Physica A 483 (2017) 227–238

Contents lists available at ScienceDirect

Physica A

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

Human mobility in space from three modes of public transportation

Shixiong Jiang [a], Wei Guan [a,*], Wenyi Zhang [a], Xu Chen [b], Liu Yang [a]

[a] MOE Key Laboratory of Urban Transportation Complex System Theory and Technology, Beijing Jiaotong University, Beijing 100044, China
[b] State Key Lab of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China

HIGHLIGHTS

- We use trip displacements of three modes to explore human mobility.
- Trip displacements of taxi and bus follow the exponential distribution.
- Trip displacements of subway follow the gamma distribution.
- Fusion trip displacements follow the power-law with an exponential cutoff.

ARTICLE INFO

Article history:
Received 29 November 2016
Received in revised form 28 April 2017
Available online 5 May 2017

Keywords:
Human mobility
Exponential distribution
Power-law
Displacement
Travel time

ABSTRACT

The human mobility patterns have drew much attention from researchers for decades, considering about its importance for urban planning and traffic management. In this study, the taxi GPS trajectories, smart card transaction data of subway and bus from Beijing are utilized to model human mobility in space. The original datasets are cleaned and processed to attain the displacement of each trip according to the origin and destination locations. Then, the Akaike information criterion is adopted to screen out the best fitting distribution for each mode from candidate ones. The results indicate that displacements of taxi trips follow the exponential distribution. Besides, the exponential distribution also fits displacements of bus trips well. However, their exponents are significantly different. Displacements of subway trips show great specialties and can be well fitted by the gamma distribution. It is obvious that human mobility of each mode is different. To explore the overall human mobility, the three datasets are mixed up to form a fusion dataset according to the annual ridership proportions. Finally, the fusion displacements follow the power-law distribution with an exponential cutoff. It is innovative to combine different transportation modes to model human mobility in the city.

1. Introduction

The understanding of human mobility is important to many fields, such as transportation [1,2], urban planning [3] and epidemics [4]. The study of mobility was started by analyzing animal trajectories from the statistical physics perspective [5], in which the trajectories followed the power-law distribution, also known as Lévy flights [6]. Following the previous studies, Edwards et al. [7] utilized a high-resolution dataset of wandering albatross flights to find out that flight times were gamma

* Corresponding author.
E-mail address: weig@bjtu.edu.cn (W. Guan).

http://dx.doi.org/10.1016/j.physa.2017.04.182
0378-4371/

distributed with an exponential decay. These studies help people to learn about the mobility of nature. Then, researches extend to understand, model and predict human mobility.

Brockmann et al. [8] found that the travel distances of bank notes followed a power-law distribution and can be modeled as a continuous-time random walk. By analyzing the mobile phone data, González et al. [9] indicated that human trajectories followed a truncated power-law with a high degree of temporal and spatial regularity. And the space–time structure can be captured well by remarkably few temporal or spatial predictable periodic patterns [10]. However, Kang et al. [11] illustrated that the distribution of human's intra-urban travel followed the exponential law with various exponents from a city to another city. And Wang et al. [12] observed that the trip duration distributions can be approximated by lognormal distributions. Furthermore, some models are proposed to study human mobility based on mobile phone data, such as the communication model [13], the external trip model [14] and the kernel density estimation model [15].

The Global Positioning System (GPS) data are widely used as a common tool to analyze human mobility. Human walks performed in outdoor settings resembled a truncated form of Lévy walks, which showed evidence of scale-freedom [16]. Yan et al. [17] discovered the absence of scaling properties in the displacement distribution at the individual level. And the aggregated displacement distribution followed a power-law with an exponential cutoff. Zhao et al. [18] showed that human mobility can be modeled as a mixture of different modes, and these movement patterns of each mode can be fitted by a lognormal distribution. Sadilek and Krumm [19] extracted significant and robust patterns in location data to predict human mobility with high accuracy. Considering the impact of activity changes, the human movement can be predicted from a spatiotemporal perspective [20].

The displacements and elapsed time of the trips by taxi in the urban areas tended to follow an exponential distribution [21]. The exponential law of intra-urban mobility was explained as a result of the exponential decrease in average population density in urban areas [22]. Furthermore, Liang et al. [23] predicted human flows with a convincing fidelity with a parameterless rank-based model. Several similar studies have been conducted in recent years [11,24,25]. Average speed and trajectory clustering were also exploited to explore human mobility [24,26].

The large-scale Oyster card database from London subway helped to reveal that the intra-urban movement was strongly heterogeneous in terms of volume [27]. Xu et al. [28] noticed that power-law can fit distributions of node throughflows, station total inflows and outflows in the subway system. For the bus trips, the variabilities of mobility patterns at individual and aggregated levels were measured [29]. Furthermore, Wang et al. [30] indicated that the travel distance distribution of bus passengers was peaked, which can be fitted by a negative binomial distribution. There are also other data sources used to explore human mobility, including the mobile ticketing data [31], the Foursquare users dataset [32], the origin–destination (OD) data [33], the Place of Interests [34], check-in data [35] and etc.

The integration of different modes of transportation is innovative for human mobility study [36]. This paper aims to model human mobility from trip displacements of taxi, subway and bus in Beijing. For each mode, several candidate models are compared to select the most fitting one. In addition, these three datasets are merged together to explore human mobility in the city. We want to supplement the previous researches which rely on data from a single mode of transportation to explore human mobility. This study can be applied in traffic modeling and transportation planning.

The rest of this paper is organized as follows. Section 2 introduces the datasets of taxi, subway and bus, which are employed in the paper. Section 3 describes the method of model selection. The results and analysis are presented in Section 4. Section 5 gives the detailed discussion. And conclusion is provided in Section 6.

2. Data description

2.1. Taxi GPS dataset

The adopted taxi GPS data were generated by more than 46,000 taxis in Beijing, China, from January 12th, 2015 to January 18th, 2015, which contained 2,915,488 trips. By the end of 2014, there were about 67,000 taxis in Beijing [37]. The taxi GPS data, which was at about 1-min intervals, covered 68.66% of taxi during that time period. The GPS equipment captured several attributes of the taxi, consisting of vehicle ID, longitude, latitude, timestamp, instantaneous velocity, operational status and so on.

2.2. Subway transaction dataset

A transaction dataset of the Beijing subway system was collected from December 23th, 2014 to December 25th, 2014, which included 13,380,969 trips. This dataset covered all trips by subway during that time period. However, the data were not the original transaction records, but the processed OD matrices reflecting the flows between any pair of subway stations. The subway trips were stored in a matrix per half hour. Besides, the location information (longitude and latitude) of each subway station was collected from a map website, Amap [38].

2.3. Bus transaction dataset

The transaction data of bus were collected from the Automatic Fare Collection (AFC) system, from March 10th, 2015 to March 12th, 2015, containing 20,105,096 records. There were totally 877 bus routes in Beijing by the end of 2014 [37]. And

the dataset included 666 routes in the bus system, including 75.94% of bus routes during that time period. Each transaction is stored as a record, which represents the travel of a passenger in a bus vehicle. Each record includes card number, boarding station ID, alighting station ID, boarding line ID, alighting line ID, boarding time, alighting time, trade type, and etc. In addition, a dataset with the location (latitude and longitude) and station ID of each bus station is gathered. The specific boarding and alighting locations for each trip are confirmed by matching the station ID.

However, there are some limitations in the original records. A station is represented by a mark in the record, which is the integral kilometer length from the station to the first station in the route. Thus, some neighboring stations in a route might share the same mark. To handle this problem, stations with the same mark in a route are regarded as one station, and its location is the geometric center of these stations.

3. Methodology

3.1. Trip definition

A taxi trajectory is a sequence of time-stamped points recorded for a taxi, which contains longitude, latitude, instantaneous velocity, operational status and etc. When a taxi does not have any passenger, it is marked as vacant; or occupied otherwise. In this study, a taxi trip refers to a sub-trajectory with the occupied status. A trip is represented by $(p_1, p_2, \ldots, p_n)$, in which $p_i = (lon_i, lat_i, t_i)$. p_i denotes the ith record in the trip. lon_i, lat_i and t_i are the longitude, latitude and time of the record p_i. The pick-up and drop-off points are presented by $p_o = p_1$ and $p_d = p_n$, respectively.

A subway trip is from the boarding station ($p_o = (lon_o, lat_o)$) to the alighting station ($p_d = (lon_d, lat_d)$). The origin and destination of a subway trip refer to the boarding station and alighting station.

A bus trip is presented by a record in the dataset, which includes boarding station ($p_o = (lon_o, lat_o)$), alighting station ($p_d = (lon_d, lat_d)$), boarding time (t_o), alighting time (t_d), trade type and so on. A bus trip starts at the boarding station and ends at the alighting station. Thus, the boarding station and alighting station of a bus trip denote its origin and destination.

Due to the limitation of the datasets, the three datasets of different modes are collected in different time periods. Thus, transfers are not taken into consideration in the study. In addition, the bus and subway trips have to begin and end at the stations. And the real origin and destination might be within a buffer zone of the station.

3.2. Trip measure

The displacement can be measured from the origin and destination locations as presented in Eqs. (1) and (2):

$$c = arccos(\sin(lat_o)\sin(lat_d) + \cos(lat_o)\cos(lat_d)\cos(lon_d - lon_o)) \quad (1)$$

$$\Delta l = R \times c \times \pi/180 \quad (2)$$

where c is the lengths of the great circle arcs in degrees connecting the origin and destination locations on the surface of a sphere, $R = 6371$ km denotes the average radius of the earth, Δl denotes the displacement between the origin and destination locations, lat_o and lon_o represent the latitude and longitude of the origin (p_o), lat_d and lon_d are the latitude and longitude of the destination (p_d).

For taxi, the origin and destination locations are the pick-up and drop-off points, while for subway and bus, they are the boarding and alighting stations. The travel time (Δt) of taxi and bus trips can be calculated as $\Delta t = |t_d - t_o|$. In addition, Δl and Δt are measured in kilometers and minutes respectively.

3.3. Data cleaning

The taxi trips are extracted from the original GPS data. However, some trips need to be detected and removed from the dataset. Trips missing some information, such as longitude or latitude, are removed. To eliminate the abnormal data, we remove trips with a duration $\Delta t > 120$ min or $\Delta t < 1$ min, because few passengers travel by taxi for such a long or short time [21].

The subway trip dataset is quite simple and stored in OD matrices. In this study, trips boarding and alighting at the same station are removed, because it is abnormal for passengers to make such trips.

A record in the bus transaction dataset represents a trip. To get the valid records, each transaction record must satisfy some basic requirements. First, the record must be normal trade. Then, the boarding line ID and boarding vehicle ID must be the same with the alighting line ID and alighting vehicle ID, respectively.

3.4. Model selection

Model selection is to identify the most appropriate model which is supported by the actual travel data. In this paper, each dataset is fitted by some of these commonly used distributions, including power-law, exponential, lognormal, gamma and Weibull distributions, which are defined in Eqs. (3)–(7), respectively.

The power-law distribution with exponent α is expressed as

$$P(x) = Ax^{-\alpha}. \tag{3}$$

The exponential distribution is defined as

$$P(x) = Be^{-\lambda x} \tag{4}$$

where λ denotes the rate parameter.

The lognormal distribution is presented as

$$P(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right) \tag{5}$$

where μ and σ denote the mean and the standard deviation of the natural logarithm of the variable.

The gamma probability density function is

$$P(x) = \frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-x/\theta} \tag{6}$$

where $k > 0$ denotes the shape parameter, $\theta > 0$ is the scale parameter and $\Gamma(\cdot)$ presents the gamma function.

The Weibull probability density function is

$$P(x) = \frac{a}{b}\left(\frac{x}{b}\right)^{a-1} e^{-(x/b)^a} \tag{7}$$

where $a > 0$ is the shape parameter and $b > 0$ denotes the scale parameter.

In this paper, the Akaike information criterion (AIC) method, which considers both fits with the actual data and complexity of the models, and is able to compare several models at the same time [21], is employed to compare the candidate models. A and B are normalization constants. For the power-law distribution in bounded ranges, $[m, n]$, A is the normalization constant given by $A = (\alpha - 1)/(m^{1-\alpha} - n^{1-\alpha})$. Taking $[m, n]$ to be the bounded ranges for the exponential distribution, B is the normalization constant given by $B = \lambda/(e^{-m\lambda} - e^{-n\lambda})$. The model selection procedure is showed as follows:

First, the maximum likelihood estimates (MLE) is adopted to fit the parameters [7,39,40].

Then, the AIC score for each candidate model i is calculated by

$$\mathrm{AIC}_i = -2 \log L_i + 2\, K_i \tag{8}$$

where L_i is the likelihood for model i, which is depended on the values of the parameters estimated from the MLE method. And K_i denotes the number of parameters in model i.

Finally, the relative likelihoods of the models are represented by the Akaike weights W_i. It is defined as:

$$W_i = \frac{e^{-\Delta_i/2}}{\sum_{j=1}^{N} e^{-\Delta_j/2}} \tag{9}$$

where $\Delta_i = \mathrm{AIC}_i - \mathrm{AIC}_{\min}$, $\mathrm{AIC}_{\min} = \min\{\mathrm{AIC}_i\}$, N is the number of candidate models, i and j ($i, j \in \{1, 2, \ldots, N\}$) are indices of models. The model i with the largest Akaike weights is most likely to be selected.

4. Results and analysis

The displacement is utilized to explore human mobility in the study. The origin and destination reflect the human movement directly, which can be presented by the displacement. Contrary to the actual path traveled from the origin to the destination, the displacement, which is the length of a straight line connecting the origin and destination, is not influenced by the geographic environment and the urban road network. Furthermore, the displacement makes the comparison between different cities much easier.

4.1. Taxi trip displacement distribution

Taxi is the most flexible mode of the public transportation, which provides a 24 h door-to-door service. The pick-up and drop-off places of the taxi trips are more probable to be, or near the real origin and destination locations, comparing with the boarding and alighting stations of subway and bus. Hence, the trips of taxi are most suitable to study human mobility.

The probability of the displacement is measured from the taxi GPS dataset (see Fig. 1). The average displacement of taxi trips is 6.53 km. The probability increases from the beginning, and arrives the peak at 1.9 km. The reason of the rise in the first interval is that displacements of taxi trips are seldom very short. Because it requires time and fare to hail a taxi. In China, taxi fare is generally calculated as follows: the minimum fare applied to the taximeter when a passenger gets on (flag fall), plus extra charges for additional distance and waiting time, plus applicable surcharges [12]. Besides, the flag fall in

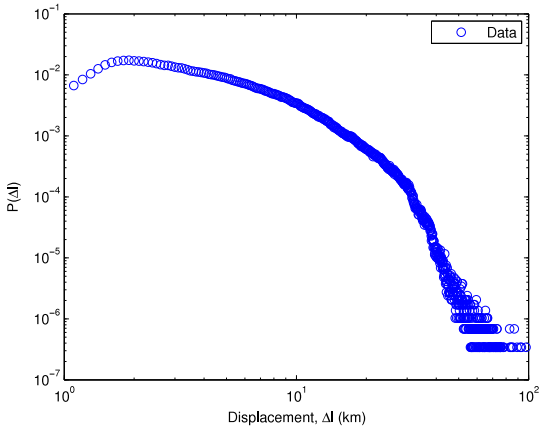

Fig. 1. Displacements distribution of taxi trips.

Table 1
Results of model selection for displacements of taxi trips.

Part	Model	Item	Value
First part	Power-law	MLE for α	1.3016
		(95% CI)[a]	(1.2997, 1.3036)
		R^2_{pow} [b]	0.9662
		W_{pow} [c]	0
	Exponential	MLE for λ	0.1917
		(95% CI)[a]	(0.1914, 0.1921)
		R^2_{exp} [b]	0.9985
		W_{exp} [c]	1
Latter part	Power-law	MLE for α	5.3238
		(95% CI)[a]	(5.2988, 5.3487)
		R^2_{pow} [b]	0.9510
		W_{pow} [c]	0
	Exponential	MLE for λ	0.1768
		(95% CI)[a]	(0.1758, 0.1778)
		R^2_{exp} [b]	0.9896
		W_{exp} [c]	1

[a] A 95% confidence interval.
[b] Coefficient of determination to measure the goodness of fit.
[c] Akaike weights representing relative likelihoods of models.

Beijing covers the first 3 km travel distance. It is found that the flag fall distance is approximated to the product of the peak displacement ($\Delta l = 1.9$ km) and the detour factor, which is about 1.5 [41]. According to the statistical results, about 96.04% trips have a displacement less than 20 km, which suggests that most trips occur in the urban areas. And 3.96% of the trips are more than 20 km. Owing to the high fare, passengers prefer to take subway or other public transportations rather than taxi for long trips.

Considering about the distribution of displacements, we partition the distribution into two parts according to the displacement of 20 km [12,21]. As shown in Fig. 1, $P(\Delta l)$ increases in the first interval. However, both candidate models are monotone decreasing functions. It is apparent that the first interval of taxi displacement is not consistent with the candidate distributions. After removing the first interval, the travel distance of taxi can be represented by an exponential distribution better [42].

Thus, displacements in the ranges, 2–20 km and 20–100 km, are fitted to select the best model respectively. Most movements of the first part happen in the urban areas, while for the latter part, trips tend to occur between urban and suburbs [21]. The power-law and exponential distributions are compared using the AIC method for the two parts. Table 1 illustrates the detailed results for the model selection. From the table, it is concluded that the displacements are well fitted by the exponential distribution because of $W_{exp} > W_{pow}$ in the both parts. We also notice that the exponential exponent for the latter part is less than that of the first part, which is consistent with the previous study [21], suggesting that the short trips are more sensitive. In addition, the both fitting results can be observed in Fig. 2.

4.2. Subway trip displacement distribution

Subway is a fundamental mode of public transportation in Beijing. For its punctuality and economy, a large number of passengers prefer to take subway, especially the commuters. The subway trips are indispensable parts to reveal human mobility in the city. Subway trips take place in the fixed subway network and passengers must enter and exit the system

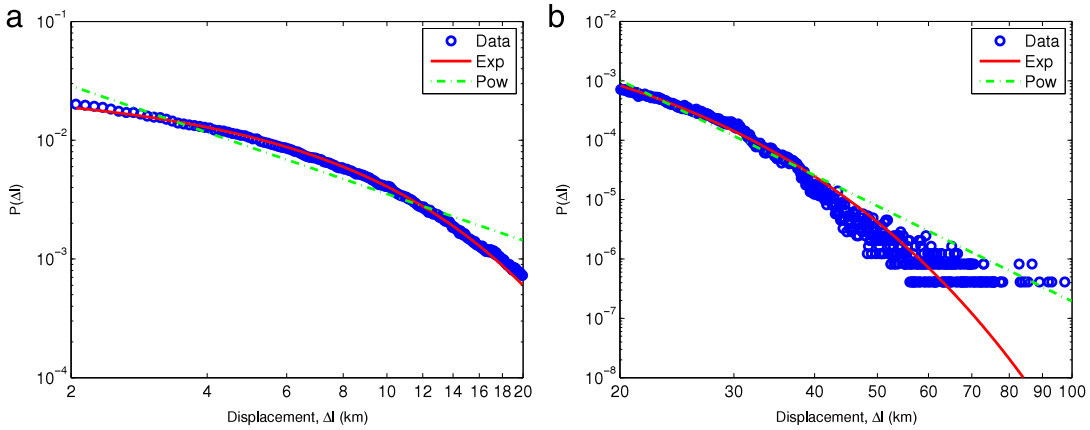

Fig. 2. Distributions of taxi trip displacements: (a) 2–20 km. The red solid line represents an exponential with exponent $\lambda = 0.1917$. The green dot-dashed line indicates a power-law with exponent $\alpha = 1.3016$. (b) 20–100 km. The red solid line represents an exponential with exponent $\lambda = 0.1768$. The green dot-dashed line indicates a power-law with exponent $\alpha = 5.3238$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

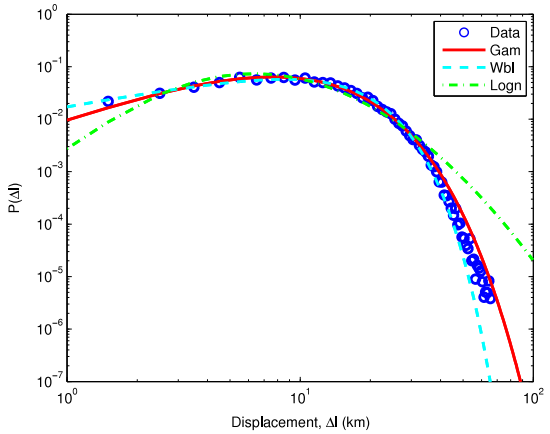

Fig. 3. Distributions of subway trip displacements. The red solid line indicates the gamma distribution ($k = 2.6450$, $\theta = 4.6318$), the blue dashed line represents the Weibull distribution ($a = 13.7899$, $b = 1.7501$), and the green dot-dashed line denotes the lognormal distribution ($\mu = 2.3048$, $\sigma = 0.6835$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

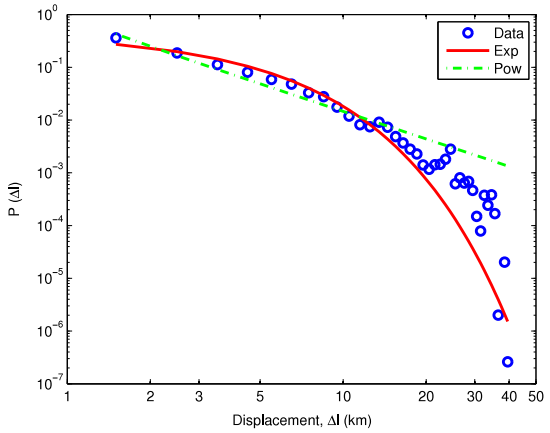

Fig. 4. Distributions of bus trip displacements. The red solid line represents an exponential with exponent $\lambda = 0.3183$. The green dot-dashed line indicates a power-law with exponent $\alpha = 1.7314$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Results of model selection for displacements of subway trips.

Model	Item	Value
Gamma	MLE for k	2.6450
	(95% CI)[a]	(2.6431, 2.6469)
	MLE for θ	4.6318
	(95% CI)[a]	(4.6282, 4.6355)
	R^2_{gam}[b]	0.9955
	W_{gam}[c]	1
Weibull	MLE for a	13.7899
	(95% CI)[a]	(13.7854, 13.7943)
	MLE for b	1.7501
	(95% CI)[a]	(1.7494, 1.7508)
	R^2_{wbl}[b]	0.9934
	W_{wbl}[c]	0
Lognormal	MLE for μ	2.3048
	(95% CI)[a]	(2.3045, 2.3052)
	MLE for σ	0.6835
	(95% CI)[a]	(0.6832, 0.6838)
	R^2_{logn}[b]	0.9622
	W_{logn}[c]	0

[a] A 95% confidence interval.
[b] Coefficient of determination to measure the goodness of fit.
[c] Akaike weights representing relative likelihoods of models.

at the subway stations. A subway trip refers to a travel from a passenger's boarding station to its alighting station, which reflects the movement of the passenger. The displacement between any pair of subway stations is calculated from their longitudes and latitudes. Combined with the OD flows, the displacement of each trip is attained.

The distribution of subway trip displacements computed from the dataset is showed in Fig. 3. The probability increases quickly from the beginning and reaches the peak when the displacement is about 9 km. The reason of the rise at first is different from that of the taxi trips. This is largely resulted from their different pricing structures and functions. At that time, the subway system adopted a flat fare of 2 RMB. However, taking subway requires more time to enter and exit the subway system, which is the time cost for passengers. On the whole, displacements of subway trips average at 15.15 km. About 14.37% of the subway trip displacements are more than 20 km, which is quite different from the displacements of taxi trips. This is in accordance with the actual experience that the taxi serves the short trips while the subway satisfies the long trips. Compared with the distribution of taxi trip displacements, the distribution of subway trip displacements also shows a long tail, but it is not as significant.

Given the distribution of subway trip displacements, it is doubtless that the power-law and exponential distributions cannot capture the distribution accurately. Thus, we select gamma, Weibull and lognormal distributions to fit the displacements. And the AIC method is used to compare the models. The detailed results of the model selection are presented in Table 2. It is indicated that the gamma distribution fits the subway trip displacements rather than the lognormal and Weibull distributions. Moreover, the fitting results are presented in Fig. 3.

4.3. Bus trip displacement distribution

As an indispensable transportation mode in the city, bus completed 4.772 billion trips in 2014, which took 51.3% of the public trips in Beijing [37]. Thus, it is necessary to take the trips of bus into account to explore human mobility in the city.

The distribution of bus trip displacements calculated from the bus transaction dataset is presented in Fig. 4. The probability is quite large at the beginning. Then, it decreases as the displacement increases. The flag fall of bus is 2 RMB in the bus system covering the first 10 km travel distance. And passengers can get a 50% discount if they pay by the smart card. In addition, it is convenient to take a bus thanks to the widespread bus stops around the city. Thus, passengers are likely to take buses for short trips. While for long trips, bus is uncompetitive due to its low speed. In the dataset, 95.82% of the trips are less than 10 km. Few trips have a displacement more than 40 km due to the limits of the route length. On the whole, the bus mainly serves for short trips.

The displacement is fitted with two distributions in a range, [1, 40] km, including the power-law and exponential distributions, to select the better one. According to the AIC method, the detailed results of the model selection are illustrated in Table 3. It is concluded that the exponential distribution approximates the bus trip displacements better than the power-law distribution. As plotted in Fig. 4, the exponential distribution captures the trend better.

4.4. Fusion trip displacement distribution

In 2014, the taxi, subway and bus completed 0.668, 3.387 and 4.772 billion trips respectively [37], taking 7.57%, 38.37% and 54.06% of the total trips in the public transportation. In the three datasets, there are 2,915,488, 13,380,969 and

Table 3
Results of model selection for displacements of bus trips.

Model	Item	Value
Power-law	MLE for α	1.7314
	(95% CI)[a]	(1.7308, 1.7321)
	R^2_{pow} [b]	0.9671
	W_{pow} [c]	0
Exponential	MLE for λ	0.3183
	(95% CI)[a]	(0.3182, 0.3185)
	R^2_{exp} [b]	0.9470
	W_{exp} [c]	1

[a] A 95% confidence interval.
[b] Coefficient of determination to measure the goodness of fit.
[c] Akaike weights representing relative likelihoods of models.

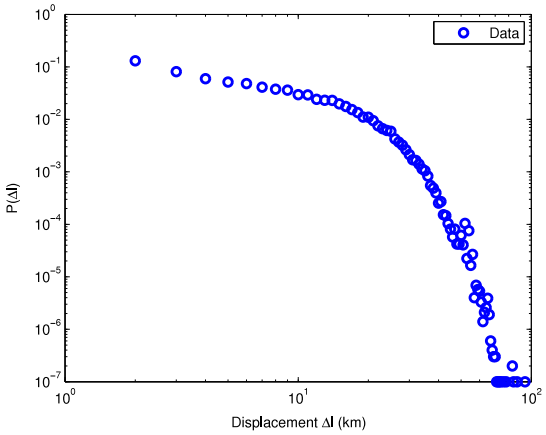

Fig. 5. Distribution of the fusion trip displacements.

20,105,096 trips for taxi, subway and bus respectively. Thus, the corresponding numbers of trips (0.757 billion taxi trips, 3.837 billion subway trips and 5.406 billion bus trips) are randomly selected from the three datasets to form a fusion dataset with 10 million trips to explore human mobility [13].

The displacement distribution of the fusion dataset is illustrated in Fig. 5. The probability is large in the first interval. Then, it decreases as the displacement increases. The fusion displacements distribution is similar with the displacements of bus trips in the frontal part and similar with the displacements of subway trips in the latter part.

There are more than 85% of trips with a displacement less than 16 km. After reaching 16 km, the probability is quite low compared with the frontal part. The displacements of the two parts show different characteristics. Thus, we partition the distribution into two parts according to the displacement of 16 km. The AIC method is utilized to compare the power-law and exponential distributions for both parts. Table 4 presents the detailed results of the model selection.

According to Table 4, it is indicated that the displacements can be well fitted by the power-law distribution in the frontal part. While the exponential distribution approximates the displacements rather than the power-law distribution in the latter part. The two fitting results are plotted in Fig. 6. It need to be noticed that the first part occupies 85.18% of the trips and the latter part takes up the other 14.82%. Referring to Eqs. (3) and (4), the fitting power-law distribution in [1, 16] km is shown in Eq. (10) and the fitting exponential distribution in (16, 100] km is presented in Eq. (11).

$$P(x) = 0.2226x^{-0.7788} \tag{10}$$

$$P(x) = 0.3192e^{-0.1618x}. \tag{11}$$

To validate the consistency and reliance of the fusion dataset, another four fusion datasets are formed and fitted respectively (see Table 5). It is indicated that the fitted distributions of these fusion datasets are similar. Since all variances are less than 10^{-5}, it is evident that the fusion is reliable and consistent.

5. Discussion

The study observes that the displacements of different modes have different distributions. It suggests that the differences should be given a sufficient consideration in urban planning and traffic forecasting. Moreover, the displacements of three modes are combined to explore human mobility in the public transportation, which follows the power-law distribution with an exponential cutoff.

Table 4
Results of model selection for the fusion trip displacements.

Part	Model	Item	Value
First part	Power-law	MLE for α	0.7788
		(95% CI)[a]	(0.7778, 0.7797)
		R^2_{pow} [b]	0.9978
		W_{pow} [c]	1
	Exponential	MLE for λ	0.1371
		(95% CI)[a]	(0.1369, 0.1373)
		R^2_{exp} [b]	0.9575
		W_{exp} [c]	0
Latter part	Power-law	MLE for α	4.2928
		(95% CI)[a]	(4.2865, 4.2992)
		R^2_{pow} [b]	0.9464
		W_{pow} [c]	0
	Exponential	MLE for λ	0.1618
		(95% CI)[a]	(0.1615, 0.1621)
		R^2_{exp} [b]	0.9916
		W_{exp} [c]	1

[a] A 95% confidence interval.
[b] Coefficient of determination to measure the goodness of fit.
[c] Akaike weights representing relative likelihoods of models.

Table 5
The fittings of the fusion trip displacements: α for power-law and λ for exponential distribution.

Test	1–16 km		16–100 km	
	MLE for α	MLE for λ	MLE for α	MLE for λ
1	0.7788	0.1371	4.2928	0.1618
2	0.7785	0.1371	4.2940	0.1618
3	0.7789	0.1371	4.2896	0.1616
4	0.7790	0.1372	4.2910	0.1617
5	0.7790	0.1371	4.2920	0.1617
Variance	4.30×10^{-8}	2.00×10^{-9}	2.83×10^{-6}	7.00×10^{-9}

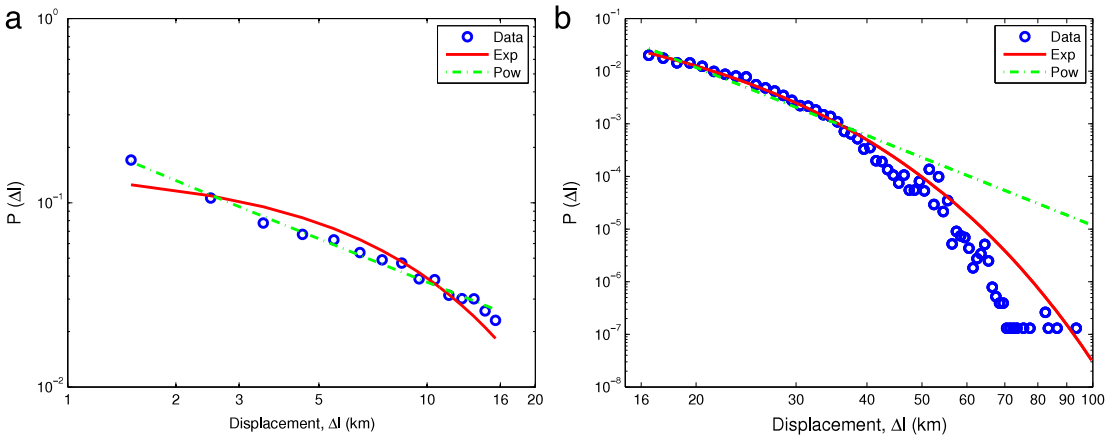

Fig. 6. Distributions of the fusion trip displacements. (a) 1–16 km. The red solid line represents an exponential with exponent $\lambda = 0.1371$. The green dot-dashed line indicates a power-law with exponent $\alpha = 0.7788$. (b) 16–100 km. The red solid line represents an exponential with exponent $\lambda = 0.1618$. The green dot-dashed line indicates a power-law with exponent $\alpha = 4.2928$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Human mobility patterns of each single mode have been studied in recent years. For the taxi trips, Jiang et al. [2] found that the distribution of trail length follows a double power-law, implying intracity and intercity movements. Many researches indicate that the distribution of taxi trip displacements follows the exponential distribution [12,21,22]. On the whole, our findings are consistent with the mentioned empirical studies and the taxi trip displacements follow an exponential distribution. For the bus trips, it is found that the travel distance distribution of bus passengers can be fitted by a negative binomial distribution [30]. Goh et al. [43] analyzed power-law correlations, characterized by algebraic behavior of the strength correlation function of bus stops. In our study, the displacements of bus trips follow the exponential distribution. For the subway trips, Roth et al. [27] revealed that the distance distribution of rides can be fitted by a negative binomial law.

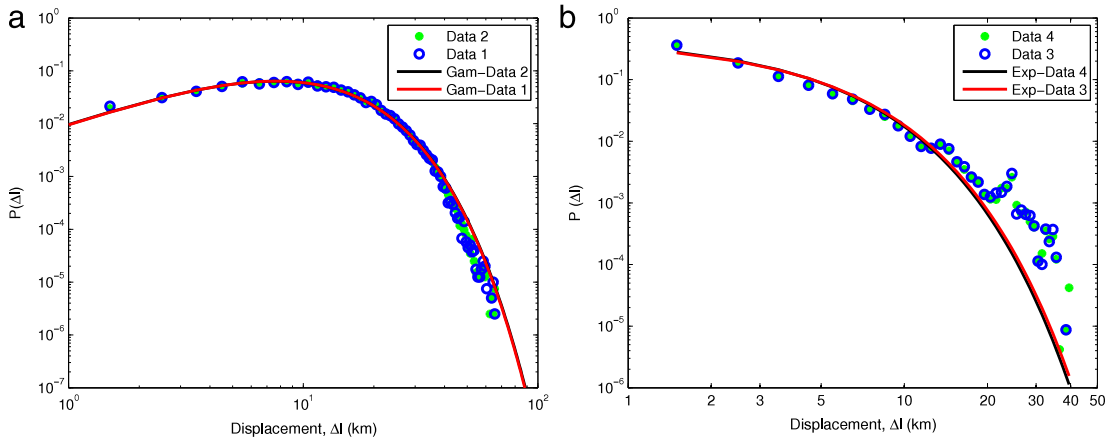

Fig. 7. Distributions of the trip displacements calculated from stations and simulated OD: (a) subway, (b) bus.

While the displacements of subway trips follow the gamma distribution in this research. Besides, Hasan et al. [44] indicated that people's most visited places were scattered over the city with some subway stations owing higher concentration. It is concluded that the geography constraints are important and the distance distribution between stations could be a major factor in the ride distribution for the bus and subway system [27].

In studies about overall human mobility, González et al. [9] speculated that the distribution of displacements over all users followed a truncated power-law from mobile phone data. Mobile phone traces of users are also utilized to investigate individual mobility patterns [13,45]. Zhao et al. [18] observed that power-law fits for overall flight from human GPS traces, which was consistent with our research. Besides, this study explores that the overall human mobility follows the power-law in the frontal part and the exponential distribution in the latter part, which is consistent with some previous researches.

The different characteristics of human mobility in each mode might attribute to the fare and service. Generally, the fare of taxi is much higher than subway and bus. From the perspective of speed, bus runs lower than taxi and subway. Besides, taxi is most flexible and subway is most punctual in the public transportation. For trips with the known OD, the displacements are fixed and passengers will select the most beneficial mode according to their budgets and demands [46]. Thus, each mode has its dominant attraction for specific trips. In all, bus tends to service the short trips, subway provides service for longer trips and taxi works as an important supplement.

There are also some limitations in the study. Transfers are an indispensable issue in the public transportation system. However, due to the limitation of the collected datasets, we do not take transfers into consideration. As a result, trips containing any transfer might be divided into two or more trips. The number of long trips will decrease, while the number of short trips will increase, compared with the real trips in consideration of transfers.

The trips of subway and bus are regarded as begin at the boarding stations and end at the alighting stations. However, the passenger usually need to travel from the real origin to the boarding station and from the alighting station to the real destination. Thus, the real displacement might be a little less or more than the displacement on station level. Generally, the widely used walking distances are 400 and 800 m for estimating the distance that people will walk to a transit stop or station [47–49]. To evaluate its impact, we randomly select 400,000 subway trips (Data 1) from the subway dataset and 400,000 bus trips (Data 3) from the bus dataset. For a subway station, the area within 800 m from the station is defined as its buffer zone. For the bus stop, it is 400 m. For a trip, the simulative real origin and destination are regarded as the randomly selected points within the buffer zones of the responding boarding and alighting stations. Then, the simulative displacements are calculated and marked as Data 2 for subway trips and Data 4 for bus trips. As shown in Fig. 7(a), the subway displacements follow the gamma distribution in both simulative and actual datasets. And the bus displacements in both datasets are fitted with the exponential distribution in Fig. 7(b). It is indicated that regarding boarding and alighting stations as origin and destination locations does not have significant effects on the fitted distribution.

As the three datasets were collected in winter, it required to discuss the seasonal differences. The work and school trips are stable in different seasons, while the leisure and shopping trips are likely to decrease with cold temperature [50–53]. In general, the travel demand in winter is less than that of warmer seasons. While for the public transport, lots of travel demands are diverted from bicycle and walking to the public transport [50,54]. Thus, the public transport holds a higher percentage of travel demand in winter. Furthermore, Liu et al. [55] found that as the temperature rises, the travel distance increases. Therefore, compared with trips in winter, the annual trips might have fewer trips in the short displacement and more trips in the long displacement.

6. Conclusions and future work

In this study, we model the distributions of trip displacements for three modes of public transportation respectively, which is based on the actual trip datasets collected in Beijing. The main contributions of this paper are as follows: (i) the

displacements of taxi trips tend to follow the exponential distribution; (ii) while the displacements of subway trips can be well fitted by the gamma distribution; (iii) the exponential distribution approximates the displacements of bus trips best; (iv) the power-law distribution fits the first part of the fusion trip displacements, but the latter part follows the exponential distribution; (v) this paper innovatively merges the displacement data from different modes together to explore human mobility.

It is expected that the empirical conclusions obtained from this study can promote understanding about human mobility and extend the research fields. Although we have taken different modes of transport to study human mobility, more modes, such as private vehicles, need to be merged together to make further exploration. It is essential to capture the datasets with connection in time and space. In recent years, smart phones with Near Field Communication function can be used to transact in the bus and subway system in Beijing. Besides, e-hailing, e-payment and check-in are widely used with the smart phone. In the future, it is possible to apply the smart phone related information to understand every movement of a person so that human mobility can be modeled and predicted in detail.

Acknowledgments

This work was supported by NSFC Project (Grant No. 71471014 and 71621001) and the Fundamental Research Funds for Central Universities (Grant No. 2015JBM060). We also greatly appreciate Dr. Zhengbing He for providing help on the data pre-processing.

References

[1] R. Kitamura, C. Chen, R.M. Pendyala, R. Narayanan, Micro-simulation of daily activity-travel patterns for travel demand forecasting, Transportation 27 (1) (2000) 25–51.
[2] B. Jiang, J. Yin, S. Zhao, Characterizing the human mobility pattern in a large street network, Phys. Rev. E (3) 80 (1) (2009) 1711–1715.
[3] H.D. Rozenfeld, H.A. Makse, Laws of population growth, Proc. Natl. Acad. Sci. 105 (48) (2008) 18702–18707.
[4] L. Mari, E. Bertuzzo, L. Righetto, R. Casagrandi, M. Gatto, I. Rodriguezituribe, et al., Modelling cholera epidemics: the role of waterways, human mobility and sanitation, J. R. Soc. Interface 9 (67) (2012) 376–388.
[5] G.M. Viswanathan, V. Afanasyev, S.V. Buldyrev, E.J. Murphy, P.A. Prince, H.E. Stanley, Lévy flight search patterns of wandering albatrosses, Nature 381 (6581) (1996) 413–415.
[6] B.B. Mandelbrot, The fractal geometry of nature, Amer. J. Phys. 51 (3) (1983) 286–287.
[7] A. Edwards, R. Phillips, N. Watkins, M. Freeman, E. Murphy, V. Afanasyev, et al., Revisiting lévy flight search patterns of wandering albatrosses, bumblebees and deer, Nature 449 (7165) (2007) 1044–1048.
[8] D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel, Nature 439 (7075) (2006) 462–465.
[9] M.C. González, C.A. Hidalgo, A.L. Barabási, Understanding individual human mobility patterns, Nature 453 (453) (2008) 779–782.
[10] J.B. Sun, J. Yuan, Y. Wang, H.B. Si, X.M. Shan, Exploring space–time structure of human mobility in urban space, Physica A 390 (5) (2011) 929–942.
[11] C. Kang, X. Ma, D. Tong, Y. Liu, Intra-urban human mobility patterns: An urban morphology perspective, Physica A 391 (4) (2012) 1702–1717.
[12] W. Wang, L. Pan, N. Yuan, S. Zhang, D. Liu, A comparative analysis of intra-city human mobility by taxi, Physica A 420 (420) (2015) 134–147.
[13] V. Palchykov, M. Mitrovi, H.H. Jo, J. Saramäki, R.K. Pan, Inferring human mobility using communication patterns, Sci. Rep. 4 (2014) 6174.
[14] L.F. Huntsinger, K. Ward, Using mobile phone location data to develop external trip models, in: Transportation Research Board Annual Meeting, 2015.
[15] T.M.T. Do, O. Dousse, M. Miettinen, D. Gatica-Perez, A probabilistic kernel method for human mobility prediction with smartphones, Pervasive Mob. Comput. 20 (C) (2014) 13–28.
[16] I. Rhee, M. Shin, S. Hong, K. Lee, S. Chong, On the levy-walk nature of human mobility, in: Proceedings - IEEE INFOCOM, Vol. 19, no. 3, 2008, pp. 924–932.
[17] X.Y. Yan, X.P. Han, B.H. Wang, T. Zhou, Diversity of individual mobility patterns and emergence of aggregated scaling laws, Sci. Rep. 3 (9) (2013) 2678.
[18] K. Zhao, M. Musolesi, P. Hui, W. Rao, S. Tarkoma, Explaining the power-law distribution of human mobility through transportation modality decomposition, Sci. Rep. 5 (2015) 9136.
[19] A. Sadilek, J. Krumm, Far out: Predicting long-term human mobility, in: AAAI Conference on Artificial Intelligence, 2012.
[20] W. Huang, S. Li, X. Liu, Y. Ban, Predicting human mobility with activity changes, Int. J. Geogr. Inf. Sci. 29 (9) (2015) 1569–1587.
[21] X. Liang, X. Zheng, W. Lv, T. Zhu, K. Xu, The scaling of human mobility by taxis is exponential, Physica A 391 (5) (2011) 2135–2144.
[22] X. Liang, J. Zhao, D. Li, K. Xu, Unraveling the origin of exponential law in intra-urban human mobility, Sci. Rep. 3 (10) (2012) 65.
[23] X. Liang, J.C. Zhao, K. Xu, A general law of human mobility, Sci. China Inf. Sci. 58 (10) (2015) 1–14.
[24] J. Tang, F. Liu, Y. Wang, H. Wang, Uncovering urban human mobility from large scale taxi gps data, Physica A 438 (2015) 140–153.
[25] H. Cai, X. Zhan, J. Zhu, X. Jia, A.S.F. Chiu, M. Xu, Understanding taxi travel patterns, Physica A 457 (2016) 590–597.
[26] J. Kim, H.S. Mahmassani, Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories, Transp. Res. C 9 (2015) 164–184.
[27] C. Roth, S.M. Kang, M. Batty, M. Barthélemy, Structure of urban movements: polycentric activity and entangled hierarchical flows, PLoS One 6 (1) (2011) e15923.
[28] Q. Xu, B. Mao, Y. Bai, Network structure of subway passenger flows, J. Stat. Mech. Theory Exp. 2016 (3) (2016) 033404.
[29] C. Zhong, E. Manley, S.M. Arisona, M. Batty, G. Schmitt, Measuring variability of mobility patterns from multiday smart-card data, J. Comput. Sci. 9 (2015) 125–130.
[30] M.S. Wang, L. Huang, X.Y. Yan, Exploring the mobility patterns of public transport passengers, J. Univ. Electron. Sci. Technol. China 41 (1) (2012) 2–7.
[31] S. Rahman, J. Wong, C. Brakewood, Use of mobile ticketing data to estimate an origin–destination matrix for new york city ferry service, Transp. Res. Rec.: J. Transp. Res. Board 2544 (2016) 1–9.
[32] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, C. Mascolo, A tale of many cities: universal patterns in human urban mobility, PLoS One 7 (5) (2012) e37027.
[33] M. Saberi, H.S. Mahmassani, D. Brockmann, A. Hosseini, A complex network perspective for characterizing urban travel demand patterns: graph theoretical analysis of large-scale origin–destination demand networks, Transportation (2016) 1–20.
[34] M. Papandrea, K.K. Jahromi, M. Zignani, S. Gaito, S. Giordano, G.P. Rossi, On the properties of human mobility, Comput. Commun. 87 (C) (2016) 19–36.
[35] Y. Liu, Z. Sui, C. Kang, Y. Gao, Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data, PLoS One 9 (1) (2014) e86026.
[36] J. Sochor, H. Strömberg, M.A. Karlsson, Implementing mobility as a service: Challenges in integrating user, commercial, and societal perspectives, Transp. Res. Rec.: J. Transp. Res. Board 4 (2536) (2015) 1–9.
[37] J. Guo, X. Li, Beijing transport annual report 2015, 2015.
[38] Amap, The coordinates picker system in amap. http://lbs.amap.com/console/show/picker (accessed 17.10.16).

[39] A. Clauset, C.R. Shalizi, M.E.J. Newman, Power-law distributions in empirical data, Ann. Appl. Stat. 51 (4) (2009) 661–703.
[40] A. Mashanova, T.H. Oliver, V.A.A. Jansen, Evidence for intermittency and a truncated power law from highly resolved aphid movement data, J. R. Soc. Interface 7 (42) (2010) 199–208.
[41] V.S. Chalasani, J.M. Denstadli, Engebretsen, K.W. Axhausen, Precision of geocoded locations and network distance estimates, J. Transp. Stat. 8 (2) (2005) 1–16.
[42] M. Veloso, S. Phithakkitnukoon, C. Bento, P. Olivier, N. Fonseca, Exploratory study of urban flow using taxi traces, in: The Workshop on Pervasive Urban Applications, 2011.
[43] S. Goh, K. Lee, M.Y. Choi, J.Y. Fortin, Emergence of criticality in the transportation passenger flow: Scaling and renormalization in the seoul bus system, PLoS One 9 (3) (2014) e89980.
[44] S. Hasan, C.M. Schneider, S.V. Ukkusuri, M.C. González, Spatiotemporal patterns of urban human mobility, J. Stat. Phys. 151 (1) (2013) 1–15.
[45] F. Calabrese, D. Mi, G.D. Lorenzo Jr., J.F. Ratti, C. Ratti, Understanding individual mobility patterns from urban sensing data: A mobile phone trace example, Transp. Res. C 26 (1) (2013) 301–313.
[46] Y. Liu, C. Liu, N.J. Yuan, L. Duan, Exploiting heterogeneous human mobility patterns for intelligent bus routing, in: IEEE International Conference on Data Mining. 2014, pp. 360–369.
[47] A.M. El-Geneidy, P.R. Tétreault, J. Surprenant-Legault, Pedestrian access to transit: Identifying redundancies and gaps using a variable service area analysis, in: Transportation Research Board Meeting, 2009.
[48] T. Kimpel, K.J. Dueker, A.M. El-Geneidy, Using gis to measure the effects of service area and frequency on passenger boardings at bus stops, 2007.
[49] D.B. Hess, Access to public transit and its influence on ridership for older adults in two u.s. cities, J. Transp. Land Use 2 (1) (2009) 3–27.
[50] M. Cools, E. Moons, L. Creemers, G. Wets, Changes in travel behavior in response to weather conditions: Do type of weather and trip purpose matter, Transp. Res. Rec.: J. Transp. Res. Board 2157 (2157) (2010) 22–28.
[51] V.W. Stover, E.D. Mccormack, The impact of weather on bus ridership in pierce county, washington, J. Public Transp. 15 (1) (2012) 95–110.
[52] P. Arana, S. Cabezudo, M. Peñalba, Influence of weather conditions on transit ridership: A statistical study using data from smartcards, Transp. Res. Part A Policy Pract. 59 (1) (2014) 1–12.
[53] M. Zhou, D. Wang, Q. Li, Y. Yue, W. Tu, R. Cao, Impacts of weather on public transport ridership: Results from mining data from different sources, Transp. Res. C 75 (2017) 17–29.
[54] A. Singhal, C. Kamga, A. Yazici, Impact of weather on urban transit ridership, Transp. Res. Part A Policy Pract. 69 (69) (2014) 379–391.
[55] C. Liu, Y.O. Susilo, A. Karlström, The influence of weather characteristics variability on individual travel mode choice in different seasons and regions in sweden, Transp. Policy 41 (2015) 147–158.