

Article

---

# Optimizing Cruising Routes for Taxi Drivers Using a Spatio-Temporal Trajectory Model

---

Liang Wu, Sheng Hu, Li Yin, Yazhou Wang, Zhanlong Chen, Mingqiang Guo, Hao Chen and Zhong Xie

## Special Issue

Geospatial Big Data and Urban Studies



Edited by

Prof. Dr. Maria Antonia Brovelli, Dr. Hussein Abdulmuttalib, Dr. Peter Mooney and Prof. Dr. Chris Pettit



## Article

# Optimizing Cruising Routes for Taxi Drivers Using a Spatio-Temporal Trajectory Model

Liang Wu <sup>1,2</sup> , Sheng Hu <sup>1</sup> , Li Yin <sup>3</sup>, Yazhou Wang <sup>2</sup>, Zhanlong Chen <sup>1</sup>, Mingqiang Guo <sup>1</sup>, Hao Chen <sup>2</sup> and Zhong Xie <sup>1,2,\*</sup>

<sup>1</sup> School of Information Engineering, China University of Geosciences, Wuhan 430074, China; wuliang@cug.edu.cn (L.W.); gishusheng@gmail.com (S.H.); chenzhanlong@mapgis.com (Z.C.); guomingqiang@mapgis.com (M.G.)

<sup>2</sup> National Engineering Research Center for GIS, Wuhan 430074, China; wangyazhou105@gmail.com (Y.W.); chen hao.ch@mapgis.com (H.C.)

<sup>3</sup> Department of Urban and Regional Planning, State University of New York, Buffalo, NY 14214, USA; liyin@buffalo.edu

\* Correspondence: xiezhong@cug.edu.cn; Tel.: +86-027-6788-3728

Received: 26 September 2017; Accepted: 13 November 2017; Published: 19 November 2017

**Abstract:** Much of the taxi route-planning literature has focused on driver strategies for finding passengers and determining the hot spot pick-up locations using historical global positioning system (GPS) trajectories of taxis based on driver experience, distance from the passenger drop-off location to the next passenger pick-up location and the waiting times at recommended locations for the next passenger. The present work, however, considers the average taxi travel speed mined from historical taxi GPS trajectory data and the allocation of cruising routes to more than one taxi driver in a small-scale region to neighboring pick-up locations. A spatio-temporal trajectory model with load balancing allocations is presented to not only explore pick-up/drop-off information but also provide taxi drivers with cruising routes to the recommended pick-up locations. In simulation experiments, our study shows that taxi drivers using cruising routes recommended by our spatio-temporal trajectory model can significantly reduce the average waiting time and travel less distance to quickly find their next passengers, and the load balancing strategy significantly alleviates road loads. These objective measures can help us better understand spatio-temporal traffic patterns and guide taxi navigation.

**Keywords:** trajectory data mining; taxi planning; spatio-temporal trajectory model; load balance

## 1. Introduction

With the development of urbanization, people place increasing demands on urban traffic and transportation. Taxis, with respect to their flexibility and convenience, have become one of the most popular modes of urban transportation [1]. Taxis are an indispensable component of the urban transportation system and meet the travel demands of a great number of people. In many cities, however, people expecting to take a taxi are used to stopping a vacant taxi via “roadside beckoning”. Therefore, the locations of taxi drivers picking up their passengers are highly random [2]. Both the people seeking to take a taxi and taxi drivers have insufficient location information on the best location for a pick-up, which causes the phenomenon that a taxi driver has difficulty finding passengers while people find it difficult to locate a vacant taxi. Vacant taxis in urban road networks not only generate superfluous energy (i.e., oil and gas) consumption but also occupy road space, causing traffic flow congestion and air pollution problems [3,4].

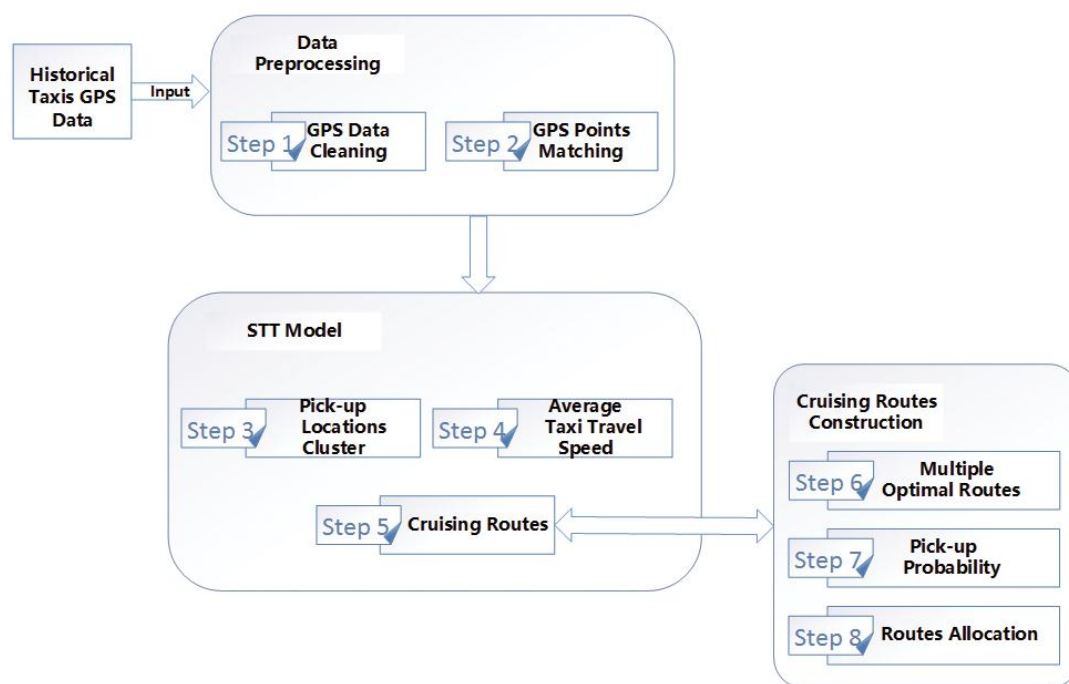
In recent years, taxis have been widely equipped with global positioning system (GPS) sensors, which are mobile devices that can monitor taxi locations and statuses at regular intervals. Therefore, a large number of GPS trajectories with spatio-temporal information have been collected. Such a large amount of tracking data provides an unprecedented opportunity to discover the implicit information and understand taxi drivers' driving behaviors, human mobility, and the dynamics of street networks [5,6]. The mining of taxi GPS trajectories has received increasing attention from the data mining, intelligent transportation, movement patterns and ubiquitous computing communities [2,7,8]. Several studies have considered the historical GPS trajectories of taxis, such as understanding human mobility [9–12], estimating traffic emissions [13–16], planning routes [17–20], and formulating taxi/passenger search strategies [3,6,21–23].

Unlike other transportation modes such as buses and the subway, which only operate along fixed lines, taxis are more flexible because drivers are able to plan their pick-up destinations and cruising routes. Historical pick-up/drop-off information can be useful for drivers to find their next passengers. Several studies have considered taxi route-planning with different approaches toward the historical GPS trajectories of taxis [1–3,6,8,21–28].

On the one hand, many studies focus on drivers' strategies for finding passengers [8,21,23,24,27,28]. Obviously, it is helpful to increase cruising efficiency and reduce unnecessary cruising time by examining the cruising patterns of experienced drivers. Zhang et al. described both the efficient and inefficient taxi service strategies based on a large-scale GPS historical database [8]. Hu et al. discussed the characteristics of urban taxi drivers' activity distributions at different temporal and spatial levels to identify taxi drivers' operation patterns and searching-behavior patterns to help taxi drivers reduce searching time [23]. Liu et al. presented a framework including a series of models to study how a taxi driver gathers and learns information in an uncertain environment; they found that drivers not only learn from their own experiences but also communicate with other drivers [28].

On the other hand, many studies tend to concentrate on determining the hot spots of pick-up locations [1,2,6,29]. Hot spots of pick-up locations exist on road networks at different times in a single day [3,22]. Experienced drivers usually know where they are more likely to quickly pick up their next passengers after dropping passengers off instead of remaining vacant on the road network. Moreira et al. presented a novel application using time-series forecasting techniques to predict the taxi-passenger demand at taxi stands at 30-minute intervals to improve taxi-driver mobility intelligence [2]. Liu et al. discussed the crowdedness of moving objects and explored hot spots from the crowdedness using historical GPS trajectories [9]. Hwang et al. proposed a grid-based clustering approach considering four factors, including waiting time, distance, average revenue and the probability of finding passengers when clustering, to recommend the next pick-up locations to taxi drivers [30].

In this paper, we consider the clustering of pick-up locations to explore the hot spots of recommended locations. In addition, we consider computing the probability of picking up the next passengers (along the routes or at pick-up locations) and how to allocate the cruising routes to more than one taxi driver in a small-scale region of neighboring pick-up locations. We integrate the abovementioned factors into a spatio-temporal trajectory (STT) model. The framework of this paper is illustrated in Figure 1.



**Figure 1.** The System Framework of this Work. STT Model = Spatial-temporal Trajectories Model.

## 2. Materials and Methods

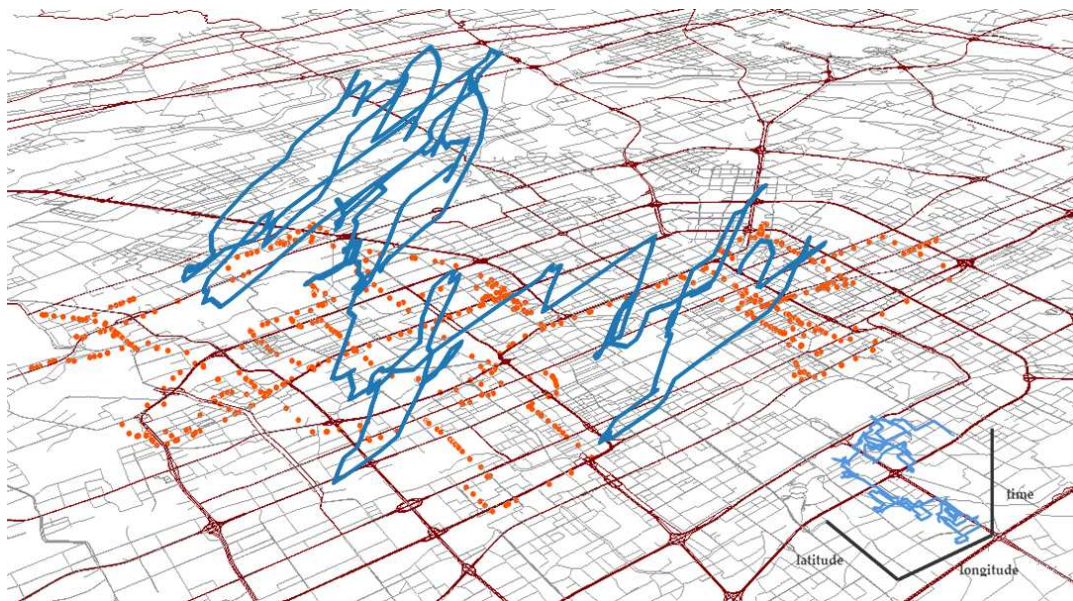
### 2.1. Data Sources

All the available GPS point data and road network data are obtained from the database of Beijing City government data resources [31], which provide access to trajectory information in Beijing. Our experimental dataset contains trajectories recorded by over 12,000 taxis in the 30 day period of November 2011 in Beijing, China. A trajectory of a taxi consists of a set of GPS points, each of which contains pick-up or drop-off information. A GPS data point consists of 7 properties: taxi ID, timestamp, latitude, longitude, speed, driving direction (i.e., the degree of deviation from north) and taxi-occupied tag (the value of this variable is binomial, 0 or 1, and indicates whether a taxi is vacant or occupied, respectively). Data columns are shown in Table 1. A GPS data point is collected every 1 min; thus, there are more than 1000 million entries collected in our dataset.

**Table 1.** Data columns of a global positioning system (GPS) point data.

TaxiID	Timestamp	Longitude (°)	Latitude (°)	Speed (km/h)	Directions (°)	Occupied Tag
001140	20121101001504	117.109	40.153	75	136	1

In our work, a taxi trajectory is a series of GPS points logged for a working taxi (Figure 2). Specifically, a taxi driver wanders the road network to search for passengers (vacant status). Then, the driver picks up passengers and drives them to their intended locations (occupied status). After passengers get out of the taxi, the driver again searches for new passengers (vacant status). In this process, the locations where passengers get into the taxi are defined as the pick-up locations, while the locations where passengers get out of the taxi are defined as the drop-off locations. In our experiments, we can obtain the complex taxi trajectories from the large-scale GPS point dataset. Furthermore, we are capable of capturing spatio-temporal traffic patterns by analyzing the pick-up and drop-off information and taxis' status change patterns.



**Figure 2.** The Continuous Trajectory of a Taxi in one Day in 3D Space. In this figure, the vertical axis represents the temporal progression of the taxi's daily movement, while the horizontal plane represents the road network in Beijing.

## 2.2. Data Preprocessing

The initial track data were disordered and not suitable for further analysis; thus, data preprocessing was conducted. This work consists of two steps: GPS data cleaning and GPS point matching. As GPS devices are difficult to make completely precise, repetitive or deflected entries might exist in our dataset. To avoid potential confounding impacts, the adverse GPS points, such as out-of-study-range points, time-repeated points and overspeed points (speed higher than 90 km/h in Beijing is considered as overspeed) were removed. In addition, due to GPS measurement errors and road geometric errors in digital maps, the GPS locations of taxis might not appear on road network links. Map-matching is a critical procedure to precisely match GPS points to network links [32]. Chen et al. [7] proposed an efficient and high-performance multi-criteria dynamic programming map-matching (MDP-MM) algorithm with a multi-criteria dynamic programming technique, wherein the objective is to map large-scale low-frequency floating car data. In our paper, considering the limitations of the traditional map-matching algorithm [33–36] for the large scale of taxi GPS data in our experiment, the same MDP-MM algorithm is employed.

## 3. Spatio-Temporal Trajectory Model

In this study, a spatio-temporal trajectory model is proposed to recommend to taxi drivers locations where they can quickly pick up passengers and suitable routes to these locations. Three factors have been considered in the STT model: the pick-up locations cluster, average taxi travel speed and cruising routes.

### 3.1. Cluster of Pick-Up Locations

When the passengers get out of the taxi or a taxi wanders on the road network, the taxi is in vacant status and a taxi driver must find the next passengers. Experienced drivers tend to go to locations around where they are more likely to pick up passengers. Especially during the off-peak periods, experienced drivers usually wait at locations such as shopping centers or bus and train stations.

Pick-up locations usually represent the hot spots where there are high taxi travel demands [2,21,28,32]. Thus, we can unite the historical pick-up information to explore high-demand areas and analyze the

distribution patterns of pick-up hot spots. Clustering is a feasible and meaningful approach to identify hot spots of moving vehicles in an urban area [9]. In particular, using a density-based clustering method, we can discover clusters with arbitrary shapes and avoid the adverse impacts of noise and unusual points. Moreover, density-based clustering has a strong ability to address geographical characteristics [37]. In this work, pick-up locations were explored by using the density-based spatial clustering of applications with noise (DBSCAN) algorithm [38], which is one of the most common clustering algorithms cited in the scientific literature [39–42]. Given a set of pick-up points in a study area, it groups together points that are closely packed together, marking points that lie alone in low-density regions as outliers. DBSCAN requires two parameters: epsilon (Eps) and minimum points (MinPts). In addition, weekdays have different patterns from weekends [8]. In our experiment, the parameters were different for weekdays and weekends (Table 2).

**Table 2.** Cluster parameters of weekdays and weekends.

	Eps (m)	MinPts
Weekdays	130	20
Weekends	140	25

### 3.2. Average Taxi Travel Speed and Travel Time

Taxis operate throughout the day and cover most road segments in the road network; therefore, we can derive the average taxi travel speed and travel time from taxi trajectory data [5]. The average travel speed was computed by averaging the travel speeds of all taxis passing through a road segment  $r$  within each specific time period  $t$ , as shown in Equation (1) [20].

$$\overline{V}_r(t) = \frac{\sum_{i=1}^m \sum_{j=1}^{n_k} V_{(i,j)}}{N_r(t)}, \quad (1)$$

where  $m$  is the total number of taxis,  $n_k$  records the total number of points of a taxi  $k$  on road segment  $r$  within time period  $t$ ,  $V_{(i,j)}$  denotes the  $j^{\text{th}}$  instantaneous travel speed, and  $n_r(t)$  denotes the total number of points of all taxis and is equal to  $\sum_{i=1}^m n_i$ . Then, the average travel time on road segment  $r$  during time period  $t$  can be computed as follows:

$$\overline{T}_r(t) = L_r / \overline{V}_r(t), \quad (2)$$

where  $L_r$  is the length of road segment  $r$ .

### 3.3. Cruising Routes

Based on the clustered pick-up locations and average taxi travel speed, we are then able to explore the cruising routes to pick-up locations. We can obtain multiplex (at least one) routes from the current point to any pick-up location within the road network. In the literature, most studies have considered the best route choice from start to pick-up locations with different approaches. In this paper, we consider how to allocate the cruising routes to more than one taxi driver in a small-scale region to the neighboring pick-up locations. For example, using our method, taxi driver  $A$  determined one route from the current location  $L_a$  to pick-up location  $P$ , while taxi driver  $B$ , not as far from  $A$ , might recommend the same route to  $L_a$ . If all taxi drivers in a small-scale region follow the same drive route, the road network may become overloaded, which would increase congestion. In the STT model, we first consider exploring multiple optimal routes using the  $k$  shortest path routing algorithm. Then, the pick-up probability of each selected route is considered. Finally, based on the cruising routes and their pick-up probabilities, we utilize load balancing technology to allocate the cruising routes to taxi drivers. The details of these steps are described in the following sections.

The  $k$  shortest path routing algorithm is an extension algorithm of the shortest path routing algorithm [43], where more than one route between two locations will be obtained in the road network. The algorithm not only finds the shortest path but also finds the  $k - 1$  of other paths in increasing order of cost [44,45]. Parameter  $k$  denotes the number of shortest paths to find.

Our STT model is designed to provide the pick-up locations service for taxi drivers and the cruising routes to these locations. However, determining how to define a “good” cruising route is an important issue. In this work, we computed the probability of picking up the next passengers using the method proposed by Yuan et al. [3]. The reason for using this method is that it considers how to support a high probability (during the routes or at the recommended location) of picking up passengers, a short waiting time, a short queue length at the pick-up location and a long distance of the next trip.

A cruising route is dependent on a certain pick-up location  $P$  and a route  $R$  (route  $R$  would be divided into a number of connected road segments, i.e.,  $R = \{r_1, r_2, r_3, \dots, r_n\}$ , where  $n$  denotes the total number of road segments) to  $P$ . Taxi drivers may pick up passengers on  $R$  or at  $P$ ; however, the worst situation is when the taxi driver fails to pick up passengers after waiting at  $P$  for a time  $t_p$  ( $t_p$  would be divided into a sequence of time segments with interval  $\tau$ , i.e.,  $t_p = m\tau$ , where  $m$  denotes the total number of time segments). Let  $E$  be the event that the driver succeeds in picking up a passenger if he/she selects route  $R$ . Then,

$$\Pr(E) = 1 - \prod_{i=1}^n (1 - p(r_i)) \times \prod_{j=1}^m (1 - p(p_j)), \quad (3)$$

where  $\Pr(E)$  is the probability of  $E$ ,  $p(r_i)$  denotes the probability of picking up passengers on road segment  $r_i$ , and  $p(p_j)$  denotes the probability of picking up passengers at  $P$  after waiting for  $j\tau$  time.

In addition, the conditional expectations of the duration (denoted by  $T$ ) and expected distance (denoted by  $D$ ) from current time  $t_0$  to the beginning of the next trip are computed for the purpose of making comparisons between the recommended routes and historical routes. Based on Bayes rules,

$$E(T|E) = \frac{\sum_{i=1}^n t_{i+1} \Pr(E_i) + t_{n+1} \times (1 - \prod_{j=1}^m (1 - p(p_j))) + \sum_{j=1}^m j\tau \Pr(E_{j+n})}{\Pr(E)}, \quad (4)$$

where  $E(T|E)$  is the conditional expectations of the duration, and  $t_i$  denotes the driving time to arrive at road segment  $t_i$ .  $\Pr(E_i)$  denotes the probability of the event that a driver succeeds in picking up

a passenger at  $r_i$ , and  $\Pr(E_i) = \begin{cases} p(r_1), & i = 1 \\ p(r_i) \times \prod_{k=1}^{i-1} (1 - p(r_k)), & i = 2, 3, \dots, n \end{cases}$ .

The conditional expectations of expected distance are as follows:

$$E(D|E) = \frac{\sum_{i=1}^n d_{i+1} \Pr(E_i) + d_{n+1} \times (1 - \prod_{j=1}^m (1 - p(p_j)))}{\Pr(E)}, \quad (5)$$

where  $d_i$  denotes the driving distance to arrive at the road segment  $r_i$ .

Load balancing is a critical method to improve the distribution of workloads across multiple computing resources in computing fields [46]. This method aims to optimize resource use, maximize throughput, and avoid the overloading of any single resource. One of the most commonly used applications of load balancing is to provide a single Internet service from multiple servers [47]. In the present study, the load balancing method was utilized to allocate the cruising routes to more than one taxi driver in a small-scale region to the neighboring pick-up locations. To be specific, the weighted round-robin scheduling algorithm [48], one of the most commonly used load balancing algorithms [49], was adopted. In this method, the servers and connections are two fundamental elements. Each server should be assigned a weight representing its processing capacity. Servers with higher weights

receive new connections before those with smaller weights. In this work, we defined vacant taxis as connections, cruising routes as servers and the probability of picking up the next passengers as the processing capacity of the servers. The pseudocode is provided in Algorithm 1.

---

**Algorithm 1:** The Weighted Round-Robin Scheduling Algorithm

---

**Inputs:** Candidate cruising routes set as  $R = \{R_0, R_1, \dots, R_{n-1}\}$ ;  $W(R_i)$  indicates the probability of picking up a new passenger at  $R_i$ ;  $i$  indicates the route selected last time, and  $i$  is initialized with  $-1$ ;  $cw$  is the current weight in scheduling and is initialized with  $0$ ;  $\max(R)$  is the maximum weight of all the routes in  $R$ ;  $\gcd(R)$  is the greatest common divisor of all route weights in  $R$ .

**Output:** The distributed server  $S_i$

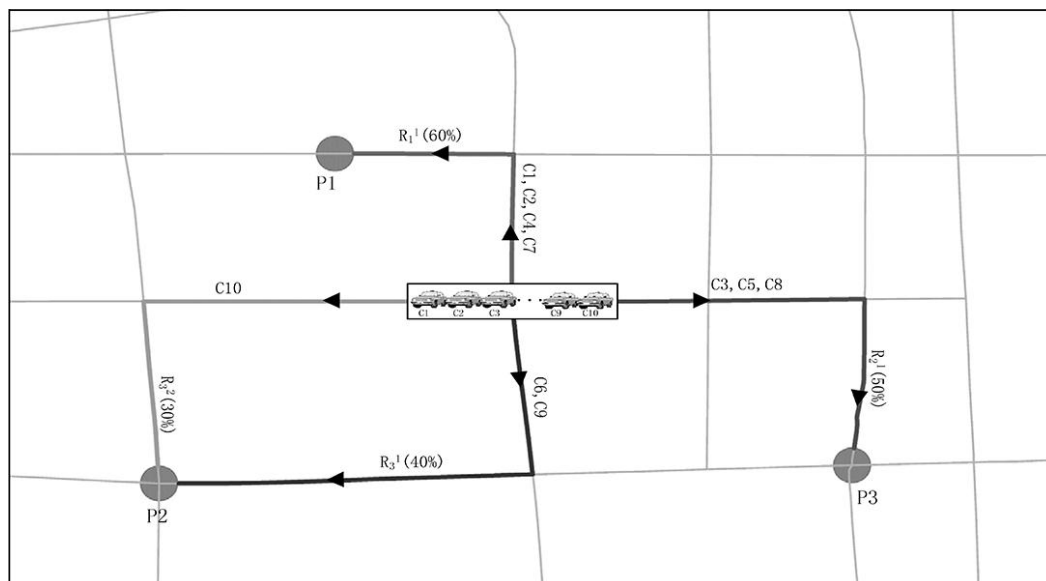
```

1: while true do
2:    $i \leftarrow (i + 1) \bmod n$ ;
3:   If  $i == 0$  then
4:      $cw \leftarrow cw - \gcd(R)$ ;
5:     If  $cw \leq 0$  then
6:        $cw \leftarrow \max(R)$ ;
7:       if  $cw == 0$  then
8:         return NULL;
9:       end if
10:    end if
11:  end if
13:  if  $W(R_i) \geq cw$  then
14:    return  $R_i$ ;
15:  end if
16: end

```

---

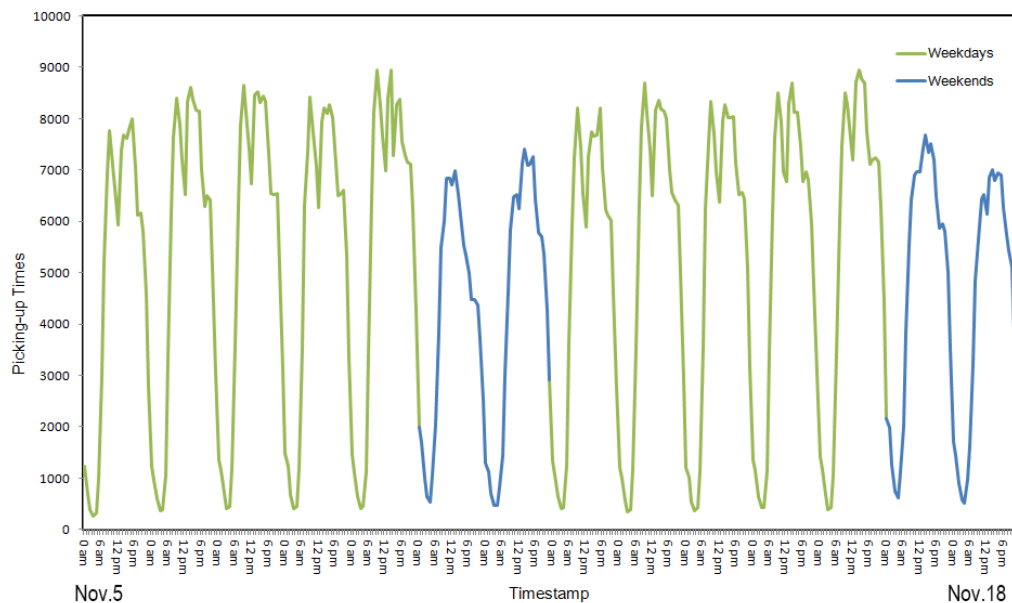
An example of this method is given in Figure 3. Assuming that there are three recommended pick-up locations ( $P_1$ ,  $P_2$ , and  $P_3$ ) and 4 recommended cruising routes  $R_1^1$ ,  $R_2^1$ ,  $R_3^1$  and  $R_3^2$  with the weights 60%, 50%, 40% and 30%, respectively, then 10 taxis ( $C_i, i \in [1, 10]$ ) are considered to be allocated. Based on our methods,  $R_1^1$  will be assigned to  $C_1, C_2, C_4$  and  $C_7$ ;  $R_2^1$  will be assigned to  $C_3, C_5$  and  $C_8$ ;  $R_3^1$  will be assigned to  $C_6$  and  $C_9$ ; and  $R_3^2$  will be assigned to  $C_{10}$ .



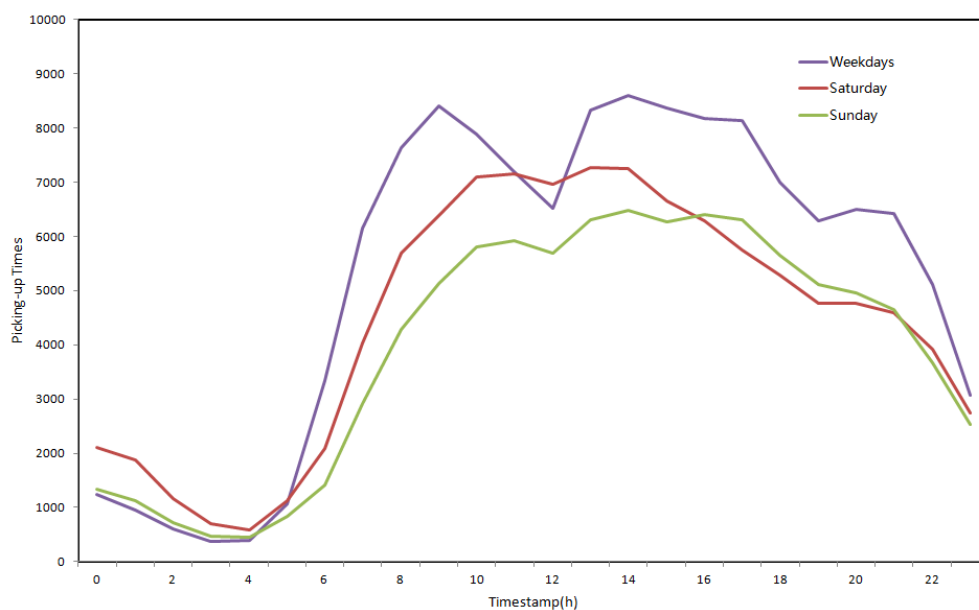
**Figure 3.** An Example of a Cruising Route Distribution Using the Weighted Round-Robin Scheduling Algorithm.

#### 4. Results

Pick-up time statistics are presented in Figures 4 and 5. Note that all the time periods or hours in this study refer to Beijing time. The pick-up patterns look similar from Monday to Friday, the patterns on weekend days are similar. No matter on weekdays or weekends, the lowest demanding time is between 4 a.m. and 5 a.m. In addition, the pick-up times on weekends are always lower than those on weekdays. With respect to peak times, there are four peak times in one day during weekdays (9 a.m. to 10 a.m., 2 p.m. to 3 p.m., 5 p.m. to 6 p.m. and 8 p.m. to 9 p.m.), while three peak times occur on weekends (10 a.m. to 12 p.m., 2 p.m. to 3 p.m., and 9 p.m. to 10 p.m. on Saturday and 5 p.m. to 6 p.m. on Sunday). Moreover, people seem to go out later on weekends than weekdays, but the duration of time that people are away is usually longer.

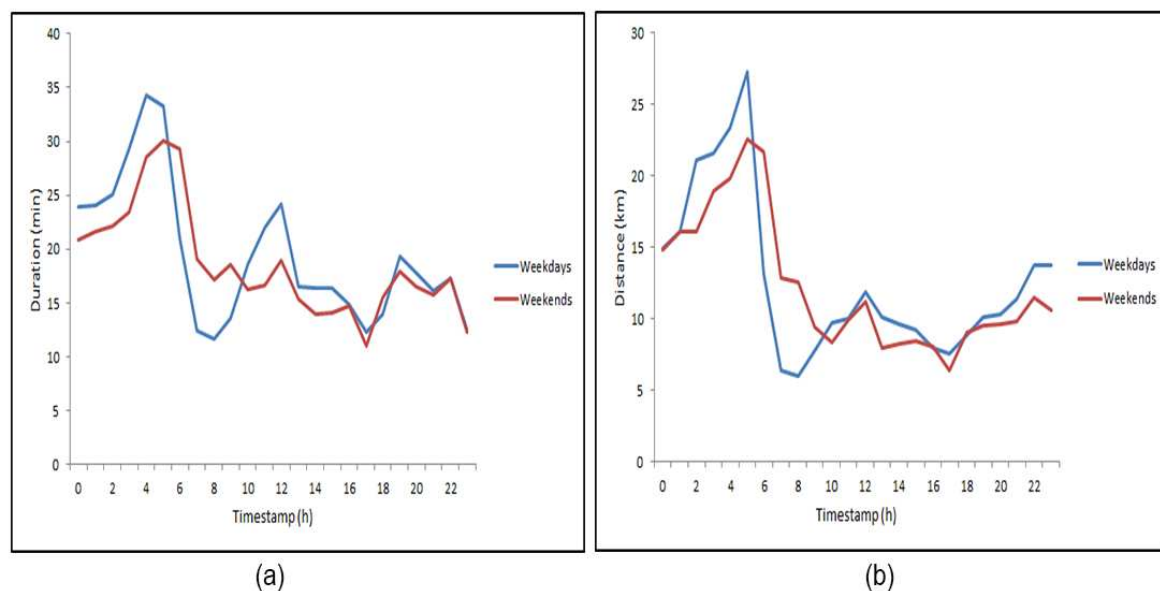


**Figure 4.** Plot of the Pick-Up Times of Taxi Drivers over 14 Days (from 5 November to 18 November).



**Figure 5.** Plot of the Average Pick-Up Times of Taxi Drivers.

The average duration and average distance from the current time to the beginning of the next trip vary at different parts of the day (Figure 6). There are rarely passengers before dawn regardless of the day of the week, and thus, drivers must spend substantial time and drive a long way to find passengers; drivers spend less time searching for passengers during high-demand periods. Compared with weekdays, drivers spend less time searching for passengers on weekends before dawn. In addition, people spent more time away from home on weekends. Furthermore, drivers spend less time searching for passengers in the weekdays before 10 a.m.



**Figure 6.** Plot of the Average Duration and Average Distance from the Current Time to the Beginning of the Next Trip. (a) shows the average duration in weekdays and weekends; and (b) shows the average distance.

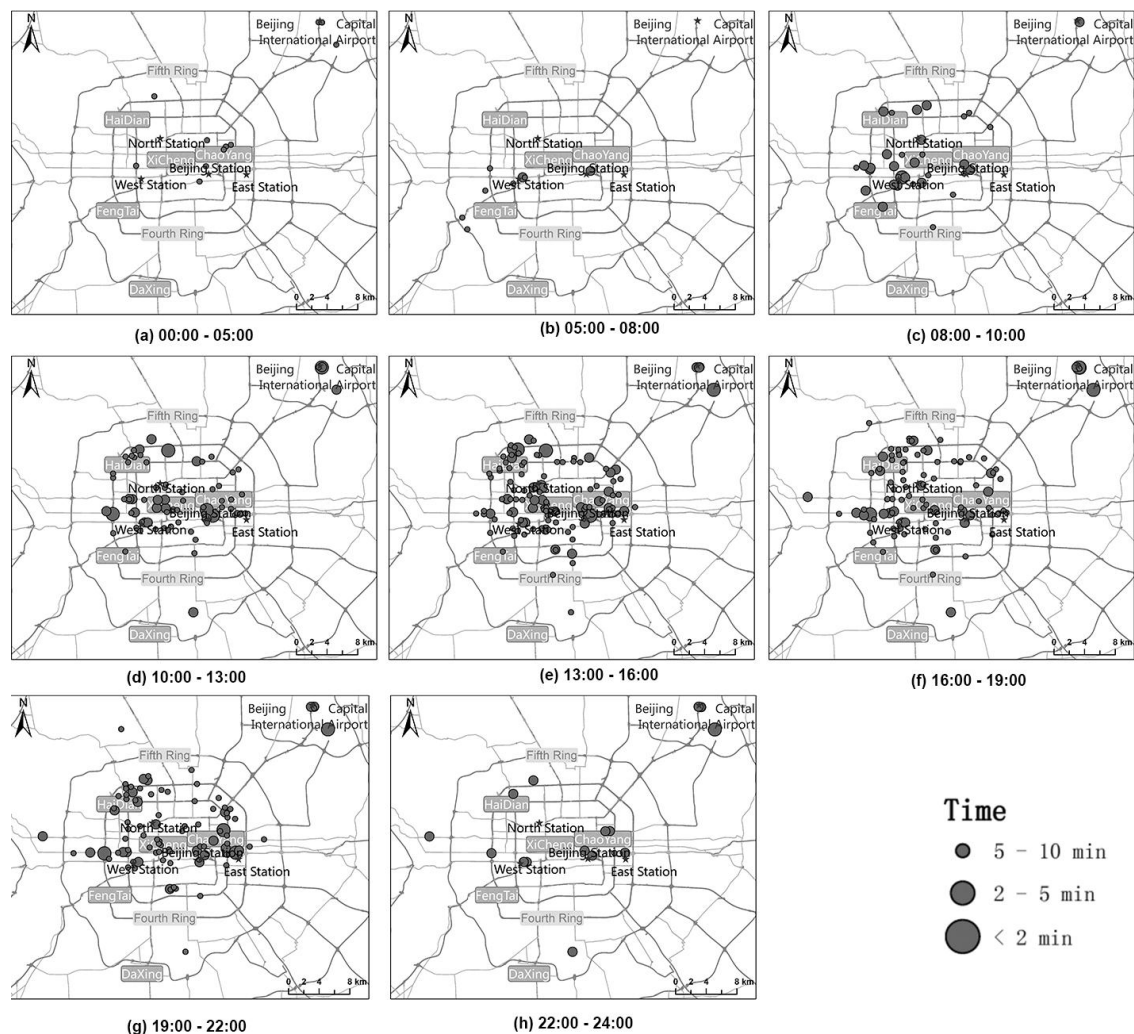
The cluster counts of pick-up locations showed variable changes at different times in one day (Table 3). We observed that in the early morning, whether on weekdays or weekends, the number of clusters was the lowest. During this time, few people went out, and there was a low demand for taxis. As the number of people going out increased, increasing taxi demand was generated, and the number of clusters began to increase. There was little difference in the number of clusters on weekends or weekdays in the morning. In the afternoon, however, the number of clusters on weekdays was more than double that on the weekend, which might be because people spend more time away from home on weekends than on weekdays. During the evening, the demand for taxis gradually reduced, while the number of hot spot demand areas began to decrease.

**Table 3.** Counts of clusters at different times in one day.

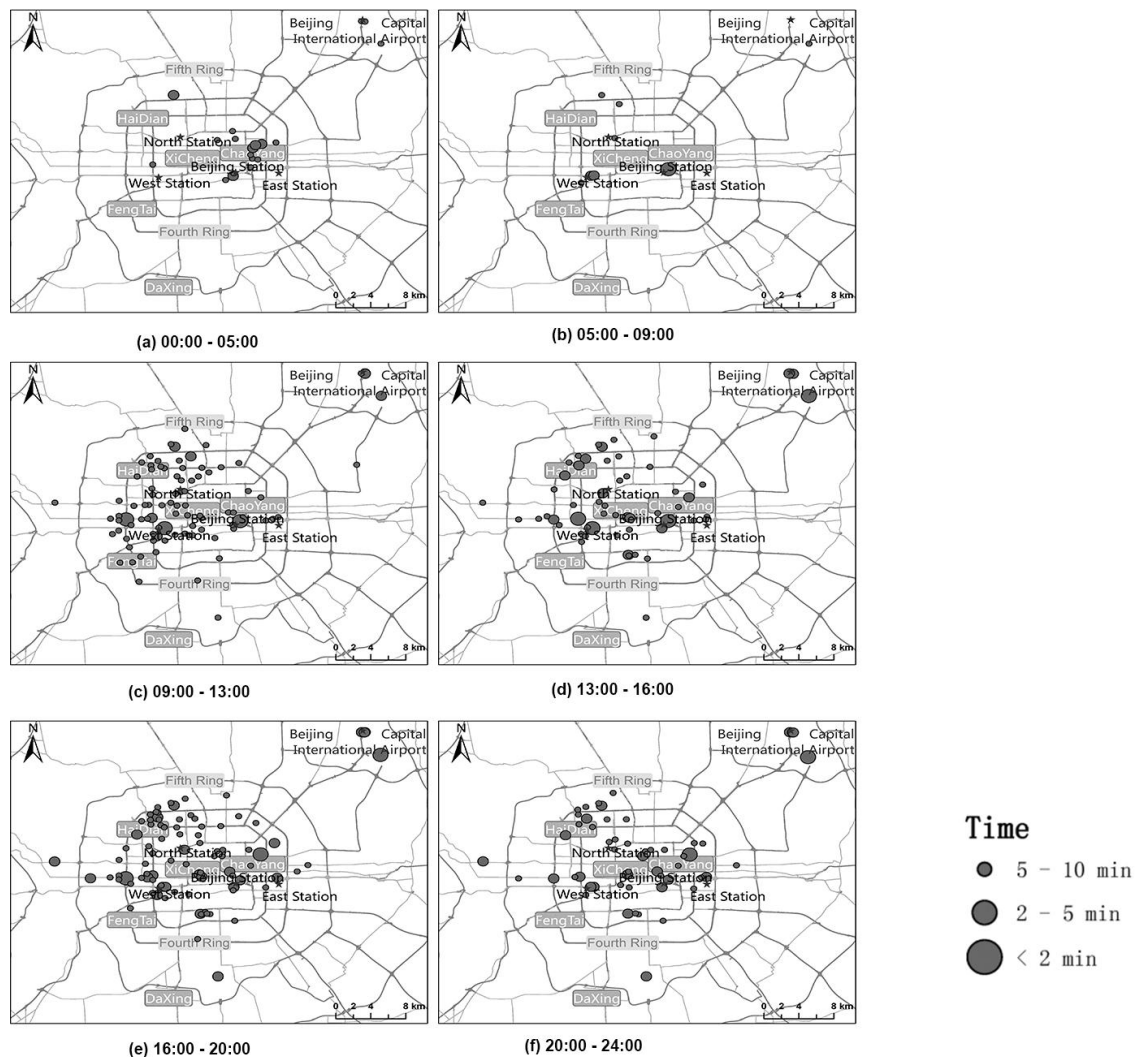
Weekdays	Count	Workdays	Count
0 a.m.–5 a.m.	12	0 a.m.–5 a.m.	20
5 a.m.–8 a.m.	8	5 a.m.–9 a.m.	8
8 a.m.–10 a.m.	33	9 a.m.–1 p.m.	90
10 a.m.–1 p.m.	75	1 p.m.–4 p.m.	56
1 p.m.–4 p.m.	127	4 p.m.–8 p.m.	91
4 p.m.–7 p.m.	101	8 p.m.–0 a.m.	55
7 p.m.–10 p.m.	95		
10 p.m.–0 a.m.	17		

The regional distribution of the cluster counts of pick-up locations at different times in one day are presented in Figure 7 (weekdays) and Figure 8 (weekends). Note that the legend time represents the average time spent picking up a passenger. We found the following:

1. In the early morning, high-demand areas were mainly in the Dongcheng District and Chaoyang District on weekdays (Figure 7a) and especially on weekends (Figure 8a). Eastern Beijing has both foreign affairs areas and business districts, and as many business centers are there, taxi demand occurred in the early morning. Moreover, as people go out for longer on weekends, there was also greater demand on weekends.
2. At 8 a.m. or 9 a.m., demand was centered around public transportation stations, such as the Beijing Station and Beijing West Station (Figures 7b and 8b).



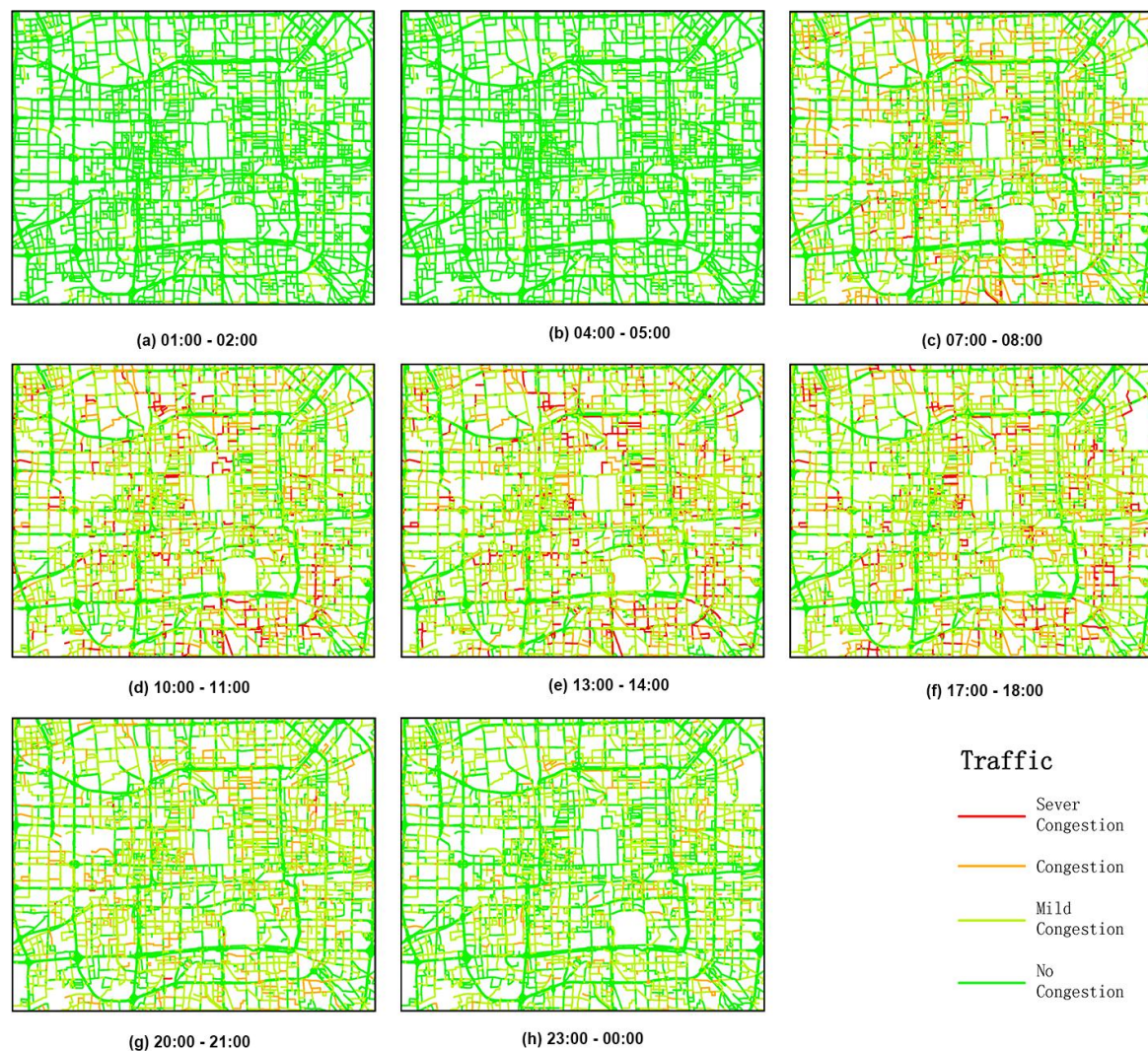
**Figure 7.** Distribution of the Cluster Counts of Pick-Up Locations in Weekdays. (a–h) correspond to the 8 time periods of one day on weekdays.



**Figure 8.** Distribution of the Cluster Counts of Pick-Up Locations on Weekends. (a–f) correspond to the 6 time periods of one day on weekends.

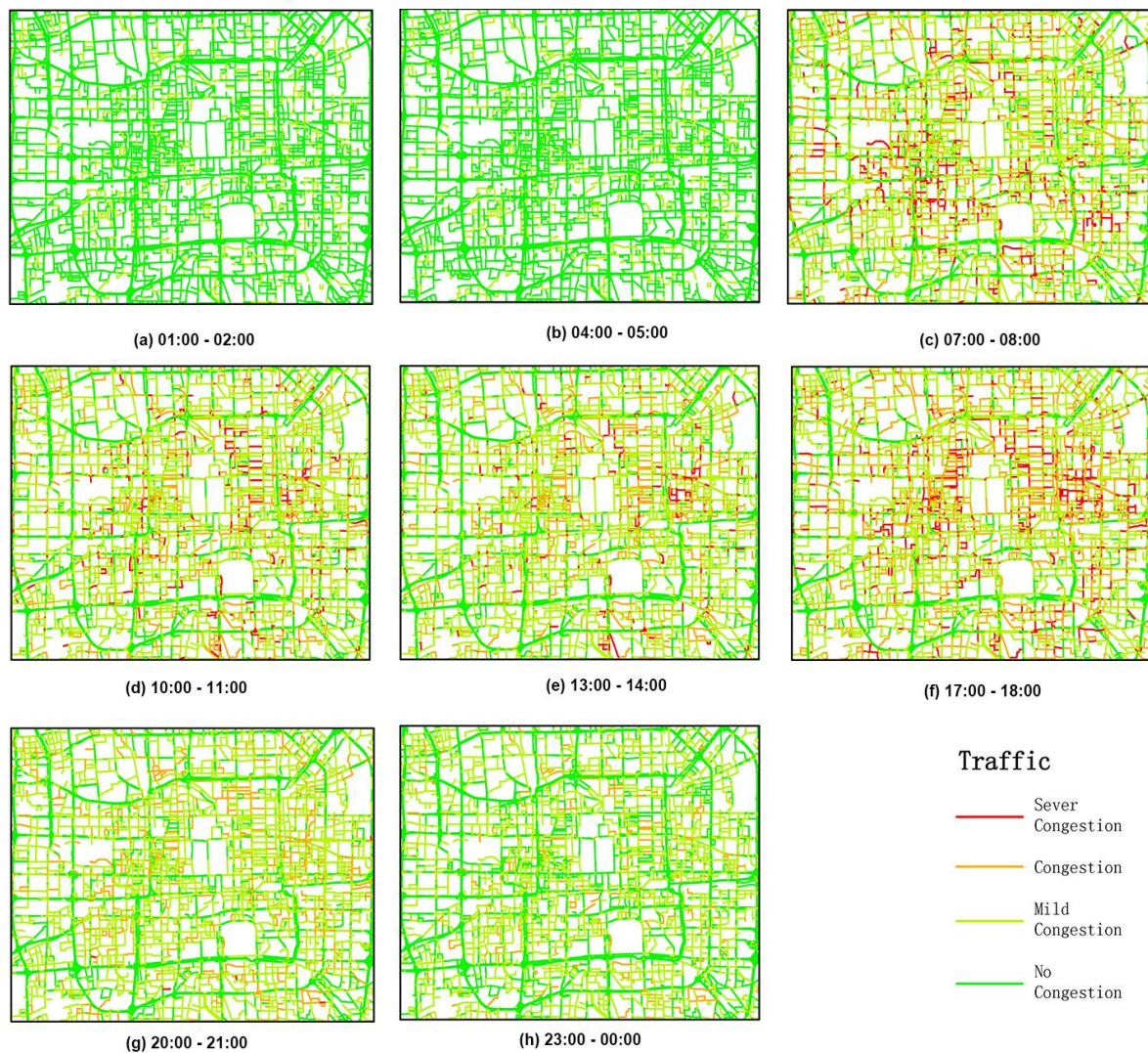
- Later in the day, the number of high-demand areas began to increase. Chaoyang is the largest district in terms of population, followed by the Haidian District and Fengtai District. Approximately 45.6% of Beijing's population lives in these districts. Figures 7 and 8 show that high-demand areas were concentrated in these areas. There are many good colleges and high schools in the Haidian District. Manufacturing and scientific research groups are concentrated between West Second Ring Road and West Third Ring Road. Moreover, most of the population is employed in the Haidian District and Fengtai District. Therefore, during the day, the number of high-demand areas in the Haidian District and Fengtai District were greater than that in the Chaoyang District; this difference was particularly pronounced on the weekends (Figures 7c–e and 8c,d).
- During the evening, the number of high-demand areas in the Chaoyang District changed more slowly than in the Haidian District and Fengtai District, particularly on the weekends. The major reason might be because there are several catering and entertainment businesses in the Chaoyang District (Figures 7f–h and 8e,f).

Traffic resistance was detected using average taxi travel speed and travel time on weekdays and weekends (Figures 9 and 10).



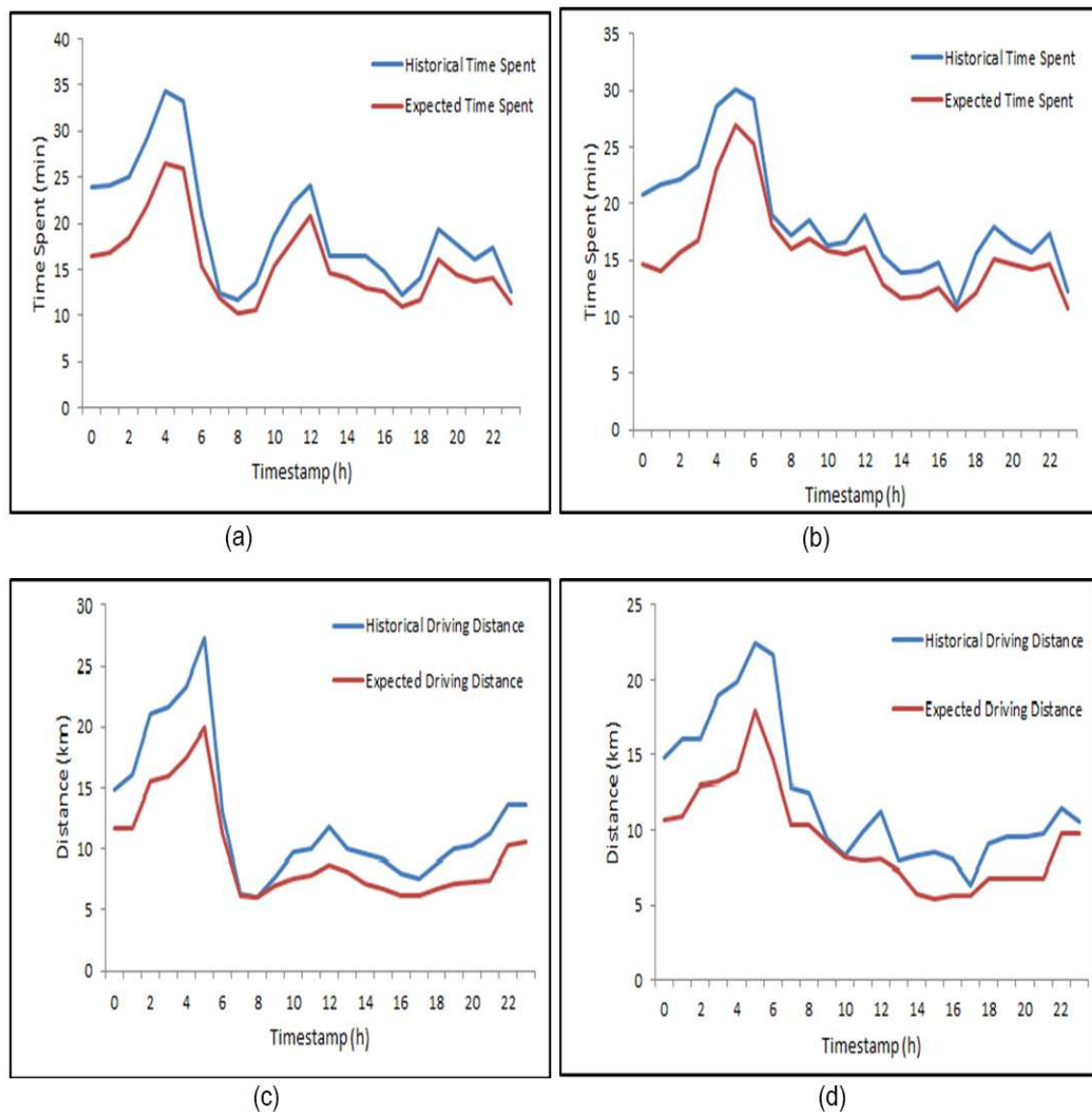
**Figure 9.** Traffic Resistance on Weekdays. (a–h) correspond to the 8 time periods of one day on weekdays.

- According to the Evaluation Index System of Urban Road Traffic Management provided by the Ministry of Public Security, a travel speed of  $>30$  km/h indicates no congestion, a travel speed of between 20 km/h and 30 km/h indicates mild congestion, a travel speed of between 10 km/h and 20 km/h indicates congestion, and a travel speed of  $<10$  km/h indicates strong congestion. We observed that before 5 a.m., traffic transportation showed the best conditions no matter which day of the week. Almost all the roads were clear (Figure 9a,b and Figure 10a,b), because during this period, there were fewer people and cars going out. Moreover, on weekdays, roads became increasingly congested from 6 a.m. onwards. During the peak morning hours, the average speed was approximately 25 km/h (Figure 9c). However, the traffic conditions were slightly better in the early afternoon (Figure 9d,e). In the period of the evening peak (5 p.m. to 7 p.m.), roads again became congested, and the congestion worsened (Figure 9f); the average speed was only approximately 22.9 km/h. After the evening peak, the traffic began to clear, and at midnight, there were no jammed roads (Figure 9g,h). On the weekends, traffic transportation showed better conditions than that on weekdays. There was no morning peak or evening peak (Figure 10c,f), but the traffic conditions were still congested during the day (Figure 10d,e). After the evening peak, the traffic began to clear. This pattern shows that people's activities have impacts on traffic transportation conditions.



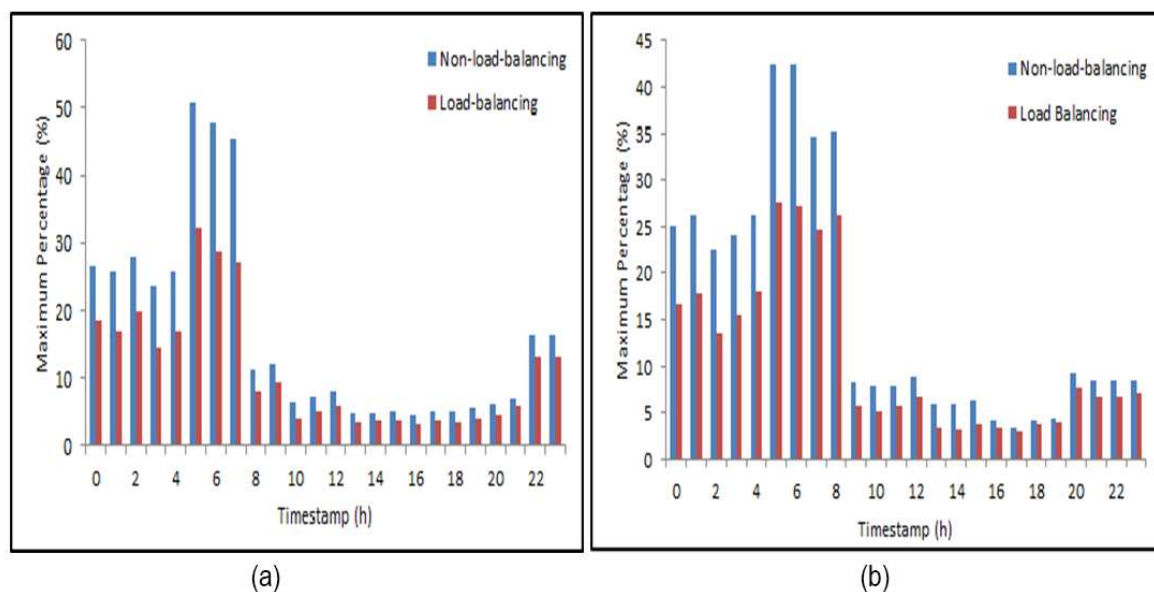
**Figure 10.** Traffic Resistance on Weekends. (a–h) correspond to the 8 time periods of one day on weekends.

In simulation experiment, separate statistical comparisons were conducted for time and distance aimed to verify the effect of the recommended cruising routes using an STT model in our work. Specifically, we divided one day (24 h) into 24 time periods. Next, we extracted a random 1000 historical taxi routes in each time period. Then, the historical time spent and driving distance and expected time spent and driving distance from drop-off location to the pick-up location of the next passenger in one route were estimated. Statistical comparisons are shown in Figure 11. We observed that the historical times spent were mostly greater than the expected times; that is, taxi drivers using our recommended route to find the next passengers can significantly reduce the time spent, while with the historical route, the taxi drivers traveled greater distances than those with our recommended route. In conclusion, taxi drivers using cruising routes recommended by our STT model can significantly reduce the waiting time (about average 2.98 min on weekdays and 3.81 min on weekends) and travel less distance (about average 2.92 km on weekdays and 2.86 min on weekends) to quickly find their next passengers.



**Figure 11.** Statistical Comparisons between Historical Routes and Recommended Routes in Time and Distance. (a) shows the comparison plot of time for weekdays; (b) shows the comparison plot of time for weekends; (c) shows the comparison plot of distance for weekdays; and (d) shows the comparison plot of distance for weekends.

In the STT model, the  $k$  shortest path routing algorithm was presented, and  $k$  cruising routes were obtained. Then, the load balancing method was presented to allocate these  $k$  cruising routes to more than one taxi driver in a small-scale region to the neighboring pick-up locations. However, if directly offering these  $k$  cruising routes and their pick-up probabilities to taxi drivers and letting them choose by themselves, the cruising route with the highest probability would be chosen by most taxi drivers. This condition would cause a large number of taxi drivers to select the same road to find their next passengers and would reduce the accuracy of the probability computed in our model. Considering this, we conducted a comparison of whether or not the load balancing method is utilized (Figure 12). We observed that the load balancing strategy significantly alleviates road loads. When the load balancing method was not utilized in the STT model, the maximum percentage would be close to 50% at the off-peak period (approximately 5 a.m. to 7 a.m.). However, the maximum percentage would reduce to just 30% when the load balancing method was utilized in our STT model.



**Figure 12.** Comparison between Load Balancing and Non-Load-Balancing. (a) shows the comparison plot for weekdays; and (b) shows the comparison plot for weekends.

## 5. Discussion

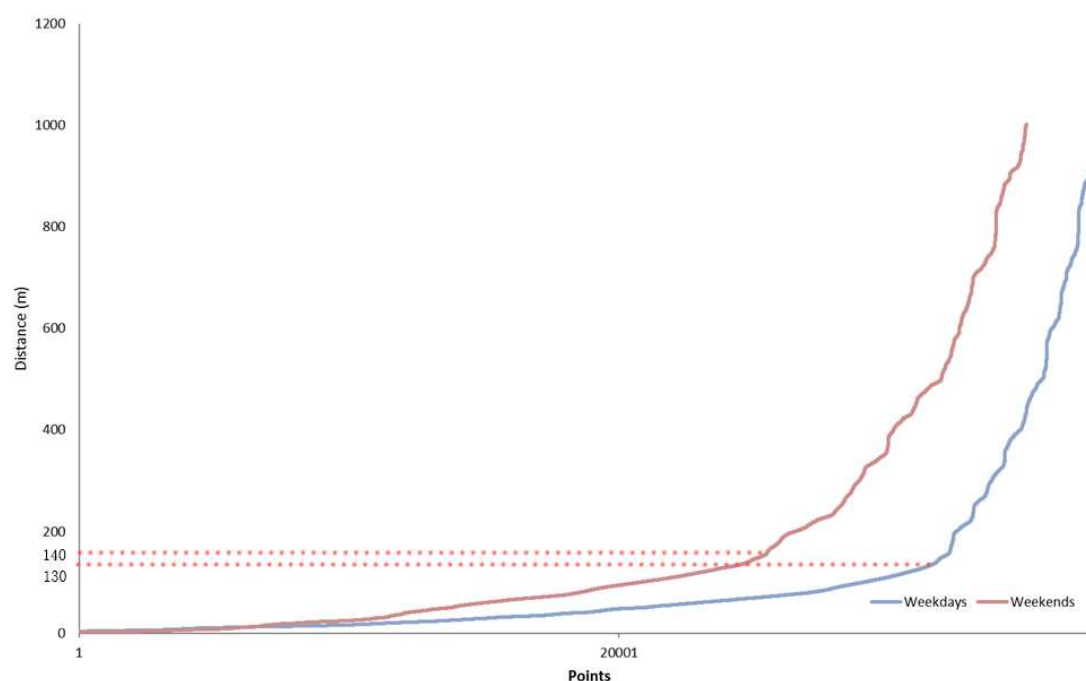
In this work, we present an STT model based on the historical GPS trajectories of taxis to explore spatio-temporal traffic patterns and guide taxi navigation. Furthermore, the STT model can provide useful information for taxi drivers to quickly pick up their next passengers, further saving time and gas and increasing profits. In addition, the taxi fleet management method in our model could help ease urban traffic problems.

Numerous previous studies have explored the optimizing cruising routes and recommendations to taxi drivers in various methods. One such study used time series forecasting techniques to predict the spatio-temporal distribution in real-time and developed an online recommendation system for the taxi stand choice in the city of Porto, Portugal [50]. The fleet equipped with their recommendation system can significantly reduce 5% of average waiting time, but they actually ignore the comparisons between weekends and weekdays. Another study propose a Time-Location-Relationship combined taxi service recommendation model, utilizing Gaussian Process Regression and statistical approaches to improve taxi drivers' profits. They compared their model with ARIMA, SVM et.al models and found their taxi service recommendation can predict more accurately than others by using the taxi GPS data in Beijing [51]. Moreover, Wong et al. emphasized that taxi drivers' cruising decisions are significantly affected by the probability of successfully picking up passengers along this cruising route [4]. Yuan et al. focused on extracting passenger waiting areas for taxi drivers and computing the probability of picking up their next passengers based on the time spent, road segment information and accessibility to waiting areas [3,22]. Qian proposed a method that transformed the taxi-routing issue into a Markov decision process of pick-up locations [52].

However, these taxi route-planning studies have a limited quantitative focus on the association of cruising routes with traffic resistance and have paid less attention to taxi fleet management. The present article fills this gap by estimating the average taxi travel speed using the historical GPS trajectories of taxis and utilizing a load balancing method, which is widely used in computing fields. Three factors, including the pick-up locations cluster, average taxi travel speed and cruising routes, have been considered in the STT model. The SST model is multi-integrated and mainly designed for macroscopically taxi fleet management but a simple instance, cruising routes recommendation for taxi drivers were depicted based on this global taxi fleet management. As a result, our study shows that taxi drivers using cruising routes recommended by our STT model can significantly reduce

the average waiting time and travel less distance to quickly find their next passengers, and the load balancing strategy significantly alleviates road loads. Our results support a growing body of recent literature underlining the superiority of combination and integration of statistical methodologies and load balancing algorithm in transportation patterns mining.

Given to clustering of pick-up locations, the Manhattan distance, rather than Euclidean distance, is used to measure the similarity between two clusters. Based on the different pick-up patterns in different periods, we divided a day (24 h) into 8 time periods on weekdays and 6 time periods on weekends (Table 3). Then, pick-up cluster analysis in the different time periods was conducted. With respect to parameters in the DBSCAN algorithm, MinPts and Eps were determined through an interactive process by examining the sorted K-dist graph [53]. Generally, we determined the value in the  $y$ -axis (distance) corresponding to the turning point as Eps [37]. The K-dist graph is shown in Figure 13. We observed that the distances corresponding to the turning point were approximately 130 m on weekdays and 140 m on weekends.

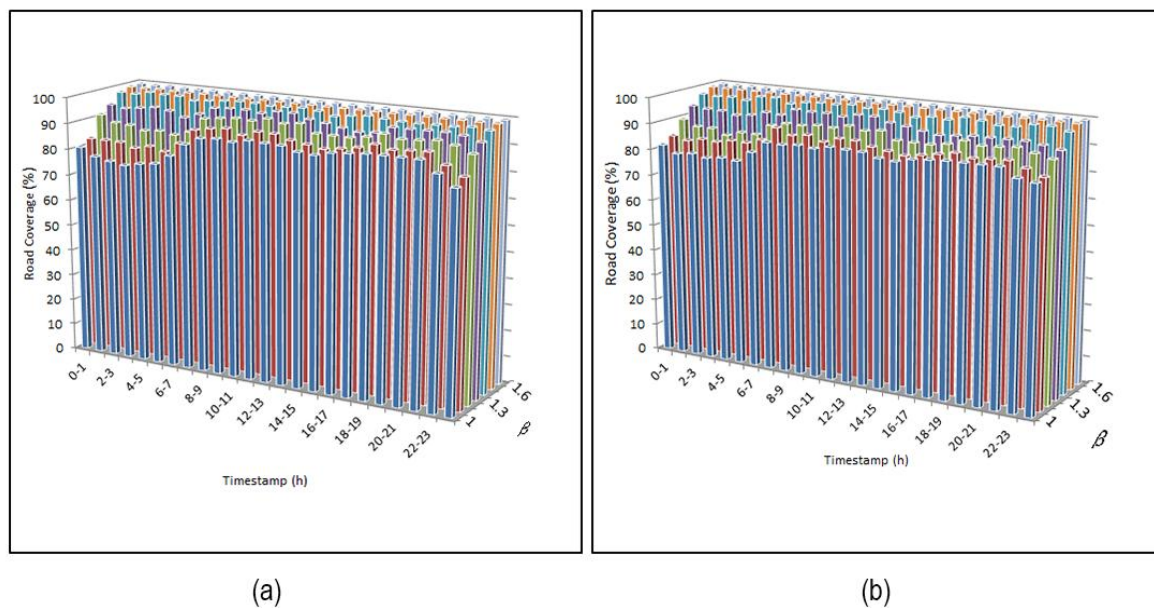


**Figure 13.** Plot of the sorted K-dist graph. In this graph,  $y$ -axis determines the Manhattan distance between points while  $x$ -axis determining the index of sorted points.

Then, average taxi travel speed and travel time were estimated to explore traffic flow patterns. Moreover, these speeds were considered a critical factor when computing the pick-up probability of recommended routes.

The K shortest path routing algorithm was presented to explore multiple optimal routes. In the STT model, parameter  $k$  was dynamically determined by estimating the travel time of finding the next passengers. Specifically, cruising routes were reserved if the travel time in this route was less than  $\beta$  times that of the shortest path routing.  $\beta$  was defined as the cruising routes threshold in our model. Considering that most people located all over the study areas are able to take taxis, the cruising routes are suggested to cover most roads. Therefore, we discussed the relationship between  $\beta$  and road coverage in different time periods (Figure 14). We observed that at any time, whether on weekends or on weekdays, if only one path is planned, it cannot cover all the main roads, especially in the early morning. When  $\beta$  is 1.5, almost all of the time intervals have a coverage of more than 99%. When  $\beta$  is 1.6, almost all of the time intervals have a coverage of more than 99.5%. Thus, when  $\beta$  changes

from 1.5 to 1.6, the coverage is not significantly improved, but the number of calculations will increase greatly. Therefore,  $\beta$  was defined as 1.5.



**Figure 14.** The Relationship between  $\beta$  and Road Coverage. (a) shows the plot for weekdays; and (b) shows the plot for weekends.

However, much improvement should be conducted in future studies. Firstly, DBSCAN clustering algorithm was utilized to explore high-demand areas from the historical pick-up information. Specifically speaking, we can discover pick-up clusters with arbitrary shapes and avoid the adverse impacts of noise and unusual points, and address geographical characteristics with DBSCAN. But the deterministic lack and adaptability of the setting of parameters isn't considered and will be examined in future studies. Secondly, due to the limitations of access to the taxi GPS trajectory data, only one-month data is collected and utilized in our experiment, the sample size used in the analysis was small. Despite this, we recognize that this analysis is only a preliminary analysis of transportations patterns mining and it has its limitations. The results of our study highlight the need for researchers to recognize the usefulness of our model as an exploratory data analysis tools for trajectories mining. Moreover, as another shortage of our this study, we does not consider certain variables, such as emergent traffic situation handling and the quantitative incidence of energy consumption because of the lack of these data. It will be very interesting to extend the approach further to consider these variables. Finally, while the data being used in this study is a little outdated in temporal granularities (i.e., 2012) and not a real-time data feed, it is possible to extend the method to enable real-time recommendations based on immediate past or historical records.

## 6. Conclusions

This work develops an STT model to explore spatio-temporal traffic patterns and guide taxi navigation. Our STT model takes advantage of a large volume of historical taxi GPS trajectories for spatial and temporal analysis, and the pick-up location clusters, average taxi travel speed and cruising routes are considered in the STT model. Specifically, average taxi travel speeds are estimated as traffic resistance, and the load balancing method is utilized for cruising route allocation. Our experimental results indicate that taxi drivers using cruising routes recommended by the STT model can significantly reduce the time spent and travel less distance to quickly find their next passengers. In addition, the load balancing utilized in our STT model could significantly alleviate road congestion.

**Acknowledgments:** This project is supported by the National Science Foundation of China (Grant No. 41671400) and The National Key Research and Development Program of China (Grant No. 2017YFB0503601). We thank the IBM Innovation Center for providing technical support and NERC GIS (National Engineering Research Center for Geographic Information System of China) for providing hardware support.

**Author Contributions:** Conceived and designed the experiments: Liang Wu, Sheng Hu, Yazhou Wang and Zhong Xie; Performed the experiments: Liang Wu, Sheng Hu, Yazhou Wang, Hao Chen and Zhanlong Chen; Analyzed the data: Liang Wu, Yazhou Wang, Zhong Xie and Li Yin; Contributed reagents/materials/analysis tools: Sheng Hu, Zhanlong Chen, Hao Chen and Mingqiang Guo; Wrote the paper: Liang Wu, Sheng Hu, Yazhou Wang, Li Yin and Zhong Xie.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Qian, X.; Ukkusuri, S.V. Spatial variation of the urban taxi ridership using gps data. *Appl. Geogr.* **2015**, *59*, 31–42. [[CrossRef](#)]
2. Moreira-Matias, L.; Gama, J.; Ferreira, M.; Mendes-Moreira, J.; Damas, L. Predicting taxi-passenger demand using streaming data. *IEEE Trans. Intell. Transp. Syst.* **2013**, *14*, 1393–1402. [[CrossRef](#)]
3. Yuan, N.J.; Zheng, Y.; Zhang, L.H.; Xie, X. T-finder: A recommender system for finding passengers and vacant taxis. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 2390–2403. [[CrossRef](#)]
4. Wong, R.C.P.; Szeto, W.Y.; Wong, S.C. A cell-based logit-opportunity taxi customer-search model. *Transp. Res. Part C Emerg. Technol.* **2014**, *48*, 84–96. [[CrossRef](#)]
5. Castro, P.S.; Zhang, D.Q.; Chen, C.; Li, S.J.; Pan, G. From taxi gps traces to social and community dynamics: A survey. *ACM Comput. Surv. (CSUR)* **2013**, *46*, 17. [[CrossRef](#)]
6. Ge, Y.; Xiong, H.; Tuzhilin, A.; Xiao, K.; Gruteser, M.; Pazzani, M. An energy-efficient mobile recommender system. In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 25–28 July 2010; pp. 899–908.
7. Chen, B.Y.; Yuan, H.; Li, Q.Q.; Lam, W.H.K.; Shaw, S.L.; Yan, K. Map-matching algorithm for large-scale low-frequency floating car data. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 22–38. [[CrossRef](#)]
8. Zhang, D.Q.; Sun, L.; Li, B.; Chen, C.; Pan, G.; Li, S.J.; Wu, Z.H. Understanding taxi service strategies from taxi gps traces. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 123–135. [[CrossRef](#)]
9. Liu, S.; Liu, Y.; Ni, L.M.; Fan, J.; Li, M. Towards mobility-based clustering. In Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 25–28 July 2010; pp. 919–928.
10. Yuan, J.; Zheng, Y.; Xie, X. Discovering regions of different functions in a city using human mobility and pois. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 186–194.
11. Jiang, B.; Yin, J.J.; Zhao, S.J. Characterizing the human mobility pattern in a large street network. *Phys. Rev. E* **2009**, *80*, 021136. [[CrossRef](#)] [[PubMed](#)]
12. Jiang, X.R.; Zheng, C.Y.; Tian, Y.; Liang, R.H. Large-scale taxi o/d visual analytics for understanding metropolitan human movement patterns. *J. Vis.* **2015**, *18*, 185–200. [[CrossRef](#)]
13. Wang, Z.C.; Lu, M.; Yuan, X.R.; Zhang, J.P.; van de Wetering, H. Visual traffic jam analysis based on trajectory data. *IEEE Trans. Vis. Comput. Gr.* **2013**, *19*, 2159–2168. [[CrossRef](#)] [[PubMed](#)]
14. Wang, X.S.; Liu, H.B.; Yu, R.J.; Deng, B.; Chen, X.H.; Wu, B. Exploring operating speeds on urban arterials using floating car data: Case study in shanghai. *J. Transp. Eng.* **2014**, *140*, 04014044. [[CrossRef](#)]
15. Liu, X.L.; Lu, F.; Zhang, H.C.; Qiu, P.Y. Intersection delay estimation from floating car data via principal curves: A case study on beijing's road network. *Front. Earth Sci.* **2013**, *7*, 206–216. [[CrossRef](#)]
16. Herring, R.; Hofleitner, A.; Abbeel, P.; Bayen, A. Estimating arterial traffic conditions using sparse probe data. In Proceedings of the 13th International IEEE Conference on Intelligent Transportation Systems, Funchal, Portugal, 19–22 September 2010; pp. 929–936.
17. Yuan, J.; Zheng, Y.; Xie, X.; Sun, G. T-drive: Enhancing driving directions with taxi drivers' intelligence. *IEEE Trans. Knowl. Data Eng.* **2013**, *25*, 220–232. [[CrossRef](#)]
18. Sun, D.; Zhang, C.; Zhang, L.; Chen, F.; Peng, Z.-R. Urban travel behavior analyses and route prediction based on floating car data. *Transport. Lett.* **2014**, *6*, 118–125. [[CrossRef](#)]

19. Chen, C.; Zhang, D.Q.; Li, N.; Zhou, Z.H. B-planner: Planning bidirectional night bus routes using large-scale taxi gps traces. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 1451–1465. [CrossRef]
20. Tang, L.L.; Li, Q.Q.; Chang, X.M.; Shaw, S.L.; Zhao, Z.L. Modeling of taxi drivers' experience for routing applications. *Sci. China Technol. Sci.* **2010**, *53*, 44–51. [CrossRef]
21. Liu, L.; Andris, C.; Ratti, C. Uncovering cabdrivers' behavior patterns from their digital traces. *Comput. Environ. Urban Syst.* **2010**, *34*, 541–548. [CrossRef]
22. Yuan, J.; Zheng, Y.; Zhang, L.; Xie, X.; Sun, G. Where to find my next passenger. In Proceedings of the 13th International Conference on Ubiquitous Computing, Beijing, China, 17–21 September; ACM: New York, NY, USA, 2011; pp. 109–118.
23. Hu, X.; An, S.; Wang, J. Exploring urban taxi drivers: Activity distribution based on gps data. *Math. Probl. Eng.* **2014**, *2014*, 13. [CrossRef]
24. Li, B.; Zhang, D.; Sun, L.; Chen, C.; Li, S.; Qi, G.; Yang, Q. Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. In Proceedings of the 2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), Seattle, WA, USA, 21–25 March 2011; pp. 63–68.
25. Yang, C.; Gonzales, E.J. Modeling taxi trip demand by time of day in new york city. *Transp. Res. Rec.* **2014**, 110–120. [CrossRef]
26. Wong, R.C.P.; Szeto, W.Y.; Wong, S.C.; Yang, H. Modelling multi-period customer-searching behaviour of taxi drivers. *Transp. B Trans. Dyn.* **2014**, *2*, 40–59. [CrossRef]
27. Zong, F. Understanding taxi driver's cruising behavior with zip model. *J. Cent. South Univ.* **2014**, *21*, 3404–3410. [CrossRef]
28. Liu, S.; Wang, S.; Liu, C.; Krishnan, R. Understanding taxi drivers' routing choices from spatial and social traces. *Front. Comput. Sci.* **2015**, *9*, 200–209. [CrossRef]
29. Yang, Y.; Yan, Z.; Qingquan, L.; Qingzhou, M. Mining time-dependent attractive areas and movement patterns from taxi trajectory data. In Proceedings of the 2009 17th International Conference on Geoinformatics, Fairfax, VA, USA, 12–14 August 2009; pp. 1–6.
30. Hwang, R.H.; Hsueh, Y.L.; Chen, Y.T. An effective taxi recommender system based on a spatio-temporal factor analysis model. *Inf. Sci.* **2015**, *314*, 28–40. [CrossRef]
31. Database of Beijing City Government Data Resources. Available online: <http://www.bjdata.gov.cn/> (accessed on 17 November 2017).
32. Quddus, M.A.; Ochieng, W.Y.; Noland, R.B. Current map-matching algorithms for transport applications: State-of-the art and future research directions. *Transp. Res. C Emerg. Technol.* **2007**, *15*, 312–328. [CrossRef]
33. Velaga, N.R.; Quddus, M.A.; Bristow, A.L. Developing an enhanced weight-based topological map-matching algorithm for intelligent transport systems. *Transp. Res. C Emerg. Technol.* **2009**, *17*, 672–683. [CrossRef]
34. Yang, D.; Zhang, T.; Li, J.; Lian, X. Synthetic fuzzy evaluation method of trajectory similarity in map-matching. *J. Intell. Transp. Syst.* **2011**, *15*, 193–204. [CrossRef]
35. Quddus, M.A.; Noland, R.B.; Ochieng, W.Y. A high accuracy fuzzy logic based map matching algorithm for road transport. *J. Intell. Transp. Syst.* **2006**, *10*, 103–115. [CrossRef]
36. Skog, I.; Handel, P. In-car positioning and navigation technologies: A survey. *IEEE Trans. Intell. Transp. Syst.* **2009**, *10*, 4–21. [CrossRef]
37. Wang, X.F.; Huang, D.S. A novel density-based clustering framework by using level set method. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1515–1531. [CrossRef]
38. Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, 2–4 August 1996; AAAI Press: Palo Alto, CA, USA, 1996; pp. 226–231.
39. Ester, M.; Kriegel, H.-P.; Sander, J.; Wimmer, M.; Xu, X. Incremental clustering for mining in a data warehousing environment. In Proceedings of the 24rd International Conference on Very Large Data Bases, New York, NY, USA, 24–27 August 1998; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1998; pp. 323–333.

40. Kalnis, P.; Mamoulis, N.; Bakiras, S. On discovering moving clusters in spatio-temporal data. In *Advances in Spatial and Temporal Databases, Proceedings of the 9th International Symposium, SSTD 2005, Angra dos Reis, Brazil, 22–24 August 2005*; Bauzer Medeiros, C., Egenhofer, M.J., Bertino, E., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 364–381.
41. Birant, D.; Kut, A. St-dbscan: An algorithm for clustering spatial-temporal data. *Data Knowl. Eng.* **2007**, *60*, 208–221. [[CrossRef](#)]
42. Erman, J.; Arlitt, M.; Mahanti, A. Traffic classification using clustering algorithms. In *Proceedings of the 2006 SIGCOMM Workshop on Mining Network Data, Pisa, Italy, 11–15 September 2006*; ACM: New York, NY, USA, 2006; pp. 281–286.
43. Cherkassky, B.V.; Goldberg, A.V.; Radzik, T. Shortest paths algorithms: Theory and experimental evaluation. *Math. Program.* **1996**, *73*, 129–174. [[CrossRef](#)]
44. Topkis, D.M. A k shortest path algorithm for adaptive routing in communications networks. *IEEE Trans. Commun.* **1988**, *36*, 855–859. [[CrossRef](#)]
45. Eppstein, D. Finding the k shortest paths. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science, Santa Fe, NM, USA, 20–22 November 1994*; pp. 154–165.
46. Iqbal, M.A.; Saltz, J.H.; Bokhart, S.H. *Performance Tradeoffs in Static and Dynamic Load Balancing Strategies*; NASA Langley Research Center: Hampton, VA, USA, 1986.
47. Cardellini, V.; Colajanni, M.; Yu, P.S. Dynamic load balancing on web-server systems. *IEEE Int. Comput.* **1999**, *3*, 28–39. [[CrossRef](#)]
48. Katevenis, M.; Sidiropoulos, S.; Courcoubetis, C. Weighted round-robin cell multiplexing in a general-purpose atm switch chip. *IEEE J. Sel. Areas Commun.* **1991**, *9*, 1265–1279. [[CrossRef](#)]
49. Saidu, I.; Subramaniam, S.; Jaafar, A.; Zukarnain, Z.A. A load-aware weighted round-robin algorithm for ieee 802.16 networks. *EURASIP J. Wirel. Commun. Netw.* **2014**, *2014*, 226. [[CrossRef](#)]
50. Moreira-Matias, L.; Fernandes, R.; Gama, J.; Ferreira, M.; Mendes-Moreira, J.; Damas, L. An online recommendation system for the taxi stand choice problem (Poster). In *Proceedings of the Vehicular NETWORKING Conference, Seoul, Korea, 14–16 November 2013*; pp. 173–180.
51. Kong, X.; Xia, F.; Wang, J.; Rahim, A.; Das, S.K. Time-location-relationship combined service recommendation based on taxi trajectory data. *IEEE Trans. Ind. Inf.* **2017**, *13*, 1202–1212. [[CrossRef](#)]
52. Qian, S.; Zhu, Y.; Li, M. Smart recommendation by mining large-scale gps traces. In *Proceedings of the 2012 IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, China, 1–4 April 2012*; pp. 3267–3272.
53. Pei, T.; Zhu, A.X.; Zhou, C.H.; Li, B.L.; Qin, C.Z. A new approach to the nearest-neighbour method to discover cluster features in overlaid spatial point processes. *Int. J. Geogr. Inf. Sci.* **2006**, *20*, 153–168. [[CrossRef](#)]

