

REVEALING SPATIAL VARIATION AND CORRELATION OF URBAN TRAVELS FROM BIG TRAJECTORY DATA

LI Xiaona^b, TU Wei^{a, c*}, Shen Shaoqing^a, YUE YANG^c, LUO Nianxue^b, LI Qingquan^{c, d}

^a Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Land and Resources, Shenzhen 518040, China

^b School of Geodesy and Geomatics, Wuhan University, Wuhan 430079, PR China -1092041224@qq.com

^c Shenzhen Key Lab of Spatial Smart Sensing and Service & Institute of Smart City, School of Architecture and Urban Planning, Shenzhen University, Shenzhen 518060, China - (tuwei, yuyang, liqq)@szu.edu.cn

^d State Key Lab of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, PR China

Commission IV, WG IV/4 and Commission IV, WG IV/10

KEY WORDS: Human travel, Big data, Trajectory, Spatial Variation, Smart card data, Correlation

ABSTRACT:

With the development of information and communication technology, spatial-temporal data that contain rich human mobility information are growing rapidly. However, the consistency of multi-mode human travel behind multi-source spatial-temporal data is not clear. To this aim, we utilized a week of taxis' and buses' GPS trajectory data and smart card data in Shenzhen, China to extract city-wide travel information of taxi, bus and metro and tested the correlation of multi-mode travel characteristics. Both the global correlation and local correlation of typical travel indicator were examined. The results show that: (1) Significant differences exist in urban multi-mode travels. The correlation between bus travels and taxi travels, metro travel and taxi travels are globally low but locally high. (2) There are spatial differences of the correlation relationship between bus, metro and taxi travel. These findings help us understanding urban travels deeply therefore facilitate both the transport policy making and human-space interaction research.

1. INTRODUCTION

Human travel plays an important role in transportation planning, traffic management, and urban spatial structure analysis. Traditionally, human travel studies rely on questionnaire to acquire enough and accurate travel information, which usually require long time and large cost. For the sake of the development of spatial information technology, the acquisition of large-scale individual spatial-temporal data are constantly becoming reality. Alternative data sources, such as smart card data (SCD) (Zhou et al., 2017), mobile phone positioning data (Yue et al., 2016; TU et al., 2017), vehicle GPS data (TU et al., 2010; TU et al., 2016), even SNS check-in data from web crawlers has been emerging. Using such multiple sourced data to reveal human travel pattern has become a hot topic. Nowadays, there are lots of research using spatial-temporal data to analyze human travel. Researches related to travel includes human mobility pattern analysis (Rhee et al., 2008; Wu et al., 2014), urban traffic flow simulation (Wu et al., 2012), critical transportation infrastructure (Fang et al., 2012), urban spatial structure analysis (Zhong et al., 2013), and urban land use (Liu et al., 2012).

The integration of information and communication technology (ICT) and spatial information technology helps to collect big trajectory data, which imply massive city-wide human travels. For example, TU et al (2012) extracted all taxi trajectories across the Yangtze River and evaluated the vulnerability of critical transportation service. Liu et al (2013) uncovered the association between human travel from taxi data and land use. These studies depended on one type taxi trajectories thus made an insufficient use of data. However, current human travel researches just utilize one source of trajectory data. On the other hand, big trajectory data not only challenge the existing approaches using single data

but also bring new research opportunities (Kang et al., 2013). One valuable question is whether urban travels from multi sourced trajectories are consistent. The answer of this question will provide us with a comprehensive understanding of urban travel. Recently, Cao et al (2016) investigated the relationship between taxi travel and human travel from mobile phone posing data. However, the relationship between each mode, like buses, metro, and taxi are not investigated.

This study aims to reveal the spatial pattern and the correlation of urban travel from multi-sourced trajectory data. By fusing with bus GPS trajectory, smart card data, and metro GIS data, bus travel (B) or metro travel (M) of an individual were extracted. Taxi travel (T) were summarized from massive taxi GPS trajectories. Then, a correlation theory based evaluation framework was developed to investigate the consistent relationship between them. The results indicate that there are spatial differentiation of the correlation relationship between bus, metro and taxi travel. It provides valuable insights on the trajectory driven urban research and applications.

2. STUDY AREA AND DATA

2.1 Study Area

The study was conducted in Shenzhen, China. It includes six administrative regions including Futian, Luohu, Nanshan, Yantian District, Baoan, Longgang, and four functional areas containing Guangming, Longhua, Pingshan, Dapeng (Figure 1). The transportation commission of Shenzhen Municipality (TCSM) provides bus, metro, taxi service for people. Up to Year

* Corresponding author

2014, Shenzhen metro system has five subway lines and 118 stations (including the 13 transferring stations) with the total length of about 178 km, covering six districts, including Luohu, Futian, Nanshan, Baoan, Longgang and Longhua. The bus system has 874 bus lines and 5265 bus stations. The taxi system has 16135 taxi cruising on the road to serve people's travel demand. According to the annual report of TCSM, the three transport systems complete more than 50% travels in the city averagely.

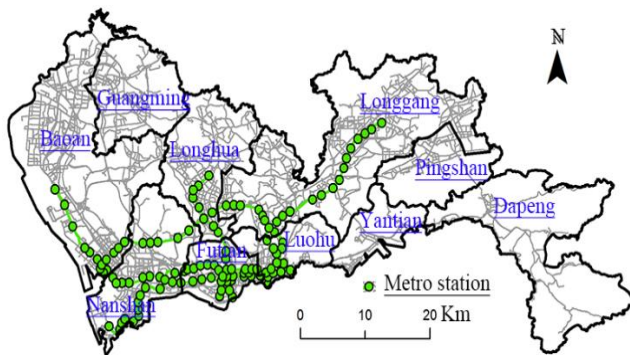


Figure 1. Study area

2.2 Dataset

The used dataset includes taxi GPS trajectories, bus GPS trajectories and smart card data. All of them were collected by TSCM during the period from September 24 to 30 in 2014. The taxi GPS trajectories record the locations of all taxis and their service statuses with a sampling interval of about 40 seconds. The bus GPS trajectory stores all buses' locations with a sampling interval of about 40 seconds. The smart card data indicates when an individual gets in or gets out a bus or a metro station. Totally, there are 9107567 taxi GPS records, 10757142 bus GPS records, and 2466341 smart card records.

Using these massive multiple sourced trajectory data, we extracted an individual travel by using one or two type of trajectory data. Each travel record is a 7-tuple as equation (1).

$$T=(id, x_o, y_o, t_o, x_d, y_d, t_d) \quad (1)$$

Where id = the record id,

x_o = the longitude of the origin place of a travel,
 y_o = the latitude of the origin place of a travel,
 t_o = the departure time of a travel,
 x_d = the longitude of the destination place of a travel,
 y_d = the latitude of the destination place of a travel,
 t_d = the arrival time of a travel.

The taxi travel are directly inferred from the taxi GPS data according to the changes of the service statuses. The metro travel are retrieved from smart card data. The bus travel is much difficult as the smart card data only contain the swapping time, instead of the location. By fusing bus trajectory and the swapping time, the real boarding bus station are interpolated. All in all, we got a week of taxi, metro, and bus travel in the Shenzhen city. Table 1 summarizes the total count of three ridership. It shows that there are 2.82 million taxi travel records, 4332914 bus travel records and 9.84 million metro travel records in the studied week.

Travel mode	Ridership
	millions
Taxi	2.82
Bus	28.4
Metro	9.84

Table 1. Summary of daily public ridership in Shenzhen, China

The scale will affect the reliability of spatial analysis result. Without loss of generality, we used traffic analysis zone (TAZ) as the basic spatial unit, which is composed of the adjacent land parcels with similar nature. TAZ can better reflect the spatial distribution of traffic travels. The Shenzhen are divided into 491 TAZs.

2.3 Travel indicators

According to previous literature (Deng et al., 2000; Yao et al., 2006; Huang et al., 2008; Chen et al., 2009; Jang et al., 2012), transportation domain usually takes travel volume, travel distance, travel time, travel direction, travel purpose as indicators to describe popular human travel. Considering the extracted travel information, we chose four typical indicators to investigate the relationship between taxi, metro, and bus travels, including travel volume, travel time, and travel distance. Using the got travel records in Section 2.2, we summarized hourly travel indicators of taxi, bus, and metro respectively. In other words, TAZ based time series taxi, bus, and metro information were gathered for the further correlation analysis.

3. METHODOLOGY

The main idea of this study is to investigate the relationship of the multi-mode urban using the Pearson correlation analysis approach, that is, the correlation level of two mode human travels are measured. Furthermore, considering the spatial effect, both global and local correlation analysis are conducted. Figure 2 displays the workflow. Using spatial or temporal travel indicators, both spatial and temporal correlation level are tested. Then, using TAZ based travel indicators, local correlation is tested. Following them, significance level is tested.

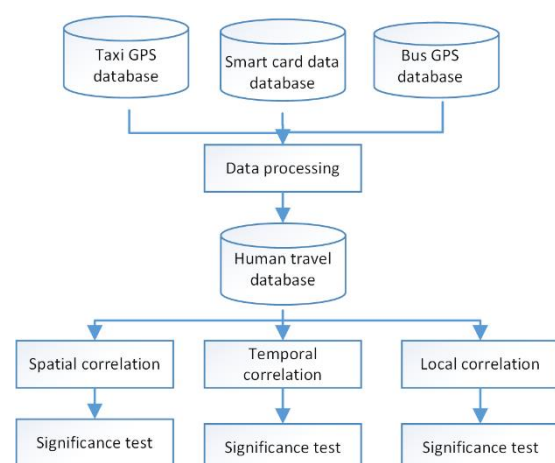


Figure 2. Workflow of methodology

3.1 Correlation analysis

Correlation analysis is a statistical method to study the relationship between random variables (Li et al., 2012). It refers

to refers to the extent to which two variables have a linear relationship with each other. The correlation can be expressed by the correlation coefficient r , which is in the range of $[-1, 1]$. The correlation coefficient value within the range $(0, 1]$ indicate a positive correlation.

There are many approaches to measure correlation level, such as Pearson correlation product, Spearman's rank correlation, and Kendall tau rank correlation. The Pearson correlation coefficient is widely used measure of the degree of correlation between two variables. The specific formula is as follows:

$$r_{xy} = \frac{\sum_{i=1}^n (x - \bar{x})(y - \bar{y})}{\sqrt{\sum_{i=1}^n (x - \bar{x})^2 \sum_{i=1}^n (y - \bar{y})^2}} \quad (2)$$

Where x = a random variable,
 y = a random variable,
 \bar{x} = the expected value of the random variable x ,
 \bar{y} = the expected value of the random variable y ,
 n = the number of samples.

In this study, by assuming the independence of travel indicators, the random variables are the selected travel indicators in Section 2.3.

3.2 Global correlation

Considering the different to the difference of variables, the global correlation is divided into two categories: (1) Spatial correlation using the travel indicator statistics at each TAZ as variables; (2) Temporal correlation using the travel indicator statistics at a hour as variables.

The measure of global correlation is based on the travel volume (tv), travel time (tt), travel distance (td), that is, $T_{global} = (tv, tt, td)$ as travel indicators. Given two travel modes, the correlation coefficient $r_{tv}, \delta_{tt}, \delta_{td}$ of each indicator is firstly calculated with the Pearson correlation approach. Therefore, the correlations of travel indicators of bus, metro and taxi are as follows:

$$r_s = \begin{bmatrix} r_{sB} & r_{sB^sM} & r_{sB^sT} \\ r_{sM^sB} & r_{sM} & r_{sM^sT} \\ r_{sT^sB} & r_{sT^sM} & r_{sT} \end{bmatrix} \quad (3)$$

Where B = bus ,
 M = metro,
 T = taxi.

The elements on the diagonal of this matrix are 1. The elements on the non-diagonal are calculated using the Pearson correlation coefficient method. It should be noted that, assuming that the travel indicators have the same effect on travel, the products of each travel indicator' correlation coefficient can be equally weighted summed to get the global correlation coefficient matrix δ_{GB-M-T}^2 , which describes the global correlation relationship between bus, metro and taxi travel.

$$\delta_{GB-M-T}^2 = \frac{1}{3} \sum \delta_s^2 \quad (4)$$

Where s = the selected travel indicator (tt, td, tv)

3.3 Local correlation

Local correlation focuses on the travel at TAZ scale. The local correlation calculation makes use of five travel indicators,

including travel time (tt), travel distance (td), travel volume of origins (tvo), travel volume of destinations (tvd), and travel directions(tdr), where tt, td are the same as above section. The additional tvo, tvd, tdr describe the travel in/out characteristics of a TAZ.

For each selected travel indicator, the correlation coefficient of two travel modes in a TAZ is firstly calculated. Following the philosophy of above section, correlation matrices of a TAZ is formatted by the equation (5). Also, the diagonal element is 1. The non-diagonal element is calculated using a travel indicator of two travel modes. Finally, the correlation matrix δ_{iB-M-T}^2 of the bus, metro and taxi travel in a TAZ is obtained by equation (6).

$$r_{iB-M-T} = \begin{bmatrix} r_{iB} & r_{iB^sM} & r_{iB^sT} \\ r_{iM^sB} & r_{iM}^2 & r_{iM^sT} \\ r_{iT^sB} & r_{iT^sM} & r_{iT}^2 \end{bmatrix} \quad (5)$$

Where B = bus ,
 M = metro,
 T = taxi.
 i = the index of a TAZ,
 s = the selected travel indicator.

$$\sigma_{iB-M-T}^2 = \frac{1}{5} \sum \sigma_{i_s}^2 \quad (6)$$

Where s = the selected travel indicator (tt, td, tvo, tvd, tdr)

3.4 Significance test

In order to evaluate the correlation level, it is necessary to test the significance of the correlation. When the sample size is bigger than 20, the t statistic can be used to test the significance of the correlation coefficient (Box, 1987). The t statistic is formatted as equation (7).

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \sim t(n-2) \quad (7)$$

Where r = the Pearson coefficient
 n = sample amount

In the study, regarding temporal correlation and local correlation, the sample number of each travel indicator are 24, then the degree of freedom $df = n-2 = 22$. According to the significance of the correlation coefficient, we know that at the 0.05 significance level, the threshold value $r_{0.05} = 0.40$. With respect to spatial correlation, the sample number is equal to the count of TAZs so the degree of freedom $df = 491-2 = 489$. At the 0.05 significance level, the threshold value $r_{0.05} = 0.09$.

4. RESULT

4.1 Global correlation results

From the spatial perspective, in general, the bus, metro and taxi travel distribution are different. Totally, the spatial correlation coefficients of bus-metro, bus - taxi and metro - taxi travel are 0.27, 0.06, 0.09 respectively. Table 3 reports the correlation coefficients of two travel models. It suggests that bus and metro, metro and taxi are significant correlated at level 0.05, but they fall into a weak correlation. Metro and taxi are not significant correlated. Table 4 gives the Pearson correlation coefficients of

two travel modes in for each indicator. The travel indicator level correlation results indicates that, comparing to travel time and travel volume, bus, metro, and taxi travel are is weaker correlated in travel distance.

	Bus	Metro	Taxi
Bus	1	0.27	0.06
Metro	0.27	1	0.09
Taxi	0.06	0.09	1

Table 3. Spatial correlation coefficients

r_{td}				r_{tt}				r_{tv}			
	B	M	T		B	M	T		B	M	T
B	1	0.03	-0.00	B	1	0.43	-0.22	B	1	0.35	0.50
M	0.03	1	-0.18	M	0.43	1	-0.15	M	0.35	1	0.64
T	-0.00	-0.18	1	T	-0.22	-0.15	1	T	0.50	0.64	1

Table 4. Spatial correlation coefficients of each travel indicator (B=bus, M= metro, T= taxi)

Figure 3 displays the temporal distribution of travel indicator. From the temporal perspective, the correlation coefficients are higher correlated than spatial correlation coefficients. The Pearson correlation coefficients for metro and bus, taxi and bus, taxi and metro, are 0.96, 0.27, and 0.29 respectively. But the coefficients of taxi and bus, taxi and metro is not significant. The correlation coefficients of each travel indicator between bus, metro and taxi are presented in Table 6. This result indicates the correlations of travel time for bus-taxi and metro-taxi are not significant. All three travel indicators of bus and metro are significantly correlated.

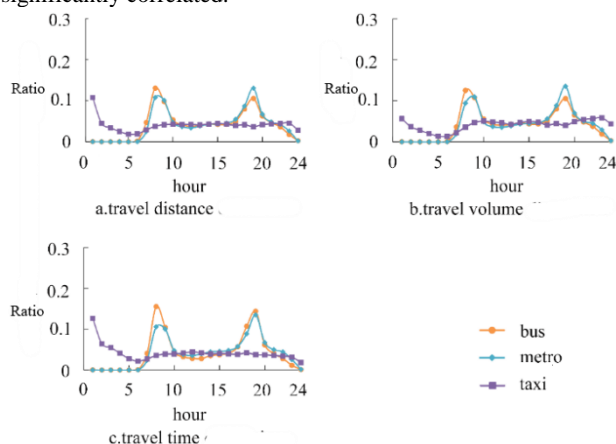


Figure 3. The temporal distribution of travel indicator

	Bus	Metro	Taxi
Bus	1	0.96	0.27
Metro	0.96	1	0.29
Taxi	0.27	0.29	1

Table 5. Temporal correlation coefficients

δ_{td}^2				δ_{tt}^2				δ_{tv}^2			
	B	M	T		B	M	T		B	M	T
B	1	0.97	0.38	B	1	0.96	-0.02	B	1	0.96	0.37
M	0.97	1	0.42	M	0.96	1	0.01	M	0.96	1	0.43
T	0.38	0.42	1	T	-0.02	0.01	1	T	0.37	0.43	1

Table 6. Temporal correlation coefficients of each travel indicator (B=bus, M= metro, T= taxi)

4.2 Local correlation results

The results show that the local correlations of the multi - mode travels are different from the global correlations. The TAZ scale correlation grades are different. As Table 7 shows, 86 TAZs are significantly correlated for bus and metro. Regarding bus and taxi, only 12 TAZs are significant correlated. However, no TAZs are with significant relationship between metro and taxi.

Figure 3 displays the distribution of local correlation coefficients. This result indicates following suggestions: (1) The TAZs with high correlation coefficients of bus and metro appear along metro line, most of which are in line 1 and line 3. It is because that bus and metro cooperate to serve daily commuting travel. Both them are mainly driven by home-working cyclical rhythm. (2) Regarding the relationship of metro and taxi, bus and taxi, there are a few TAZs with high correlation values, although the global correlation is low as Section 4.1. (3) Most of highly correlated TAZs appear at the city center, Futian and Luohu, as Figure 5 shows.

TAZ	$r \neq 0$	$ r > r_{0.05}$
B-M	491	87
B-T	491	465
M-T	491	87

Table 7. The summary of local correlation coefficients

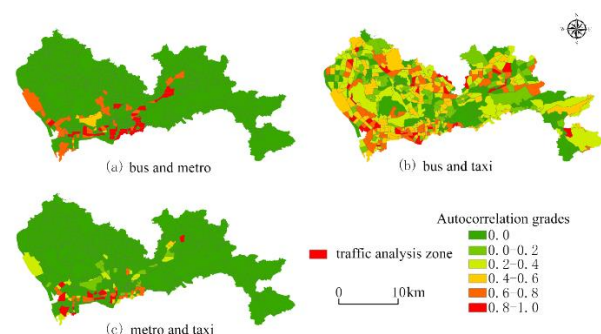


Figure 4. The spatial distribution of correlation coefficient

5. CONCLUSION

In the era of big data, current studies using single source of spatial-temporal trajectory to investigate urban travel pattern face great challenges. Understanding the relationship of multi-mode urban travel is an essential to deepen advance studies. In this paper, we presented a data-driven framework to test the correlation level between bus, metro and taxi travel. Global and local correlation were explored. The study shows that there are significant differences of different travel modes. Detailed conclusions are listed as below:

(1) There are significant differences of different travel modes. The global correlation of bus-taxi and metro-taxi are low, while the temporal global correlation of bus-metro is high but the spatial global correlation is low. The temporal global correlation coefficients and the spatial global correlation coefficients of bus-taxi, metro-taxi and bus-metro are 0.25, 0.28, 0.96 and 0.09, 0.11, 0.27 respectively.

(2) The correlation level of bus-metro, bus-taxi and metro-taxi are different in different TAZs. The areas with high correlation coefficients of metro-taxi and metro-bus is along the metro lines. The areas with high correlation coefficients of bus-taxi are scattered throughout the city.

ACKNOWLEDGEMENTS

This research was supported in part by the National Science Foundation of China (#41401444, #41671387, #91546106), Natural Science Foundation of SZU (#2016065), Shenzhen Scientific Research and Development Funding Program (#CXZZ20150504141623042), and the open fund of key laboratory of urban land resources monitoring and simulation, ministry of land and resources (KF-2016-02-009, KF-2016-02-010).

REFERENCES

- Box, J. F. Guinness, Gosset, Fisher, and Small Samples. *Statistical Science*, 1987, 2(1): pp. 45-52.
- Cao, R., Tu, W., Cao, J., and Li, Q. Comparison of urban human movements inferring from multi-source spatial-temporal data, *International. Archive of Photogrammetry and Remote Sensing. Spatial Information Science*, XLI-B2, pp.471-476, doi:10.5194/isprs-archives-XLI-B2-471-2016, 2016.
- Chen, Y. Trip Characteristics Analysis and Transportation Development Countermeasures Discussion of Hangzhou Residents. *Urban Roads Bridges & Flood Control*, 2009(12):44-48.
- Deng M., Xie, L. Trip Characteristics Analysis and Transportation Improvement Strategy of Guangzhou Residents. *Urban Planning Forum*, 2000, 24(11), pp.45-49.
- FANG, Z., SHAW, S.L., TU, W., LI, Q., LI, Y. Spatio temporal analysis of critical transportation links based on time geographic concepts: a case study of critical bridges in Wuhan, China. *Journal of Transport Geography*. 2012, 23(1), pp. 44-59.
- Huang, S., Song, R., Tao, Y. Research of Citizen Travel Mode Choosing Behaviour and Its influencing factors-A case study of Beijing. *Communications Standardization*, 2008(9), pp.124-128.
- Jang, G., Zhang, C., Wu, Z. The Influence Analysis Review of Traffic Information on Travel Behaviour. *Journal of Wuhan University of Technology (Transportation Science & Engineering)*, 2012, 36(3), pp. 450-453.
- Kang, C., S. Sobolevsky, Y. Liu, and C. Ratti. 2013. Exploring Human Movements in Singapore: A Comparative Analysis Based on Mobile Phone and Taxicab Usages. In *UrbComp'13 Proceedings of the 3rd ACM SIGKDD International Workshop on Urban Computing* pp.330 – 339.Chicago, IL: ACM.
- Li, C., He, Y., Zhang, M. Reasonable Application of Consistency Test Method. *Shanghai Archives of Psychiatry*, 2000, 12(4), pp.228-230.
- Li, X., Wang, J. Application of Correlation Analysis for Coherence Test. *Tactical Missile Technology*, 2012(3), pp.60-62.
- Liu, Y., Wang, F., Xiao, Y., et al. Urban land uses and traffic 'source-sink areas': Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning*, 2012, 106 (1):73-87.
- Rhee, I, Shin, M., Hong, S., et al. On the Levy-Walk Nature of Human Mobility. *Proceedings-IEEE INFOCOM*, 2008, 19(3), pp. 924-932.
- TU, W., FANG, Z., LI, Q. Exploring time varying shortest path of urban OD pairs based on floating car data. *18th International Conference on Geoinformatics, Geoinformatics 2010*. 2010. Beijing. IEEE.
- TU, W., LI, Q., FANG, Z., Shaw, S., ZHOU, B., CHANG, X. Optimizing the locations of electric taxi charging stations: A spatial – temporal demand coverage approach. *Transportation Research Part C: Emerging Technologies*. 2016, 65, pp. 172-189.
- TU W., Cao, J., Yue, Y., Shaw, S.L, Zhou, M., Wang, Z., Chang, X., Xu, Y., Li. Q. Coupling mobile phone and social media data: a new approach to understanding urban functions and diurnal patterns. *International Journal of Geographic Information Science*. 2017. <https://doi.org/10.1080/13658816.2017.1356464>
- Wu J., Huang L., Liu Y., et al. Urban Traffic Flow Simulation based on Mobile phone data. *Acta Geographica Sinica*, 2012, 67(12), pp.1657-1665.
- Yao, Y., Guan, H., Sun, L, et al. Effect Factors Analysis of Trips with Public Transpotation. *The International Conference of Transportation Professionals*. 2006.
- Yue, Y., Zhuang, Y., Yeh G. O. A. Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *International Journal of Geographical Information Science*. 2016, DOI: 10.1080/13658816.2016.1220561
- Zhou, M., Wang, D., Li, Q., Yue, Y., Tu, W., Cao, R. Impacts of weather on public transport ridership: Results from mining data from different sources. *Transportation Research Part C: Emerging Technologies*. 2017, 75, pp.17 – 29.