

Understanding Spatial and Temporal Change Patterns of Population in Urban Areas Using Mobile Phone Data

Tianwen Liang¹, Huan Liu² and Zheng Zhang^{3*}

^{1,2} Research Institute of Highway Ministry of Transport, Beijing, 100086, China

³ College of Metropolitan Transportation, Beijing University of Technology, Chaoyang District, Beijing, China

Abstract. The wide application of information computing technology has allowed for the emergence of big data on tracing human activities. Therefore, it provides an opportunity to explore temporal profile of population changes in geographical area subdivisions. In this paper, we present a multi-step method to characterize and approximate temporal changes of population in a geographical area subdivision using eigen decomposition. Datasets in weekday and weekend are decomposed to obtain the principal temporal change profiles in Xiamen, China. The Principal Components are common patterns of temporal population changes shared by most geographical area subdivisions. Its corresponding elements in eigenvectors could be regard as a coefficient to principal components. Then, a measure, which is the similarity of each eigenvector to a basis vector, that could characterize the temporal population change is established. Based on this, the coupling interaction between population changes and land use characteristics is explored using this measure. It shows that it is restricted by land use characteristics and also is a reflection of population changes over time. These results provided an insight on understanding temporal population change patterns and it would help to improve urban planning and establish a job-housing balance.

1 Introduction

With the rapid development of China's economy and society, the frequency at which Chinese citizens travel for business activities, official activities, tourism and entertainment becomes higher. Inevitably, the travel demand is becoming more personalized and diversified. To our knowledge, various travel demand considerably influences urban traffic conditions (Zhou, Murphy, & Long, 2014). For example, the job-housing spatial imbalance forces people to travel long distance from residential area to workplace, which results in excessive travel time and fuel consumption (van Acker & Witlox, 2011). At the moment, temporal changes of population in a specific geographical area subdivision vary on a regional basis. In other words, population changes are coupled with local built environment. Insights on population changes and interactions with built environment in a geographical area subdivision are essential for developing traffic management strategies for distinctive travel demand. It can also provide efficient, safe, environment-friendly and equitable transportation services to the public.

Temporal changes of population in a geographical area subdivision along the time have something related to human mobility and mobility patterns. Previous studies on human mobility relied on travel surveys. These kinds of dataset present detailed descriptions of demographics, place of job or housing, and travel attributes at an

individual or household level. However, these datasets are not without weakness (e.g. small sample, low update frequency and limited spatial-temporal scale). Over the past few decades, the pervasive technologies like cellular networks, GPS devices, and Wi-Fi hotspots have experienced an explosion of development. These datasets are a way to capture human mobility in a higher spatial and temporal granularity (Bagrow & Lin, 2012; Kim, 2018; Yang et al., 2018). Indeed, these kinds of wealth information reflect various aspects of urban life from the perspective of mobility, consumption and environmental impact. With the lower collection cost, large sample size, higher update frequency, and broader spatial and temporal coverage, the mobile phone data provides urban planners new opportunities to explore individual mobility. Moreover, mobile phone data represents a reasonable proxy for individual mobility and robust to the substantial biases in phone ownership across various geographical and socioeconomic groups (Calabrese, Diao, Di Lorenzo, Ferreira, & Ratti, 2013; Wesolowski, Eagle, Noor, Snow, & Buckee, 2013).

The built environment involves urban design, land use and the transportation system (Handy, Boarnet, Ewing, & Killingsworth, 2002). Current evidences exemplify many of the built environment attributions that affect human physical activities (Ding, Wang, Tang, Mishra, & Liu, 2018; Mendiola & González, 2018; M. Zhang & Zhao, 2017). Although many studies have examined the influences of various built environment factors on travel behavior, energy consumption, and travel mode choice.

*Corresponding author's e-mail: 740442287@qq.com

Discusses are far away from reaching a consensus due to the practical contexts, different geographical scale, and methodologies (Antipova, Wang, & Wilmot, 2011; Park, Ewing, Scheer, & Tian, 2018; Wang & Lin, 2017).

This study seeks to investigate the effects of the built environment on temporal changes of population in geographical area subdivision scale by using a principal component analysis (PCA) approach. Moreover, a measure that is explaining these changes within each geographical area subdivision is presented. This information could be used by planners and decision-makers to plan for future transit systems, modifying areas around metro station, promoting compact urban forms, and encouraging non-motorized travel around BRT stations. The main contributions are as follows:

- We investigate the temporal mobility profile of human mobility on a geographical area subdivision based view. Some basic population changes patterns are obtained.
- We establish a measure that is explaining population changes within each geographical area subdivision.
- We validate the measure by exploring the interaction of built environment and this measure within each geographical area subdivision in a metropolitan scale.

The rest of the paper is organized as follows. The next section we reviewed the existing literature, Then, we describe our study area and dataset, followed by the methodology and results. Finally, we conclude and explore the implication for planners, and policy-makers.

2 Literature review

Understanding the dynamics of human mobility at different spatial scale is of fundamental importance as it is related to problems such as traffic forecasting, demand management, and energy consumption (Deville et al., 2014; Jiang, Guan, Zhang, Chen, & Yang, 2017; Mucelli Rezende Oliveira, Carneiro Viana, Sarraute, Brea, & Alvarez-Hamelin, 2016; D. Zhang et al., 2014; D. Zhang, Zhao, Zhang, & He, 2015). Studies on human mobility in the past decades have revealed a number of general scaling laws at both individual and population levels (Brockmann, Hufnagel, & Geisel, 2006; Marta C. Gonza'lez, 2008; Song, Koren, Wang, & Barabási, 2010; X.-Y. Yan, Wang, Gao, & Lai, 2017) and captured some human mobility routines (Liu, Gong, Gong, & Liu, 2015; Thuillier, Moalic, Lamrous, & Caminada, 2018; X. Y. Yan, Han, Wang, & Zhou, 2013; Yong, Ni, Shen, Chen, & Ji, 2018). However, temporal changes of population, which have received much less attention comparatively, are as important as regularities. Jones and Clarke discuss the question of day-to-day variability in behavior (Jones & Clarke, 1988). They present three measures of variability and argue that understanding of variability could help the assessment of policy impacts. Another work conducted by Ahas, analyze the diurnal rhythms of city life and its spatial differences using mobile phone data. Most residents have a similar temporal rhythm concerning with work, school, services and leisure in the city (Rein Ahas, Anto Aasa, Siiri Silm, 2010), and their activity locations

show a modest monthly change while there are great changes in the size of individual activity spaces (Järv, Ahas, & Witlox, 2014). However, existing empirical studies are mostly driven by data (e.g. smart card data, taxi GPS data, and mobile phone data) from an individual or population view. Limited studies focus on population changes based on geographical area subdivisions in a metropolitan scale.

Evidences from existing studies suggest that knowledge on built environment can be used to understand travel behavior and demand management across space and time (García-Palomares, Salas-Olmedo, Moya-Gómez, Condeço-Melhorado, & Gutiérrez, 2018; Vergel-Tovar & Rodriguez, 2018; Wang & Zhou, 2017). As for the relationship between human mobility and built environment, studies are often conducted under a specific context (e.g. BRT network or metro station) at a certain aggregated spatial unit such as Traffic Analysis Zones (TAZs) and grids. The associations between BRT station level demand and built environment attributes are examined in (Vergel-Tovar & Rodriguez, 2018), which suggest the extent of developable land around the station determines BRT ridership. After mining regular boarding and alighting patterns of passengers among all metro stations, Gong find temporal profile of check in and out passenger flow are highly related to the land use structure around metro stations (Gong, Lin, & Duan, 2017). Since built environment have an effect on people's travel behavior, three types of land use are examined in terms of their effects on individual residents' travel energy consumption. The results suggest that high land-use diversity and a good jobs-housing balance significantly reduced commuting trips (M. Zhang & Zhao, 2017). However, we couldn't investigate the temporal changes within each geographical area subdivision quantitatively. To solve this problem, a measure that could reflect the temporal changes within each geographical would be needed. Under different spatial scale, application context, and dataset, a substantial body of literature has examined the impacts of the built environment on human mobility from an individual or population view. As a result, relationship between this measure and temporal changes of population are necessary to check whether the measure is reasonable or not. This paper explores and measures population changes in each geographical area subdivision using mobile phone data. The aim of this study is twofold: Firstly, a measure that explains temporal changes of population within each geographical area subdivision is presented. Secondly, empirical study on measuring population changes by the proposed methods and correlation analysis between built environment and temporal population changes in the same TAZ using mobile phone data in Xiamen, China is presented.

3 Study area and dataset

Xiamen, facing Taiwan across the sea, is an important window and base for foreign contacts in China's southeast coastal area. It is made up of Xiamen Island and some coastal areas, as shown in Figure 1. As one of the five earliest Special Economic Zones (SEZ), it has been a city

growing in strength. Despite its fame as an industrial powerhouse, this port city has not lost much of its charm, and as a sightseeing heaven has become one of the best areas to visit.

The mobile phone data used in this study consists of anonymous location estimation in Xiamen generated once a device connects to the cellular network. The dataset automatically records the location, the start and end time of each communication with cellular network and individual identification. We adopt 1 million anonymous users' records from 30 days in June 2015 for our analysis. To explore temporal changes of population, the difference of population between consecutive time window for each day and geographical area subdivision is summarized. After filtering out noise records, the number of records in the dataset is reduced to 2,050,936,966 records made by 3.6 million users. Traditionally, geographical area subdivisions like TAZs or grids are often used to conduct the spatial and temporal analysis. In this paper, traffic analysis zones (TAZs) are considered as geographical area subdivisions to conduct the analysis. We split a day into 24 time windows. Theoretically, we can obtain 24 samples for each geographical area subdivision in a day.

In general, population in a TAZ along the day is consecutive time series with obvious tendency while temporal changes of population are the difference between consecutive time windows, which is also a time series. The first task we should do it to estimate the population within a TAZ in each time slot. We follow the step in (Calabrese, Ferrari, & Blondel, 2014) to summarize population in every TAZ. (1) For each base station, estimating population communicated with this base station during an hour; (2) Mapping base stations to TAZ. In this case, a base station corresponds to a unique TAZ; (3) Summing up population fall into this TAZ and time interval.

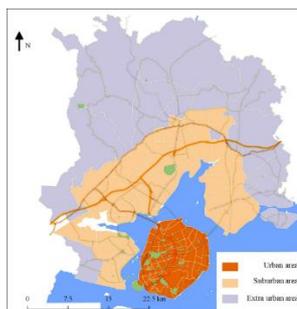


Figure 1. Study area.

4 Methodology and results

Human activities can reflect the land use development. Although human activity data being difficult to obtain directly, we can infer this information from the human mobility information because of human activities are coupled with land use characteristics. For example, people gathering around in a TAZ during morning peak hour on a weekday means this TAZ is a work-related land use type. People departing from a TAZ during a morning peak hour on a weekday implies it has a dominance of residential land use type. Therefore, we will explore the relationship between human activities and built

environment under the assumption of the mobility-activity land use relationship in the remainder of this paper.

Temporal changes of population in a TAZ are restricted by built environment around this TAZ but differ from zone to zone. We, therefore, attempt to explore the underlying patterns of population changes over time in a metropolitan area and characterize these change patterns in TAZs using low dimensional structures.

5 Principal Components Analysis and Singular Value Decomposition

Principal component analysis (PCA) investigates the intrinsic structure of complex original data by employing a coordinate transformation method that maps the original data with a new set of axes. These axes are called the principal axes or components. Each principal component has the property that it points in the direction of maximum variation or energy remaining in the data. Meanwhile, principal components are mutually orthogonal. Because the principal axes are ordered by the amount of energy in the data they capture, it appears that the structure of the complicated original data can be captured well by few principal components. This is how the dimension reduction is done by applying PCA. Calculating the principal components is equivalent to solving the symmetric eigenvalue problem for the matrix $X^T X$, and is closely related to Singular Value Decomposition (SVD)(Series & Algebra, n.d.). In the rest of this subsection, we briefly introduce SVD.

6 Spatiotemporal matrix decomposition and approximation

Let m indicates the number of TAZs in a metropolitan area while n is the number of successive time windows. Let X be a $n \times m$ measurement matrix with rank r , where column j indicates the time series of population changes in the j -th TAZ, and row i represents an instance of all population changes at time i . Because eigenvectors of $X^T X$ are principal components of measurement X (Series & Algebra, n.d.), it can be drawing that:

$$X^T X v_i = \lambda_i v_i \quad i = 1, \dots, r \quad (1)$$

Where λ_i is the eigenvalue corresponding to the $m \times 1$ eigenvector v_i . Furthermore, because $X^T X$ is a positive semidefinite matrix. Hence, its eigenvalues $\{\lambda_i\}_{i=1}^r$ that capture the energy along each principal component is positive, and the eigenvectors $\{v_i\}_{i=1}^r$ are orthogonal with each other. The definition of singular values is:

$$\sigma_i = \sqrt{\lambda_i} \quad (i = 1, \dots, r) \quad (2)$$

$$u_i = \frac{X v_i}{\sigma_i} \quad (i = 1, \dots, r) \quad (3)$$

Equation (3) shows that when population changes are weighted by v_i , producing one dimension of the transformed data. Thus, vector u_i captures the temporal variation to all changes along principal axis i . Because the set of $\{u_i\}_{i=1}^r$ captures population changes over time,

termed as Eigen Geographical area subdivision in this paper, marked with eigen G_{u_i} , $i=1, \dots, r$. The set of eigenvectors $\{v_i\}_{i=1}^r$ can be arranged in order as columns of a matrix V , which has size $m \times m$. Likewise, we can form the $n \times n$ matrix U in which column i is u_i . Then, V , U , and σ_i can be arranged to reconstruct as:

$$X = U \Sigma V^T \quad (4)$$

Where U has $\{u_i\}_{i=1}^r$ as its columns and V has $\{v_i\}_{i=1}^r$ as its columns, Σ is a $n \times m$ diagonal matrix with singular values on the diagonal. The principal component of matrix X can be obtained by $U \Sigma$. what's more, each column X_i can be written as:

$$X_i = \sum_{j=1}^r v_{ij} \cdot (\sigma_j u_j) \quad i = 1, \dots, m \quad (5)$$

Where v_{ij} represents the i -th row and the j -th column. Equation (5) indicates that each column of X is a linear combination of $\{u_i\}_{i=1}^r$. The truncated form of SVD is:

$$\hat{X} = U_t \Sigma_t V_t^T \quad (6)$$

Where \hat{X} is an approximation of X . Only the first t columns of U , the first t rows of V^T , and the top-left part of Σ size t are involved in the calculation, with the remaining entries of the matrices abandoned. The variable t can be interpreted as the intrinsic dimensionality.

How many principal components adequately explain the total variation in X is an important but controversial problem in principal component analysis. There are two types of methods: the first one is the rules-of-thumb method such as cumulative percentage of total variation, size of variances of principal components, and scree graph; and the other one is more formal. For example, rules based formal tests of hypothesis and statistically based rules. Here, we determined the number of principal components based on cumulative percentage with its total variance should reach 90%.

To decompose the spatiotemporal matrix of population changes along time in a metropolitan area, TAZs are taken as the variable and the number of population in a time slot is taken as a sample. Considering the number of population in TAZs varies from day to day, the population change patterns in weekdays and weekends would also differ. Therefore, we categorize days in June 2015 into two datasets: weekdays and weekends. In general, the first several principal components with the largest proportion of the variance will reflect the most common patterns of population changes. According to equation (6), temporal changes of population in a TAZ can be characterized with the first few elements of the eigen geographical area subdivision.

7 Measurement on temporal patterns of population changes using Eigen Geographical area subdivision

Two issues exist when the eigen geographical area subdivision is used to explore the temporal patterns of population in a TAZ along the day. The first problem is the ability of the Eigen Geographical area subdivision to approximate the temporal changes of population at both the aggregation and individual TAZ levels. To solve this problem, using the cumulative proportion of variance

curves, we examine the extent of how many principal components and Eigen Geographical area subdivision in explaining the variance of the original dataset and each TAZ respectively. The second problem is to characterize population changes within each TAZ quantitatively. According to equation (5), what determine the structure of population changes within a TAZ depends on two parts: eigenvectors and Eigen geographical area subdivision. Since eigen geographical area subdivisions are common change patterns shared by all TAZs, eigenvectors can be regarded as a unique coefficient for each corresponding TAZ. Based on this thought, we need to find a vector that could be used as a medium, we named E here, to make a comparison between eigenvectors. For a TAZ, population changes equal the total of all principal components if its corresponding eigenvectors is a vector with all element of being one, in that case, population changes in other TAZs can be got on this TAZ basis.

For a TAZ T_a , if its corresponding eigenvector is a vector with all elements being 1, the total population change in T_a is equal to the sum of all principal components of spatial and temporal matrix X , in which case temporal population changes in another TAZ can be obtained by calculating the similarity of the eigenvector of T_a to its corresponding eigenvectors. To this end, we adopt the ω_i as a measure on explaining the deviation between a TAZ and the sum of all principal components, where ω_i is the similarity between vectors $v_i = (v_{i1}, \dots, v_{im}, \dots, v_{ir})$ and $E = (1, \dots, 1, \dots, 1)$, E is an r -dimensional vector with an element of 1.

$$\omega_i = \arccos \left(\frac{v_i \cdot E}{|v_i| \cdot |E|} \right) \quad (7)$$

8 Exploring temporal changes of population

Two datasets, temporal changes of population in TAZs on weekdays (TAZD), temporal changes of population in TAZs on weekends (TAZE), are decomposed to obtain the principal components and eigenvectors. Then, the cumulative proportion of variance is used to examine the extent to which the number of principal components approximates original dataset and each TAZ. As shown in Figure 2. Figure 2(a) suggests the ability in explaining the original dataset while the rest in Figure 2 implies the approximation to each TAZ. Because the service area of a base station varies from urban areas to outskirts. Thus, we summarized that how many principal components would make a good estimation to a TAZ in urban, suburban and extra urban areas separately. In Figure 2, we can see that it needs more principal components to approximate a TAZ compared with approximating the original dataset. The number of principal components to make a good estimation of a TAZ would gradually grow from urban areas to outskirts. That exhibited the regular human activities on urban areas, i.e., going to work from home during the morning peak and leaving for home during the evening peak. Human activities are more stochastic and harder to capture on extra urban areas. A common change pattern of population on urban areas would make a better reflection of the relationship between

TAZs and built environment. Taken together, the number of principal components is taken to characterize temporal changes of population in this paper, which its total

variance can illustrate 90% of the average variance of TAZ on suburban areas.

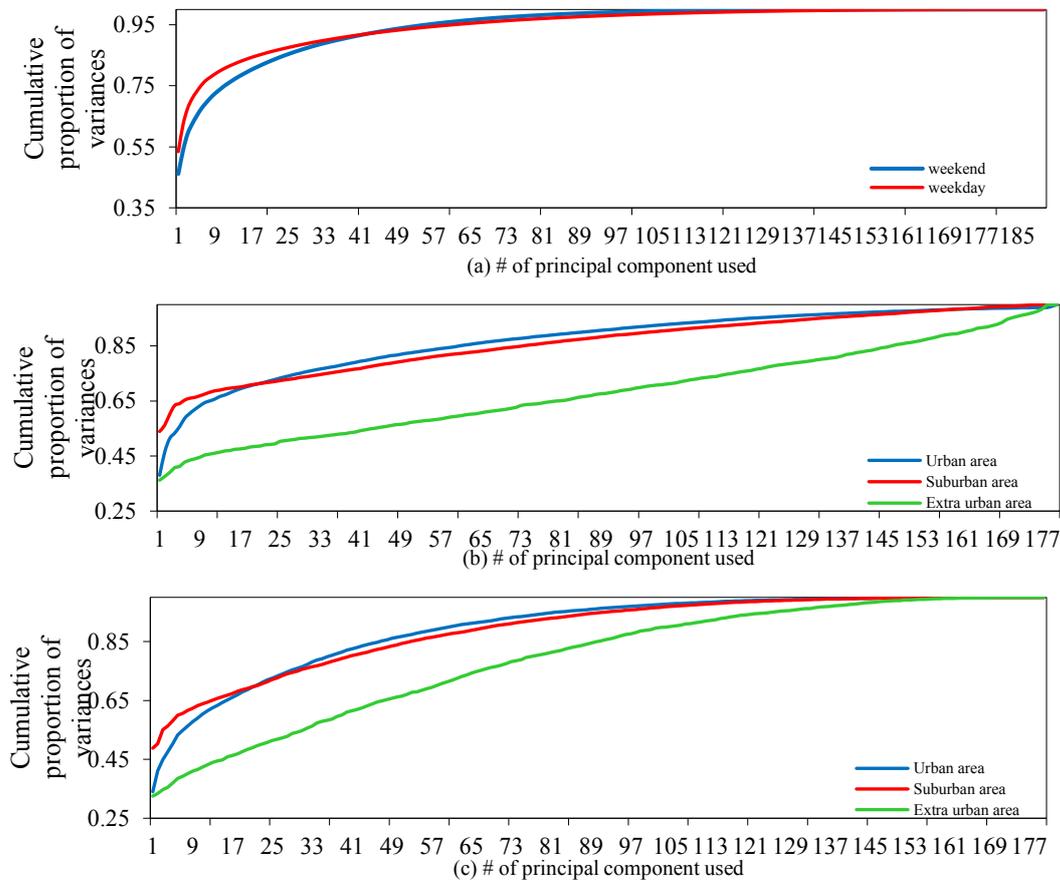


Figure 2. Cumulative proportion of variance: (a)explaining the variance of the original dataset; (b),(c) explaining the variance of each TAZ.

As described in previous sections, principal components indicate common population changes at the TAZ level, the corresponding eigenvectors can label the characteristic of each TAZ. Figure 3 is a total of all principal components when its corresponding eigenvector is equal to E . The positive values indicate the crowd gathering in a TAZ while the negative values imply population going down. Thus, the similarity between E and each eigenvector implies the structure of temporal changes of population for the corresponding TAZ. The element in eigenvector is positive indicating that dataset had the characteristics of the corresponding principal components more or less in the positive direction, whereas the negative element shows the opposite patterns of the corresponding principal components.

The temporal change profile of TAZD is shown in Figure 3(a). They had very large positive values during the morning and small peaks during afternoon and evening. Figure 4 concludes the simplified patterns of TAZD when ω_i varies from 0° to 180° . According to the physical implication, $\omega \approx 0^\circ$ suggests that temporal changes of population are similar with the city rhythm

while $\omega \approx 180^\circ$ implies that temporal changes of population are opposite to this temporal profile. The value of ω varies approximately from 72° to 105° in both datasets (Figure 4), indicating that a larger volume of people going out during the same time period in the morning peak but prefer to coming in more dispersion during the afternoon and evening, i.e., going to work from home during the morning peak, and leaving for home during the evening peak.

Similarly, change patterns of TAZE is investigated in Figure 3(b). Meanwhile, the simplify patterns of ω are concluded in Figure 5. People prefer to go out during the morning, i.e. this represents the most common patterns of people during the weekend to participate in outdoor activities. The value of ω varies approximately from 70° to 110° (Figure 5). Compared with weekdays, the similarity illustrates a larger fluctuation because of the activity pattern of people during weekend becoming more stochastic.

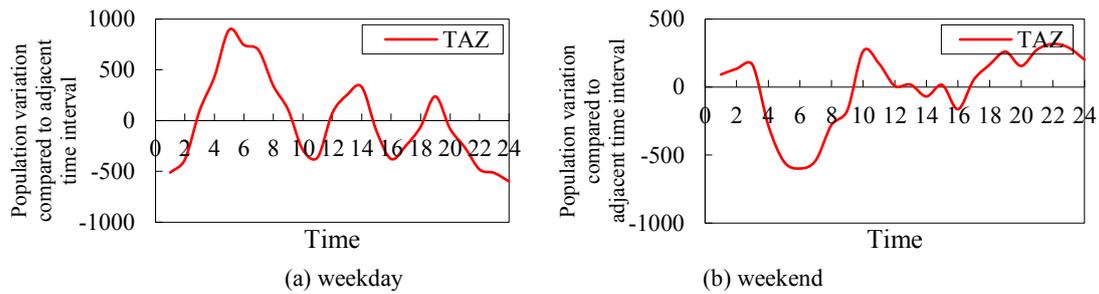


Figure 3. Total of selected principal components.

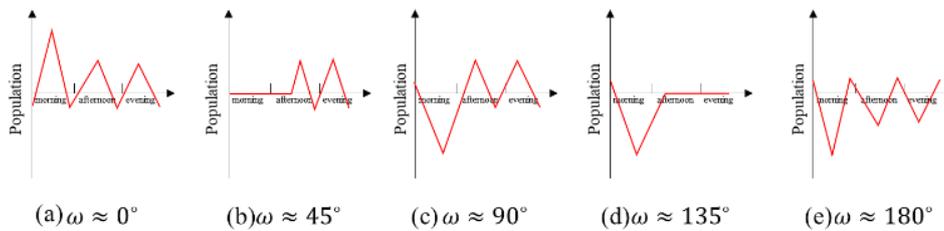


Figure 4. Simplified patterns of TAZD when ω varies from 0° to 180° .

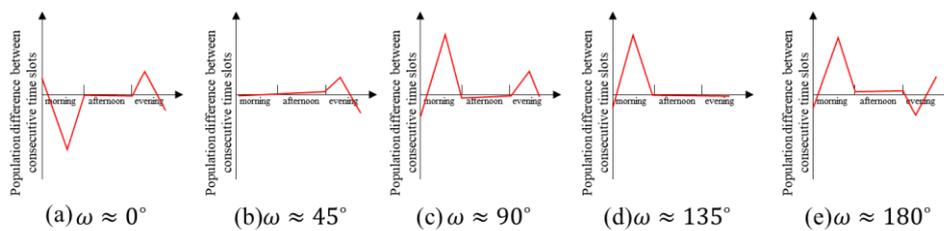


Figure 5. Simplified patterns of TAZE when ω varies from 0° to 180° .

9 Relationship of population changes with land use structure

The measure, ω , is used to characterize a temporal population changes, and population changes are coupled with the land use type within TAZs. On weekdays, the most common activities in the urban system are going to work from home at approximately 8:00 AM and leaving the workplace and going home at approximately 6:00 PM. Whereas, during the weekend, people generally go out for entertainment in the morning and go home at night. Therefore, TAZD dataset could exhibit the structure of residential land use and workplaces such as industry and commercial land use. When the ω shifted from 90° to the maximum (or minimum), the people decreases (increases) during the evening (or morning) peak, showing the dominance of workplaces over residential land use (or vice versa). When ω of TAZD is 90° , it indicates the dominance of residential land use over workplaces.

However, on the weekend the workplace as well as businesses and entertainment land use attract people in the morning and are a source of people going back in the evening. When ω of a TAZEi approaching 0° , the TAZ has a dominance of residential land use. The TAZ is characterized by entertainment land use when ω of TAZE is 90° . When ω of TAZE is 180° , implying that the TAZ has a dominance of workplaces.

Thus, the ω values of TAZD and TAZE appear to be highly related to the built environment around TAZs. As one ω describes TAZs from one aspect, two ω could combined multiple aspects to reveal the characteristics of the TAZ in a more subtle and precise manner. Considering that the development strength around the same TAZ has little difference among different types of land use. the ratio of different land use areas is adopted to represent the structure of land use around a TAZ. The residential land use, service land use, and industrial land use are extracted from the current land use map of Xiamen from 2015, which is collected by Xiamen municipal commission of urban planning, to construct the independent variables. In general, residential land use generated trips during the morning peak hour and attracted people during the evening peak hour, while the service and industrial land use are the opposite. Therefore, the ratio of residential land use to service land use and industrial land use is used as the explanatory variable. To avoid invalid values for the explanatory variables, the following equation is used to represent the ratio of land use a to land use b:

$$avb = a / (a + b + 1) \quad (8)$$

Here, a and b may be residential land use, service land use, or industrial land use. The population changes are likely related not only to the land use around the TAZ but also to the public transportation system. Therefore, two variables are selected to represent the public

transportation system. The dummy variable Public Transport is used to indicate whether the TAZ is linked to an external public port, such as a checkpoint, railway station, or airport, and the variable PublicTransport Num is used to represent the number of bus stations or metro station within the TAZ. The variables and descriptive statistics are listed in Table 1. Ordinary Least Squares (OLS) regression model analysis is conducted on a TAZ basis. The model is calibrated to obtain the coefficients of each variable, as listed in Table 2. The independent variables PublicTransport and PublicTransportNum are not significant, which indicates that the inner or outer transportation system have little effects on the population change patterns. In other words, the population changes of a TAZ result mainly from its catchment area but not from other public transportation systems.

The coefficients of rvs are significant at the 0.000 level, and rvi is not significant. Therefore, the ratio of residential land use to service land use correlate more with population changes compared with the ratio of residential land use to industrial land use.

When TAZD ω is the dependent variable, rvs and the intercept are 86.93 and 76.36, respectively. Therefore, when rvs approaching 0, the TAZD ω is 76.36°, indicating a low volume of passengers during the morning and a high volume of passengers during the afternoon. When rvs increasing by 0.01, the ω of TAZD increases by 1.37°. Accordingly, we can figure out the rvs when residential land use and service land use are balanced using TAZD ω .

Table 1. Descriptive Statistics for All the Independent and Dependent Variables.

Variable Names	Explanations	Min.	Max.	Mean	SD
Dependent variables					
TAZD ω	The ω of TAZD	72.654	105.382	89.875	5.973
TAZE ω	The ω of TAZE	76.260	109.085	91.127	6.775
Independent variables					
rvs	The ratio of residential land use to service land use	0	0.98	0.63	0.22
rvi	The ratio of residential land use to industrial land use	0	0.97	0.54	0.23
PublicTransport	1 if the geographical unit is linked to a public transport system	0	1	0.05	0.22
PublicTransportNum	The number of bus stations or metro stations within the geographical unit	0	134	34.16	25.95

Table 2. Regression coefficients

	TAZD ω	TAZE ω
rvs	86.93**	83.03**
rvi	-1.96*	-42.40*
PublicTransport	-12.52	-7.97
PublicTransportNum	-0.094	-0.042
Intercept	76.36*	78.55*
Adjust R2	0.6311	0.6024

Significant codes: 0.01***, 0.05**, 0.1*.

When TAZE ω is used as a dependent variable, the situation is much like that of TAZD ω , when rvs is 0, i.e., service land use is dominant over residential land use, the TAZE ω is approximately 78.55, which indicates that a large volume of people gathering around during the morning peak. When rvs increasing by 0.01, the ω of TAZE decreases by 0.8°.

10 Discussions and conclusions

The broad application of pervasive computing technology provides us an opportunity to trace human activities. Insights on human mobility patterns would help urban

planners and transportation operators to make reasonable and efficient transport management policy. In this paper, we present a work to characterize and approximate temporal changes of population in a geographical area subdivision using PCA. Temporal changes of population in a geographical area subdivision are presenting a regular temporal profile over a long period of time. Besides, the population change patterns are both restricted by land use characteristics and are also a reflection of them. So, exploring the interaction between population changes and land use characteristics would help to obtain job-housing balance.

Since mobile phone trace data can be regarded as a reasonable proxy for individual mobility. We construct a population changes matrix using mobile phone data, where the row indicates a time instant for all geographical area subdivisions while the column indicates temporal population changes profile in a specific geographical area subdivision. Based on this spatial and temporal matrix, an eigen decomposition method is applied to this matrix. Common population change patterns shared by most of geographical area subdivisions are extracted. After that, a measure that could be used to characterize temporal change profile for each geographical area subdivision is constructed. In order to exemplify whether this measure is reasonable or not. The linkage between this measure and land use characteristics with each geographical area subdivision are investigated. It can help to understand the urban system from a new perspective.

This method could also help improve the planning of urban space. Under the context of land mixed-use in a metropolitan area, the primary premise of establishing a job-housing balance is that we should acquire a good knowledge of the relationship between land use and population changes in a geographical area subdivision. The values of measure we establish here could indicate the dominant land use type within a geographical area subdivision. Through this, we can use the measure to check if this geographical area subdivision could obtain job-housing balance.

Although we can reveal the temporal population changes profile using PAC. Several directions are worth noting for future research. For example, population in a geographical area subdivision in a time slot is actually static, population change is generated because of travel demand. So, alternative information like trips, trips by bus and trips by taxi may be more convincing to describe the temporal change profile influenced by land use characteristics. The analysis can also be adapted for different purpose, such as clustering temporal population changes into several classes, figuring out the relationship of typical temporal change patterns with land use characteristics.

References

1. Antipova, A., Wang, F., & Wilmot, C. (2011). Urban land uses, socio-demographic attributes and commuting: A multilevel modeling approach. *Applied Geography*, 31(3), 1010–1018.
2. Bagrow, J. P., & Lin, Y.-R. (2012). Mesoscopic structure and social aspects of human mobility, 7(5), 3–8.
3. Brockmann, D., Hufnagel, L., & Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439(7075), 462–465.
4. Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira, J., & Ratti, C. (2013). Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. *Transportation Research Part C: Emerging Technologies*, 26, 301–313.
5. Calabrese, F., Ferrari, L., & Blondel, V. D. (2014). Urban Sensing Using Mobile Phone Network Data: A Survey of Research. *ACM Computing Surveys*, 47(2), 1–20.
6. Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., ... Tatem, A. J. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, 111(45), 15888–15893.
7. Ding, C., Wang, Y., Tang, T., Mishra, S., & Liu, C. (2018). Joint analysis of the spatial impacts of built environment on car ownership and travel mode choice. *Transportation Research Part D: Transport and Environment*, 60, 28–40.
8. García-Palomares, J. C., Salas-Olmedo, M. H., Moya-Gómez, B., Condeço-Melhorado, A., & Gutiérrez, J. (2018). City dynamics through Twitter: Relationships between land use and spatiotemporal demographics. *Cities*, 72(September 2017), 310–319.
9. Gong, Y., Lin, Y., & Duan, Z. (2017). Exploring the spatiotemporal structure of dynamic urban space using metro smart card records. *Computers, Environment and Urban Systems*, 64, 169–183.
10. Handy, S. L., Boarnet, M. G., Ewing, R., & Killingsworth, R. E. (2002). How the built environment affects physical activity: Views from urban planning. *American Journal of Preventive Medicine*, 23(2 SUPPL. 1), 64–73.
11. Järv, O., Ahas, R., & Witlox, F. (2014). Understanding monthly variability in human activity spaces: A twelve-month study using mobile phone call detail records. *Transportation Research Part C: Emerging Technologies*, 38, 122–135.
12. Jiang, S., Guan, W., Zhang, W., Chen, X., & Yang, L. (2017). Human mobility in space from three modes of public transportation. *Physica A: Statistical Mechanics and Its Applications*, 483, 227–238.
13. Jones, P., & Clarke, M. (1988). The significance and measurement of variability in travel behaviour. *Transportation*, 15(1–2), 65–87.
14. Kim, Y. L. (2018). Seoul's Wi-Fi hotspots: Wi-Fi access points as an indicator of urban vitality. *Computers, Environment and Urban Systems*, (October 2017), 0–1.
15. Liu, X., Gong, L., Gong, Y., & Liu, Y. (2015). Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography*, 43, 78–90.
16. Marta C. González, C. A. H. A.-L. B. (2008). Understanding individual human mobility patterns. *Nature*, 453(5), 779–782.
17. Mendiola, L., & González, P. (2018). Temporal dynamics in the relationship between land use factors and modal split in commuting: A local case study. *Land Use Policy*, 77(July 2017), 267–278.
18. Mucelli Rezende Oliveira, E., Carneiro Viana, A., Sarraute, C., Brea, J., & Alvarez-Hamelin, I. (2016). On the regularity of human mobility. *Pervasive and Mobile Computing*, 33, 73–90.

19. Park, K., Ewing, R., Scheer, B. C., & Tian, G. (2018). The impacts of built environment characteristics of rail station areas on household travel behavior. *Cities*, 74(August 2017), 277–283.
20. Rein Ahas, Anto Aasa, Siiri Silm, M. T. (2010). Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C*, 18(1), 45–54.
21. Series, H., & Algebra, L. (n.d.). Singular Value Decomposition and Least Squares Solutions *, 420, 403–420.
22. Song, C., Koren, T., Wang, P., & Barabási, A. L. (2010). Modelling the scaling properties of human mobility. *Nature Physics*, 6(10), 818–823.
23. Thuillier, E., Moalic, L., Lamrous, S., & Caminada, A. (2018). Clustering Weekly Patterns of Human Mobility Through Mobile Phone Data. *IEEE Transactions on Mobile Computing*, 17(4), 817–830.
24. van Acker, V., & Witlox, F. (2011). Commuting trips within tours: How is commuting related to land use? *Transportation*, 38(3), 465–486.
25. Vergel-Tovar, C. E., & Rodriguez, D. A. (2018). The ridership performance of the built environment for BRT systems: Evidence from Latin America. *Journal of Transport Geography*, (July 2017), 0–1. <https://doi.org/10.1016/j.jtrangeo.2018.06.018>
26. Wang, D., & Lin, T. (2017). Built environment, travel behavior, and residential self-selection: a study based on panel data from Beijing, China. *Transportation*, 1–24.
27. Wang, D., & Zhou, M. (2017). The built environment and travel behavior in urban China: A literature review. *Transportation Research Part D: Transport and Environment*, 52, 574–585.
28. Wesolowski, A., Eagle, N., Noor, A. M., Snow, R. W., & Buckee, C. O. (2013). The impact of biases in mobile phone ownership on estimates of human mobility The impact of biases in mobile phone ownership on estimates of human mobility, (February).
29. Yan, X.-Y., Wang, W.-X., Gao, Z.-Y., & Lai, Y.-C. (2017). Universal model of individual and population mobility on diverse spatial scales. *Nature Communications*, 8(1), 1639.
30. Yan, X. Y., Han, X. P., Wang, B. H., & Zhou, T. (2013). Diversity of individual mobility patterns and emergence of aggregated scaling laws. *Scientific Reports*, 3, 1–5.
31. Yang, Z., Franz, M. L., Zhu, S., Mahmoudi, J., Nasri, A., & Zhang, L. (2018). Analysis of Washington, DC taxi demand using GPS and land-use data. *Journal of Transport Geography*, 66(January), 35–44.
32. Yong, N., Ni, S., Shen, S., Chen, P., & Ji, X. (2018). Uncovering stable and occasional human mobility patterns: A case study of the Beijing subway. *Physica A: Statistical Mechanics and Its Applications*, 492, 28–38.
33. Zhang, D., Huang, J., Li, Y., Zhang, F., Xu, C., & He, T. (2014). Exploring human mobility with multi-source data at extremely large metropolitan scales. *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking - MobiCom '14*, 201–212.
34. Zhang, D., Zhao, J., Zhang, F., & He, T. (2015). coMobile: Real-time Human Mobility Modeling at Urban Scale Using Multi-View Learning. *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '15*, 1–10.
35. Zhang, M., & Zhao, P. (2017). The impact of land-use mix on residents' travel energy consumption: New evidence from Beijing. *Transportation Research Part D: Transport and Environment*, 57(October), 224–236.
36. Zhou, J., Murphy, E., & Long, Y. (2014). Commuting efficiency in the Beijing metropolitan area: An exploration combining smartcard and travel survey data. *Journal of Transport Geography*, 41, 175–183.