Exploring Human Spatio-Temporal Travel Behavior Based on Cellular Network Data: A Case Study of Hangzhou, China

Chen, H.,[1] Xu, C.,[2] Raghavan, V.[3] and Song, X.[1,4*]
[1]College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 10049, China
Email: chenhao214@mails.ucas.edu.cn, xfsong@ucas.ac.cn
[2]Chinese Academy of Surveying and Mapping, Beijing 10083, China
[3]Graduate School for Creative Cities, Osaka City University, Osaka 558-5858, Japan
[4]State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China
** Corresponding author*

Abstract:
Rapid urbanization has burdened urban infrastructures and traffic. Understanding the travel behavior is a potential solution to traffic problems. The traditional travel surveys are trapped in the cost and update frequency. The advent of ubiquitous cellular network data offers opportunities to uncover the travel behavior in a novel way. This study presents the implement of cellular network data on the travel behavior study. Specifically, we proposed the method to detect individual stay points in different time and spatial scales. The inbound and outbound travel flows of every urban district were estimated to reveal its travel characteristic. The results are reached based on the real-world data of Hangzhou, China. The distribution of detected homes is compared with census data and show a consistency of 0.88. About 61.1% of Hangzhou people commute less than 5 km, and urban periphery areas travel longer than urban center areas. The inbound-dominated and outbound-dominated sub-districts is characterized based on the proportion of inbound and outbound trips. The results proved the potential of deriving knowledge from cellular network to serve urban infrastructure planning and management.

1. Introduction

With a development of urbanization, metropolitan cities are experiencing a steady growth in the share of population and vehicles in China. As a result, various traffic problems, such as traffic congestion, traffic accidents, have emerged gradually. Understanding human travel behavior is an essential and primary task to solve road traffic issues. Travel surveys, which contain rich semantic information, are the most adopted data source for travel behavior studies, but they are often plagued by high costs and low update rate. With the development of information and communication technologies (ICT), we have witnessed an unexpected outbreak of pervasive systems, such as WiFi, Bluetooth, GPS and cell phones. These systems allow us to capture the spatio-temporal trajectory of human mobility at any time, which provided alternative opportunities for the research of travel behavior. The most widely used spatio-temporal data is GPS trajectory data, and applications that evaluated with these data have been conducted in traffic information estimation (Work et al., 2010, Keler et al., 2017 and Wang et al., 2017) and travel state detection (Zhang et al., 2011, Gidófalvi and Yang 2015 and Houbraken et al., 2017).

Compared to traditional travel survey data and GPS trajectory data, cellular network data provides another important chance to monitor individual mobility due to the high penetration in population and large coverage in geographic areas. The significance of this data has firstly been highlighted on human mobility laws and predictability. Based on cellular network data, studies have proven that human mobility behaviors show a high degree of temporal and spatial regularity (González et al., 2008), which can be predicted with a high

Online ISSN 2673-0014/

probability (Song et al., 2010). Meanwhile, cellular network data have shown the potential for travel behavior studies. As the most significant places in individual travel behavior, stay points can be extracted with various trajectories mining method (e.g. frequency method, spatial clustering and density estimation) based on cellular network data (Isaacman et al., 2011, Hoteit et al., 2014 and Jiang et al., 2016). When considering the dynamic behavior of mobile phone users, the travel behavior studies fall into two categories. The first one is to mine the most path for collective users based on the trip assignment approach (Toole et al., 2014, 2015 and Gundlegård et al., 2016). Another one is to accurately predict individual mobile path utilizing map matching methods (Algizawy et al., 2017 and Luo et al., 2017). To explore the origin of heterogeneity in travel patterns identified from cellular network data, some scholars examine the influencing factors that govern human movement. For example, Yuan et al., (2012) have proved that explanatory factors of age and gender, social temporal orders and built environment impact individual activity behavior. Similarly, some other factors have been investigated in the studies of Calabrese et al., (2013), Csáji et al., (2013) and Yang et al., (2016).

Constructing high-quality origin-destination (OD) information (White and Wells, 2002, Calabrese et al., 2011 and Morency et al., 2015) is a basic step for the applications of travel behavior. Tolouei et al., (2017) demonstrated that the OD matrix generated from cellular network data does not seem to be either biased or less accurate than conventional methods once systematically refined or adjusted. Furtherly, travel flow distribution can be predicted on the aggregated level serving for traffic planning and management (Mario B. Rojas et al., 2016, Sadeghvaziri et al., 2016 and Ni et al., 2018). Besides, travel behavior analysis also results applications in job-housing spatial distribution analysis (Yuan and Raubal, 2016), land-use classification (Pei et al., 2014), and tourists travel behavior evaluation (Phithakkitnukoon et al., 2015). These studies and applications provide useful insights into human travel behavior, but they were mostly conducted from macroscopic views, and detailed human behavior explorations are rare.

This study intends to explore the potential of using cellular network data to infer human travel behavior from a microscopic view. We first detected individual trips by identifying stay points and then analyzed urban travel flows by the statistics on trips. For stay points detection, our main contribution lay in the involvement of temporal information, which leads to exploring travel behavior on multiple spatial and temporal scale. By estimating the inbound and outbound travel flow, we are able to reveal the travel characteristic for urban space. A case study was finally conducted in Hangzhou, China to validate the proposed approach.

2. Dataset and Case Study Area

2.1 Dataset

The dataset used in this study are collected from Hangzhou mobile cellular network during a sampling period of a whole week in August 2014 (4^{th} Aug-10^{th} Aug). The network has 10 thousand base transceiver stations (BTSs) (Figure 1). The registered users in this cellular network are beyond 8 million. Each data record consists of time and location information for an anonymous user, including:

- The user's identification number (ID), which is encrypted for the privacy issue;
- The timestamp of communication event;
- The occupied BTS, which is used for user localization; and
- The type of communication event, e.g. paging, short message service (SMS), location update (LU), handover (HO) and so on.

Individual trajectory is a chronological sequence of a user's records. The location information in the trajectory is assigned with the coordinate of the occupied BTS. Indeed, the user could be anywhere of signal propagation area around the occupied BTS. Therefore, there will be a location error between user actual location and BTS location. The location error will vary from several hundred meters in downtown area to a few kilometers in suburban and rural area, depending on the spatial distribution of BTSs.

2.2 Study Area

The downtown area of Hangzhou is selected as study area, which is consisted of 8 districts, i.e. YuHang (YH), GongSu (GS), XiaCheng (XC), ShangCheng (SC), BinJiang (BJ), JiangGan (JG), XiaoShan (XS) and XiHu (XH), or 97 sub-districts

(Figure 1). According to the results of 2010 Chinese Census taken by the National Bureau of Statistics, the residential population of this region is 5.98 million, 68.7% of the total population in Hangzhou.

3. Methodology

Our goal is to understand individual and collective travel behavior based on individual trajectory. We are interested in analyzing individual trajectory utilizing the space-time method to detect the stay points of user movements. Once we got the stay points, we can identify the origins and destinations and extract the trips between them. For travel behavior analysis, we need to aggregate trips of collective users and obtain inbound and outbound trips for regions.

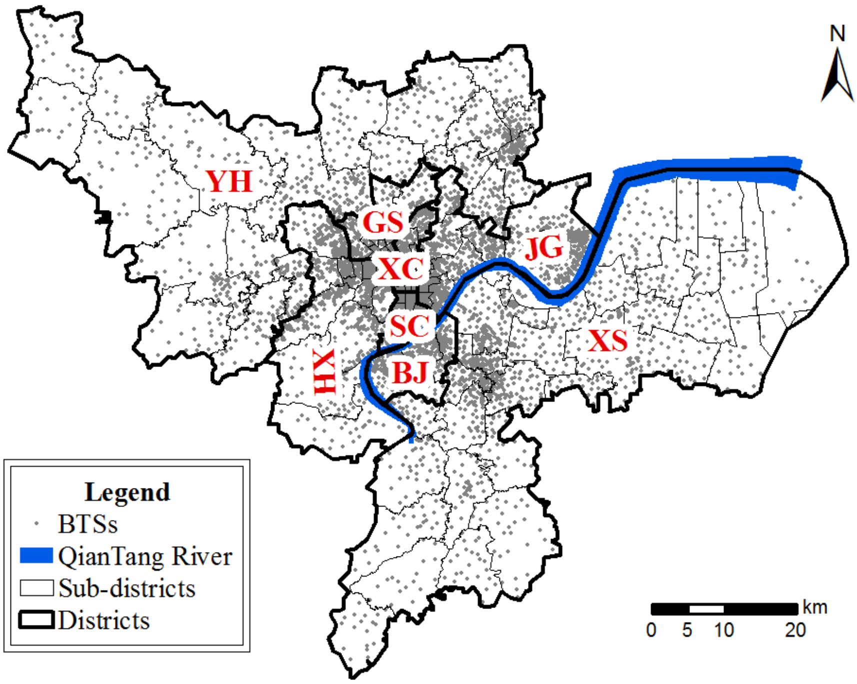


Figure 1: BTSs distribution in Hangzhou

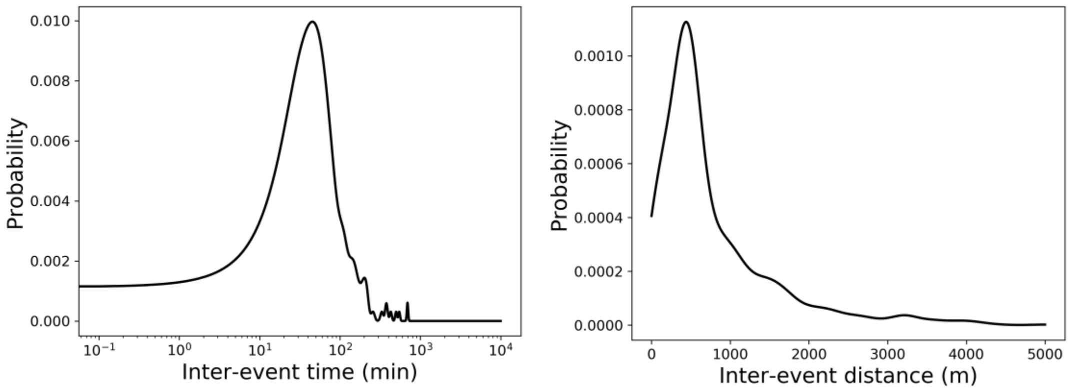
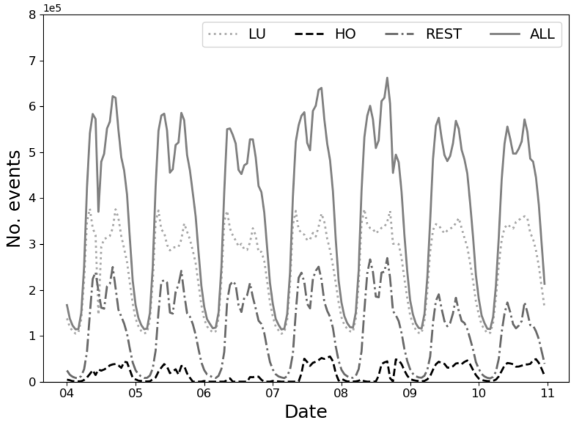


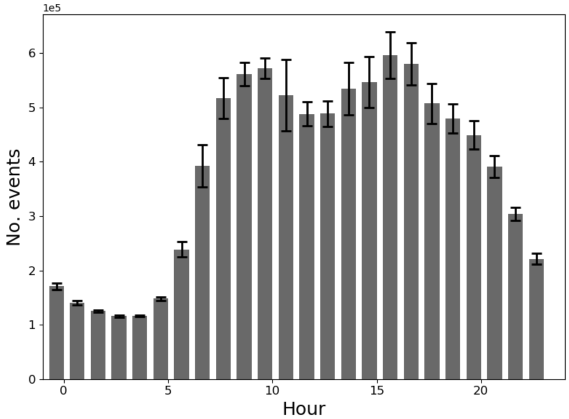
Figure 2: Distribution of inter-event time and inter-event distance

3.1 Characteristics of Dataset

We first explore the temporal and spatial characteristics of dataset. The inter-event time and inter-event distance, i.e. the time and distance interval between two consecutive trajectory points, are calculated. Figure 2 shows the distribution of inter-event time and inter-event distance of all users. The average value of inter-event time and inter-event distance are 42 min and 627 meters, which allow the detection of change in locations where the user stops for as little as 0.7 hours and 0.6 kilometers. The temporal resolution of dataset is much higher than that of Calabrese et al., (2013) and Gundlegård et al. (2016) because our dataset has more kinds of communication events. The time bias is also needed to be considered when using cellular network data. We count the number of events by hour and day. Figure 3a shows the change in quantity for different kinds of events during the sampling period. It is depicted that, no matter on workdays or at weekends, the quantities of all types of events presents the similar variation tendencies. Figure 3b shows the average number of total events at one-hour interval, together with the error bars showing one standard deviation. It can be seen that the mobile phones are more activate during 8am-8pm than 8pm-8am. Since the available users satisfied for trip extraction are biased at different time, we will scale up to the whole travel behavior from the detectable users.



(a)



(b)

Figure 3: Distribution of mobile phone activity by (a) daily and (b) hourly

3.2 Stay Points Detection

Stay points are the significant and meaningful places where a person spends most time or visits frequently, such as home, workplace, restaurant and hospital. A common method for the detection of these locations is 2D kernel density estimation (KDE). However, the generated 2D heat map is inconvenient for time related analysis. Considering the indispensability of temporal information in human travel behavior analysis, we adopt space-time kernel density estimation (STKDE) algorithm to detect individual stay points. STKDE is the time extension of tradition kernel density estimation. Since the temporal information is involved in STKDE, the travel pattern can be identified from a spatio-temperal perspective. The space-time density of individual trajectory is estimated as:

$$\hat{f}(x,y,t) = \frac{1}{nh_s^2h_t}\sum_i K_s\left(\frac{x-x_i}{h_s},\frac{y-y_i}{h_s}\right)K_t\left(\frac{t-t_i}{h_t}\right)$$

Equation 1

where $\hat{f}(x,y,t)$ is the density of voxel (x,y,t). The computation domain of (x,y,t) is a space-time cylinder, of which the radius and half height is defined by the two bandwidth parameter h_s and h_t. The density value of each cell is calculated based on the number of trajectory points within the cylinder, i.e., more points will cause a high density. n is the number of trajectory, and h_s and h_t are the spatial and temporal bandwidths respectively.

Each trajectory point in the vicinity is weighted on the kernel density function (K_s, K_t). There are several types of kernel function, for example, Gaussian kernel and quartic kernel. In this study, a common type of multivariate kernel function of Epanechnikov kernel is used:

$$K_s(u,v) = \begin{cases} \frac{2}{\pi}\left(1-(u^2+v^2)\right), & (u^2+v^2) < 1 \\ 0, & otherwise \end{cases}$$

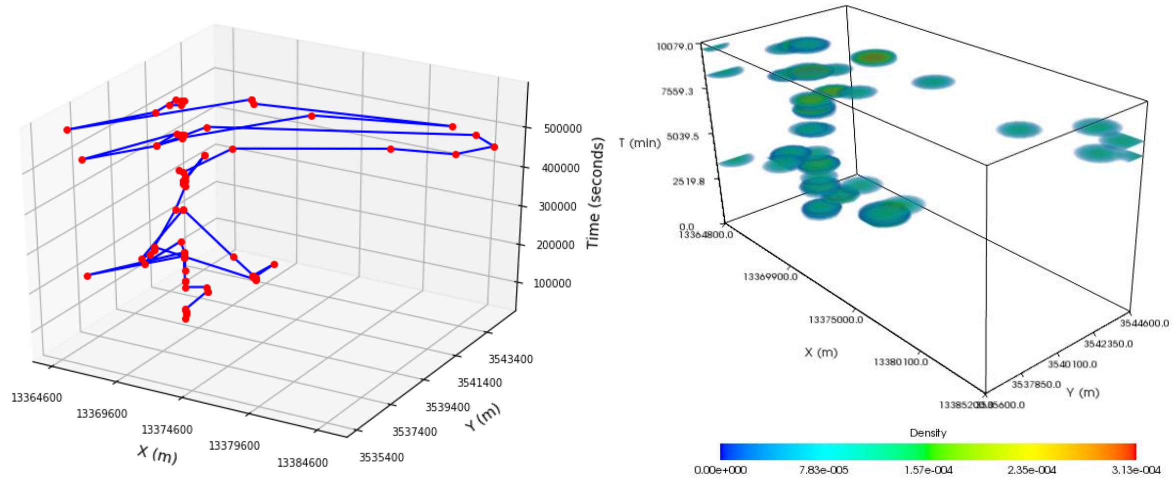
Equation 2

$$K_t(w) = \begin{cases} \frac{3}{4}(1-w^2), & w^2 < 1 \\ 0, & otherwise \end{cases}$$

Equation 3

It is well recognized that bandwidth is a non-trivial parameter that controls the degree of smoothing of estimated density surface. A larger bandwidth will generate a smoother surface, but it may cause the loss of characteristics of true density distribution, while a smaller one leads to dispensable spatio-temporal fluctuations. We adopt the data-driven approach of cross-validation to calibrate both spatial and temporal bandwidths of STKDE based on the experiments with actual datasets. The result of STKDE is a volume data of 4-demensional values consisting of a spatio-spatial vector of (x, y, t) and a scalar of density d. To visualize the volume data, a technique of volume rendering is most used. Several commercial rendering productions are popular in some disciplines, such as medical science, geology, and computer graphics. But there are few of generic software for geography. In this study, Mayavi, a python visualization module, is utilized to display the 3D volumetric objects (Figure 4). To estimate the density, the entire space-time domain of individual trajectory is divided into a number of 3D grid cells. From the result of STKDE, we have a density value for each 3D grid cell (x_g, y_g, t_g).

Stay points will be identified from the high density areas if we aggregate the density values along the time axis. Especially, the space-time domain division make it possible to extract the stay points at different time scale by purpose, such as sub-daily, daily, weekly or monthly. Home and workplace are two typical stay points. In order to further detect home and workplace from all stay points, a common approach is to associate a stay point with nighttime and daytime based on predefined period of time (e.g. 10pm-7am and 9am-5pm) considering the rhythm of city life. This time association method has been applied by Phithakkitnukoon et al., (2012), Demissie et al., (2016) and Gundlegård et al. (2016) and showed its reliability.



(a) (b)

Figure 4: Visualization of (a) individual trajectory and (b) the estimated space-time densities

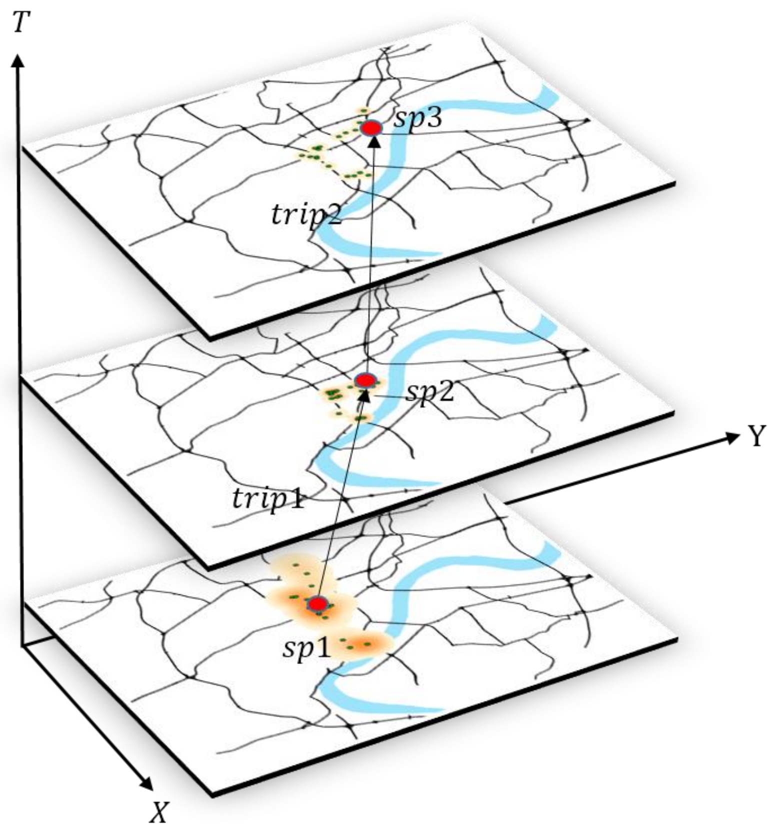


Figure 5: Demonstration of trips between stay points

3.3 Trips Generation

A trip is the individual movement between two stay points. Mathematically, a trip is a vector that originated from origin stay point to destination stay point. The trip length is the Euclidean distance from origin to destination. Figure 5 shows a demonstration of individual trips from the stay points detected with STKDE. The trips can be sub-daily, daily, weekly, etc., based on the time scale of stay points sp_1, sp_2 and sp_3 . Trips can be categorized from time perspective based on human travel behavior. For example, commuting trips usually happen at rush hours per day, while tourist trips may be at monthly, quarterly, or even yearly time intervals. Trips can also be expressed on different spatial scale according to the scale of stay points. Geographically, a typical type of stay point could be a point, e.g. a POI (Points of Interest). A stay point may also be polygon type, for example, residential area, sub-districts, districts, cities, even countries, serving different analysis requirement.

Generally, the direction of a trip can be unidirectional or bidirectional. The direction of a unidirectional trip starts from the origin stay point to the destination stay point. Commuting trip is a typical bidirectional trip, which is made back and forth between workplace and home. Thus, for a place in urban space, it may have both inbound trips (trips that end in there) and outbound trips (trips that originate from there). In order to characterize the place, we extract the trips by direction, and define inbound proportion, outbound proportion and the ratio value as following:

$$p_{out}^{i} = \frac{N_{out}^{i}}{N_{out}^{i} + N_{stay}^{i}}$$

Equation 4

$$p_{in}^{i} = \frac{N_{in}^{i}}{N_{in}^{i} + N_{stay}^{i}}$$

Equation 5

$$ratio^i = \frac{N_{out}^i}{N_{in}^i}$$

Equation 6

where p_{out}^i and p_{in}^i are the outbound proportion and inbound proportion for stay point i relatively. N_{out}^i and N_{in}^i are the number of outbound and inbound trips for stay point i, and N_{stay}^i is the number of trips that share the same origin and destination which is exactly stay point i. For a stay point, if the outbound proportion is greater than the inbound proportion, i.e., the ratio is greater than 1, it may be recognized as an outbound-dominated stay point, otherwise, it is inbound-dominated.

4. Results and Discussion

4.1 Spatial Distribution of Homes

First, we are interested in obtaining the spatial distribution of homes, the most important stay points in human movement, for collective users. We will assume that all of the users are local residents. Because not every user has enough records to infer the trip, only the users whose trajectory duration exceeds half of the sampling days are used for study, which account for 41.6% of total users. The home place of each user have been detected from stay points utilizing time association method. The study area is divided into grid cells of 2km by 2km, and for each cell, we accumulate the number of detected homes located there. Figure 6 shows the spatial distribution of homes. The density values are classified using the Jenks natural breaks methods. It seems that most users stay in the center part of the study area. When overlaying the point map layer of spatial distribution of residential areas investigated in 2014, a quite well spatial matching is depicted between the detected homes and the residential area locations.

For each home place, the sub-district it belongs to can be determined by comparing its position with the extent of every sub-district. This process makes it possible to discern individual home place at sub-districts level. The selected users are expanded with detection rate of 0.416 and MNO's market penetration rate of 0.7 to obtain the home places distribution of total users. We validate the estimated results with population census data from 2010 Chinese Census (Figure 7). 78 sub-districts have detected users, accounting for 80% of all sub-districts. The Spearman correlation coefficient between the estimation and census population is 0.88, which prove a high consistency of estimated results. It is presented that the population growth mainly occurs in the districts of Shangcheng and Xiaoshan, corresponding to the center and periphery areas of Hangzhou respectively. The phenomenon of population change implies the patterns of infilling and edge-expansion of Hangzhou during the urban sprawl process (Yue et al., 2013).

4.2 Distribution of Commuting Distance

Commuting between home and workplace is a most important travel of city life. Individual home in the nighttime of 4 August and workplace in the following daytime of 5 August are extracted respectively to generate individual commuting trip from the former to the latter. The lengths of commuting trips are calculated and counted by the distance interval of 1km (Figure 8). It can be seen that the users who travel less than 5 km, account for 61.1% of the all users.

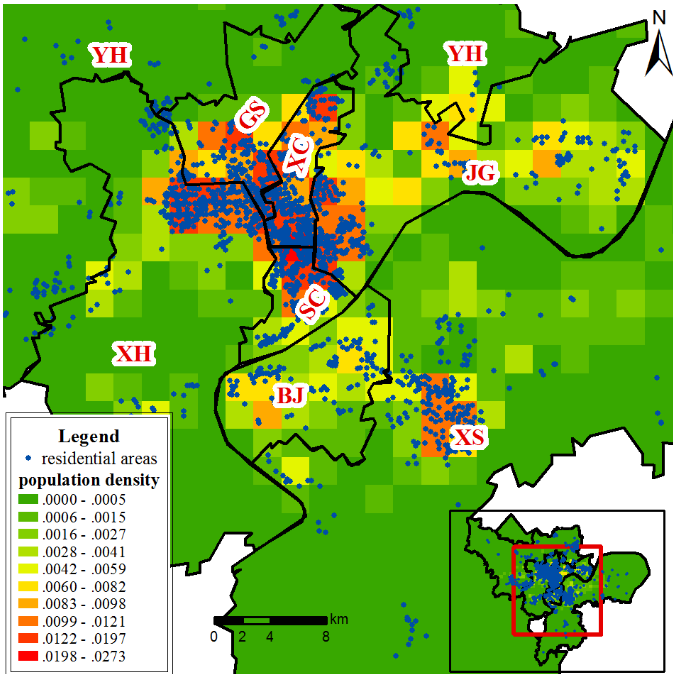
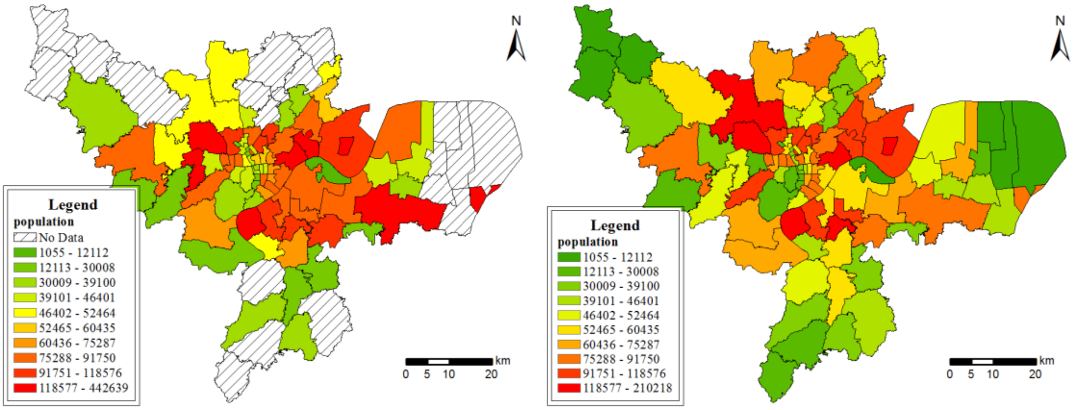


Figure 6: Spatial distribution of detected homes and residential areas



(a) (b)

Figure 7: Density distribution of (a) estimation and (b) population census data at sub-districts level

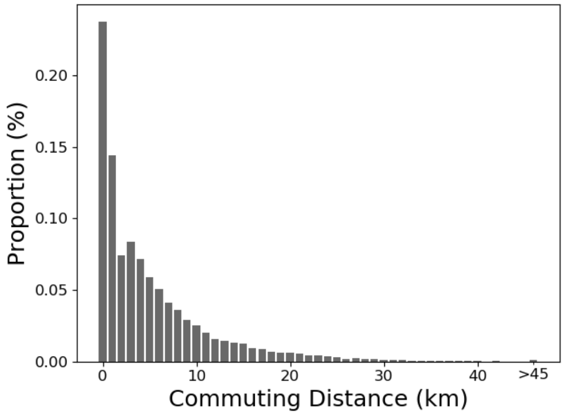


Figure 8: Distribution of commuting travel distance

The average length of all commuting trips is 5.59 km, larger than 4.41 km, which is the survey result from the 2010 Annual Report of Hangzhou Transportation Development. The estimated results show the increase of commuting distance in 4 years for Hangzhou residents. Also individual commuting trip could be generated at sub-districts level by associating home and workplace with the sub-districts they belong to, which allows us to study the difference in residents' commuting travel behavior for different sub-districts. Figure 9 shows the average length of outbound commuting trips, i.e. the commuting trips that origin from a sub-district, at sub-districts level. It seems that the trips origin from the city center is shorter than that from the urban periphery. It is because that the employment concentrates in the area of city center. Besides, the residents of city center prefer to work nearby, while the people in urban periphery need to commute longer to work. The phenomenon reflects the employment agglomeration effect in the core zone of Hangzhou.

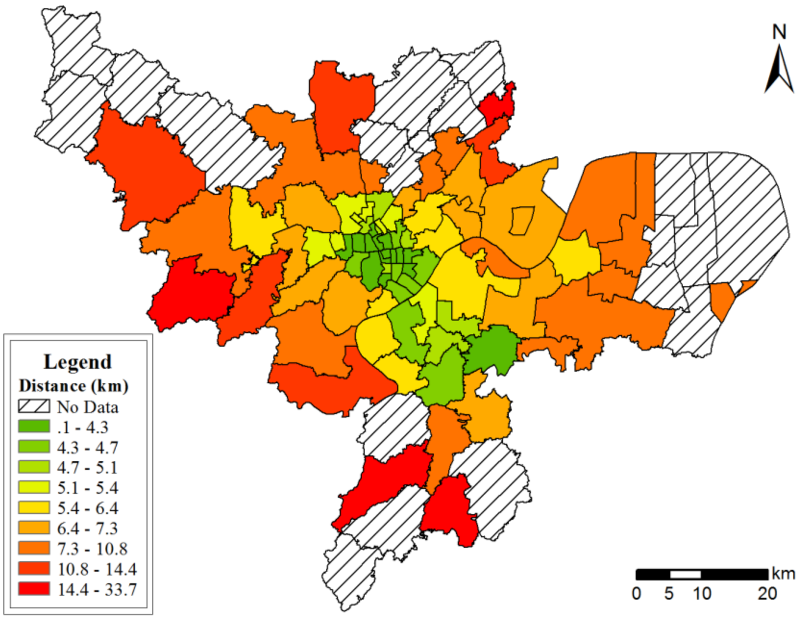


Figure 9: Distribution of outbound commuting trips length at sub-districts level

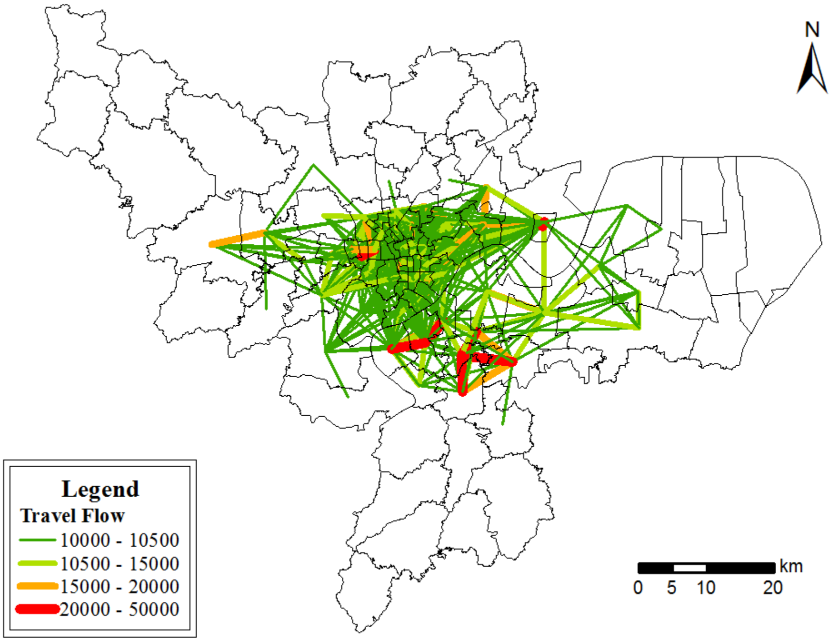


Figure 10: Spatial distribution of travel flow across sub-districts

4.3 Spatial Distribution of Travel Flow

To reveal the travel flow pattern of residents, OD matrix is constructed from the commuting trips at sub-districts level. Each item of the OD matrix denotes the number of trips from a sub-district to another one. Figure 10 is the travel flows from sub-districts to sub-districts, and only the trips which number is more than 10 thousand is presented. The link width represents the trips amount. It can be seen that most trips are made between urban periphery areas and the center areas, while the flows among urban periphery areas are sparse. Table 1 shows the statistics results of inbound and outbound proportion for all sub-districts, from which the travel flow of every sub-district can be characterized. For example, Beishan, whose outbound proportion and inbound proportion are 0.4 and 0.56 respectively, is an inbound-dominated sub-district. Figure 11 shows the distribution of ratio values for all sub-districts. It is depicted that, most sub-districts have no obvious distinction between inbound proportion and inbound proportion. Sub-districts in city centers are mostly inbound-dominated, while the outbound-dominated districts mainly are located in urban periphery.

Table 1: Inbound and outbound trips proportion of sub-districts

sub-district	p_{out}	p_{in}	*Ratio*	sub-district	p_{out}	p_{in}	*Ratio*
Baiyang	0.29	0.36	0.91	Puyan	0.23	0.15	1.1
Beigan	0.27	0.3	0.96	Qiaosi	0.18	0.14	1.05
Beishan	0.4	0.56	0.73	Qingbo	0.36	0.46	0.84
Bengbu	0.29	0.24	1.07	Renhe	0.1	0.1	1.01
Caihe	0.25	0.32	0.91	Shangtang	0.37	0.3	1.11
Changhe	0.24	0.29	0.93	Shiqiao	0.29	0.19	1.14
Changqing	0.42	0.48	0.9	Suoqian	0.41	0.34	1.12
Chaohui	0.33	0.39	0.92	Tianshui	0.44	0.61	0.71
Chaoming	0.41	0.39	1.04	Wangjiang	0.29	0.25	1.05
Chengxiang	0.29	0.32	0.96	Wenhui	0.47	0.42	1.09

Cuiyuan	0.31	0.47	0.76	Wenxin	0.39	0.24	1.26
Daguan	0.42	0.36	1.11	Wenyan	0.19	0.1	1.11
Dinglan	0.46	0.19	1.57	Wuchang	0.26	0.45	0.65
Dongxin	0.32	0.23	1.13	Wulin	0.46	0.57	0.79
Gudang	0.38	0.36	1.04	Xiangfu	0.3	0.28	1.03
Hemu	0.44	0.28	1.28	Xianlin	0.51	0.37	1.31
Hezhuang	0.25	0.23	1.02	Xiaohe	0.42	0.35	1.11
Hubin	0.37	0.61	0.62	Xiaoying	0.3	0.43	0.81
Husu	0.45	0.47	0.97	Xiasha	0.16	0.14	1.02
Jiangcun	0.4	0.23	1.28	Xihu	0.27	0.37	0.85
Jianqiao	0.37	0.24	1.22	Xinjie	0.21	0.19	1.03
Jinjiang	0.18	0.13	1.06	Xinqiao	0.2	0.25	0.93
Jiupu	0.24	0.18	1.08	Xintang	0.26	0.23	1.03
Kaixun	0.37	0.33	1.06	Xixi	0.36	0.48	0.82
Kangqiao	0.43	0.27	1.29	Xixing	0.27	0.28	0.99
Liangzhu	0.67	0.5	1.53	Yaqian	0.31	0.25	1.25
Linyin	0.33	0.43	0.85	Yipeng	0.34	0.27	1.11
Liuxia	0.27	0.19	1.11	Yuhang	0.15	0.1	1.05
Nanxing	0.32	0.31	1.01	Zhanongkou	0.38	0.23	1.65
Nanyuan	0.25	0.16	1.12	Zhuantang	0.17	0.12	1.07
Ningwei	0.18	0.17	1.01	Ziyang	0.34	0.22	1.17

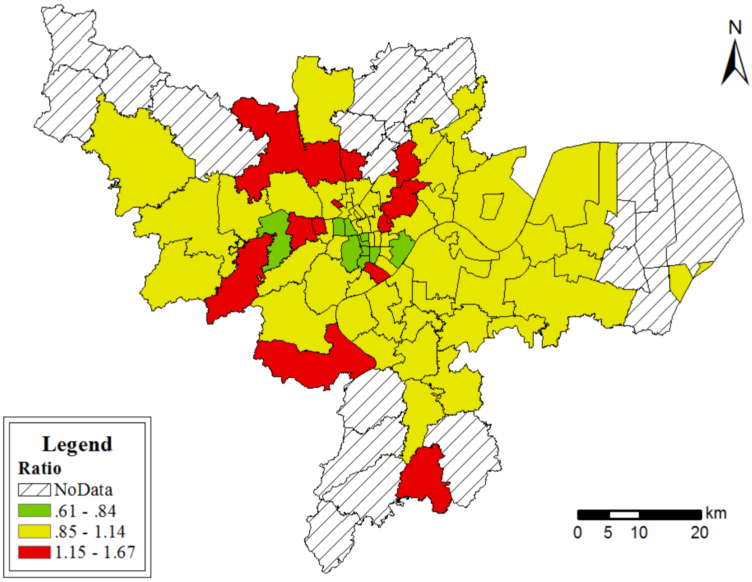


Figure 11: Ratio values of all sub-districts

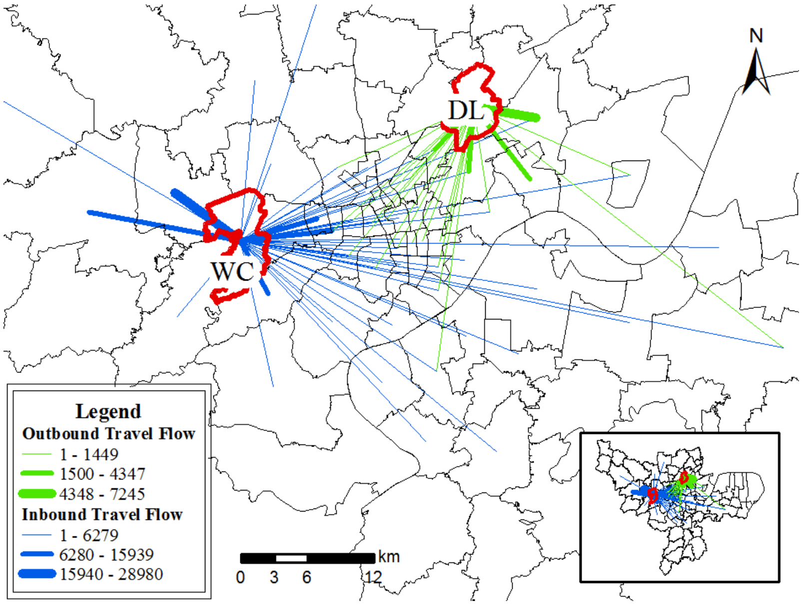


Figure 12: Outbound travel flow of Dinglan and inbound travel flow of Wuchang

We focus on two typical sub-districts of Dinglan (DL) and Wuchang (WC). Dinglan is an outbound-dominated sub-districts, because its outbound proportion (0.46) is greater than inbound proportion (0.19). Actually, Dinglan is a sub-district which has more floating population than local population according to the population census. The preference of working outside of floating population lead to the high value of outbound proportion. Wuchang, on the contrary, is an inbound-dominated sub-district with outbound proportion 0.26 versus inbound proportion 0.45. According to the city layout, Wuchang is a sub-district with more office blocks such as Alibaba and tourism zones such as Xixi National Wetland Park, which lead to more inflow of office works and tourists. Figure 12 shows the outbound travel flow of Dinglan and inbound travel flow of Wuchang, respectively. We can find that the users in Dinglan are prefer to work in the nearby sub-districts, while people who work in Wuchang are from more sub-districts, including those of nearby and further urban center areas.

5. Conclusion

Understanding the urban travel behavior is an essential and primary task for traffic planning and management. The traditional travel behavior patterns modeling often rely on travel surveys, which are limited by high costs or infrequent update rates. The cellular network data have showed a wealth resource for human travel behavior understand. In this paper, we propose a method to detect individual stay points and extract trips for spatio-temporal travel behavior analysis utilizing the cellular network data. First, scaled homes are compared with the census data and show a quite well consistency of 0.88. Next. The distribution of commuting travel distance presents that 61.1% users travel less than 5 km, while those who live in the urban periphery areas travel longer than urban center areas. Final, we can reveal the travel flow across regions by counting the amount of trips. We calculated the inbound and outbound trip proportion for all sub-districts. From which, the inbound-dominated and outbound-dominated sub-districts can be characterized. The results show the potential in applying cellular network data to traffic study. In the further, we plan to extend the current study to specific transportation implementations. The ultimate aim is to derive knowledge from cellular network data to target specific urban areas that served urban infrastructure planning and management improvement.

Acknowledgements

This work was supported by the National Key Research and Development Foundation of China (No. 2017YFB0503702 and No.

2017YFB0503605), 973 Program (No. 2013CB733402) and the National Natural Science Foundation of China (No.41601486 and No. 40771167).

References

Algizawy, E., Ogawa, T. and El-Mahdy, A., 2017, Real-Time Large-Scale Map Matching Using Mobile Phone Data. *ACM Transactions on Knowledge Discovery from Data*, Vol. 11(4), 1-38. doi: 10.1145/3046945.

Calabrese, F., Di Lorenzo, G., Liang, L. and Ratti, C., 2011, Estimating Origin-Destination Flows Using Mobile Phone Location Data. *IEEE Pervasive Computing*, Vol. 10(4), 36-44. doi: 10.1109/MPRV.2011.41.

Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, Jr. J. and Ratti, C., 2013, Understanding Individual Mobility Patterns from Urban Sensing Data: A Mobile Phone Trace Example. *Transportation Research Part C: Emerging Technologies*, Vol. 26(0), 301-313. doi: 10.1016/j.trc.2012.09.009.

Csáji, B. C., Browet, A., Traag, V. A., Delvenne, J. C., Huens, E., Van Dooren, P., Smoreda, Z. and Blondel , V. D., 2013, Exploring the Mobility of Mobile Phone Users. *Physica A: Statistical Mechanics and its Applications*, Vol. 392(6), 1459-1473. doi: 10.1016/j.physa.2012.11.040.

Demissie, M. G., Phithakkitnukoon, S., Sukhvibul, T., Antunes, F., Gomes, R. and Bento, C., 2016, Inferring Passenger Travel Demand to Improve Urban Mobility in Developing Countries Using Cell Phone Data: A Case Study of Senegal. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 17(9), 2466-2478.

Gidófalvi, G. and Yang, C., 2015, Scalable Detection of Traffic Congestion from Massive Floating Car Data Streams. *Proceedings of the 1st International ACM SIGSPATIAL Workshop on Smart Cities and Urban Analytics*, 114-121.

González, M. C., Hidalgo, C. A. and Barabási, A. L., 2008, Understanding Individual Human Mobility Patterns. *Nature*, Vol. 453(7196), 779-782.

Gundlegård, D., Rydergren, C., Breyer, N. and Rajna, B., 2016, Travel Demand Estimation and Network Assignment Based on Cellular Network Data. *Computer Communications*, Vol. 95, 29-42. doi: 10.1016/j.comcom.2016.04.015.

Hoteit, S., Secci, S., Sobolevsky, S., Ratti, C. and Pujolle, G., 2014, Estimating Human Trajectories and Hotspots Through Mobile Phone Data. *Computer Networks*, Vol. 64(0), 296-307. doi: 10.1016/j.comnet.2014.02.011.

Houbraken, M., Logghe, S., Schreuder, M., Audenaert, P., Colle, D. and Pickavet, M., 2017, Automated Incident Detection Using Real-Time Floating Car Data. *Journal of Advanced Transportation*, Vol. 2017, 1-13.

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J. and Varshavsky, A., 2011, Identifying Important Places in People's Lives from Cellular Network Data. *International Conference on Pervasive Computing*, 133-151.

Jiang, S., Ferreira, J. and Gonzales, M. C., 2016, Activity-Based Human Mobility Patterns Inferred from Mobile Phone Data: A Case Study of Singapore. *IEEE Transactions on Big Data*, Vol. 3(2), 208-219.

Keler, A., Krisp, J. M. and Ding, L., 2017, Detecting Vehicle Traffic Patterns in Urban Environments Using Taxi Trajectory Intersection Points. *Geo-spatial Information Science*, Vol. 20(4), 333-344. doi: 10.1080/10095020.2017.1399672.

Luo, A., Chen, S. and Xv, B., 2017, Enhanced Map-Matching Algorithm with a Hidden Markov Model for Mobile Phone Positioning. *ISPRS International Journal of Geo-Information*, Vol. 6(11), 327. doi: 10.3390/ijgi6110327.

Mario B. Rojas, I., Sadeghvaziri, E. and Jin, X., 2016, Comprehensive Review of Travel Behavior and Mobility Pattern Studies that Used Mobile Phone Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2563, 71-79. doi: 10.3141/2563-11.

Morency, C., Munizaga, M., Greaves, S., Raimond, T., Bonnel, P., Hombourger, E., Olteanu-Raimond, A. M. and Smoreda, Z., 2015, Passive Mobile Phone Dataset to Construct Origin-destination Matrix: Potentials and Limitations. *Transportation Research Procedia*, Vol. 11, 381-398.

Ni, L., Wang, X. and Chen, X., 2018, A Spatial Econometric Model for Travel Flow Analysis and Real-World Applications with Massive Mobile Phone Data. *Transportation Research Part C: Emerging Technologies*, Vol. 86, 510-526. doi: 10.1016/j.trc.2017.12.002.

Pei, T., Sobolevsky, S., Ratti, C., Shaw, S. L., Li, T. and Zhou, C., 2014, A New Insight Into Land Use Classification Based on Aggregated Mobile Phone Data. *International Journal of Geographical Information Science*, Vol. 28(9), 1988-2007. doi: 10.1080/13658816.2014.913-794.

Phithakkitnukoon, S., Smoreda, Z. and Olivier, P., 2012, Socio-Geography of Human Mobility: A Study Using Longitudinal Mobile Phone Data. *PLoS ONE*, Vol. 7(6), e39253. doi: 10.1371/journal.pone.0039253.

Phithakkitnukoon, S., Horanont, T., Witayangkurn, A., Siri, R., Sekimoto, Y. and Shibasaki, R., 2015, Understanding Tourist Behavior Using Large-Scale Mobile Sensing Approach: A Case Study Of Mobile Phone Users in Japan. *Pervasive and Mobile Computing*, Vol. 18, 18-39. doi: 10.1016/j.pmcj.2014.07.003.

Sadeghvaziri, E., Rojas, I. V. M. B. and Jin. X., 2016, Exploring the Potential of Mobile Phone Data in Travel Pattern Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2016(2594), 27-34.

Song, C., Qu, Z., Blumm, N. and Barabási, A. L., 2010, Limits of Predictability in Human Mobility. *Science*, Vol. 327(5968), 1018-1021. doi: 10.1126/science.1177170.

Tolouei, R., Psarras, S. and Prince, R., 2017, Origin-Destination Trip Matrix Development: Conventional Methods versus Mobile Phone Data. *Transportation Research Procedia*, Vol. 26, 39-52. doi: 10.1016/j.trpro.2017.07.007.

Toole, J. L., Çolak, S., Alhasoun, F., Evsukoff, A. and González, M. C., 2014, The Path Most Travelled: Mining Road Usage Patterns from Massive Call Data. *CoRR*, Vol. abs/1403.0636.

Toole, J. L., Colak, S., Sturt, B., Alexander, L. P., Evsukoff, A. and González, M. C., 2015, The Path Most Traveled: Travel Demand Estimation Using Big Data Resources. *Transportation Research Part C: Emerging Technologies*, Vol. 58, 162-177. doi: 10.1016/j.trc.2015.04.022.

Wang, J., Yu, L., Gao, Y., Zhang, J. and Song, G., 2017, Analysis of Speed Differences in Data from Remote Traffic Microwave Sensors and Floating Car Data Systems on Expressways: Beijing Case Study. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2017(2643), 112-120.

White, J. and Wells, I., 2002, Extracting Origin Destination Information from Mobile Phone Data. *Eleventh International Conference on Road Transport Information and Control (RTIC 2002)*, 30-34.

Work, D. B., Blandin, S., Tossavainen, O. P., Piccoli, B. and Bayen, A. M., 2010, A Traffic Model for Velocity Data Assimilation. *Applied Mathematics Research eXpress*, Vol. 2010(1), 1-35. doi: 10.1093/amrx/abq002.

Yang, X., Fang, Z., Xu, Y., Shaw, S. L., Zhao, Z., Yin, L., Zhang, T. and Lin, Y., 2016, Understanding Spatiotemporal Patterns of Human Convergence and Divergence Using Mobile Phone Location Data. *ISPRS International Journal of Geo-Information*, Vol. 5(10), 177.

Yuan, Y., Raubal, M. and Liu, Y., 2012, Correlating Mobile Phone Usage and Travel Behavior – A Case Study of Harbin, China. *Computers, Environment and Urban Systems*, Vol. 36(2), 118-130. doi: 10.1016/j.compenvurbsys.2011.-07.003.

Yuan, Y. and Raubal, M., 2016, Analyzing the Distribution of Human Activity Space from Mobile Phone Usage: An Individual and Urban-Oriented Study. *International Journal of Geographical Information Science*, Vol. 30(8), 1594-1621.

Yue, W., Liu, Y. and Fan, P., 2013, Measuring Urban Sprawl and its Drivers in Large Chinese cities: The case of Hangzhou. *Land Use Policy*, Vol. 31, 358-370. doi: 10.1016/j.landusepol-.2012.07.018.

Zhang, Y., Zuo, X., Zhang, L. and Chen, Z., 2011, Traffic Congestion Detection Based On GPS Floating-Car Data. *Procedia Engineering*, Vol. 15, 5541-5546. doi: 10.1016/j.proeng.2011.08-.1028.