City-wide examining transport network accessibility using taxi GPS data

Peer-reviewed author version

Cui, Jianxun; LIU, Feng; JANSSENS, Davy; Shi, An & WETS, Geert (2015)
City-wide examining transport network accessibility using taxi GPS data. In: Zhang, Yu; Yan, Xuedong; Yin, Yafeng (ed.). PROCEEDINGS OF THE 15TH COTA INTERNATIONAL CONFERENCE OF TRANSPORTATION PROFESSIONALS, p. 3588-3599.

DOI: 10.1061/9780784479292.332
Handle: http://hdl.handle.net/1942/20393

DOI: 10.1061/9780784479292.332
Handle: http://hdl.handle.net/1942/20393

City-wide examining transport network accessibility using taxi GPS data

Jianxun Cui[1] , Feng Liu[2,3], Davy Janssens[2] , An Shi[1] and Geert Wets[2]

[1] School of Transportation Science and Engineering, Harbin Institute of Technology (HIT), 150001, Harbin, China; PH (86) 451-86412866; FAX: (86) 451-86412866; email: cuijianxun@hit.edu.cn
[2]Transportation Research Institute (IMOB), Hasselt University, Wetenschapspark 5, bus 6, B-3590, Diepenbeek, Belgium; PH (0032) 1126-9125; FAX: (0032) 1126-9199; email: feng.liu@uhasselt.be
[3]Corresponding author

Abstract

This study has developed a method to city-wide identify all regions with poor transport accessibility, using GPS data generated by taxis. This approach is composed of four major steps, including travel pattern modeling, accessibility measure building, region with poor accessibility detecting, and specific problem analyzing.

Using GPS data collected from all taxis operating in the Chinese city of Harbin, the performance of this approach is tested. In total, 10 worst regions have been identified with accessibility measures reaching only 81.3% of the overall average across the city. A serious accessibility problem to filling stations has also been discovered, in which drivers from nearly 92.6% of the residential areas have to travel longer than 30 min to refill their vehicles. The experimental results demonstrate the potential and effectiveness of the proposed method in detecting accessibility problems and assisting policy makers in improving accessibility situations across the city.

1. INTRODUCTION

Accessibility is defined as the ease and extent to which land-use and transport systems enable individuals to reach activities and destinations by means of certain transport modes, e.g., the number of jobs accessible within 30 min by car or bus. While traditional transport network measures, e.g. congestion, capture how quickly people can move from one place to another; accessibility indicates how well people are able to get to desired destinations. Accessibility takes into account not only travel efficiencies but the distribution of land-use and activity locations across the transport network. It has been considered as a key dimension of quality of life and a priority for sustainable urban transport management and planning.

Over the last decade, there has been a growing number of studies on the development of accessibility measures; the typical process is as follows (e.g. Karou & Hull, 2014). (i) A geographical area is chosen as the *study region* for accessibility analysis. (ii) A set of relevant activity regions are represented with the size or number of the associated activity opportunities. (iii) The physical separation between the study region and each of the activity regions is measured, in terms of certain travel characteristics, mostly travel times. (iv) Based on the inferred travel characteristics, accessibility measures in various forms are computed. (v) The obtained measures can be used to evaluate how land use and transport network changes have affected accessibility situations during different time periods in history. Alternatively, accessibility measures have also been estimated based on various policy scenarios, providing insights into how different transport plans would lead to varied impacts on accessibility and inequity (in terms of accessibility) in the city.

However, despite the development of accessibility analysis techniques and their importance in supporting land use and transport decision-making, the data source which is used to derive travel times, raises concerns about the accuracy of the obtained travel characteristics, particularly regarding car-based travel. Typically, *travel surveys*, which document full daily activities and travel of a small sample of individuals during a time frame of one or several days, are utilized, regardless of the limitations inherent to the data collection method. Alongside, travel times have also been drawn from sensors, e.g. loop detectors and video cameras, which are installed in a road network to monitor traffic flow. However, due to the high installation cost, the sensors are usually set up on highways, as it is expensive to instrument a whole city with such static devices. Consequently, the traffic data is only limited to the high-capacity roads, and sheds little light on the traffic flow in the rest of the city.

Thus, due to the data constraints, the accuracy of the existing accessibility measures has been limited, leading to a certain level of deviation between what is revealed by the measures and what the actual situation is. In addition, the existing analysis is often restricted to only a statistical average day and a relatively small area as well as to a subset of the population, because of the lack of a large dataset that is spatially and temporally detailed and extensive as well as involves more individuals. Consequently, the results are difficult to be applied to evaluate accessibility at different time periods of a day (e.g. accounting for congestion), in various types of days (e.g. weekdays and weekend), at a higher geographical scale (e.g. an entire city), or in a whole population group. For a long time, data problems have been one of the essential challenges of the current research on transport accessibility analysis.

The advancement of the Global Positioning System (GPS) has created the opportunity to use this technology as a new data collection method to overcome the lack of reliable travel data. For travelers who carry GPS devices in pockets or with their vehicles, the accurate travel routes and travel times can be monitored automatically, providing detailed spatial and temporal travel information and near real time traffic conditions in the road network. Particularly, in many major cities around the world today, GPS devices are already installed in taxis originally for real-time dispatch systems to optimize client orders, thus no additional cost is incurred for the data collection. With the combined advantages, GPS-equipped taxis have the potential of collecting travel data across a large sample of the population and uncovering the traffic dynamic of the entire city. So far, the data has been explored, among others, for urban computing and travel demand modelling. Particularly, travel times derived from the data have been applied to dynamic routing and navigation tools. In a typical study (Gühnemann et al., 2004), GPS data is first collected from 132 and 212 taxis in Berlin and Vienna, respectively, over a period of more than two years. Based on the data, the average driving speed and travel times in each hour of the day at each major road link across the entire network of these two cities are then derived, using a routing algorithm, e.g. Dijkstra. The derived velocities are further compared against the speed drawn from static sensors (e.g. Coifman, 2002). For high-capacity roads, the GPS-based average velocities are 6.3 km/h lower than the sensor-based ones; while on low-level urban roads, the difference reaches more than 20 km/h. Meanwhile, the GPS-based velocities show considerable variation over different hours of the day and among road links, suggesting that the average speed and travel times derived from sensors deliver an unrealistic picture for each specific link and during specific hours of the day.

Nevertheless, despite the potential and multitude applications of taxi GPS data, the data has not been explored for accessibility analysis. Instead, travel times derived from GPS data collected from a few individuals' private cars, have been utilized for accessibility studies. In the research (Huang & Levinson, 2011), GPS data is used to analyze the impact of a location's accessibility on people' choices for non-work activities on the location. During the process, GPS data of 478 trips conducted by 4 individuals is used to infer travel times and distance, from which accessibility measures are derived. The study demonstrates the capacity of GPS data to record precise trips, and the effectiveness of using the data to estimate accessibility. However, the study solely focuses on how accessibility affects individuals' destination choices, and only the accessibility of a limited number of locations is analyzed. A method, which is based on GPS data while systematically examining the accessibility situations in the entire transport network as well as concentrating in the identification and investigation of all problematic regions, has so far been lacking.

This paper extends the current research on transport network analysis by means of taxi GPS data, and particularly addresses the above mentioned limitations with respect to the development of reliable methods for the city-wide examination of accessibility conditions. Specifically, a set of measures is developed for each of the regions of the city, based on taxi GPS data. The obtained measures can then be used to identify regions with poor accessibility, and thus assisting policy makers in seeking optimal solutions that best address the accessibility problems in these areas. Compared to traditional techniques, the proposed method offers the following advantages. (i) It analyzes transport accessibility situations across an extensive area of the city. (ii) It generates more objective and up-to-date measures, catching up with the fast pace of urban land development and population growth. (iii) This method accounts for varied traffic conditions (e.g. congestion), across different time periods of a day. (iv) It is a cost-effective approach and easily transferable to cities where taxis are installed with GPS devices. (v) Particularly in this study, GPS data recorded from all taxis licensed in the Chinese city of *Harbin*, is explored. The data provides a unique opportunity for the analysis on a significant share of individuals' trips across the city as well as for the examination into the accessibility situations across the entire urban network.

The remainder of this paper is organized as follows. Section 2 describes the taxi GPS data and Section 3 details the proposed methodology. A case study is carried out in Section 4 and Section 5 further compares the experimental results against a baseline model. Finally, Section 6 ends this paper with major conclusions and discussions.

2. DATA DESCRIPTION

The GPS data was collected between July and September in 2013 from all taxis (i.e. more than 16,000 in total) licensed in *Harbin*. All of the taxis are equipped with GPS devices as a part of security measures to protect drivers from being attacked. The devices record the positions of the vehicles every 30 seconds during the day and 2 min (minutes) at night, generating data of 1.6G in size each day. Apart from the GPS data, the geographic positions and types of all activity locations in the city are also utilized, and all the locations are classified into 18 types.

According to the GPS records, the average number of passenger trips for a taxi is 30/day. All the 16,000 taxis thus generate a total of 0.48 million passenger trips daily in the urban area. In comparison, the city has roughly 1.0 million private cars; if we assume of 2.41 trips per day for each private car according to the statistics ("The change in travel behavior in WuHan", 2009), all the private vehicles produce 2.41 million trips. As a result, the taxi passenger trips account for 16.6% of the total personal trips made by both taxis and private cars within the urban area. Although the total number of taxis is much less than that of private cars, taxi trips represent a significant share of individuals' trips, and they can thus be considered as a good indication of the real travel demand patterns in the city.

3. METHODOLOGY

3.1. Overview of the approach

The method is composed of 4 major steps, including (i) modeling city-wide taxi passenger travel patterns, (ii) building accessibility measures for each region in the city, (iii) identifying regions with poor accessibility, and (iv) further examining the specific transport situations of the problematic regions. Prior to the steps, a preliminary step is conducted for raw GPS data processing.

3.2. GPS data processing

A GPS trajectory from a taxi during a day can be represented as $p_1\ (l_1,\ t_1,\ s_1)\ldots p_n\ (l_n,\ t_n,\ s_n)$, where n is the *length* of the sequence, i.e. the total number of the GPS points. Each $p_k(k=1,\ldots,n)$ denotes a point, consisting of a time stamp t_k, a latitude and longitude coordinate set $l_k=\{x_k,\ y_k\}$, and a status message s_k indicating if the taxi is 'occupied' by clients or 'empty' when the taxi driver is looking for clients. The GPS data are first processed to remove error records, and passenger trips are then identified based on s_k. Let $p_b\ (l_b,\ t_b,\ s_b)\ldots p_e\ (l_e,\ t_e,\ s_e)$ as a passenger trip, the travel time of the trip, i.e. D_OD_{Trip}, is obtained as $D\_OD_{Trip} = t_e - t_b$.

3.3. City-wide taxi passenger travel pattern modeling

The entire study city is divided into disjoint regions using a grid-based partition method, with the number of grids as *Grid_X* and *Grid_Y* along the latitude and longitude directions, respectively, generating a total of $Grid\_X \times Grid\_Y$ regions. Each region r_i ($i=1,\ldots,\ Grid\_X \times Grid\_Y$) can also be formulated as a set of double numbers specifying the grid positions of the region, denoted as $Reg(i_x,\ i_y)$ ($i_x=1,\ldots,Grid\_X$ and $i_y=1,\ldots,Grid\_Y$). A day is classified into different types (i.e. weekdays, weekend and public holidays) and various time slots. Under the spatial and temporal division, two matrices are constructed. The first one is a travel demand model, represented as $OD(r_i, r_j, TimeP, Day, DayT)$, with each matrix element accommodating all trips that leave from region r_i, end in region r_j, and start within time period *TimeP* on day *Day* with type *DayT*. Two features are then extracted for each of the matrix elements, including the total number of the *OD* trips, i.e. m_OD_{ij}, and the average travel time over the trips, i.e. u_OD_{ij}. The variable u_OD_{ij} is computed as

$$\mathrm{u\_OD_{ij}} \hat{=} \mathrm{u\_OD_{ij}}(\mathrm{r_i}, \mathrm{r_j}, TimeP, Day, DayT) = \frac{\sum_{Trip=1}^{m\_OD_{ij}} (D\_OD_{Trip})}{m\_OD_{ij}} \tag{1}$$

The second matrix, referred as $OP(r_i, r_j, TimeP, Day, DayT)$, features the travel path of the trips, with each matrix cell containing all intermediate trips that only pass regions r_i and r_j but do not start or end there. Two variables are also extracted for each of the cells, including the total number of the intermediate *OP* trips, i.e. m_OP_{ij}, and the average travel time over the trips, i.e. u_OP_{ij} , as

$$\mathrm{u\_OP_{ij}} \hat{=} \mathrm{u\_OP_{ij}}(\mathrm{r_i}, \mathrm{r_j}, TimeP, Day, DayT) = \frac{\sum_{Trip=1}^{m\_OP_{ij}} (D\_OP_{Trip})}{m\_OP_{ij}} \tag{2}$$

Where, D_OP_{Trip} refers to the travel time for an *OP* trip. Specifically, let $p_b\ (l_b, t_b, s_b)$ $\to \dots\ p_{k1}\ (l_{k1}, t_{k1}, s_{k1})\ \dots\ p_{k1+m1}\ (l_{k1+m1}, t_{k1+m1}, s_{k1+m1}) \dots\ p_{k2}\ (l_{k2}, t_{k2}, s_{k2})\ \dots\ p_{k2+m2}\ (l_{k2+m2}, t_{k2+m2}, s_{k2+m2})\ \dots \to p_e\ (l_e, t_e, s_e)$ represent a passenger trip, in which $\{l_{k1}, \dots, l_{k1+m1}\} \in r_i$, $\{l_{k2}, \dots, l_{k2+m2}\} \in r_j$, $m1>=0$, and $m2>=0$. D_OP_{Trip} for the intermediate trip passing r_i and r_j is estimated as $D\_OP_{Trip}=(t_{k2}+t_{k2+m2})/2-(t_{k1}+t_{k1+m1})/2$.

Based on the matrix for each day, the average travel time between two regions can be derived from all the matrices corresponding to all the survey days. Let $M\_OD_{ij}(r_i, r_j, TimeP, DayT)$ represent the total number of days of *DayT* when the number of trips, i.e. m_OD_{ij} , is higher than a certain threshold value, defined as TH_{MOD}. If $M\_OD_{ij}>0$, the average travel time, i.e. U_OD_{ij}, can be computed as

$$U\_OD_{ij} \hat{=} U\_OD(r_i, r_j, TimeP, DayT) = \frac{\sum_{Day=1}^{M\_OD_{ij}(r_i, r_j, TimeP, \mathrm{DayT})} u\_OD_{ij}(r_i, r_j, TimeP, Day, DayT)}{M\_OD_{ij}(r_i, r_j, TimeP, \mathrm{DayT})} \tag{3}$$

On the contrary, if $M\_OD_{ij} = 0$, none of the days when m_OD_{ij} is higher than TH_{MOD}, have been observed from r_i to r_j in *TimeP* throughout the survey period, the travel time is approximated based on the *OP* trips, i.e. U_OP_{ij}, as

$$U\_OP_{ij} \hat{=} U\_OP(r_i, r_j, TimeP, DayT) = \frac{\sum_{Day=1}^{M\_OP_{ij}(r_i, r_j, TimeP, \mathrm{DayT})} u\_OP_{ij}(r_i, r_j, TimeP, Day, DayT)}{M\_OP_{ij}(r_i, r_j, TimeP, \mathrm{DayT})}. \tag{4}$$

Where, $M\_OP_{ij}(r_i, r_j, TimeP, DayT)$ denotes the total number of survey days when m_OP_{ij} is higher than TH_{MOD}.

In the travel time estimation process, we assume that, although a few of taxi drivers might give passengers a roundabout trip, most of them are honest and typically find out the fastest route to send passengers to destinations based on their knowledge. The travel time between two regions consists of the actual driving time on road links and the waiting time at traffic lights. All the times are dependent on the actual driving routes. When choosing a route, besides the distance of the route, drivers also consider other factors, such as the time-variant traffic flows on the roads, traffic signals and direction changes. Thus, the travel times of the passenger trips can properly represent the actual duration of trips in the network. In addition, parameter TH_{MOD} is defined to select the days when sufficient number of OD trips have been observed, in order to ensure the representation and accuracy of the derived travel times. For region pairs that do not undertake enough OD trips on any day across the entire survey period, the travel times are estimated with the OP trips that only pass the regions but do not originate or destine there.

3.4. Accessibility measure building

To detect accessibility problems, we first identify regions that generate and attract trips in the morning and night periods, respectively, more than a threshold value, defined as TH_M. The obtained places would represent high density of residential areas and are thus chosen as the study regions. The large volume of travel also ensures the estimated travel times more accurate and better representative of the general travel conditions. Alongside, all the activity locations in the city are projected into regions based on the positions of the locations, and the obtained regions are used as activity regions.

Various methods have been developed to compute accessibility measures (e.g. Geurs & Wee, 2004). The location-based method including the *contour measure* and *potential measure*, which incorporates the spatially distribution of activities and travel time constraints, is adopted in the current study. A contour measure describes the total number of activities that could be reached from a region within a certain time. Specifically, the measure for region r_i within travel time TH_T over all activity types, i.e. AC_i, or for each individual activity type c, i.e. AC_{ci}, can be obtained as

$$
\begin{aligned}
AC_i &= \sum_{c} AC_{ci} \\
AC_{ci} &= \sum_{U_{ij} < TH_T,\, j=1}^{\mathrm{TotalOfActReg}} (a_{cj}).
\end{aligned}
\tag{5}
$$

Where, U_{ij} is the average travel time between r_i and r_j, and U_{ij}={U_OD_{ij} or U_OP_{ij}}; a_{cj} is the number of activities of type c in region r_j; *TotalOfActReg* is the total number of activity regions in the city.

A potential measure uses an impedance function, i.e. $f(U_{ij})$, to reflect the declining attractiveness of activities at a destination region as travel times between the two corresponding regions increase. Let AP_i denote as the measure for region r_i and AP_{ci} as the measure for each type c in region r_i; they can be computed as

$$
\begin{aligned}
AP_i &= \sum_{c} AP_{ci} \\
AP_{ci} &= \sum_{j=1}^{TotalOfAct\mathrm{Re}g} a_{cj}\, f\left(U_{ij}\right) \\
f\left(U_{ij}\right) &= e^{-\beta U_{ij}}
\end{aligned}
\tag{6}
$$

Where β is a sensitivity parameter to travel times. With values ranging from 0 to 1, β controls the effect of travel time changes on the attractiveness of the activities.

Both contour and potential measures evaluate accessibility at regional levels. They consider all people living in a region as a whole, and assume that people have a set of social and economic activities which need to be met at different destinations. The measures are expressed with the quantity of the activity locations which can be approachable within a certain time limit, and they are determined by the travel distance between the study region and the activity regions as well as the quality of the transport infrastructure linking these places. A low value of the measures signals a problem of long time travel in order to reach activity destinations, due to long travel distance and/or bad traffic conditions, e.g. congestion.

3.5. Regions with low accessibility detection

Based on the obtained measures, the regions, which have low accessibility to activities at a specific time period of a day, are ultimately identified.

3.6. Specific land use and transport problem examination

In order to further investigate the specific accessibility problems of the detected regions, an in-depth examination into the land use and transport situations surrounding the regions is conducted.

4. CASE STUDY

In this section, adopting the proposed approach and using both the GPS and land use data, we carry out a case study. In this process, travel pattern models are first constructed and accessibility measures are then developed. Next, the poor accessibility regions are detected and the specific problems are analyzed.

4.1. Travel pattern models

During the model construction, the entire city is divided into regions; the parameters Grid_X and Grid_Y decide the total number of the study units. The larger these values are, the higher the spatial resolution reaches, but the less the number of observed trips between the regions is. In order to derive results that are statistically sound and representative, these two variables are set as 40 respectively, resulting in a total of 1600 regions with each being 1.87 km2 in size. The day is segmented into 4 periods including 7-8:30am, 8:30am-16pm, 16pm-18pm and 18pm-7am, i.e. morning, daytime, evening and night periods, respectively, according to the distribution of travel speed observed from the GPS data. In this case study, only the accessibility in the morning period of weekdays is analyzed; the similar process can be applied to other periods as well as to the weekend.

In the morning of each weekday, two matrices including *OD* and *OP* are constructed. Fig. 1(a)-1(b) describe the distribution of the travel times of all the 99 *OD* and 281 *OP* trips from regions *Reg(23, 24)* to *Reg(24, 25)* on a certain day. The mean travel times for the *OD* and *OP* trips are 13.4 min and 12.3 min, respectively, with the *OP* trips being 1.1 min shorter. This reflects the fact that people usually take highways when passing a region while have to go through low capacity roads inside a region in order to leave or reach a specific location in the area.

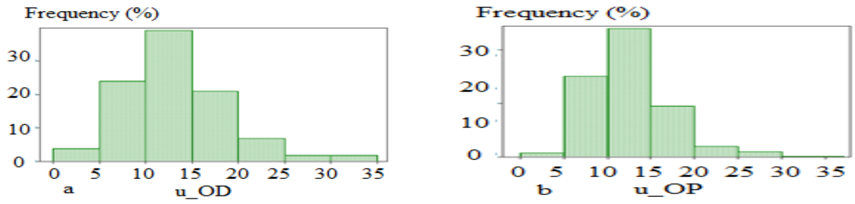

Fig. 1. The distribution of travel times of OD trips (a) and OP trips (b)

4.2. Building accessibility measures

Out of all the 1600 regions of the city, 241 (15.1%) original and 433 (27.1%) destination regions generate and attract at least one trip per day in the morning and night periods, respectively. Among these regions, 233 undertake both types of functions, from which 108 are chosen as the study regions under the setting of $TH_M=20$. Alongside, 241 regions accommodate at least one activity location, with the average number of activities per region as 16.8. Fig. 2(a)-2(c) describe the geographic distribution of the original regions that generate morning trips and the destination regions that absorb night travel, as well as the distribution of the activity regions, respectively. Similar spatial distribution among the three types of regions is observed. The correlation coefficient between the number of the morning or night trips a day and that of activity locations in a region is 0.85 and 0.83, respectively. The high coefficients imply a high level of association between the passenger travel demand of a region and the number of activities it accommodates, further suggesting that the passenger travel models, derived from all taxis running in the city, can sufficiently represent mobility demand patterns for activities across the city.

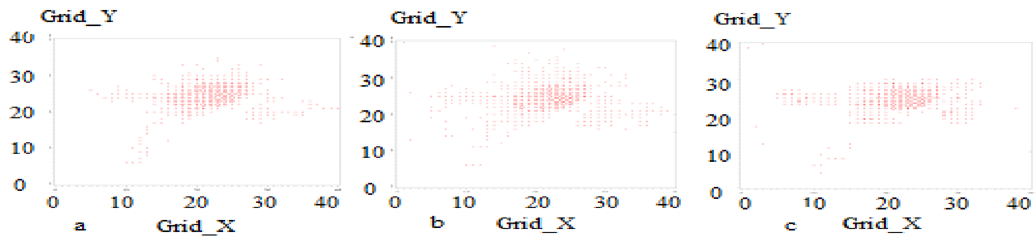

Fig. 2. The distribution of original regions (a), destination regions (b) and activity regions (c)

Out of all the 26,028 pairwise combinations formed by the 108 study and 241 activity regions, 18,904 pairs (i.e. 72.7%) are filtered out from the *OD* matrix by TH_{MOD} specified as 3. U_OD_{ij} is used for the travel time between the two corresponding regions. Regarding the remaining pairs, the travel times are estimated with U_OP_{ij}; all these region combinations meet the condition of $M\_OP_{ij} > 3$.

In the calculation of accessibility measures, different cut-off values for TH_T and β have been used, depending on the type of considered activity types and travel modes (e.g. Anderson et al., 2013). In this experiment, TH_T is chosen as 30 min and β as 0.1. Fig. 3(a)-3(b) describe the distribution of the obtained contour and potential measures respectively. Considerable variation across different regions is noted. For example, while the average contour measure is 14,779, the minimum and maximum are 9,372 and 15,735, respectively. The top 10 and 20 low measure regions have the average as 12,012 and 12,690, i.e. 81.3% and 85.9% of the overall average across all regions, reaching merely 74.9% and 79.1% of the total activity locations in the city in 30 min, respectively. In addition, the correlations between the contour and potential measures as well as between the two contour measures with TH_T as 30 min and 40 min respectively, for a same region, are 0.88 and 0.92, implying that the respective pair of measures leads to comparable results.

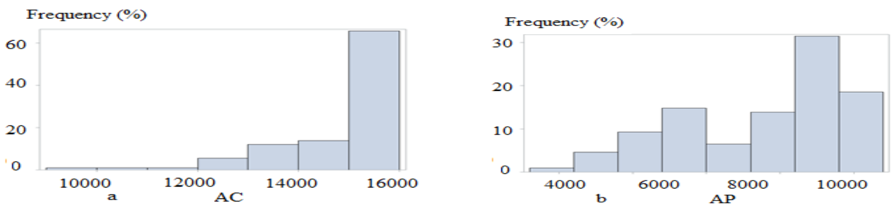

Fig. 3. The distribution of contour measures (a) and potential measures (b) across regions

4.3. Identifying regions with poor accessibility

Fig. 4(a) shows the spatial distribution of the first 20 poor accessibility regions, demonstrating that most of these regions are located outside the city center. The detailed information on the top 10 of these regions are also presented in Table 1.

Table 1. The 10 worst regions in the morning period of weekdays[a]

Rank	Reg	N	Contour	Potent	School/ university	Shop	Hospital	Factory/ company	Filling station
1	8,24	30	9372(58.4)	2052	1 (49.8)	1 (46.6)	1(53.5)	1(61.4)	1(0)
2	14,20	20	10895(67.9)	2733	2 (63.3)	2(66.9)	3(68.2)	5(69.6)	1(0)
3	16,19	20	11765(73.3)	3022	3(77.8)	4(78.9	2(79.9)	3(73.2)	1(0)
4	17,19	21	12341(76.9)	3550	4(80.3)	3(81.5)	4(71.2)	2(74.5)	1(0)
5	28,19	29	12347(77.0)	3867	5(84.0)	7(77.7)	5(73.6)	4(76.6)	1(0)
6	16,21	29	12493(77.9)	3982	6(74.3)	10(89.1)	6(76.9)	6(68.9)	1(0)
7	27,29	23	12496(77.9)	4179	7(78.1)	5(70.0)	7(87.2)	8(79.3)	1(0)
8	20,19	21	12518(78.0)	4026	9(72.3)	9(74.2)	10(82.0)	7(80.5)	1(0)
9	23,29	21	12822(78.0)	5048	11(72.9)	8(74.7)	9(73.3)	11(81.7)	1(0)
10	22,29	24	13073(81.5)	4856	12(75.7)	6(77.3)	8(74.6)	13(75.6)	1(0)

[a] The columns from left to right represent the following features of a region: the rank, the position, the number of average trips generated in the morning per day, the contour measure and potential measures, the contour measures to individual activity types of school/university, shop, hospital, factory/company and filling station, respectively. The number in bracket indicates the percentage of the contour measure relative to the total number of activities of all types (for the 4th column) or the corresponding individual type (for the 6th -10th columns).

4.4. The specific land use and transport situations of the problematic regions

To further analyze the specific problems of the detected regions, we examine a region with the typical accessibility problem. This region, i.e. *Reg(8,24)*, located in the western part of the city as indicated with the *rank* as *1* in Fig. 4(b), generates average 30 trips in the morning per day. It is the region with the lowest measure, and reaches only 58.4% of the total activities of the city within 30 min. Distance is observed between the accessible area within 30 min from this region, as indicated by the black line, and the city center where most activities are established, as delimited by the red oval in Fig. 4(b). The average driving speed from this region towards its eastern direction is only around *16.9* km/h, which hampers people quickly reaching the high density of activity area. According to the Euclidian distance between this region and the city center, in order to reach the activity area within 30 min, a travel speed that must be higher than 24.64 km/h is thus required.

The accessibility problem of this region is also manifested by the difficulties to reach most of the individual types of activities. For instance, among all 18 types, it has the lowest measure to 16 classes, except ‘financial center’ and ‘public place’. In particular, the measure is 0 under both 30 and 40 min thresholds to filling stations, suggesting that no filling stations are reachable from this region within 40 min. A further investigation discloses that, among all 147 filling stations provided by the city, 139 (i.e. 94.6%) are outside the urban area and only 8 are scattered on the edge of the city, as demonstrated by small black rectangular in Fig. 4(b). This leads to general poor accessibility to filling stations across the city, with only 8 regions (i.e.7.4%) capable of getting to a certain station in 30 min by car. Drivers in most parts of the city have to travel a long time in order to refill their vehicles.

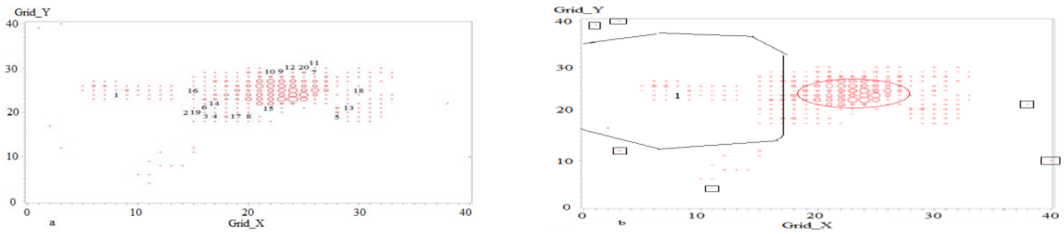

Fig. 4. The distribution of the first 20 poor accessibility regions (a) and the specific situation of the worst region (b)

Note: the number represents the rank of the regions in the decreasing order of the measures.

5. COMPARING THE RESULTS WITH A BASELINE MODEL

To examine the accuracy of this approach, we compare the results derived from the study with a baseline model which only considers the spatial direct distance and the typical travel speed. Specifically, we calculate the expected travel time between 2 regions, i.e. U_OE_{ij}, as $ED(r_i, r_j)/Speed_E$. Where, $ED(r_i, r_j)$ is the Euclidian distance between the centroids of regions r_i *and* r_j, and $Speed_E$ is the typical travel speed in the morning of weekdays which is set as 21.42 km/h as derived from the GPS data. Based on U_OE_{ij}, the contour measure, i.e. AE_i, is calculated according to Formula (5). Fig. 5 demonstrates the compared results derived from the baseline model and the proposed method.

In contrast to the proposed method, the baseline model uses the shortest distance between two regions, the obtained travel times are anticipated to be shorter, thus leading to a higher measure. The characteristics are well reflected in this figure. For instance, in Fig. 5(a), the travel times U_OD_{ij} are longer than the expected times U_OE_{ij} for all the region pairs, with the average difference between them as 8.95 min. While in Fig. 5(b), AE_i is higher than AC_i for all study regions, with the coefficient between these two measures as 0.73. However, although $U\_OD_{ij} > U\_OE_{ij}$ and $AE_i > AC_i$ for all regions, the level of the differences varies across regions, reflecting the fact that the actual travel route and travel speed are region-dependent. A universal travel speed and theoretical distance used in the baseline model would not capture the region specific characteristics. Instead, the proposed method based on GPS data is able to account for the actual travel routes that are decided based on various factors, such as route directness and traffic conditions.

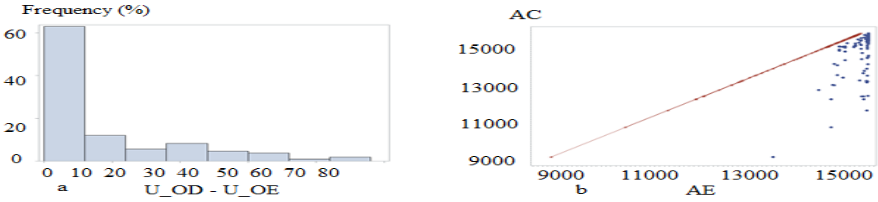

Fig. 5. The distirbution of the differences between U_OD_{ij} and U_OE_{ij} (a), and the correlation between AE_i and AC_i (b)

6. CONCLUSIONS AND DISCUSSIONS

The achievement of good transport accessibility and equity in the distribution of urban services is one of the supreme goals for transport managers and urban planners. As a city grows and attracts more immigrants, the accessibility issue has become more important than ever, particularly in many emerging countries. To meet the growing challenge, we have developed an approach which measures accessibility to different urban services and activities of the city by auto. This method responds to several limitations in existing accessibility analysis techniques, and is particularly tailored to the emerging world. It offers an alternative and practical tool to help policy makers in identifying the current accessibility problems and designing transport infrastructure and service plans towards better accessibility and reduced inequity across the city.

The significance of this method mainly lies in the use of the massive taxi GPS data as well as the results derived from this utilization. As previously described, both conventional travel diaries and traffic monitoring systems using sensors are unable to produce reliable information on travel times within the network. Thus, the floating car data technique (FCD) has moved into the focus of current research, in which travel data are obtained from monitoring vehicles which flow with the traffic. However, the FCD technique raises two major challenges, including: (i) requiring a sufficiently large number of sample vehicles, which is estimated at about 1% of the total fleet (burr & Simmons, 2002); (ii) the high communication costs for data transmission. In order to overcome these problems, an alternative FCD technique by using data from taxis has been proposed, which presents a solution to both the critical mass problem of probe vehicles and the financial concern. To further explore the potential of the taxi-based FCD technique, this current study has used GPS data collected from all taxis operating in a city, for the development of an accessibility analysis method.

The experiment on the designed method shows the major strength of this approach. (i) All 108 residential regions across the city are ranked by the obtained measures, from which the 10 worst regions are revealed. (ii) The specific problems of the detected regions are examined. For instance, for the city's poorest accessibility region, it is found that the assessable area in 30 min of driving does not reach the city center where most of the activities are established. (iii) The detailed examination into each individual activity categories discloses that only 7.4% of the study regions are able to get to a certain filling station in 30 min by car.

Nevertheless, despite the effectiveness of this approach, there are still certain areas for future improvement. First, while the current study uses the average travel time for the estimation of accessibility measures, it ignores the probabilistic distribution of the travel times (*see* Fig. 1). Instead of using a single average value, the distribution of the travel times could be integrated into the calculation of the measures. Secondly, the settings of several parameters (e.g. TH_M and TH_{MOD}) defined in this method affect the statistical significance and representative of the derived results; in the meantime they are subject to the constraint of the GPS sample size. A detailed examination into the correlation between different parameter settings and the resultant measures would be important to tune up the best parameter combinations. Thirdly, this study solely focuses on car travel, targeting a specific group of people. Accessibility can also be analyzed for other types of modes, or for a multi-modal transport system in which the choices between various modes are considered. The accessibility analysis using bus GPS data or combined GPS data gathered from a variety of modes would definitely improve the current method.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the Urban China Initiative (UCI) through the UCI Grant 2013. This research is also supported by the National Natural Science Foundation of China (Project No. 71203045), Heilongjiang Natural Science Foundation (Project No. E201318) and the Fundamental Research Funds (Grant No.HIT.KISTP.201421).

REFERENCES

Anderson, P., Levinson, D., and Parthasarathi, P. (2013). “Accessibility Futures.” *Transactions in GIS*, 17(5), 683–705.

Burr, J., and Simmons, N. (2002). “Floating Vehicle Data – Number of Probes for Real Time Coverage.” *ITS World Congress*, Chicago.

Coifman, B. (2002). Estimating travel times and vehicle trajectories on freeways using dual loop detectors. Transportation Research A, 36A(4):351–364.

Geurs, K. T. and Wee, B. V. (2004). “Accessibility evaluation of land-use and transport strategies: review and research directions.” *Journal of Transport Geography*, 12, 127–140.

Gühnemann, A., Schäfer, R. P., and Thiessenhusen, K. U. (2004). “Monitoring Traffic and Emissions by Floating Car Data.” *Institute of Transport Studies*, Working Paper, Issue ITS-WP-04-07.

Huang A., and Levinson, D. (2011). “Accessibility, network structure, and consumers destination choice: a GIS analysis of GPS travel data.” *University of Minnesota: Nexus Research Group*, Working Papers 000095.

Karou, S., and Hull, A. (2014). “Accessibility modelling: predicting the impact of planned transport infrastructure on accessibility patterns in Edinburgh, UK.” *Journal of Transport Geography*, 35, 1–11.

“The change in travel behavior in WuHan”. 武汉市民出行方式显著改变. (2009). http://news.sohu.com/20090213/n262220111.shtml.