Journal of Transport Geography 61 (2017) 72–86

Contents lists available at ScienceDirect

Journal of Transport Geography

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

# Revealing intra-urban travel patterns and service ranges from taxi trajectories

Shen Zhang[a], Jinjun Tang[b,*], Haixiao Wang[c], Yinhai Wang[d], Shi An[a]

[a] School of Transportation Science and Engineering, Harbin Institute of Technology, No. 73, Huang-He Street, 150090 Harbin, China
[b] School of Traffic and Transportation Engineering, Central South University, Changsha 410075, China
[c] School of Energy and Transportation Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China
[d] Department of Civil & Environmental Engineering, University of Washington, Seattle, WA 98195-2700, United States

ARTICLE INFO

*Keywords:*
Taxi trajectory
Spatial point pattern
Utilization distribution
Movement-based kernel density estimation

ABSTRACT

As an important transport tool, taxi plays a significant role to meet travel demand in urban city. Understanding the travel patterns of taxis is important for addressing many urban sustainability challenges. Previous research has primarily focused on examining the statistical properties of taxi trips to characterize travel patterns, while it may be more appropriate to explore taxi service strategies on seasonal, weekly or daily time scale. Therefore, intra-urban taxi mobility is investigated by examining taxi trajectory data that were collected in Harbin, China, while 12-week corresponding to 12-month is chosen as the sampling period in our study. The multivariate spatial point pattern analysis is firstly adopted to characterize and model the spatial dependence, and infer significant positive spatial relationships between the picked up points (PUPs) and the dropped off points (DOPs). Secondly, the points of interest (POIs) are identified from DOPs using the emerging hot spot detection technique, then the taxi services and movement patterns surrounding POIs are further examined in details. Moreover, our study builds on and extends the existing work to examine the statistical regularities of trip distances, and we also validate and quantify the impacts posed by airport trips on the distance distributions. Finally, the movement-based kernel density estimation (MKDE) method is proposed to estimate taxis' service ranges within three isopleth levels (50, 75 and 95%) between summer/weekday and winter/weekend from taxi driver's perspective, and season as well as temperature factors are identified as the significant effect within certain service range levels. These results are expected to enhance current urban mobility research and suggest some interesting avenues for future research.

## 1. Introduction

As an important component of the public transportation sector, taxis provide flexible and convenient mobility solutions for urban residents. Locating the taxi-hailing demand of urban residents and optimizing the resource allocation of taxi service is critical to improving the quality of public transit services. However, the limitations of travel survey data reduce the quality of traditional travel demand studies and can result in misleading conclusions. Therefore, alternative data sources are needed to more accurately and comprehensively understand the spatial-temporal characteristics of travel demand.

The widespread adoptions of location-based services (LBS) provide unprecedented opportunities to study various mobility patterns from trillions of trails and footprints (Liang et al., 2012). A careful analysis of these digital footprints from taxi GPS localizers can provide an innovative strategy to facilitating urban public transit planning and operational decision-making, and thus have attracted considerable attentions from researchers.

Jiang et al. (2009) verified the scaling properties of taxi trip length and suggested that such property is attributed to the underlying street topology. Liu et al. (2012a, 2012b) introduced a new method to explore intra-urban human mobility and land use variations based on taxi trajectory data from Shanghai city. Castro et al. (2013) proposed an overview of mechanisms for using taxi GPS data to analyze people's movements and activities, which includes three main categories: social dynamics, traffic dynamics and operational dynamics. Liang et al. (2012) found the taxis' traveling displacements and elapsed time follow an exponential distribution instead of a power-law. Veloso et al. (2011a, 2011b) used taxi data collected in Lisbon city to study urban mobility, spatiotemporal variation of taxi services, relationships be-

[*] Corresponding author.
*E-mail addresses:* shenzhang@hit.edu.cn (S. Zhang), jinjuntang@csu.edu.cn (J. Tang), teacherwhx@126.com (H. Wang), yinhai@u.washington.edu (Y. Wang), anshi@hit.edu.cn (S. An).

http://dx.doi.org/10.1016/j.jtrangeo.2017.04.009
Received 14 November 2016; Received in revised form 18 April 2017; Accepted 20 April 2017
Available online 05 May 2017

tween pick-up and drop-off locations and drivers' behaviors. Wang et al. (2015) investigate statistical distributions of trip displacements, trip durations and interevent time by exploring large amounts of GPS traces collected in five metropolitan cities. Zhu and Guo (2014) proposed a hierarchical method to deal with the problem of how to extract clusters from similar flows in taxi trips. Liu et al. (2015) used a two-level hierarchical polycentric city structure to study spatial interaction perspective in Shanghai city with large scale taxi data. Wu et al. (2014) introduced a novel method to explore urban human mobility based on social media check-in data, in which they constructed transition probability to model travel demand distribution. Liu et al. (2010) analyzed taxi drivers' spatial selection behavior, spatio-temporal operation behavior, route choice behavior, and operation tactics with taxi GPS traces. Pan et al. (2013) worked on urban land-use classification using taxi GPS traces, and found pick-up/set-down dynamics exhibited clear patterns corresponding to the land-use classes of these regions. Tang et al. (2015, 2016a, 2016b) analyzed urban travel mobility, OD distribution, hotspots identification in Harbin city using large-scale taxi GPS trajectories. Since taxi travel pattern may also relate to vacant search behavior, Wong et al. (2014a, 2014b) further proposed a cell-based model to predict local customer-search movements of vacant taxi drivers.

These studies mentioned above have provided valuable insights into the patterns and mechanisms of taxi mobility, they tend to combine data on the location of origins and destinations, trip purpose, trip length, trip duration, departure and arrival times and travel modes in their analyses. Moreover, it is worth noting that the aforementioned studies generally focused on understanding travelers' behaviors on ordinary weekdays. Until recently, few studies have examined the variations of travel patterns between seasons, while the spatial-temporal properties of seasonal transport trips still tend to be regular and periodic. As a result, a purpose of this study is to model the travel patterns of taxi passengers on seasonal, weekly or daily time scale, and assess potential climate change effects on travel behavior. Additionally, most current studies merely focus on partial statistical properties of taxi trips (i.e., the trip displacements and durations). However, trip is not the only representative measure of taxi's travel pattern, while it may be more appropriate to estimate service coverage on an individual level from the perspective of public transport. Therefore, this research also aims to fill this gap by measuring taxis' service range sizes referred to home range studies (Braham et al., 2015; Dürr and Ward, 2014). The service range of a taxi is described in terms of a probabilistic model in our study, while the calculation of range size is required from the estimation of the utilization distribution (UD: the name given to the probability distribution of individual spatial location). Therefore, the service range model not only provides a useful global representation of space coverage patterns of taxis, but also represents the probability of taxi-hailing opportunities for passengers in a defined area, which is of great importance in public transport service studies. Such results can also help guide traffic modeling, public transportation planning, urban planning, and infrastructure development.

Our on-going work is focused on the analysis of taxi-GPS traces acquired in the city of Harbin, China, to better understand urban mobility from both driver's and passenger's perspectives. The contribution of this work lies on the following aspects: we investigate the multivariate spatial point pattern between pick-up and drop-off locations, analyze the local mobility patterns around POIs, examine the statistical distributions of trip distances, and explore the possibility of estimating the service range and identifying the seasonal effects from spatial area perspective.

## 2. Materials and methods

### 2.1. Study area

The case study is carried out in Harbin city, which locates in the northeast of China. The study area is restricted to Harbin urban area (126°57′–126°73 N lat and 45°68′–45°80E long) compared to its administrative area, which is rather large with latitude spanning 44° 04′–46° 40′ N, and longitude 125° 42′–130° 10′ E.

The area is characterized by a sub-humid continental monsoon climate with four distinct seasons, including a cold and dry winter (December–February) and a hot and rainy summer (June–August). The annual mean temperature is + 4.25 °C (39.6 °F), and extreme temperatures have ranged from − 42.6 °C (− 45 °F) to 39.2 °C (103 °F).

### 2.2. Data collection and study design

We conduct the study for the whole year of 2014. Since there are over 10 million GPS records in one day, 12-week corresponding to 12-month is chosen as the sampling period. Data is collected from 1000 distinct taxis, which account for nearly 12.5% of taxis in Harbin area. The sampling rate is 30 s, and total samples come to 2880 a day. Each data sample contains information of taxi id, timestamp, latitude, longitude, instantaneous velocity, and taximeter state (vacant or occupied). A data cleaning process is applied, removing trips with $<$ 200 m and $>$ 60 km (the realistic longest trips from one side of the city to the other could be around 55 km).

For illustration purposes, we have mapped the distribution of all sample taxi records in one day (Fig. 1) using open street map. Fig. 1 clearly shows that the majority of sample records are in the inner city area. Then, we classify taxi trips into two parts based on their status: (1) pick up passengers from origins to destinations. (2) Roam on the road to find next passenger. Fig. 1a demonstrates a one-day trajectory of a taxi, where red dotted lines denote the occupied trip paths, and green dotted lines indicate the unoccupied trips. The spatial overlapping parts are caused by temporal factors. Based on the data, we can also identify the locations where passengers were picked up and dropped off (Fig. 1b and c), and thus the origin and destination of a completed occupied trip. PostgreSQL 9.4 and its spatial extension PostGIS are proposed for managing GPS data sets in our study, while ArcGIS 10.3 is adopted as the GIS interface for manipulating and visualizing the PostGIS data. Besides, the core of the analytical inspection and graphics are fully executed in the R 3.2.5 environment.

### 2.3. Trips classification

Taxi trip is a very important part of human beings movements in urban areas. In this section, three parameters including travel distance, time and average speed are used to explore taxi mobility. As we mentioned in above section, taxi drivers always exhibit different driving behaviors at different status: load up passengers and vacant. Thus, the trips can be classified into two parts. Dataset for occupied taxi $k$ at time period $\tau$ can be expressed as: $R^{o} = (k, l^{o}, \tau^{o})$, in which $l^{o} = (x^{o}, y^{o})$ denotes utm32 geometry coordinates in the specific Universal Transverse Mercator (UTM) zone. Similarly, the dataset of non-occupied taxi can be defined as $R^{n} = (k, \mathrm{l}^{n}, \tau^{\ n})$ and $\mathrm{l}^{n} = (x^{n}, y^{n})$. So, the travel distance is defined as:

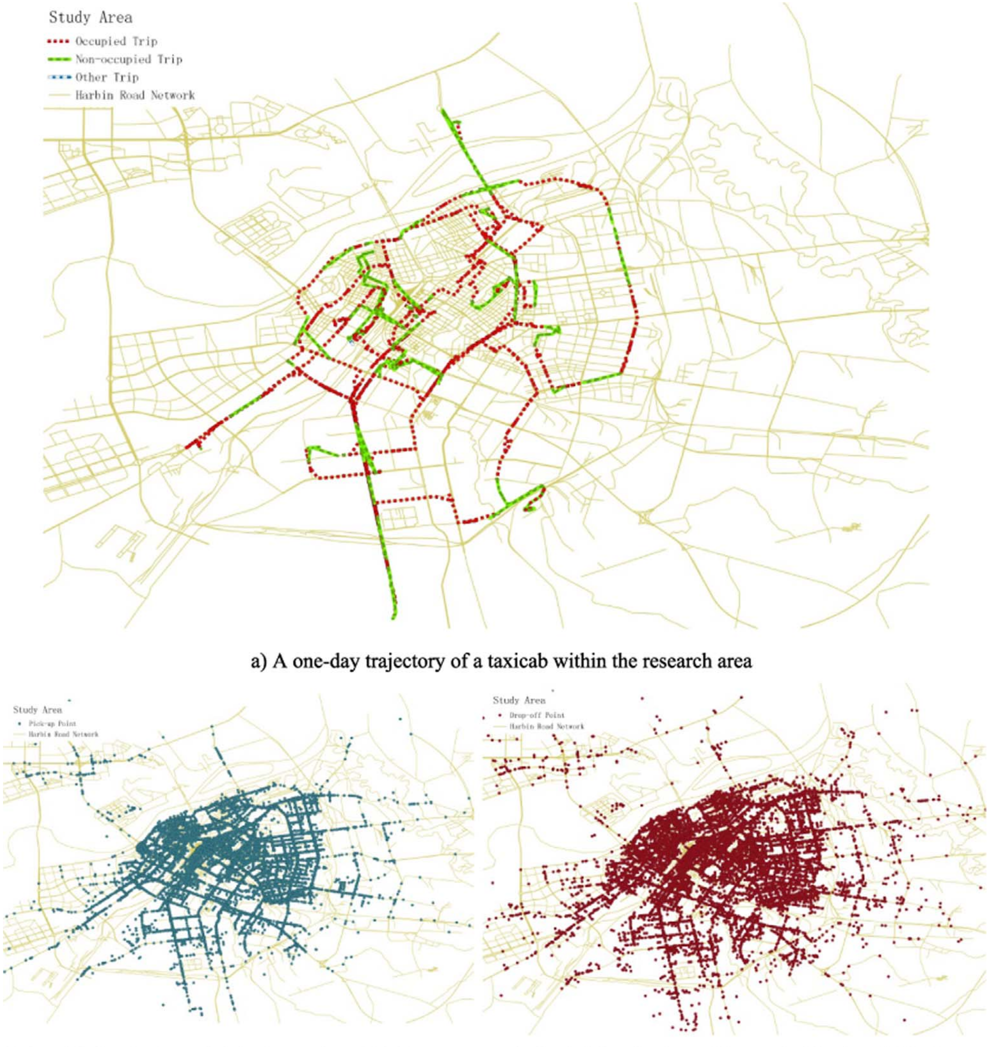
$$d = \sum_{i=1}^{N-1} \left| l_i - l_{i-1} \right| \tag{1}$$

where $N$ is the total number of data samples in an unique trip with status of occupied or non-occupied, |.| means the real distance in meters between two adjacent locations in geographic coordinates.

Finally, occupied and non-occupied trips are extracted for weekday/weekend and winter/summer respectively.

### 2.4. Multivariate spatial point pattern of origins and destinations

Since there are multiple types of spatial points in our study, which indicate the origin/destination points of occupied and unoccupied trips,



a) A one-day trajectory of a taxicab within the research area

b) Spatial distribution of picked up points (PUPs)

c) Spatial distribution of dropped off points (DOPs)

**Fig. 1.** An overview of the study area and trajectory data.

it is also important to get a deep insight into the spatial dependence pattern of these points. It should be noted that picked up points (PUPs) and dropped off points (DOPs) are considered to be more important and chosen for this study. Therefore, a multi-type point process test based on K-cross function (Gelfand et al., 2010), which is an extension of the K-function in spatial point pattern theory (Illian et al., 2008), is proposed for exploring the potential spatial associations between PUPs and DOPs. Given a multivariate process with constant intensity for PUPs and DOPs, the K-cross function in an observed in region A, is given by the expression:

$$\widehat{K}_{ij}(r) = \frac{1}{\widehat{\lambda}_i \widehat{\lambda}_j |A|} \sum_k \sum_l w_{x_k x_l}^{-1} I(\| x_k - x_l \| \le r) \tag{2}$$

where $w_{x_k} x_l$ is edge effect correction and $\widehat{\lambda}_i$ and $\widehat{\lambda}_j$ are the empirical estimators of the intensity for PUPs and DOPs. This is the maximum likelihood estimator for a homogeneous Poisson process.

A Monte Carlo K-cross test is implemented as follows. Firstly, the marginal intensities $\widehat{\lambda}_i$ and $\widehat{\lambda}_j$ are estimated, PUPs and DOPs are further combined to generate a single process. Then this procedure is repeated s-1 times, and $\widehat{K}_{ij,k}(r)$ is computed for each simulation. The maximum and minimum values of $\widehat{K}_{ij,k}(r)$, k = 2, …,s are the upper and lower envelopes, respectively. If $\widehat{K}_{ij}(r)$ is greater than the upper envelope or lower than the lower envelope for the observed process, a respective attraction or inhibition will be detected between PUPs and DOPs.

### 2.5. Service range estimation

In our study, the service range of a taxi is estimated from the computation of the utilization distribution (UD: the name given to the probability distribution of individual spatial location). The utilization distribution takes the form of a two dimensional probability density function that represents the probability distribution of individual spatial location.

UDs are commonly computed through a static approach involving location-based kernel density estimations (LKDE) (Silverman, 1986; Worton, 1989), which considers locations as a set of independent points, can however be questioned when recorded locations are to a large extent serially correlated (Kie et al., 2010). As a result, a number of improvements or alternative approaches have been recently developed (Getz et al., 2007; Horne et al., 2007; Keating and Cherry, 2009). Movement-based kernel density estimation (MKDE) (Benhamou and Cornélis, 2010; Benhamou, 2011) (Benhamou and Cornélis, 2010; Benhamou, 2011), which takes advantage of movement information to improve UD estimation, constitutes a significant progress and hence is proposed for estimating taxi service range in our study.

In brief, the MKDE method estimate spatial probability density **U** at the center of any quadrat $\boldsymbol{Z_Q}$ of a virtual grid overlaid on the whole environment using a circular bivariate Gaussian kernel function as follows:

$$\mathrm{U}(Z_Q) = \frac{1}{2\pi N}\sum_{i=1}^{N}\frac{1}{h_i^2}\exp\left[-\frac{||Z_Q - z_i||^2}{2h_i^2}\right] \tag{3}$$

where $z_i$ (with $i = 1, 2, \ldots, N$) is $N$ recorded locations within the quadrat $\boldsymbol{Z_Q}$. Besides, the associated smoothing parameter $h_i$ for the $i$th location $z_i$ can be calculated with:

$$h_i^2 = h_{\min}^2 + 4p_i(1 - p_i)(h_{\max}^2 - h_{\min}^2)T_i/T_{\max} \tag{4}$$

where $h_{min}$ and $h_{max}$ are the minimum and maximum smoothing parameters respectively, while $T_{max}$ is the maximum duration (in seconds) allowed for a step built by successive relocations. The choice of these values is more an empirical than mathematical matter.

For the upper time limit $T_{max}$, although the time interval between records was set to a fixed value in this study, the exact interval between consecutive records could vary slightly – due to temporal noise of the GPS unit – or considerably – e.g. when the taxi was lying under a bridge for a long time period, hampering the ability of the GPS unit to locate satellite fixes. Therefore, we set $T_{max}$ to 3 min, assuming that the taxis' behavior is approximately homogenous (i.e. the taxi moved according to a biased random walk) during this period. Then all the GPS fixes above $T_{max}$ are excluded from the UD calculation.

Secondly, the minimum smoothing parameter $h_{min}$, has to be specified to incorporate the GPS unit's classical noise (i.e. mean distance of the fixes and their centroid), because the actual location of the taxi is somewhere within a cloud of the recorded GPS fixes rather than at the fix location itself. We set $h_{min}$ to 384 m based on the results of the static test.

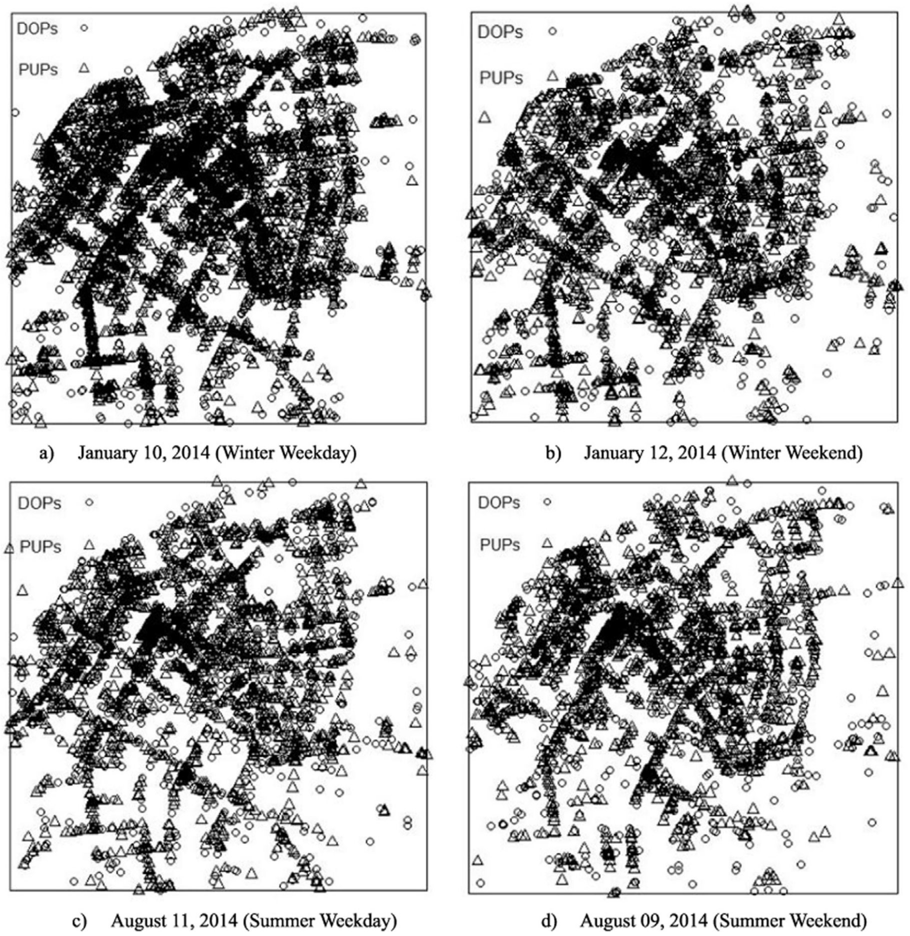
Finally, because the smoothing parameter acts as a standard deviation in the kernel function, the maximum smoothing parameter $h_{max}$ should not be larger than half the mean distance the tracked taxi is expected to move for time $T_{max}$ when fully active. Intermediate values of the mean cosine of turning angles should translate into intermediate $h_{max}$ values. Therefore, we estimated that we could set the upper limit of the smoothing parameter to $h_{max} = 900$ m.

## 3. Results analysis and discussion

### 3.1. Global spatial patterns of PUPs and DOPs

PUPs and DOPs play different roles in traffic generation and attraction from different perspectives. For example, PUPs represent generation of travel demand from passenger's perspective, while they also reflect the attraction for taxi drivers to look for riders. Therefore, it is critical for us to explore their spatial patterns and dependences in depth. A global spatial measure, multivariate point pattern is proposed to examine the interactions and validate the existence of cluster pattern, which is also a prerequisite for hot spot detection in next section.

For each day, the PUPs and DOPs are extracted from the trips within an observe window, which is constructed based on a rectangle boundary of Harbin inner city (126°57′–126°73 N lat and 45°68′–45°80E long). In order to guarantee the Monte Carlo simulation efficiency, the numbers of PUPs and DOPs are around 3000 by limiting taxi sample size. Then the PUPs and DOPs are integrated in pairs for further multivariate point pattern analysis. The pairwise integration results of the picked up points (PUPs) and the dropped off points (DOPs) are visualized in four subgraphs (Fig. 2). Then, in the multivariate point pattern analysis, we examined spatial dependence pattern using K-cross function, which returns $1/\lambda$ times the number of point $A$ within a distance $r$ of a point $B$. The Monte Carlo K-test, which is demonstrated over a large number ($n = 99$) of simulation envelopes



a) January 10, 2014 (Winter Weekday)

b) January 12, 2014 (Winter Weekend)

c) August 11, 2014 (Summer Weekday)

d) August 09, 2014 (Summer Weekend)

**Fig. 2.** Pairwise integration results of PUPs and DOPs.

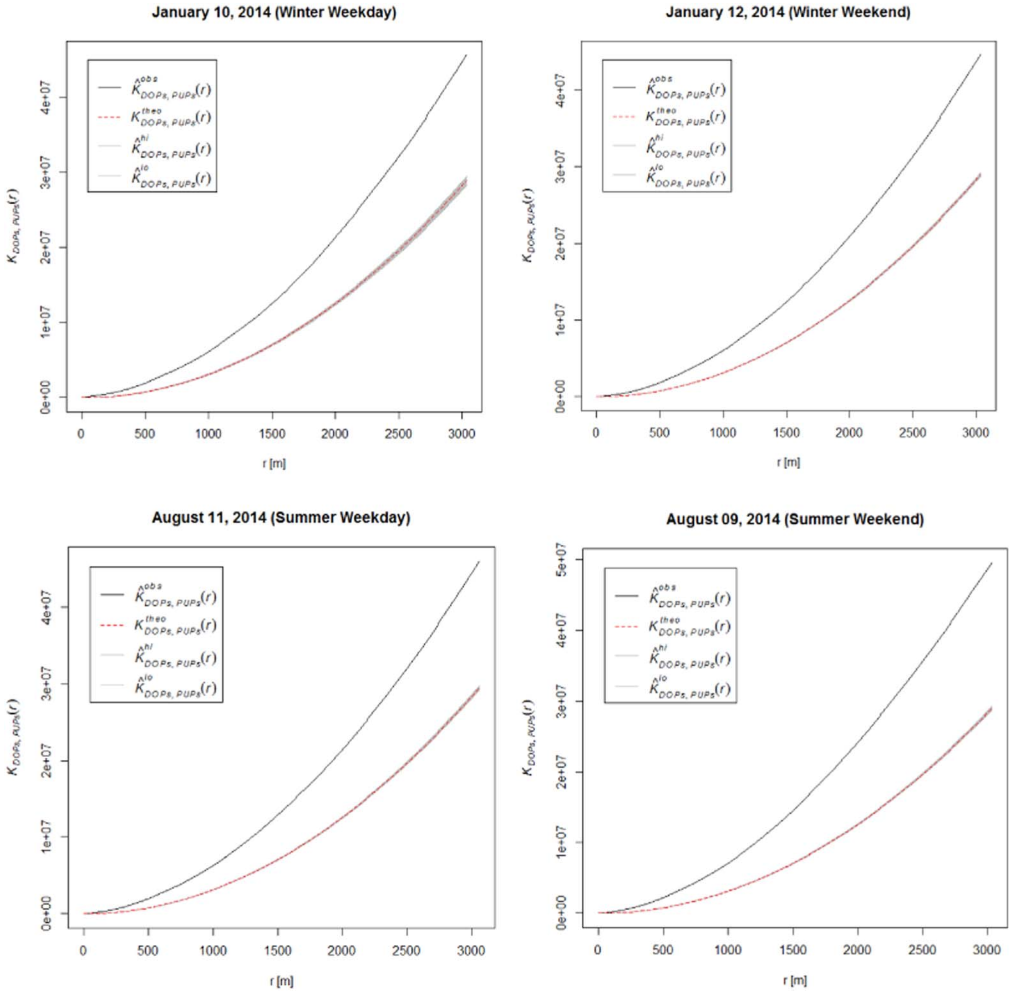


**Fig. 3.** Monte Carlo tests based on the K-cross function for spatial dependence between PUPs and DOPs.

under homogeneity hypotheses, is applied to the OD point process to test the spatial dependence. Fig. 3 shows the results of the Monte Carlo tests based on the K-cross function for spatial dependence between OD points of occupied trips in four representative random days.

All the observed K-cross values of four subgraphs remain above the upper envelope and up to 3000 m, evidencing positive dependence between every pair of the PUPs and DOPs. The four groups of simulation results are extremely similar and further validate the stability of these dependence relationships. Moreover, the results also provide evidence of clustered patterns for both PUPs and DOPs.

We further investigate the dependencies between PUPs and DOPs based on spatial distance and intensity. Briefly, at any spatial location $u$, let $\lambda(u)$ be the intensity of the point process PUPs, and the covariate Z ($u$) represents distance from $u$ to the nearest DOP. Then the function ρ in $\lambda(u) = \rho(Z(u))$ is estimated to determine how the PUPs intensity depends on the distance to the nearest DOP. In our study, kernel smoothing is used to estimate the function ρ, using methods of relative distribution, as explained in (Jones, 1991).

Fig. 4 illustrates the estimation results, and similar slope peaks can be observed at an approximate distance of around 50 m. The estimated standard deviation exhibits a generally increasing function of distance before these peaks, while the spatial distribution intensity of the distance covariate is sharply decreasing after these peaks. These results indicate that the optimal roaming distance retains within 100 m, which implies the best picking-up opportunities for taxi drivers as well as the best taxi-hailing opportunities for passengers. It is worth noting that this is still the theoretical conclusion, and taxis can hardly roam for the specific passenger successfully in most cases. Besides, a long-tail phenomenon can be observed over a distance of 400 to 600 m, which indicates that taxis seldom fail to pick up passengers over such a distance. Furthermore, Fig. 7 reveals the function ρ approximately represents a mixture of exponential or power law distributions. Since ρ varies across daily datasets as shown in subgraphs (Fig. 4), for the sake of space, this study didn't dive more into the parameter-estimation process and the fitted results.

### 3.2. Local mobility around places of interest

As stated above, we have validated the dependence relationships and clustered patterns of PUPs and DOPs. Then we can further investigate the local mobility patterns with the OD points, where places of Interest (POI) are often chosen as local observation site with high travel demand. Traditional user's POI refers to different significant places across multiple dimensions, such as the home, office, park or shopping mall. In the literature, the relationship between travelers' movement behaviors and POIs has attracted considerable research interest due to the large range of applications (Bhattacharya et al., 2015). Prior literature mainly focuses on extracting and predicting POIs from users' check-in points or trajectories. However, we found that the previous detection method was not so exact, lacking of temporal considerations. In comparison, the emerging hot spot detection technique is proposed to identify the POIs from PUPs or DOPs, then taxi services and movement patterns surrounding POIs are examined in our study.

The procedure can be summarized as follows: Firstly, two space time cubes are built for weekday and weekend separately, then we

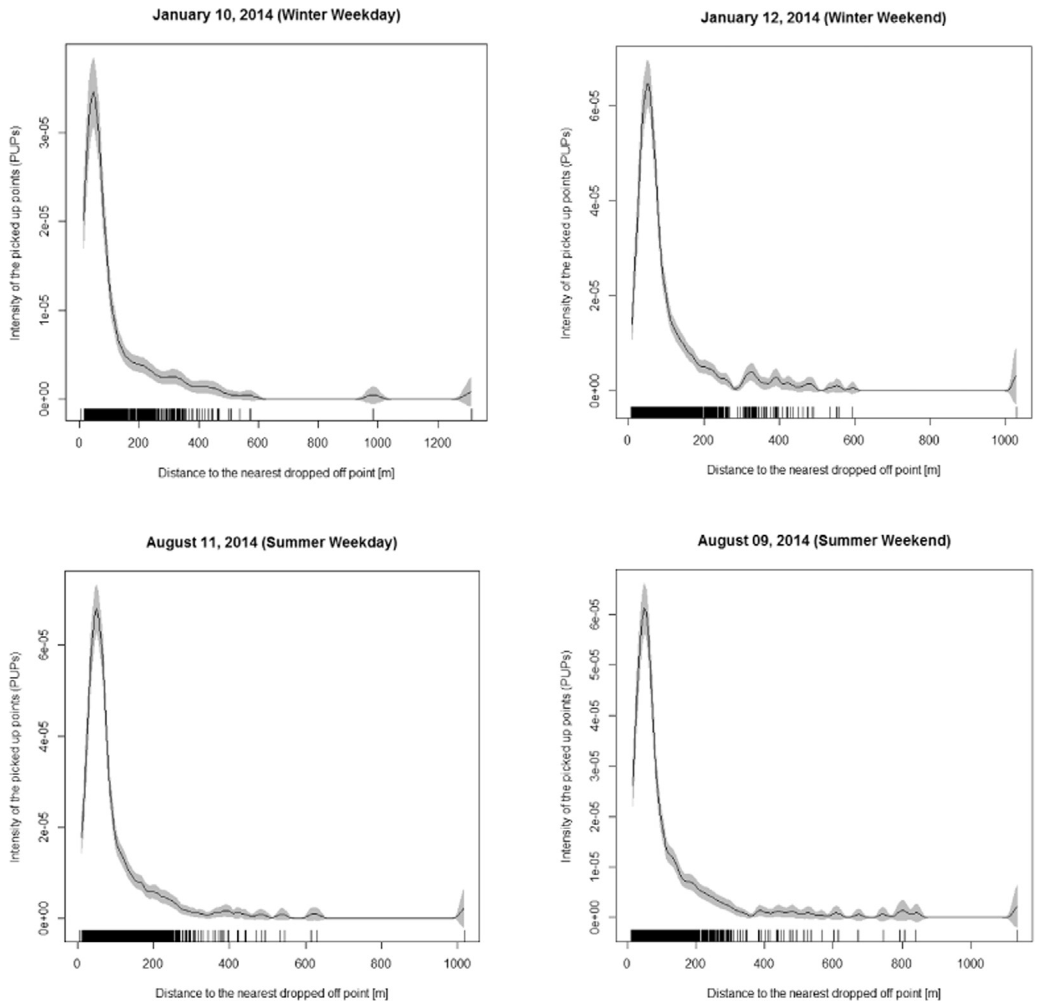
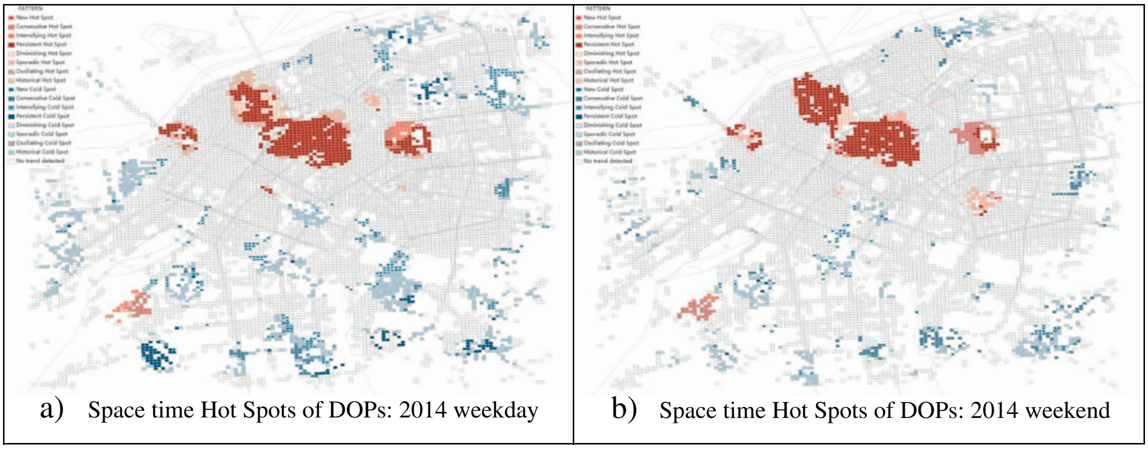


**Fig. 4.** Estimates of ρ for PUPs intensity as a function of distance to nearest DOP. Solid lines are estimates of ρ. Grey shading indicates ± 2 standard deviation (nominally 95% pointwise confidence) intervals.



a) Space time Hot Spots of DOPs: 2014 weekday

b) Space time Hot Spots of DOPs: 2014 weekend

**Fig. 5.** Space time Hot Spots of DOPs: 2014.

summarize all the destinations into the space-time cubes by aggregating them into space-time bins. In the space time cube, the time step interval is one month, which means that we choose month as time scale but not day or hour. Next, the Getis-Ord Gi* statistic (Getis and Ord, 1992) is calculated for each space-time bin. Once the space-time hot spot analysis is completed, each bin in the space time cubes has an associated z-score, *p*-value, and hot spot bin classification added to it. That means a trend over time or space can be identified. Finally, these hot and cold spot trends are evaluated using the Mann-Kendall trend test, and the results are shown in Fig. 5. Since an evidencing positive dependence between PUPs and DOPs has been recognized from Section 3.1, only POIs of DOPs are chosen for further study from passenger's perspective for sake of space.

As shown in the following figures, two typical areas have been identified as obvious persistent hotspots, which mean these locations have been statistically and spatially significant for > 10 months with

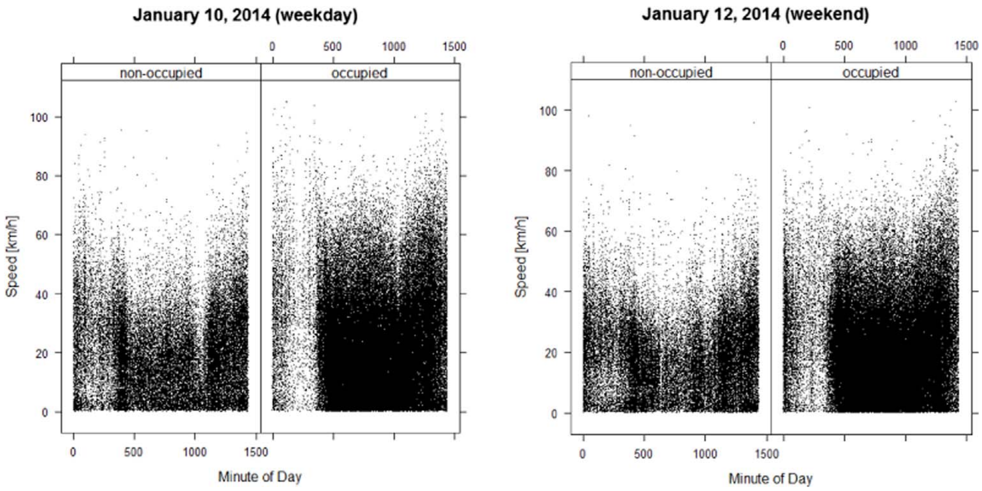


**Fig. 6.** Temporal distribution of speed within 8 km area around POIs over 1440 min (solar day).

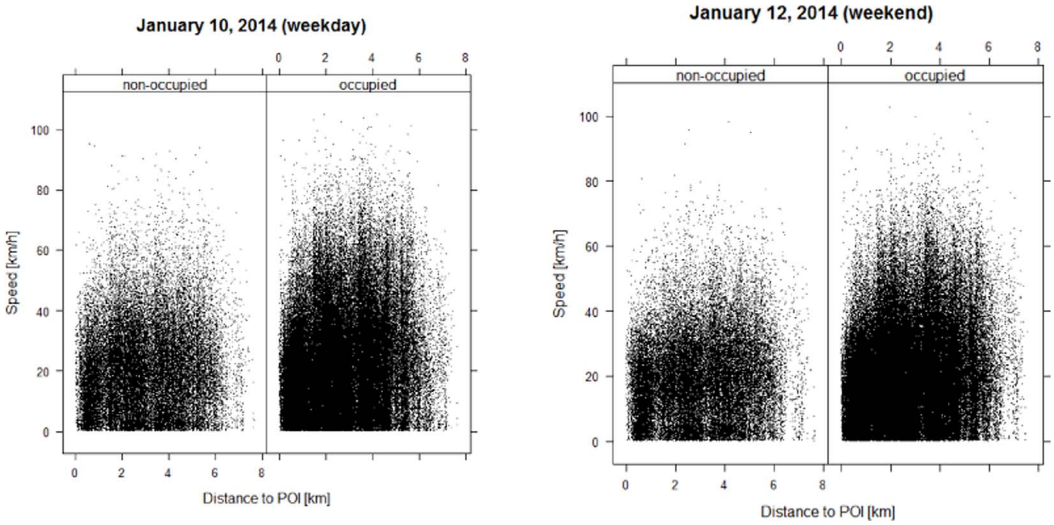


**Fig. 7.** Relationships between speed and distances to nearest POIs.

no discernible increase or decrease trend over time. These persistent hotspots are: 1) Qiulin CBD and Harbin Railway Station; 2) Daoli CBD. Both of them should be monitored with more attention and used in the following analysis.

Fig. 6 illustrates the temporal distribution of speed around POI areas over weekday and weekend. The data points of occupied trips accounted for 70%–80%, whereas the unoccupied trip data accounted for 20%–30% of our sample size. Obviously, the average speed of unoccupied trips is generally lower than that of occupied trips, which is probably due to the roaming process for passengers by unoccupied taxis. Moreover, the average speed of both trips was obviously higher in the midnight and the early hours of the morning, which conforms closely to the regular pattern of the travelers. Furthermore, there is a significant drop in the average speed at the evening rush hour in weekday, while the trend is not very significant in weekend. This phenomenon can be preliminarily explained by the unstable travel demand in weekend compared to weekday. It can be concluded that the commuting trips can also lead to road congestion around POIs. But beyond that, there is no obvious speed pattern difference between weekday and weekend, which indicates that average speed around POIs during off-peak periods is not so sensitive to the weekday and weekend.

Furthermore, this study would like to explore the relationship between taxi speed and distances to nearest POIs, and a positive relationship on weekday/weekend can be directly observed in Fig. 7. A hypothesis test (*t*-test and F-test) can also provide strong evidence that distance have a significant effect and can be a good predictors for speed ($p < 0.001$).

### 3.3. Trip distance distribution

In the literature, the trip displacement has been treated as an alternative measure and applied in a comparative analysis to observe the travel behavior (Wang et al., 2015). Therefore, our study builds on and extends the existing work to explore the taxi mobility, while a more precise measure "trip distance" is adopted instead of the trip displacement. Besides, since the complementary cumulative distribution function (CDF), which is defined to be $P(x) = \Pr(X \geq x)$, is considered to be more useful to characterize the heavy-tailed cases (Clauset et al., 2009), this study propose CDF instead of the probability density to capture the distributions. As a result, we compare and fit several commonly used heavy-tailed distributions to our datasets, including power law, log-normal, and exponential.

Fig. 8 shows the CDFs of occupied trip distances computed from four example datasets, and the red, green, blue solid lines denote the tested distribution models (power-law, log-normal, and exponential respectively). It is clear that the exponential distribution is not appropriate for the trip distances. However, the log-normal and power law distribution both provide reasonable fits to the datasets, but no model appears to appropriately capture the distribution tails. Then Vuong's test is adopted to estimate the test statistic and select better-fitting model

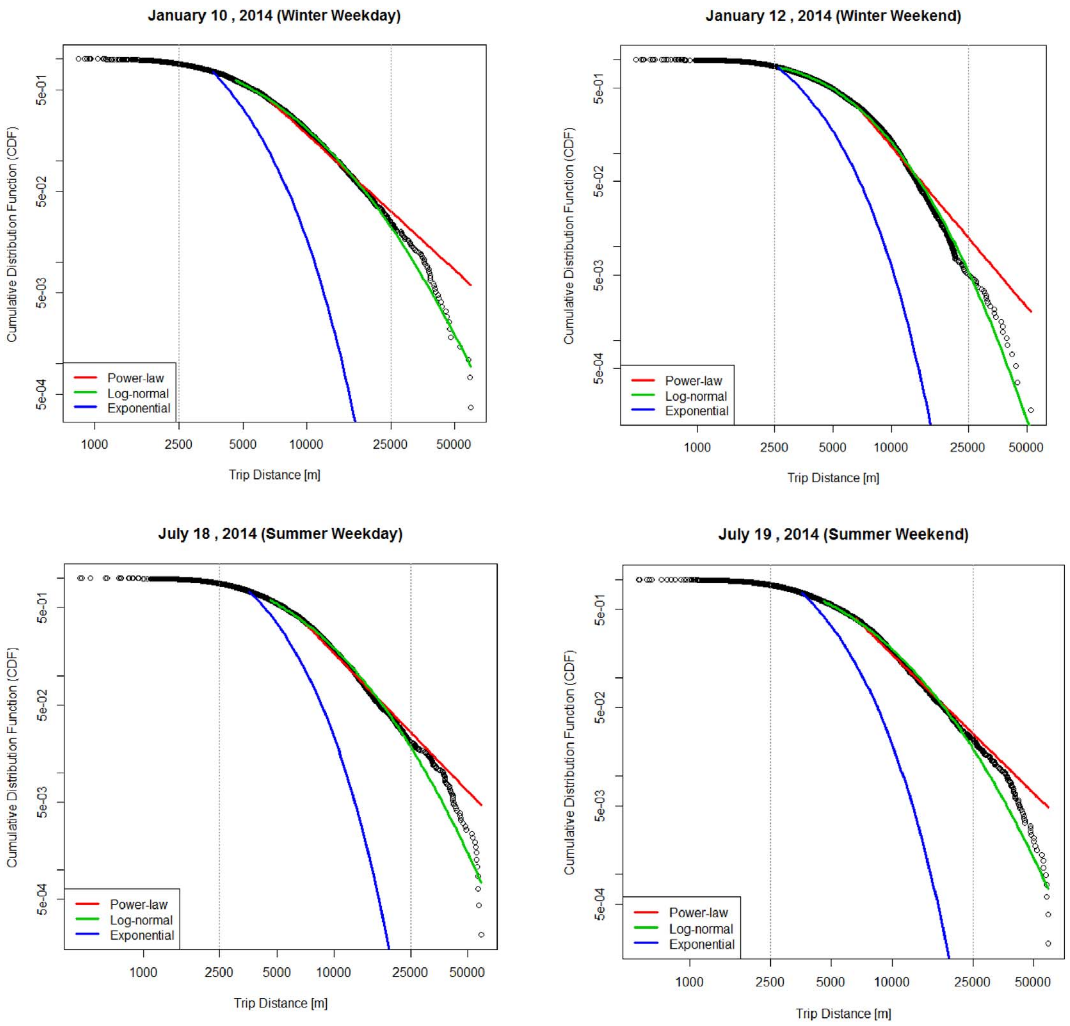


**Fig. 8.** Cumulative distribution functions of trip distances.

using the Kullback-Leibler criteria (Vuong, 1989), and it is found that trip distance distributions fit better to log-normal distribution for this continuous cases, which is defined as

$$P(x) = C\frac{1}{x}\exp\left[-\frac{(lnx-\mu)^2}{2\sigma^2}\right] \tag{5}$$

where $\mu$ and $\sigma$ denote, respectively, the mean and the standard deviation of the natural logarithm of the variable.

In Fig. 8, common trends are observed in the four subgraphs. For each dataset, the full CDF can be divided into three parts by the lines at 0.25 km and 25 km. The first part maintains a relatively steady state with subtle changes, and the second part gradually falls down as the trip distance continues to increase, but the final part varies from day to day.

Similar observations were also reported by Liang et al. (2012), Wang et al. (2015) and Noulas et al. (2012). We note that approximately 11% of occupied trips are within 2.5 km in all the studied datasets. Furthermore, there are < 1% of occupied trips with displacements < 1 km in Harbin. This is fairly consistent with our daily experience. Especially, since the starting price for a taxi ride in the Harbin covers the first 3 km, individuals rarely take taxis for very short trips, but turn to non-motorized modes due to economic consideration. This pricing strategy may well explain the threshold of 2.5 km observed in Fig. 8.

The second dividing line at 25 km can be also attributed to the travel cost. As stated in Wang et al. (2015), people typically use subway and other public transportation systems more than taxis for longer distances and more routine intra-city travel purposes. However, long distance taxi trips do happen for many reasons. For example, clients who plan to catch the flights or trains for inter-city travels or return from usually exhaustive inter-city travels via airports and train stations, especially when carrying big and/or heavy luggage, tend to use taxis, which quite often result in long distance intra-city trips.

Generally, the above results are partly consistent with the empirical observations reported in Refs. (Liang et al., 2012; Veloso et al., 2011a, 2011b; Wang et al., 2015). Liang et al. (2012) analyzed two taxi-trace datasets in Beijing during different periods. They illustrated the full distribution of displacements with two ranges of exponential fits. Veloso et al. (2011a, 2011b) explored the taxi GPS traces in Lisbon. They considered the trips within 25 km and fitted the trip displacements with a Gamma distribution. Wang et al. (2015) reported the trip displacement distributions fit best to log-normal and exponential distributions in six city datasets. To summarize, our study shows that the taxi trip displacements in Harbin tend to follow log-normal distributions. In practice, the examined heavy-tailed distributions (power law and log-normal) are subexponential (Embrechts et al., 1997).

Moreover, some previous studies state that the airports, which are always located far from the heart of the city, attract large amounts of taxi traffic. As a result, the relatively high portion of taxi airport trips with long distances may lead to a disturbance to the whole distribution. They also found that there were shifts in the spike regimes of distance distributions before and after the removal of taxi airport trips (Wang et al., 2015). Since there is a lack of airport ground access modes (e.g.,

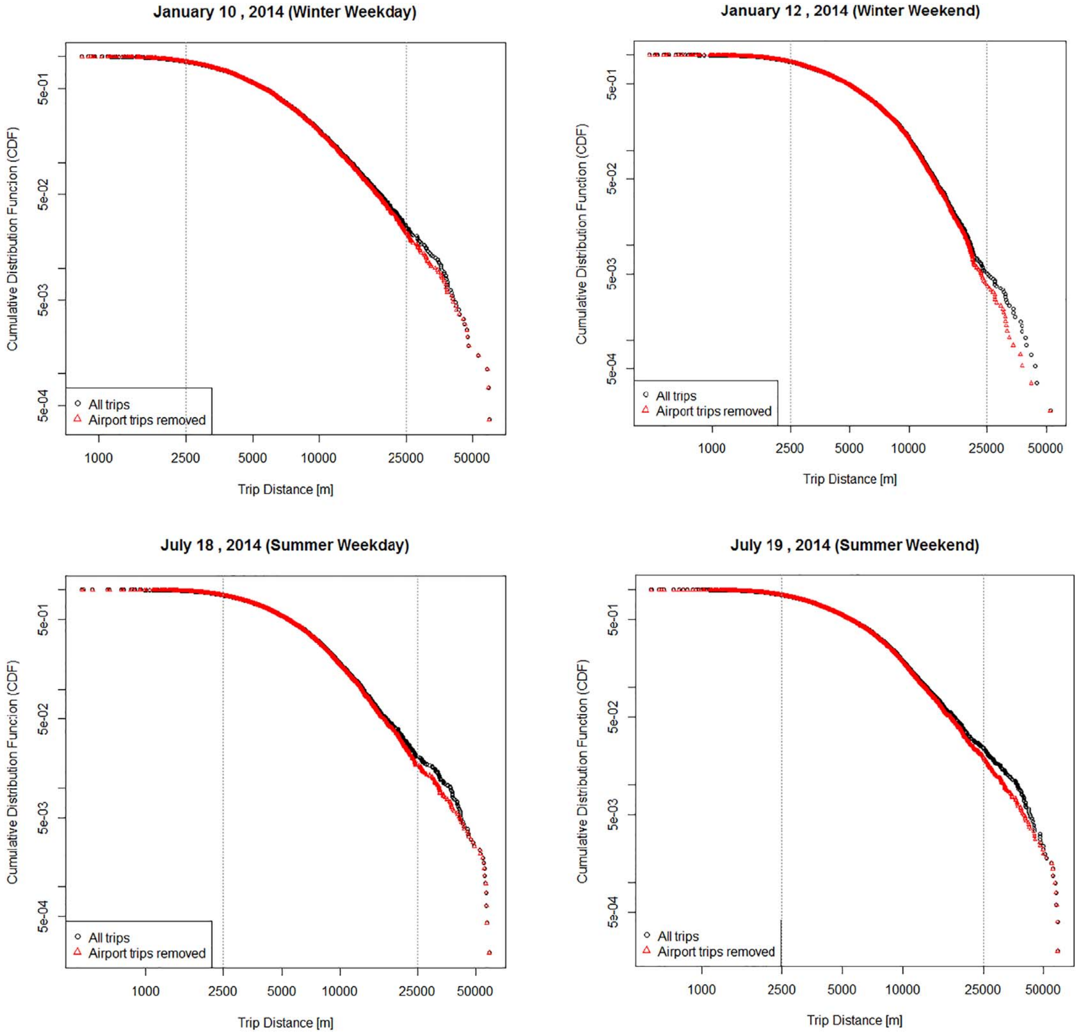


**Fig. 9.** Cumulative distribution functions of trip distances for all the trips and non-airport trips.

airport express) in Harbin when compared to other Chinese cities, it is of great value in examining whether or not taxi airport trips also impact the distance distribution in our study.

Fig. 9 compares the distance distributions for all the taxi trips with the corresponding non-airport parts. As can be seen from Fig. 9, there are intuitive shifts in the tail parts of distance distributions before and after the removal of taxi airport trips. This reasonably suggests that the appearance of tail spikes in Fig. 8 is due to a portion of taxi airport trips in Harbin, and the removal of airport trips is also significantly helpful in smoothing distribution tails. These findings are consistent with the empirical observations reported by Wang et al. (2015).

Furthermore, we try to quantify the impact posed by airport trips on our datasets. Hence we perform a hypothesis test suggested by Clauset et al. (2009) using a goodness-of-fit test, via a bootstrapping procedure. Such tests are based on measurement of the "distance" between the distribution of the empirical data and the hypothesized model. This test generates a $p$-value that can be used to quantify the plausibility of the hypothesis. If the $p$-value is large, then any difference between the empirical data and the model can be explained with statistical fluctuations, while if$p \simeq 0$, then the model does not provide a plausible fit to the data and another distribution may be more appropriate. Generally, a lager $p$-value denotes a better fitting result.

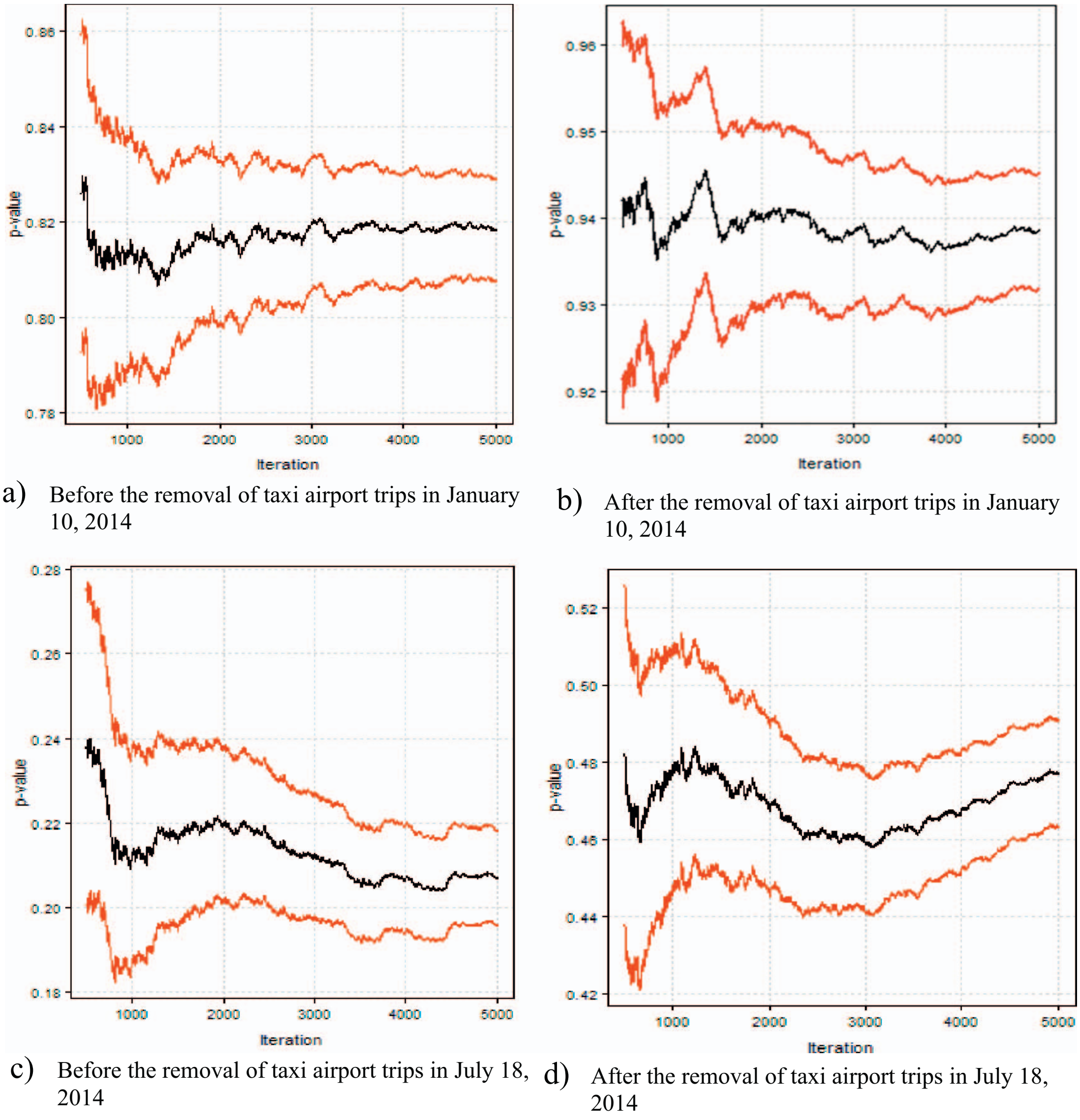
Fig. 10 shows the results from the bootstrap procedure for determining the plausibility of the log-normal hypothesis for random two days. The top row shows the cumulative estimate of the $p$-value in January 10, 2014, while the bottom shows the $p$-value in July 18, 2014. The left and right subgraphs for each row denote the results before and after removing the taxi airport trips respectively. As can be seen from the figure, there is a significant increase in $p$-values by removing the airport trips. These quantitative results suggest that the airport trips could play an important role in shaping the distance distribution of taxi trips, especially in cities like Harbin, which lacks other airport ground access modes (e.g., airport express).

### 3.4. Taxi service range estimation and effect analysis

The mechanism of the MKDE methods is that the frequency probabilities with which the taxi used specific areas within their service ranges can be estimated by defining different isopleth levels. Therefore, three types of service range were calculated using MKDE within the 50%, 75% and 95% isopleth levels between summer/weekday and winter/weekend with the following definitions.

- Results within 50% isopleth level represent the core service range
- Results within 75% isopleth level represent the common/extended service range
- Results within 95% isopleth level represent the more extended service range

Then plausible results for the size and shape of the taxi service range were illustrated in this section (see the example of three taxis in Fig. 11 for the winter/weekday and summer/weekend, respectively).



a) Before the removal of taxi airport trips in January 10, 2014

b) After the removal of taxi airport trips in January 10, 2014

c) Before the removal of taxi airport trips in July 18, 2014

d) After the removal of taxi airport trips in July 18, 2014

**Fig. 10.** Hypothesis test results for log-normal model before and after the removal of taxi airport trips. The dashed-black line shows the cumulative mean estimate of the *p*-value after 5000 iterations. The dashed-red lines give approximate 95% confidence intervals. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

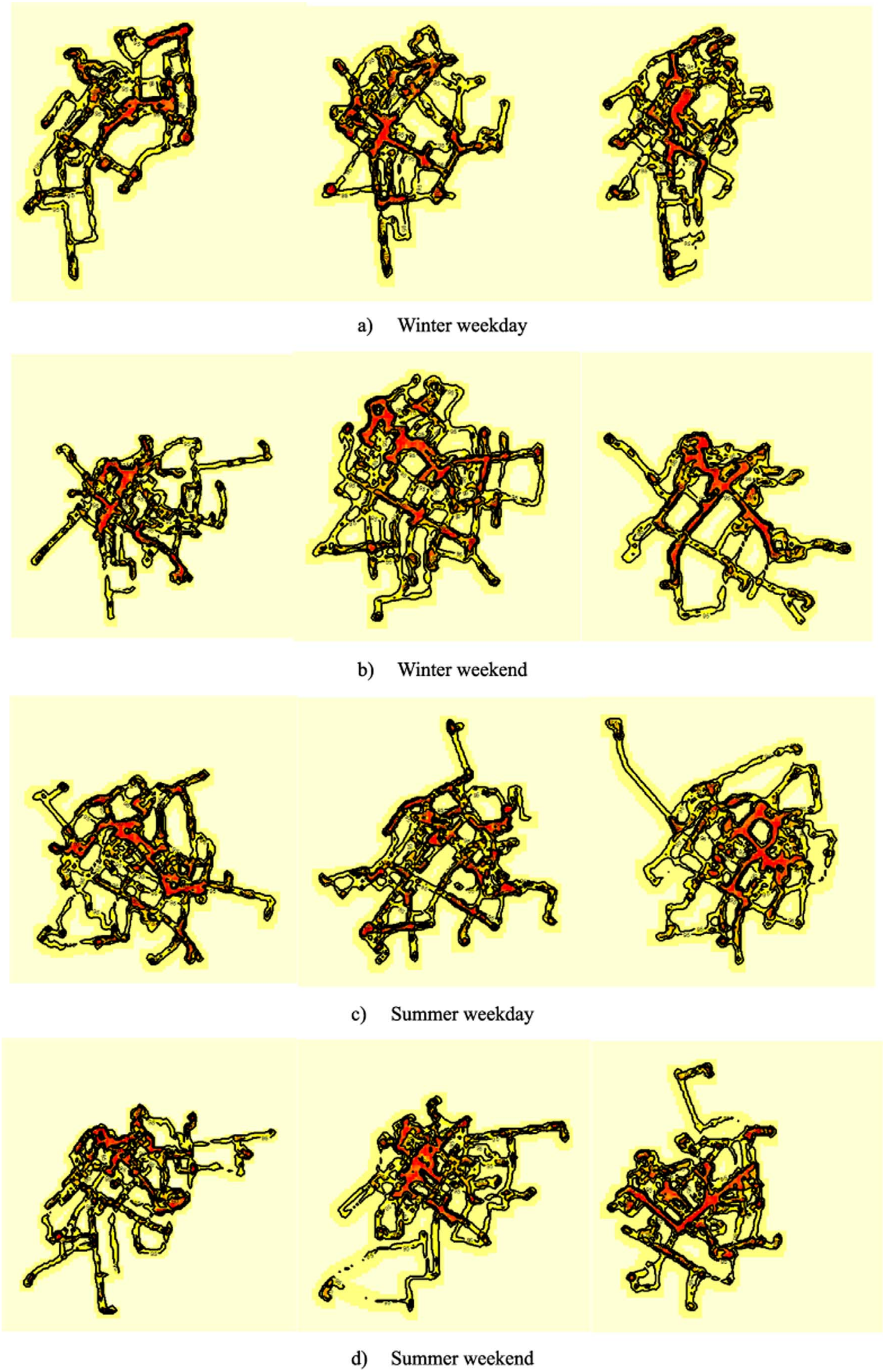
The core service range size derived from the UD (i.e. within 50% isopleth) ranged from IQR 8.25–9.92 km$^2$ with a median of 9.22 km$^2$ (Fig. 12 and Table 1) for summer weekday and IQR 7.13–8.42 km$^2$ (median 7.77 km$^2$) for winter weekday. When focusing on weekend and excluding the 50% of GPS fixes that were farthest away from the centroid, the service size ranged from IQR 7.59–9.75 km$^2$ (median 8.86 km$^2$) for summer and IQR 7.10–9.67 km$^2$ (median 8.13 km$^2$) for winter (Fig. 13 and Table 2).

The more extended service range size (area within the 95% isopleth of the UD) was estimated to range from IQR 40.68–46.11 km$^2$ with a median of 43.12 km$^2$ (Fig. 12 and Table 1) and IQR 37.27–45.36 (median 40.41 km$^2$) for summer and winter weekday, respectively. Using the MKDE method excluding the 5% farthest fixes for weekend (Fig. 13 and Table 2), the service range was estimated to be in the range IQR 36.76–45.73 km$^2$ (median 41.54 km$^2$) for summer and 35.50–45.71 km$^2$ with a median of 40.30 km$^2$ for winter.

Generally, the taxi service ranges are broader in summer than in winter. One possible reason may be that the taxi drivers may narrow down their service ranges through the freezing winter or rainy days. Moreover, the service ranges differ more significantly between summer and winter in weekday, while the seasonal difference in service range is relatively small in weekend. However, according to the comparison of Tables 1 and 2, we can hardly identify a directional difference between weekday and weekend.

The MKDE estimation results are further cross-classified by the season characteristic (summer and winter) and the type of day (weekday and weekend), which can be referred as a 2 × 2 factorial ANOVA design. The 2 × 2 factorial design, which is also known as factorial ANOVA or two-way ANOVA (Wonnacott and Wonnacott, 1997), will be used to distinguish the main effects and interaction effects on taxi service range. An interaction between Factor A and Factor B means that the effect of Factor A is different, depending on the level of Factor B. The hypothesis test results are shown in Table 3.

Since the *p*-value of 0.001 for season is less than the 0.01 significance level (within the 50% kernel), we rejected the null hypothesis that the mean service range sizes (within the 50% kernel) of different seasons are all equal, and observed a highly significant effect of season ($F = 11.013, p = 0.001$) on taxi service range. Then, a similar significant effect of season can also be concluded from F-statistic with a *p*-value under 0.05 (within the 75% kernel). Moreover, all the other *p*-values for season and day type comparison are > 0.05 significance level. It shows there is no difference in taxi service range


a) Winter weekday

b) Winter weekend

c) Summer weekday

d) Summer weekend

**Fig. 11.** Service range calculated by MKDE method for three example taxis. Three isopleth levels (50%, 75%, 95%) are presented and labeled with contour and shades of color.

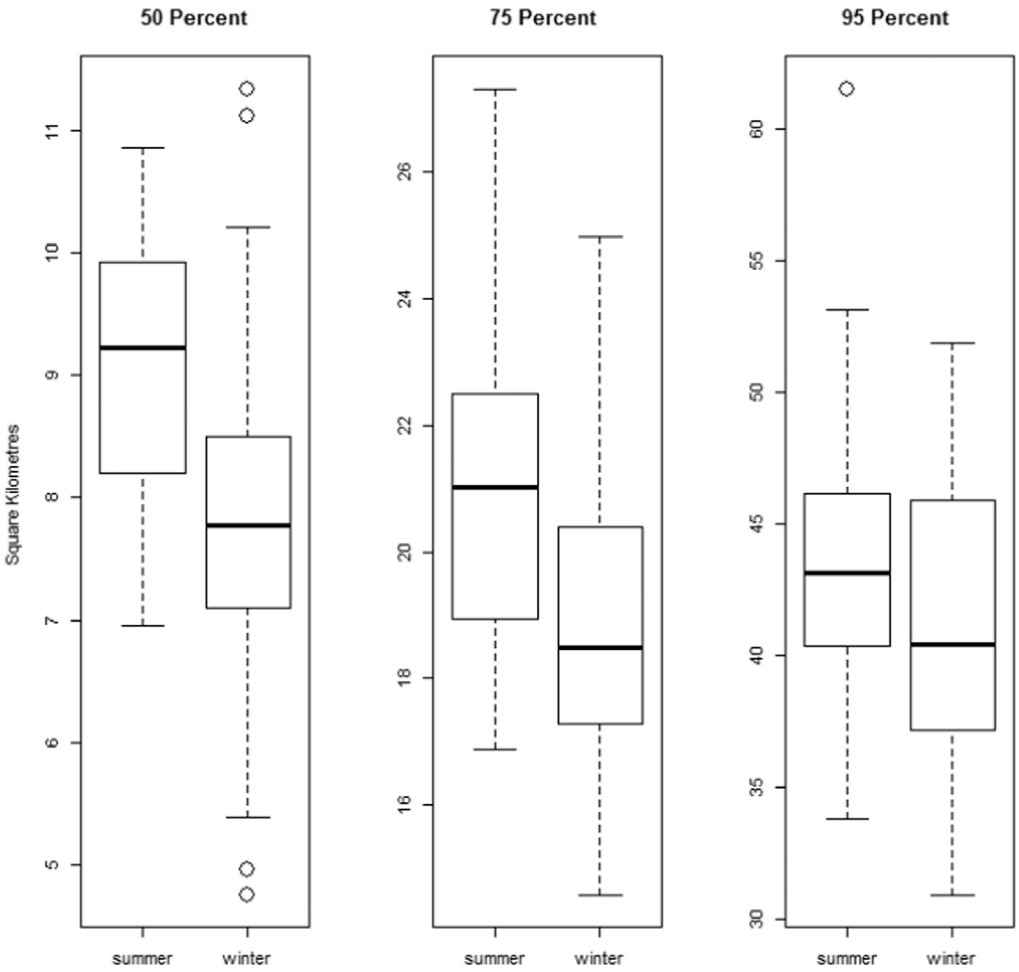


**Fig. 12.** The boxplots of the areas (in km$^2$) of the weekday seasonal ranges from left to right the 50, 75 and 95% kernel contours. Each panel depicts summer on the left, and winter on the right.

**Table 1**
Weekday home range sizes (km$^2$) of 100 taxis recruited in a GPS study in Harbin in 2014, estimated for three isopleth levels (50, 75 and 95%) by the MKDE method.

| | MKDE (Summer) | | | MKDE (Winter) | | |
|---|---|---|---|---|---|---|
| | 50% | 75% | 95% | 50% | 75% | 95% |
| Median | 9.22 | 21.02 | 43.12 | 7.77 | 18.48 | 40.41 |
| IQR[a] | 8.25–9.92 | 19.08–22.39 | 40.68–46.11 | 7.13–8.42 | 17.29–20.36 | 37.27–45.36 |
| 5–95 Percentile | 7.36–10.72 | 17.15–24.33 | 35.09–51.43 | 5.16–10.70 | 14.87–23.90 | 32.60–49.51 |

[a] Interquartile range.

based on other comparisons. Finally, the interaction $p$-values ($> 0.05$) indicate that there is a no interaction between the season and day type, i.e., different day type from different seasons have the same effect on taxi service range.

Since only season factor has been identified as the significant effect under certain conditions, we further attempted to quantify the relationship between season-related characteristics and service range size as well as service number. Temperature and precipitation are two common predictors characterizing season attributes, so daily mean temperature is chosen as the predictor, while daily mean service range size is treated as the outcome to estimate linear regression models. Moreover, since precipitation from rain and snow exhibits different measurement scales (e.g. < 10 mm/day for light rain, < 1 mm/day for light snow), the daily precipitation is encoded as a factor with 3 levels (light, moderate, heavy) for the regression models. The hypothesis test results are shown in the following tables.

In Table 4, The F-test results show that the model performs significantly better than expected ($p < 0.001$). Besides, according to the value of coefficient of determination $R^2$, the predictors explain 87% of the variance in the outcome (taxi service number). Furthermore, the $t$-tests for each of the individual coefficients provide strong evidence that both temperature and precipitation have significant effect.

In Table 5, the F-test results show that the performance of the models is quite different at various isopleth levels. The models perform significantly better at 50% and 75% isopleth levels ($p < 0.05$), while the last model can hardly exhibit predictable performance (within 95% isopleth). Besides, the values of $R^2$ illustrate that the predictors explain 55% and 43% of the variance in the outcome (daily mean service range) within 50% and 75% isopleth respectively, while only 25% variance in the outcome variable can be accounted for by the predictor within 95% isopleth. Furthermore, the slope value of *Temperature* shows that a 1°c increase would cause 0.02 and 0.04 km$^2$ increase in taxi service range size within 50% and 75% isopleth respectively. However, the $t$-tests for each of the individual coefficients provide strong evidence that the

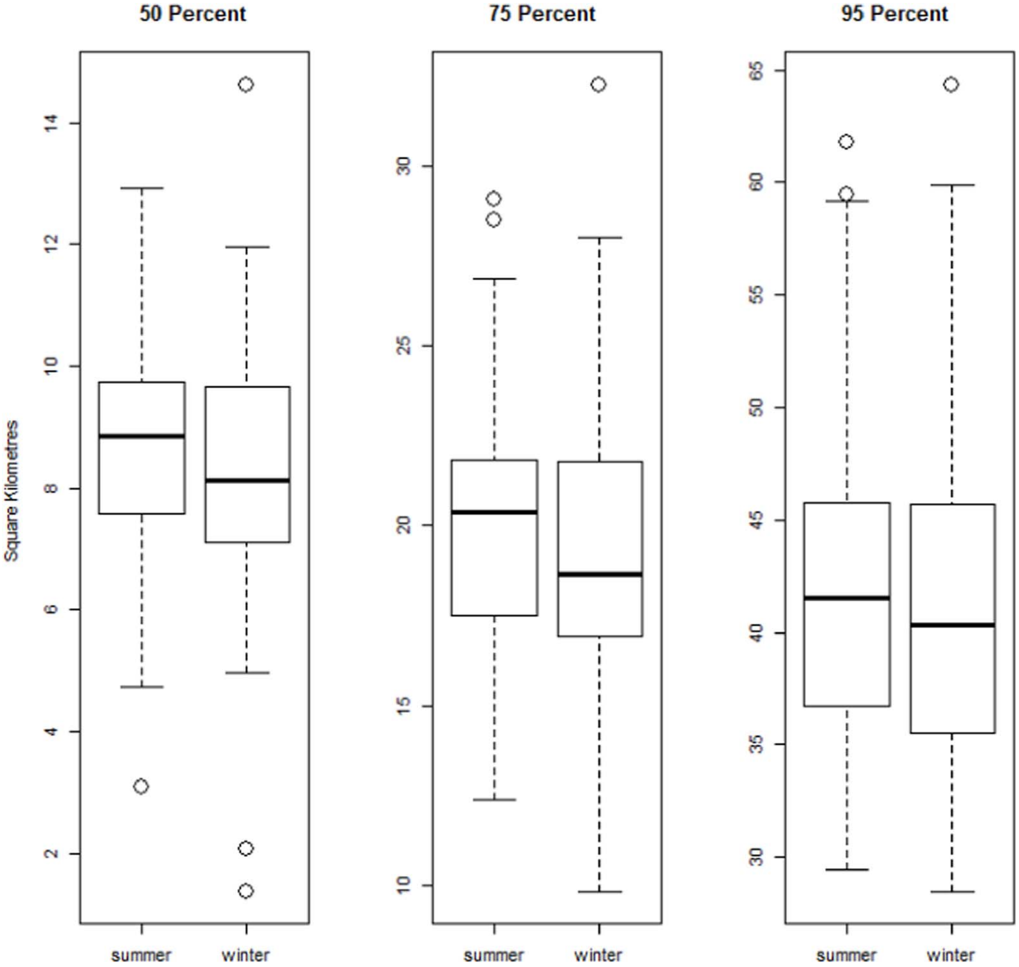


**Fig. 13.** The boxplots of the areas (in km$^2$) of the weekend seasonal ranges from left to right the 50, 75 and 95% kernel contours. Each panel depicts summer on the left, and winter on the right.

**Table 2**
Weekend home range sizes (km$^2$) of 100 taxis recruited in a GPS study in Harbin in 2014, estimated for three isopleth levels (50, 75 and 95%) by the MKDE method.

| | MKDE (Summer) | | | MKDE (Winter) | | |
|---|---|---|---|---|---|---|
| | 50% | 75% | 95% | 50% | 75% | 95% |
| Median | 8.86 | 20.38 | 41.54 | 8.13 | 18.64 | 40.30 |
| IQR[a] | 7.59–9.75 | 17.49-21.83 | 36.76–45.73 | 7.10–9.67 | 16.94–21.79 | 35.50–45.71 |
| 5–95 Percentile | 6.29–11.50 | 15.53–27.04 | 32.87–59.21 | 4.67–11.80 | 12.87–26.88 | 31.42–59.27 |

[a] Interquartile range.

**Table 3**
Effect analysis using factorial ANOVA.

| Effect | 50% kernel | | 75% kernel | | 95% kernel | |
|---|---|---|---|---|---|---|
| | F-statistic | *P*-value | F-statistic | *P*-value | F-statistic | *P*-value |
| Season | 11.013 | 0.001 | 6.587 | 0.012 | 2.463 | 0.119 |
| Day | 0.002 | 0.967 | 0.157 | 0.692 | 0.403 | 0.527 |
| Interaction | 0.761 | 0.385 | 0.660 | 0.418 | 0.509 | 0.477 |

*Precipitation* has no significant effect. A variable selection procedure based on the Akaike information criterion (AIC) can also validate that only *Temperature* should be chosen as the predictor for regression model.

In fact, it's hard to distinguish the impacts of precipitation on traffic between summer and winter in Harbin, and the reason is probably that the slippery road surfaces due to accumulated snow can impair travel in winter, even if there is no precipitation or only precipitation occurred prior day, which is totally different from rainfall conditions in summer.

**Table 4**
Hypothesis test for the correlation significance between predictors and taxi service number.

| | Estimation | T-statistic | *p*-value |
|---|---|---|---|
| Temperature | − 6.341 | − 6.171 | 6.990e-05 |
| Precipitation | − 45.767 | − 3.126 | 0.964e-02 |
| | R$^2$ | F-statistic | *p*-value |
| Model | 0.876 | 46.880 | 4.137e-06 |

Besides, since relative few rainy and snowy days are included to build the sample dataset due to the dry-climate characteristics in

**Table 5**
Hypothesis test for the correlation significance between predictors and service range (Intercept omitted).

| Service range level | | 50% isopleth | 75% isopleth | 95% isopleth |
|---|---|---|---|---|
| Temperature | Estimation | 0.022 | 0.036 | 0.052 |
| | T-statistic | 3.026 | 2.585 | 1.866 |
| | *p*-Value | 0.007 | 0.0178 | 0.077 |
| Precipitation | Estimation | − 0.117 | − 0.361 | − 0.771 |
| | T-statistic | − 0.424 | − 0.701 | − 0.746 |
| | *p*-Value | 0.676 | 0.491 | 0.464 |
| Model | $R^2$ | 0.550 | 0.436 | 0.253 |
| | F-statistic | 12.230 | 7.713 | 3.377 |
| | *p*-Value | 0.006 | 0.019 | 0.096 |

Harbin, it may also intensify the instability of experimental results.

## 4. Conclusions

Taxi GPS traces are valuable resources to investigate taxi drivers' service behaviors and passenger dynamics. Despite its great potential, the utility of taxi GPS data has yet to be fully exploited. For example, fewer studies assess potential season change effects on taxi service coverage. In recognition of this limitation, we endeavor to reveal spatial-temporal dynamics of collective taxi mobility at intra-urban scale from a novel travel perspective, which considers not only the trips to satisfy passenger needs, but also the service ranges of taxis to satisfy the geographical travel needs of the drivers on multi-time scale. The main contributions of this paper can be summarized as follows:

- We examine the global spatial dependence pattern between PUPs and DOPs using multivariate point pattern analysis, and draw significant positive spatial relationships. Then we further investigate the dependencies based on spatial distance and intensity, and present an intuitive estimation of the dependency function.
- We identify POIs from DOPs using emerging hot spot detection technique, and investigate the local mobility patterns around POIs over weekday and weekend, suggesting some common travel patterns of taxis. Then, we also explore the relationship between taxi speed and distances to nearest POIs, a positive relationship can be directly observed.
- We follow and extend the existing work to examine the statistical regularities of trip distances, suggesting a log-normal distribution. Besides, we also validate and quantify the impacts posed by airport trips on the distance distributions. The results show that the airport trips could play an important role in shaping the distance distribution, especially in cities like Harbin, which lacks other airport ground access modes (e.g., airport express).
- We estimate the service range sizes of taxis based on MKDE method within three levels between summer/weekday and winter/weekend, and we also attempt to distinguish the seasonal effects and day-type effects on taxi service ranges, the results show that only season factor can exert a significant effect within 50% isopleth level. A further hypothesis test validates the daily mean temperature can be a good predictor for taxi service range under the same MKDE estimation condition.
- Findings of this study can help better understand the spatial–temporal demand coverage, which will be essential for designing differential traffic management strategies to effectively relieve traffic problems within urban areas. Some suggestions for public transport policy can be further drawn as follows:
- Dynamic price policy: Dynamic price model should be proposed to simulated taxi drivers' enthusiasm to balance the supply-demand. Unlike private cars, the taxi service number fluctuates slightly, so surge pricing at peaking hours contributes little to increase on-the-road supply. However, as shown in Section 3.4, taxis are sensitive to seasonal variation, for example, taxi service ranges are broader in summer than in winter. Therefore, a distance fee in winter may help to promote drivers' willingness to pick up long-distance taxi riders.
- Traffic restriction policy: As stated in Section 3.1, there is a significant drop in the average speed around POIs at the evening rush hour in weekday, and the positive relationship between taxi speed and distances also validates the POIs' attraction. Comparing with the traffic flow analysis and prediction (Tang et al., 2017; Yan et al., 2017; Zou et al., 2017), location-based services (LBS) from taxi GPS trajectories can provide more abundant spatio-temporal information of traveling behavior. Therefore, it will help to promote public transport efficiency by restricting the use of private vehicles around POIs at peaking hours.
- City planning for POIs: It is found that taxis tend to cluster around POIs, and thus choose POIs as the centers of their service ranges. However, there are relatively highly concentrated POIs in Harbin, which may be the causes for spatial supply-demand imbalance. As a result, city planning for new POIs by equalizing their distribution in space will help in balancing taxi services throughout the city. The method proposed in Section 3.1 can also be used to evaluate the plan implementation results.

Further studies may expand the data source to include private car, bus, and metro trips. This combination of diverse data could also shed light on comprehensive patterns of urban spatial interaction, providing a multi-faceted picture of urban dynamics. In addition, the results of this study may not be applicable for holiday, as travel demands, trip purposes, and travel decisions could be very different. The comparison analysis of travel behaviors between normal weekends and holidays could be considered in our future research, and we will also validate whether our findings hold in other cities.

## Acknowledgements

This research was supported by the National Natural Science Foundation of China (Grant No. 71403068, 51138003, 51329801 and 51578198), the Hi-Tech Research and Development Program of China (Grant No. 2014BAG03B04).

## References

Benhamou, S., 2011. Dynamic approach to space and habitat use based on biased random bridges. PLoS One 6 (1), e14592.

Benhamou, S., Cornélis, D., 2010. Incorporating movement behavior and barriers to improve kernel home range space use estimates. J. Wildl. Manag. 74 (6), 1353–1360.

Bhattacharya, T., Kulik, L., Bailey, J., 2015. Automatically recognizing places of interest from unreliable GPS data using spatio-temporal density estimation and line intersections. Pervasive Mob. Comput. 19, 86–107.

Braham, M., Miller, T., Duerr, A.E., Lanzone, M., Fesnock, A., LaPre, L., Driscoll, D., Katzner, T., 2015. Home in the heat: dramatic seasonal variation in home range of desert golden eagles informs management for renewable energy development. Biol. Conserv. 186, 225–232.

Castro, P.S., Zhang, D., Chen, C., Li, S., Pan, G., 2013. From taxi GPS traces to social and community dynamics: a survey. ACM Comput. Surv. 46 (2) (article 17).

Clauset, A., Cosma, R., Mark, E., 2009. Power-law distributions in empirical data. SIAM Rev. 51 (4), 661–703.

Dürr, S., Ward, M.P., 2014. Roaming behaviour and home range estimation of domestic dogs in Aboriginal and Torres Strait Islander communities in northern Australia using four different methods. Prev. Vet. Med. 117, 340–357.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events, Stochastic Modelling and Applied Probability. Springer, Berlin, Heidelberg.

Gelfand, A.E., Diggle, P.J., Guttorp, P., Fuentes, M., 2010. Handbook of Spatial Statistics. CRC Press, Boca Raton.

Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24 (3), 189–206.

Getz, W., Fortmann-Roe, S., Cross, P.C., Lyons, A.J., Ryan, S.J., Wilmers, C.C., 2007. LoCoH: nonparameteric kernel methods for constructing home ranges and utilization distributions. PLoS One 2 (2), e207.

Horne, J.S., Garton, E.O., Krone, S.M., Lewis, J.S., 2007. Analyzing animal movements using Brownian bridges. Ecology 88, 2354–2363.

Illian, J., Penttinen, A., Stoyan, H., Stoyan, D., 2008. Statistical Analysis and Modelling of Spatial Point Patterns. John Wiley and Sons, Chichester, United Kingdom.

Jiang, B., Yin, J., Zhao, S., 2009. Characterizing the human mobility pattern in a large street network. Phys. Rev. E 80, 021136.

Jones, M.C., 1991. Kernel density estimation for length-biased data. Biometrika 78 (3), 511–519.

Keating, K.A., Cherry, S., 2009. Modeling utilization distributions in space and time. Ecology 90, 1971–1980.

Kie, J.G., Matthiopoulos, J., Fieberg, J., Powell, R.A., Cagnacci, F., Mitchell, M.S., Gaillard, J.M., Moorcroft, P.R., 2010. The home-range concept: are traditional estimators still relevant with modern telemetry technology? Philos. Trans. R. Soc. B 365, 2221–2231.

Liang, X., Zheng, X., Lu, W., Zhu, T., Xu, K., 2012. The scaling of human mobility by taxis is exponential. Phys. A 391, 2135–2144.

Liu, L., Andris, C., Ratti, C., 2010. Uncovering cabdrivers' behavior patterns from their digital traces. Comput. Environ. Urban. Syst. 34 (6), 541–548.

Liu, X., Gong, L., Gong, Y., Liu, Y., 2015. Revealing travel patterns and city structure with taxi trip data. J. Transp. Geogr. 43, 78–90.

Liu, Y., Kang, C., Gao, S., Xiao, Y., Tian, Y., 2012a. Understanding characteristics of intra-urban trips using taxi trajectory data. J. Geogr. Syst. 14 (4), 463–483.

Liu, Y., Wang, F., Xiao, Y., Gao, S., 2012b. Urban land uses and traffic 'source–sink areas': evidence from GPS-enabled taxi data in Shanghai. Landsc. Urban Plan. 106 (1), 73–87.

Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., Mascolo, C., 2012. A tale of many cities: universal patterns in human urban mobility. PLoS One 7 (5), e37027.

Pan, G., Qi, G., Wu, Z., Zhang, D., Li, S., 2013. Land-use classification using taxi GPS traces. IEEE Trans. Intell. Transp. Syst. 14 (1), 113–123.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall/CRC, Boca Raton, Florida, USA.

Tang, J., Jiang, H., Li, Z., Li, M., Liu, F., Wang, Y., 2016b. A two-layer model for taxi customer searching behaviors using GPS trajectory data. IEEE Trans. Intell. Transp. Syst. 17 (11), 3318–3324.

Tang, J., Liu, F., Wang, Y., Wang, H., 2015. Uncovering urban human mobility from large scale taxi GPS data. Phys. A 438 (15), 140–153.

Tang, J., Liu, F., Zou, Y., Zhang, W., Wang, Y., 2017. An improved fuzzy neural network for traffic speed prediction considering periodic characteristic. IEEE Trans. Intell. Transp. Syst. 1–11, 99. http://dx.doi.org/10.1109/TITS.2016.2643005.

Tang, J., Zhang, S., Liu, F., Zhang, W., Wang, Y., 2016a. Statistical properties of urban mobility from location-based travel networks. Phys. A 461 (16), 694–707.

Veloso, M., Phithakkitnukoon, S., Bento, C., 2011a. Sensing urban mobility with taxi flow. In: Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks. 41–44.

Veloso, M., Phithakkitnukoon, S., Bento, C., 2011b. Urban mobility study using taxi traces. In: Proceedings of the 2011 International Workshop on Trajectory Data Mining and Analysis. 23–30.

Vuong, H.Q., 1989. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica 57, 307–333.

Wang, W., Pan, L., Yuan, N., Zhang, S., Liu, D., 2015. A comparative analysis of intra-city human mobility by taxi. Phys. A 420, 134–147.

Wong, R.C.P., Szeto, W.Y., Wong, S.C., 2014b. A cell-based logit-opportunity taxi customer-search model. Transp. Res. C 48, 84–96.

Wong, R.C.P., Szeto, W.Y., Wong, S.C., Yang, H., 2014a. Modeling multi-period customer-searching behavior of taxi drivers. Transportmetrica B 2 (1), 40–49.

Wonnacott, T.H., Wonnacott, R., 1997. Introductory Statistics. John Wiley & Sons, Chichester, UK.

Worton, J.B., 1989. Kernel methods for estimating the utilization distribution in home range studies. Ecology 70, 164–168.

Wu, L., Zhi, Y., Sui, Z., Liu, Y., 2014. Intra-urban human mobility and activity transition: evidence from social media check-in data. PLoS One 9 (5), e97010.

Yan, Y., Zhang, S., Tang, J., Wang, X., 2017. Understanding characteristics in multivariate traffic flow time series from complex network structure. Phys. A 477, 149–160.

Zhu, X., Guo, D., 2014. Mapping large spatial flow data with hierarchical clustering. Trans. GIS 18 (3), 421–435.

Zou, Y., Yang, H., Zhang, Y., Tang, J., Zhang, W., 2017. Mixture modeling of freeway speed and headway data using multivariate skew-t distributions. Transportmetrica A: Transport Sci. http://dx.doi.org/10.1080/23249935.2017.1318973.