, Gregor (2013) reports several extreme commuters in London, UK and contends that the number of such commuters is growing. Per some reporters, the rise of extreme commuters are composite results of the tough labor market, dual-worker problem and/or personal choices, e.g., better education for children rather than proximity to workplace (e.g., Moss, 2012; Jones, 2012; Gregor, 2013).

Based on traditional data such as census data, nevertheless, extreme travel has been paid much attention by both governmental agencies and individual scholars. Moss and Qing (2012) argue that work hours, locations and styles have greatly changed due to the rise of modern Internet and Communication Technologies (ICT) and the increasingly globalized economy. These changes contribute to the emergence and growth of "super commuters", who work in the central county of a given metropolitan area, but live beyond the boundaries of that metropolitan area, commuting long distance by air, rail, car, bus, or a combination of modes once or twice per week. Among the top 5 USA counties for super-commuting, each county has 7.3 to 13.2 percent of the workforce that belongs to super commuters. But even before the emergence or growth of super commuters, "extreme commuters" have been profiled and quantified by USA Census (2005), which defines extreme commuters as those workers who travel 90 min or more to work, one-way and collects related data about them. Based on the above, Rapino and Fields (2013) propose three definitions regarding long commuting: (1) Extreme Commuting: Traveling 90 or more minutes to work. (2) Long-distance Commuting: Traveling 50 or more miles to work. (3) Mega Commuting: Traveling 90 or more minutes and 50 or more miles to work. In this study, we assume that human being's tolerance of long travel is similar across countries. Therefore, the above thresholds were used to categorize extreme commuting in Beijing as well.

Why long commuting or extreme travel have been given so much attention and should be given more attention? Existing literature has rarely directly dealt with this. But several streams of literature have provided some clues. The first stream argues that travel time is disutility and long travel time impairs people's access to opportunities that increase their utility (e.g., see Lyons & Urry, 2005 for a good review). Therefore, extreme travelers would have to face more disutility and have worse access to opportunities. Scholars and decision-makers should examine who they are, where they are and what we can and should do to alleviate their suffering from the public-policy perspective.

The second stream considers travel time as disutility and at the same time contends that long travel (including extreme travel) can result from a host of factors such as 'motoring poor', automobile mismatch, spatial mismatch, centralization of services or facilities, suburbanization of low-skilled jobs but not corresponding labor force, race discrimination and low household income (e.g., Froud, Johal, Leaver, & Williams, 2002; Gobillon, Selod, & Zenou, 2007; Kenyon

et al., 2002; Taylor & Ong, 1995). In light of this, one can conclude that long travel is not just a transportation problem and it can also be a symptom or indication of serious social, economic and racial problems. Examining extreme travel thus can help us better qualify, diagnose and even cure those problems.

The third stream relates commuting to issues such as car dependence, energy waste, air pollution, traffic congestion and urban sprawl. Long travel, long commutes in particular, can exacerbate those issues. Shortening commuting trips and/or jobs-housing balance thus are much needed (California Planning Round Table, 2008; Cervero, 1991; Horner, 2004; Weitzel, 2003).

In addition to long commuting, the other three types of extreme travelers: early birds, night owls, and frequent travelers among transit riders should also be given much more attention. But few existing studies have looked at this. This may be simply because transit riders in most cases only account for a small percentage of all the travelers in developed countries and the other three types of extreme travelers therefore account for an even smaller percentage. But in developing countries like China, the numbers of transit riders and early birds, night owls, and frequent travelers among transit riders could still be significant. In Beijing, for instance, 29% of all the travelers use transit and therefore there are millions of them. For early birds and night owls, they have to break their normal biological clock to make their trips. In the long run, this could not only be harmful to their health but also to their social life (e.g., see Mintz, 2014). For frequent travelers, like the long commuters, they have to suffer from lots of disutilities. Therefore, all the four types of extreme travelers do matter to people's welfare and thus have important policy implications.

To sum up, although extreme commuting patterns have important policy implications, little has been done in the scholarly work on extreme travelers and ever less has been done in China. Existing exploration also rarely adopts emerging data sources. This study attempts to fill some existing gaps by examining extreme transit behaviors based on smart card data. We will benchmark our results from smart card data (SCD) with those derived from the more conventional travel survey. This benchmarking exercise would also provide some clues to the socioeconomic backgrounds of extreme travelers, as recorded in the travel survey data.

3. Study area and data

3.1. Beijing and its transit system

Our study area – Beijing Metropolitan Area (BMA), China – covers an area of 16,410 km² and has a population of more than 20 million in 2010. The BMA lies in northern China, to the east of the Shanxi altiplano and south of the Inner Mongolian altiplano. The southeastern part of the BMA is a flatland, extending east for 150 km to the coast of the Bohai Sea. Mountains cover an area of 10,072 km², 61% of the whole study area. Gaining momentum from China's recent economic success, Beijing, as the capital city, is becoming one of the world's most populous and fastest growing metropolis. (See Yang, Cai, and Ottens (2013) for more background information about Beijing).

Beijing's public transit system consists mainly of buses and subway lines. Thanks to the continuous expansion of subway lines as well as subsidies for public transit, the share of subway and bus trips among all commutes in BMA has risen steadily, reaching 38.9% in the end of 2010, making Beijing the largest public transit system in terms of daily ridership in China (Beijing Transportation Research Center, 2011). The car was still the most dominant mode of travel in Beijing, largely due to recent growth in car ownership and usage.

Contrary to many developed cities where the share of biking and walking is reviving and increasing, the trend in Beijing was a decline since 2010.

As already implied, GIS data about bus lines & stops, and subway routes & stations are essential for geocoding and mapping SCD. In 2010, there were 1928 bus routes (Fig. 1a) and 21,372 bus stops (Fig. 1b) in the BMA. Note that a bus route could have two directions, e.g., the bus No. 113 has two routes, one from Dabeiyao to Qijiahuozi and the other from Qijiahuozi to Dabeiyao. These routes are counted separately in this paper. The average distance between consecutive bus stops is around 231 m. As of April in 2010, there are 9 subway lines, including the airport express line, and 147 subway stations associated with (Fig. 1c).

We use Beijing TAZ data to aggregate the analytical results for better visualization. In total, 1911 TAZs are defined (Fig. 1d) according to the administrative boundaries, main roads, and the planning layout in the BMA.

3.2. Beijing bus/metro smart card data in 2010

Since 2005, over 90% of bus/metro riders in Beijing use anonymous smart cards to pay their fare. This high rate of smart card usage among bus/metro riders is largely due to governmental subsidies for public transit. For example, riders using smart card data enjoy 60% discounts on any routes in the local bus system, and the discount rate for students is at 80%. Smart cards also enable cardholders to pay for other services such as taxi, electricity, and sewage that are offered by the local government or government linked companies.

When cardholders use their smart cards to pay for bus/metro services, card readers installed on the bus and in the station automatically record information about trip origin and/or destination stops/stations, boarding and/or alighting time, smart card numbers, as well as the card type (e.g. student cards versus regular cards).

Information collected for bus trips needs some further elaboration. There exist types of bus fares in 2010 (all bus lines have been turned into distance based fare since January 2015). The first fixed fare is for short routes most of which are within the fifth ring road, while the second distance based fare is associated with long routes. For the first type, 0.4 Chinese Yuan (CNY; approximately 0.06 USD) is charged for individual bus rides in 2010, and the corresponding SCD contains only the departure (tap-in) time and no departure stop ID, arrival (tap-off) time or arrival stop ID. Cardholders' spatiotemporal information is incomplete for this kind of route. Our strategy for tackling this incomplete information issue of fixed-fare records have been elaborated in Long and Thill (2015), and the strategy proposed by Ma, Wang, Chen, and Liu (2012) could be used in future for solving the incomplete information of Beijing bus SCD. For the distance-based fare type, the amount charged depends on the route ID and trip distance, and the SCD contains both tap-in and tap-off information. Both types of bus SCD records are used in this paper for analyzing extreme bus/metro riders and profiling their mobility patterns. Still, one subway ride costs 2 CNY, regardless of the trip distance and time, and the subway records have complete spatiotemporal information.

We collected one-week bus/metro SCD in 2010 (5–11 April, Monday to Sunday) from Beijing Municipal Administration & Communications Card Co. (BMAC). The dataset records 97.9 million trips (59.3 m fixed fare bus trips, 23.4 m distance based fare bus trips, and 15.2 m metro journeys) for 10.5 million cardholders. SCD have been geocoded using bus/metro routes and the locations of stops/stations (Fig. 1). Only SCD records in weekdays (5–9 April) are used in

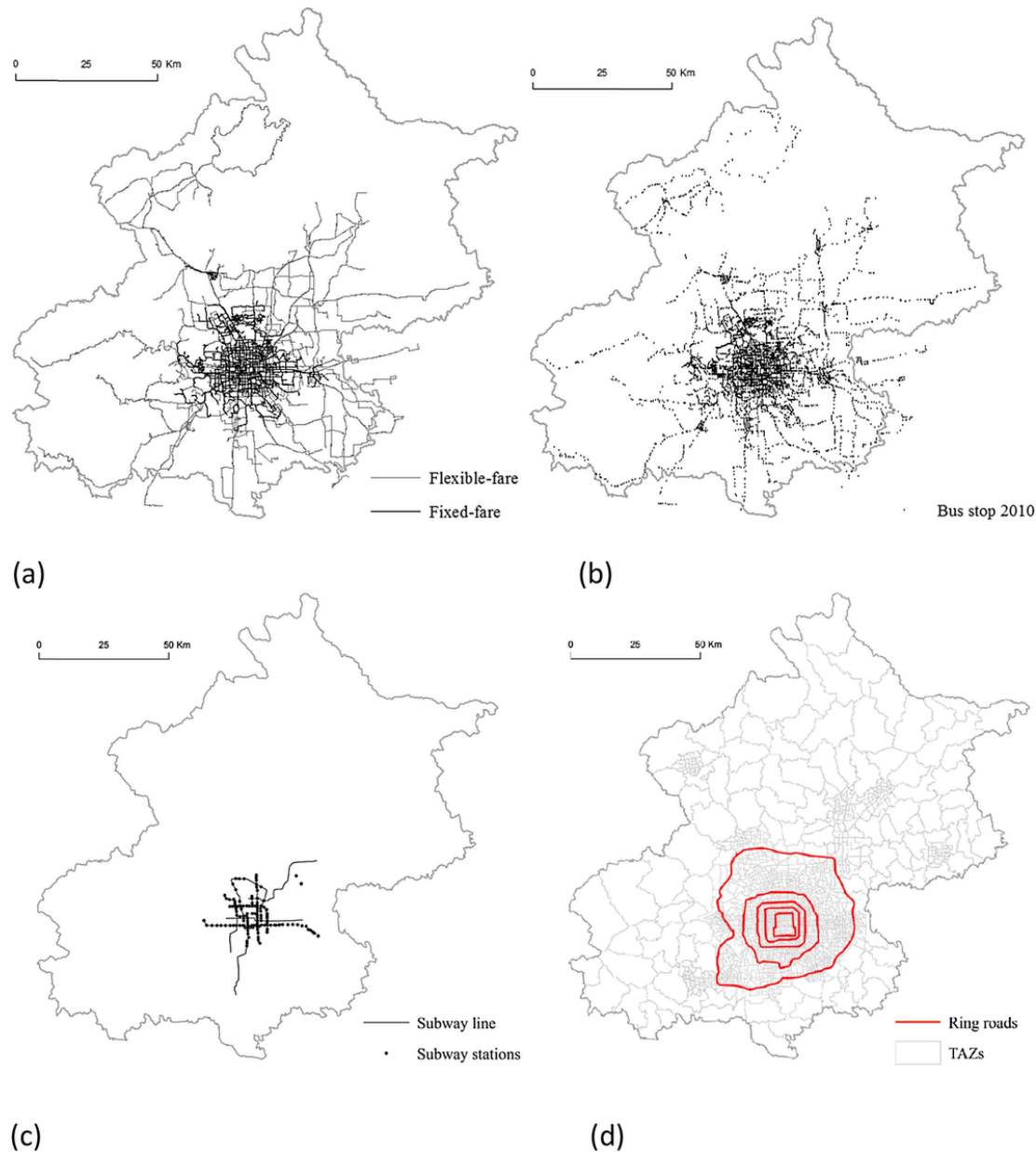


Fig. 1. Bus routes in 2010 (a), bus stops in 2010 (b), subway lines and stations in 2010 (c), and traffic analysis zones in 2010 (TAZs) (d) of the BMA. Note: All maps are from the Beijing Institute of City Planning. Some bus routes and stops are outside the BMA, as shown in a-d since some residents live outside the BMA and in adjacent towns in Hebei province.

the subsequent analysis, and then there are 9.4 million active cardholders.

3.3. Household travel survey of Beijing in 2010

The 2010 Beijing Household Travel Survey (the 2010 survey hereafter) is used to profile individual extreme travelers, or more specifically, extreme transit riders in the city. The 2010 survey represents the conventional approach to understanding transit behaviors. This survey adopts a multistage sampling strategy with a targeted 1% sampling rate. 1085 out of 1911 TAZs in the entire BMA are featured in this survey, with the sparsely populated TAZs excluded. In each TAZ, 10 to 50 households are selected to take a face-to-face inter-

view. The final sample consists of 46,900 households (116,142 residents) in the BMA. The 2010 survey provides one-day travel diaries of all respondents, which gives travel time for each employee (distance unavailable except the TAZs of Origin-Destination). The 2010 survey documents the travel diaries in a typical day for each respondent. The travel mode is identified as the mode of the trip segment with the longest duration. We choose the mode of the highest mobility when there are two or more segments that have the same, longest duration. The 2010 survey presents also household information including household structure, income, and residential location at the TAZ level, as well as personal information including gender, age, occupation, industry of employment, etc. Like SCD, only records surveyed in weekdays are pertained in the subsequent analysis.

4. Methods

Our analytical approach can be outlined as follows: Firstly, we set working definitions for different types of extreme travel behaviors and consequently identify extreme travelers from the SCD dataset. Secondly, we characterize the spatiotemporal trajectories of extreme travelers. Lastly, we supplement the SCD data with the household survey to profile the socioeconomic conditions of identified extreme travelers.

4.1. Defining and identifying extreme travelers from the SCD

We define four types of extreme travelers in Beijing according to their transit behaviors in weekdays (Table 1). These working definitions draw upon the 2010 survey, existing literature, as well as researchers' own experiences of living in Beijing. For instance, the regular working hour starts on 8:30 or 9:00 a.m. in Beijing, and therefore boarding public transit before 6:00 a.m. would be considered as unusually early.

4.2. Extracting jobs-housing locations and commuting trips of extreme travelers from the SCD

To analyze the mobility pattern of identified extreme travelers, we construct commuting journeys based on commuters' job and home locations, which are in turn inferred using the following procedure:

- The card type is not a student card;
- $D_j \geq 6$ h, where D_j is the duration that a cardholder stays at place j , which is associated with all bus stops within 500 m of one another. Note that the benchmark of 6 h are set based on the analysis on full-time job in the 2010 survey;
- $j > 1$, which means that j is not the first place in a weekday that the server records;
- The place where a cardholder visited most frequently in five weekdays will be defined as the final workplace of the cardholder in this study.

Similarly, we deduced from the data queries that a place would be a cardholder's housing place if the data meet these conditions:

- The card type is not a student card;
- The place where a cardholder's first boarding bus stop/subway station trip in a day most frequently in five weekdays will be defined as the final housing place of the cardholder in this study.

Table 1
Working definitions of extreme travelers together with Abs (the SCD).

| Type | Definition |
|----------------------------|--|
| Early birds (EBs) | First trip < 6AM, more than two days in a week (60% of weekdays) |
| Night owls (NOs) | last trip (boarding time) > 10PM, more than two days in a week (60% weekdays) |
| Tireless itinerants (TIs) | \geq one and a half hours for one-way commuting (from the home location to job location), more than two days in a week |
| Recurring itinerants (RIs) | \geq 30 trips in weekdays of a week (\geq 6 trips per day) |
| Average Beijingers (ABs) | The "average" cardholders in the SCD |

Second, commuting trips were then identified based on the identified job and housing places of a cardholder.

More details about identifying housing and job locations and commuting trips are available in Long and Thill (2015).

4.3. Defining and identifying extreme travelers from the survey

Big data do not inform us about individual extreme travelers and reasons behind their traveling behaviors. We probed into these issues by using the 2010 survey samples mentioned above. The counterpart of each identified extreme travelers from the SCD was extracted from the survey with the same rules as we mined SCD for identifying extreme travelers. The temporal range of the SCD and the 2010 survey was different, in that we slightly adapted the definition of extreme travelers for accommodating the 2010 survey (see Table 2). Using the 2010 survey and the definitions, we analyze the first bus/metro trip of EBs, the last bus/metro trip of NOs, the bus/metro commuting trip of TIs, all bus/metro trips of RIs, and all trips of surveyed residents (Average Beijingers, ABs). Each type of extreme travelers can then be identified and their socioeconomic information can be understood (see Table 3).

We admit the gap between definitions of extreme travelers by using the SCD and the 2010 survey. The two datasets we analyzed in this paper were also concerned with two different samples. We believed that the following two aspects would mitigate the potential bias in the circumstance. First, the 2010 survey contains the most regular daily travel diaries that were reported in Beijing. Second, we integrated the findings from both datasets for each type of extreme travelers at the group level, not at the individual level. This, nevertheless, could be alleviated by the availability of week-long travel surveys in the future.

5. Results

5.1. Extreme travelers from the SCD

Based on our working definitions of extreme travelers, 188.9 thousand or 2.0% of all active cardholders in weekdays are associated with one or more extreme travel behaviors (the percentage is consistent with the findings of Rapino and Fields (2013) in the U.S.). Totally 14.2 thousand extreme travelers use student smart cards. We also note that 4890 cardholders fit two types of extreme travel behav-

Table 2
Working definitions of extreme travelers together with ABs (the 2010 survey).

| Type | Definition |
|----------------------------|--|
| Early birds (EBs) | First trip < 6AM |
| Night owls (NOs) | Last trip (boarding time) > 10PM |
| Tireless itinerants (TIs) | \geq one and a half hours for one-way commuting (from the home location to job location) |
| Recurring itinerants (RIs) | \geq 6 trips in the surveyed day |
| Average Beijingers (ABs) | The "average" travelers in the 2010 survey |

Table 3
The inventory of four types of identified extreme travelers from the SCD.

| Type | # cardholders ($\times 1000$) |
|------|---------------------------------|
| EBs | 42.8 |
| NOs | 70.6 |
| TIs | 6.7 |
| RIs | 73.7 |

iors, and 40 are three types. We find that there are more RIs and NOs than the other two types of extreme travelers, and TIs are the fewest in 2010.

The mobility patterns of four types of extreme travelers are summarized in Table 4 (to reword according to the new maps in the first two columns of Table 4). For housing, residents in the three well-known suburb residential areas (Tongzhou, Huilongguan and Tiantongyuan) are associated with more percentage to be TIs. In addition to the three areas, housing ratios of EBs, NOs and RIs exhibit similar polycentric pattern within the sixth ring road of Beijing. For jobs, most of extreme travelers' job locations are within the fifth ring road and are in the northern part of the city, which is more developed compared with the southern part. Employees in the Yizhuang Industrial Park prone to be RIs. According to the survey, most of them reside in the central city of Beijing. A notable number of TIs work at the Shangdi Information Technology cluster and around Tiantongyuan. For commuting trips, the popular destinations of EBs' commuting are in Xizhimen area. Many TIs commute a long distance from Tongzhou to the northern part of the central city. Most TIs commute from outside the central city to the central city. Many RIs commute a long distance, although only 4.0% of all identified RIs have an identified commuting trip. For typical trips of extreme travelers, there are only commuting trips for EBs and NOs, and not all EBs and NOs commute a long distance. In addition, RIs visit several places and travel a long distance in a day. In addition, we analyze all identified extreme travelers' typical trips, EBs, NOs and TIs have very few non-commuting trips (27.4%, 25.3% and 36.8% respectively), implying that they are busy with working activities.

5.2. Extreme travelers from the 2010 survey

Overall 1569 or 7.2% extreme travelers are identified from all 21,771 travelers with at least one bus/metro trip in weekdays as documented in the 2010 survey. There are 676 EBs, 236 NOs, 627 TIs, and 100 RIs (70 travelers fit two types of extreme travel behaviors). As shown in Table 5, we found that 60.2%, 11.8% and 10.9% EBs' first trips are to workplace, school and recreation place, respectively. These are significantly different from those of ABs (21.9%, 3.9% and 6.4%). Most of NOs' last trips are to home (96.2%), and 2.1% are to workplace at night. Almost one third of RIs trips are for dining (31.2%), and a considerable amount of trips are for pick-up and drop-off as well as business errands (all greatly than ABs).

We further analyzed socioeconomic characteristics of different extreme travelers and their households. Table 6 summarizes our main findings.

- (1) Classified households with 100 k CNY as high-income ones, we found that all households consisting of extreme travelers have a lower ratio as compared to AB households. This is consistent with our previous understandings of underprivileged residents in Beijing with SCD, in which we found the more one travel by bus, (s)he is with a greater probability to be economically underprivileged (Long, Liu, Zhou, & Gu, 2014).
- (2) There are more residents who rent housing in Beijing in the groups of NOs and TIs, comparing with ABs. Those renting housing in Chinese cities tend to be new comers or those who cannot afford expensive housing prices. This finding further proves that NOs and TIs are prone to be economically underprivileged.
- (3) All four types of extreme travelers on average owned fewer cars than ABs. In Beijing, most residents who could afford private cars would commute by driving, instead of by taking oftentimes

over-crowded public transit.¹ Economic-wise, car ownership is both a goods reflecting social status as well as a life necessity in the increasingly car-dominant living in Beijing. This finding reveals that all types of extreme transit riders are not linked to high-income residents.

- (4) TIs and RIs have more percentage of higher education in contrast to ABs. This finding is not beyond our expectation according to our local knowledge for Beijing that economically underprivileged residents are directly associated with lower education levels. In some cases, it is the good education that provides TIs opportunities to reside and work in large cities in China.
- (5) TIs had the lowest percentage of local Beijing residences (Hukou, a household registration system in China) among all extreme travelers. Hukou is a guarantee for some public services like education and welfare in China. It is a precondition for applying to a decent job in some cases. In the increasingly difficulty to obtain a Beijing Hukou, less Hukou percentage indicates social underprivileged situations of TIs as well.²
- (6) Fewer public-sector employees, who include civil servants and employees of public institutions are NOs and RIs. In Beijing, public-sector employees tend to have more stable jobs and higher income. More third-sector workers are NOs and TIs, and more employees in private companies are TIs. In addition, very few NOs are teachers and medical staffs. There are no soldiers and policemen in NOs, TIs and RIs.
- (7) Regarding social status of all extreme travelers, 60.9% EBs are full-time workers, 20.9% retirees, followed by 12.7% fulltime students. Most of NOs and TIs are fulltime workers. There are also a significant number of retirees in NOs (5.9%). Surprisingly, of all RIs, 38.0% are retirees while 42.0% are fulltime workers. There are also 12.0% of all RIs are jobless.

We are able to draw pictures for all four types with the analysis using the 2010 survey. For instance, most EBs are working full time in the private tertiary sector, fulfilling jobs that are less well-paid and require less educational attainment. Most of TIs are those who are also busy with working, with a higher percentage for renting house, well educated, with a lower percentage for holding a Beijing Hukou, and working for private sectors. In sum, the aforementioned analysis on various dimensions of socioeconomic characteristics of extreme transit riders reveal that they (especially the first three types) tend to be the economically underprivileged.

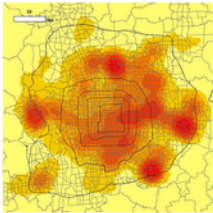
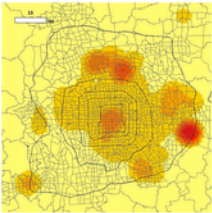

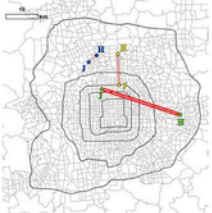
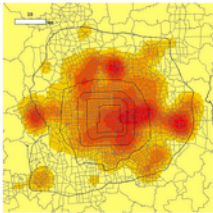
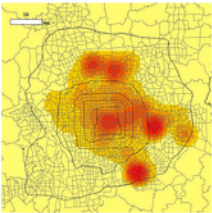

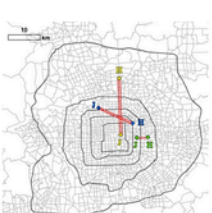
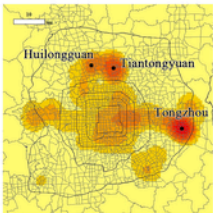
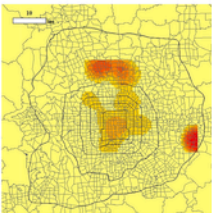
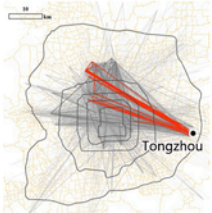
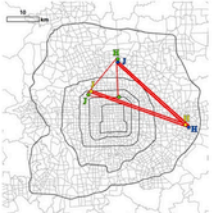
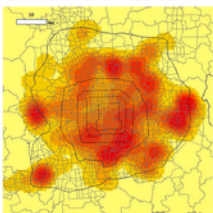
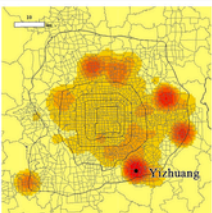

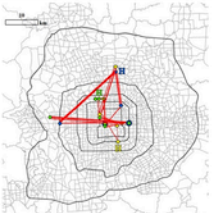


6. Conclusions and discussion

In this paper, we first extended the existing definition on extreme commuters and proposed four types of extreme transit riders, including Early Birds, Night Owls, Tireless Itinerants and Recurring Itinerants. We then identified and profiled each type of extreme travelers using both big and traditional data. Smart card data are used to visualize the overall spatiotemporal patterns of 188.9 thousand extreme travelers (2% of all active cardholders): where they reside and work, and what their mobility patterns are. In addition, we located 1568 extreme travelers in the traditional household survey data as the counterparts identified from the smart card data, and probed the travel purpose and socioeconomic attributes of extreme travelers and their households. To the best of our knowledge, this is one of the first at-

¹ As of 2010, the policy on car plate restriction was not implemented in Beijing.

² It should be mentioned that there are around 40% residents in Beijing without Hukou in 2014. The 2010 survey is with bias in terms of this attribute. While we admit this bias, we still see the possibility on benchmarking four types of extreme transit riders using the survey.

Table 4
 Mobility patterns of four types of extreme travelers from the SCD.

| Extreme travelers | Kernel density of housing ratio | Kernel density of job ratio | Commuting trips | Typical trips |
|-------------------|---|--|--|--|
| EBs |  (10.3 k) |  (9.4 k) |  (4.9 k) |  |
| NOs |  (31.6 k) |  (25.0 k) |  (17.5 k) |  |
| TIs |  (6.7 k) |  (6.7 k) |  (6.7 k) |  |
| RIIs |  (25.4 k) |  (7.8 k) |  (2.7 k) |  |
| Legend |  Low High | |  Low High | <ul style="list-style-type: none"> ● Person A ● Person B ● Person C — Ring roads — Routes |

Note that not all cardholders have a housing or job place, and a commuting trip. Housing ratio of each TAZ is the number of cardholders with a housing place in the TAZ divided by the total resident count in the TAZ. Job ratio of each TAZ is the number of cardholders with a job place in the TAZ divided by the total job position count in the TAZ. Commuting maps are prepared by using the head/tail breaks approach proposed by Jiang (2013) for data with a heavy-tailed distribution. In the maps of typical trips, “H” is for a housing place, and “J” is for a job place. Numbers in brackets are the total count of extreme travelers with corresponding information. The background layer is the TAZs in 2010. Ring roads in each map correspond to those in Fig. 1d.

tempts at understanding extreme travelers in relationship to the average Beijingers with smart card data and surveys for Chinese cities, from dimensions of the housing & job patterns, human mobility, travel motivation and demographic characteristics.

The overall contribution of this paper lies in two aspects. First, we develop a framework to generate useful aggregated information about

four types of extreme public transit riders from increasingly available big data such as smart card data. Our empirical analysis also benchmarks results from both smart card data and the more conventional travel survey. In this framework, on one hand, we identify fine-scale (bus stop and metro station level) and overall housing and job places and human mobility for each type of extreme travelers using the bus/

Table 5

Travel purposes of each identified extreme traveler from the 2010 survey.

| Type | EBs | NOs | TIs | RIIs | ABs |
|---------------------------------------|-------|-------|--------|-------|-------|
| Sleep/Rest | 0.4% | 0.4% | 0.0% | 0.2% | 0.1% |
| Shopping | 1.0% | 0.4% | 0.0% | 6.9% | 1.4% |
| Pick-up or Drop-off Others | 1.0% | 0.0% | 0.0% | 11.7% | 4.6% |
| Accompany Others | 0.0% | 0.0% | 0.0% | 0.4% | 0.2% |
| Taking Delivery of Goods | 0.6% | 0.0% | 0.0% | 1.3% | 0.4% |
| Go Home | 1.3% | 96.2% | 0.0% | 0.7% | 44.4% |
| Have Meals | 1.3% | 0.0% | 0.0% | 31.2% | 2.4% |
| Work | 60.2% | 2.1% | 100.0% | 2.4% | 21.9% |
| Official Travel | 0.9% | 0.0% | 0.0% | 10.0% | 0.9% |
| Go to Class/Study | 11.8% | 0.0% | 0.0% | 5.5% | 3.9% |
| Personal Affairs | 6.8% | 0.0% | 0.0% | 0.5% | 2.6% |
| Homework/Take Care of the Family | 0.6% | 0.0% | 0.0% | 10.8% | 0.1% |
| Recreation, Entertainment and Fitness | 10.9% | 0.4% | 0.0% | 0.2% | 6.4% |
| Visit Relatives and Friends | 1.8% | 0.0% | 0.0% | 3.1% | 10.1% |
| Others | 1.2% | 0.4% | 0.0% | 15.1% | 0.6% |

Note: The trip purpose of EBs, NOs, TIs, RIIs, and ABs is for the first trip, last trip, commuting trip, all trips, and all trips in a day, respectively.

Table 6

Selected socioeconomic characteristics of extreme travelers from the 2010 survey.

| Extreme travelers | EBs (676) | NOs (236) | TIs (627) | RIIs (100) | ABs (116,142) |
|---|--------------|--------------|--------------|---------------|------------------|
| % annual household income \geq 100 k CNY | 4.9 | 4.2 | 6.7 | 5.0 | 7.4 |
| % renting house | 11.0 | 17.8 | 20.4 | 16.0 | 16.1 |
| # average household car ownership | 0.22 | 0.21 | 0.25 | 0.22 | 0.31 |
| % higher education (undergraduate and graduate) | 14.2 | 18.2 | 33.5 | 25.0 | 21.1 |
| % Beijing Hukou | 87.0 | 82.2 | 74.8 | 83.0 | 82.4 |
| % public-sector employees | 13.5 | 7.6 | 15.8 | 7.0 | 10.4 |
| % fulltime workers | 60.9 | 84.7 | 94.4 | 42.0 | 45.9 |
| % fulltime students | 12.7 | 2.1 | 1.3 | 1.0 | 7.3 |
| % retirees | 20.9 | 5.9 | 0.8 | 38.0 | 29.1 |

Note that numbers in brackets are the total count of extreme travelers. The *t*-test between each type of extreme travelers and ABs for each indicator reveals a significant difference ($p < 0.05$).

metro smart card data, which small-scale surveys conducted at the TAZ level are not good at. On the other hand, we derive their travel purposes and socio-demographic background using the 2010 survey, which is not possible by the anonymous smart card data. The framework demonstrates how conventional and emerging data sources can be combined and complemented with each other to understand extreme travelers. This is important as most of conventional travel surveys are not conducted for studying extreme travelers. Second, we characterize the overall spatial trajectories as well as socioeconomic characteristics of extreme travelers. We found, for instance, the spatial distribution of RI, NO and EBs' homes and residences are similar and RIIs and EBs both tend to have their workplaces inside the fifth ring road. We also found how lots of TIs cluster near the east border of the city while work in inner city.

In addition to the above, our analysis could contribute to planning practices. First, our findings can inform the public transportation

management sector for its providing better public transit services to extreme transit riders. For instance, it would alleviate their extreme condition via re-scheduling morning/night buses to shorten EBs and NOs' waiting time and providing express service for commuting for TIs to avoid possible transfers. Second, social welfare is expected to be better delivered. For example, all bus/metro riding can enjoy a 50% discount in 2015 in Beijing. Considering the willingness-to-pay of all public transit riders differ a lot, we suggest public transit fare subsidization can go to these identified extreme transit riders who are in greater need of fare discount. Moreover, the identification results can be referred in the process of affordable housing application. That is, whether an applicant qualifies the social welfare standard can be double-checked by his/her public transit records, in addition to stated or reported socioeconomic indicators. Third, these findings can help local decision-makers envision and design better urban development policies and planning reactions from aspects of space and infrastructure to eliminate time poverty and social exclusion for these transit riders.

Future work could be expanded in at least three directions. First, only one week SCD in 2010 are employed in this paper. We would like to use the SCD in multiple years to gain more knowledge on the longer-term dynamics of extreme travelers. In this regards, Batty (2013a, 2013b), rightly pointed out "[T]he power of big data is that if collected for long enough then the longer term will emerge from the short term. At the moment these data are about what happens in the short term, but over ten years or longer we will have a unique focus on the longer term – in fact we will have a snapshot of urban dynamics which is unprecedented." Second, the big data and small data we used to supplement each other could represent different aspects of the same universe (sample mismatch). For instance, senior riders are not included in SCD and most respondents in the household survey are local residents with a Beijing Hukou. We therefore may have to design separate surveys to better understand extreme travelers. Otherwise, our results can be biased. In the long run, if we want to avoid related biases, we need better strategies to collect and use information based on SCD, as recommended by Pelletier et al. (2011). Third, our analysis is restricted to extreme travel behaviors of public transit riders and do not cover travelers using other modes of transportation (e.g., car). Comparing extreme travel behaviors of different modes could shed more light on underlying mechanisms regarding what contributes to the extreme and how the extreme could have burden or benefit the travelers. In addition, the specific thresholds used in defining various types of extreme travelers deserve more attention in the future research.

Uncited reference

Ma and Banister, 2006.

Acknowledgments

The authors would like to acknowledge the financial support of the National Natural Science Foundation of China (No.51408039) and the 12th Five-year National Science Supported Planning Project of China (2012BAJ05B04). We thank Ms Lingyan Wu for her data processing. Any errors and inadequacies of the paper remain solely the responsibility of the authors.

References

Barr, S., Fraszczyk, A., Mulley, C., 2010. Excess travelling—what does it mean? New definition and a case study of excess commuters in Tyne and Wear, UK. *European Transportation Research Review* 2, 69–83.

- Batty, M., 2012. Smart cities, big data. *Environment and Planning B: Planning & Design* 39 (2), 191–193.
- Batty, M., 2013a. Big data, smart cities and city planning. *Dialogues in Human Geography* 3 (3), 274–279.
- Batty, M., 2013b. Urban modelling: A progress report. In: *Symposium on applied urban modelling AUM 2013*. University of Cambridge.
- Beijing Transportation Research Center, 2011. Annual report of Beijing's transportation development (In Chinese). available at: <http://www.bjtrc.org.cn/JGJS.aspx?id=5.2&Menu=GZCG>. Accessed 01.09.15.
- Blythe, P., 2004. Improving public transport ticketing through smart cards. In: *Proceedings of the Institute of Civil Engineers, Municipal Engineers*, Vol. 157, pp. 47–54.
- California Planning Round Table, 2008. *Deconstructing jobs-housing balance*. California Planning Roundtable, Sacramento, CA.
- Cervero, R., 1991. Jobs-housing balance as public policy. *Urban Land* 10 (14), 11–14.
- Chu, K.K.A., Chapleau, R., 2010. Augmenting transit trip characterization and travel behavior comprehension multiday location-stamped smart card transactions. *Transportation Research Record: Journal of the Transportation Research Board* 2083, 29–40.
- Froud, J., Johal, S., Leaver, A., Williams, K., 2002. Not enough money: The resources and choices of the motoring poor. *Competition and Change* 6 (1), 95–111.
- Frumin, M., Zhao, J., 2012. Analyzing passenger incidence behavior in heterogeneous transit services using smartcard data and schedule-based assignment. *Transportation Research Record: Journal of the Transportation Research Board* 2274, 52–60.
- Gobillon, L., Selod, H., Zenou, Y., 2007. The Mechanism of spatial mismatch. *Urban Studies* 44 (12), 2401–2427.
- Gregor, K., 2013. The rise of the 'extreme commuter'. *BBC News Magazine* 26 December.
- Janosikova, L., Slavik, J., Kohani, M., 2014. Estimation of a route choice model for urban public transport using smart card data. *Transportation Planning and Technology* 37 (7), 638–648.
- Jiang, B., 2013. Head/tail breaks: A new classification scheme for data with a heavy-tailed distribution. *The Professional Geographer* 65 (3), 482–494.
- Jones, C., 2012. Sour economy gives rise to extreme commuters. *USA Today* 16 August.
- Kusakabe, T., Asakura, Y., 2014. Behavioural data mining of transit smart card data: A data fusion approach. *Transportation Research Part C-Emerging Technologies* 46, 179–191.
- Landsman, L., 2013. NYC's 'extreme commuting' costs getting out of control. *CNBC News*, 09 Nov. Available at: www.cnbc.com/id/101155916. Accessed 20.12.14.
- Lee, S.G., Hickman, M., 2014. Trip purpose inference using automated fare collection data. *Public Transport* 6 (1/2), 1–20.
- Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., et al., 2015. Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers* 105 (3), 512–530.
- Long, Y., Liu, X., Zhou, J., Gu, Y., 2014. Profiling underprivileged residents with mid-term public transit smartcard data of Beijing. *arXiv preprint arXiv:14095839*.
- Long, Y., Thill, J.C., 2015. Combining smart card data and household travel survey to analyze jobs–housing relationships in Beijing. *Computers, Environment and Urban Systems* (in press).
- Lyons, G., Urry, J., 2005. Travel time use in the information age. *Transportation Research A: Policy and Practice* 39 (2–3), 257–276.
- Ma, K., Banister, D., 2006. Excess commuting: A critical review. *Transport Reviews: A Transnational Transdisciplinary Journal* 26 (6), 749–767.
- Marion, B., Horner, M.W., 2007. Comparison of socioeconomic and demographic profiles of extreme commuters in several US metropolitan statistical areas. *Transportation Research Record: Journal of the Transportation Research Board* 2013 (1), 38–45.
- Ma, X.L., Wang, Y.H., Chen, F., Liu, J.F., 2012. Transit smart card data mining for passenger origin information extraction. *Journal of Zhejiang University Science C* 13 (10), 750–760.
- Meyer, M.D., Miller, E.J., 2001. *Urban transportation planning*, 2nd ed. McGraw-Hill, New York.
- Mintz, Z., January 21, 2014. Night-shift work causes 'Chrono-Chaos,' irregular sleep patterns may pose long-term health risks. *International Business Times*. Available at: <http://www.ibtimes.com/night-shift-work-causes-chrono-chaos-irregular-sleep-patterns-may-pose-long-term-health-risks>. Accessed 08.15.
- Mohamed, K., Côme, E., Baro, J., Oukhellou, L., 2014. Understanding passenger patterns in public transit through smart card and socioeconomic data. *UrbComp'*. Aug 24, 2014. New York City, NY, USA.
- Moss, M.L., Qing, C., 2012. The emergence of the super-commuter. *Rudin Center for Transportation*. New York University Wagner School of Public Service (February).
- Pelletier, M.-P., Morency, C., Trépanier, M., 2011. Smart card data use in public transit: A literature review. *Transportation Research C* 19, 557–568.
- Rapino, M.A., and Fields, A. K. 2013. *Mega commuters in the U.S.* Working paper March 2013. Presented at the Association of Public Policy Analysis and Management Fall 2013 Conference.
- Tao, S., Corcoran, J., Mateo-Babiano, I., Rohde, D., 2014a. Exploring Bus Rapid Transit passenger travel behaviour using big data. *Applied Geography* 53, 90–104.
- Tao, S., Rohde, D., Corcoran, J., 2014b. Examining the spatial–temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap. *Journal of Transport Geography* 41, 21–36.
- Taylor, B., Ong, P., 1995. Spatial mismatch or automobile Mismatch? An examination of race, residence and commuting in US metropolitan area. *Urban Studies* 32 (9), 1453–1473.
- U.S. Census Bureau, 2005. Extreme commute rankings. Available at: http://www.census.gov/newsroom/releases/pdf/2005-03-30_Commute_extremes.pdf. Accessed 20.12.14.
- Weitzel, J., 2003. *Jobs-housing balance*. American Planning Association, Chicago.
- Yang, Z., Cai, J., Ottens, H.F.L., 2013. *Cities*, Vol. 31, 491–506. Beijing: city profile.
- Zhou, J., Long, Y., 2014. Jobs-housing balance of bus commuters in Beijing. *Transportation Research Record: Journal of the Transportation Research Board* 2418, 1–10.