Journal of Transport Geography 90 (2021) 102901

Contents lists available at ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Variability in individual home-work activity patterns

Yang Zhou [a,b,c], Jean-Claude Thill [d,g,*], Yang Xu [e], Zhixiang Fang [f]

[a] Key Laboratory for Geographical Process Analysis & Simulation of Hubei Province, Central China Normal University, Wuhan 430079, China
[b] College of Urban and Environmental Sciences, Central China Normal University, Wuhan 430079, China
[c] Key Laboratory of Geospatial Big Data Mining and Application, Hunan Province, Changsha 410006, China
[d] Department of Geography & Earth Sciences and School of Data Science, The University of North Carolina at Charlotte, Charlotte, NC 28223, USA
[e] Department of Land-Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong
[f] State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
[g] Department of Applied Economics, Renmin University of China, Beijing 100872, China

ARTICLE INFO

Keywords:
Home-work patterns
Activity variability
Commuting flexibility
Smart card data

ABSTRACT

The way people allocate time across home and work activities determines their commuting patterns and frames much of the activities they undertake in the urban space. While inter-personal and intra-personal variability and repetitiveness in these activities have been documented, they remain largely underexplored. This study highlights the variations in and between individual home-work activity patterns by using information from metro smart card data as a proxy. To this end, the concept of individual space time usage matrix (STUM) is proposed and an analytical framework is developed in support of its use to depict how each rider allocates time in the vicinity of metro stations spatially and temporally. With this framework, we can classify space-time activity patterns that can be traced back to behavioral variability. By using Wuhan, China as a case study, variability in the number of home/work locations in personal activity patterns, and flexibility of work timeframes are investigated inter- and intra-personally. Our results show that about 25% of the population has a sophisticated home-work activity pattern that does not confirm to the ordinary 1-home 1-workplace pattern. Furthermore, even for this latter group, we find quite differentiated home and work timeframe patterns. The STUM is proved to be an effective and efficient concept to create a personal profile in analyzing the activity variability with big geospatial data.

1. Introduction

As an essential part of urban daily life, commuting describes the travel between one's residence and the workplace. Substantial variations are observed among individual commuting behaviors (Kitamura and Yamamoto, 2006; Buliung et al., 2008). Such variations can be imputed to the number of home or work locations, to changing longitudinal contexts, or to the travel mode or route individuals use (Shen et al., 2013). The plasticity exhibited by commuting behaviors mirrors the relations between home and work locations, as well as activity patterns that people form at these places. That is to say, the way people allocate time between home and work determines their commuting patterns. However, such behavioral complexities remain to be fully apprehended. This article studies forms of variability in individual home-work activity patterns, which contributes to the understanding of the factors of commuting behavior plasticity.

Due to urbanization and to the development of information and communication technologies, the complexity and variability of individual home and work activity patterns have intensified in urban areas (Järv et al., 2014; Kwan, 2007). A study using the 2015 China Household Finance Survey showed that more than 20% of Chinese urban households own multiple homes (Huang et al., 2020), while others may have co-living arrangements with parents on a routine basis. As a result, some people split residence between multiple homes. Some people have a single place of work all day and every day, while others may need to travel around the city to carry out business or meet clients. Some workers may have night work shifts or weekend work requirements, showing intra-personal variability in daily commuting; some workers

* Corresponding author at: Department of Geography & Earth Sciences, The University of North Carolina at Charlotte, 9201 University City Blvd, Charlotte, NC 28223, USA.

E-mail addresses: yzhou2017@mail.ccnu.edu.cn (Y. Zhou), Jean-Claude.Thill@uncc.edu (J.-C. Thill), yang.ls.xu@polyu.edu.hk (Y. Xu), zxfang@whu.edu.cn (Z. Fang).

https://doi.org/10.1016/j.jtrangeo.2020.102901
Received 2 July 2020; Received in revised form 28 October 2020; Accepted 30 October 2020
Available online 16 November 2020
0966-6923/

may have a lot of flexibility in their daily work schedule, while others may be very regimented, or be frequently called to work overtime (Long et al., 2016). Over a longer time horizon, a number of people would choose to re-locate their home or work locations to achieve a better compatibility (Gao et al., 2018; Huang et al., 2018). The plasticity exhibited by the home and work activity patterns is a critical property shaping the diversity of commuting behaviors in a population. Therefore, it is important to understand the contexts that permit people to have flexible home/work locations and what resulting variabilities may lie in different people's allocation of time at home and at work.

Seeking such variability in people's home-work activity pattern behavior, however, is not a trivial task. A major obstacle is the lack of fine-grained mobility observations for understanding how individuals allocate time at home and work. Although travel behavior surveys can capture the home-work activity pattern of participants, sample sizes are usually small, and they often cannot reveal behavioral heterogeneity across large populations. With the prevalence of location-aware devices such as GPS, smart cards, or cell phones in the last decade, large-scale mobility datasets have become increasingly available. Rule-based methods have been proposed to identify home and work activities from mobility tracking data, predicated on simple views on people's space-time behaviors in the urban environment, such as a single fixed home place and a single fixed work place, with a trip to work about 7–9 am and a return home at 5–7 pm. Nevertheless, interpersonal variations in home or work activities are largely ignored. In particular, it is still a challenge to tie this variability across time and geographies on the basis of big geospatial datasets.

In this paper, we study inter-personal variability in home-work activity patterns as well as intra-personal daily variations in work-related activities. To achieve this goal, we propose the construct of a space-time usage matrix (STUM) to accumulate the decomposed activities into uniform spatial and temporal units for each traveler. This construct is developed for use with information extracted from metro smart card data (SCD) over an extended period of time (3 months in our study) in Wuhan, China. The STUM is effective at revealing people's time use in the vicinity of each metro station. Thus, it is possible to identify all potential home or work locations, along with highly recurring activity schedules and flexible activity schedules during a set study period. We find that about a quarter of all transit riders have a home-work activity pattern that departs from an ordinary 1 home-place 1 work-place pattern. Furthermore, transit riders have quite differentiated work schedule patterns, even if their home-work activity pattern fits the ordinary 1 home-place 1 work-place variety.

It is worth pointing out that this study uses home and work stations as proxies for the actual home and work locations of transit riders because of data limitation. Also, the privacy of each traveler is well protected since all card IDs have been anonymized. The main contributions of this paper are as follows:

- We propose and describe the STUM, which is an analytic concept to study an individual traveler's intensity of usage of the urban space spatially and temporally with appropriate aggregation and decomposition strategies. This process is effective and efficient at creating a personal portrait of activities with big geo-spatial data.
- We study the inter- and intra-personal variability of home-work activity patterns with respect to the potential multiplicity of home and/or work locations, and to the flexibility of work schedules. Thus, the analysis reveals the richness and diversity of commuting behaviors in urban life.

The rest of the paper is organized as follows: Section 2 reviews recent studies on activity and travel behavior patterns from smart card data; Section 3 presents a new analytical framework, which includes the principles of STUM, how to implement it on a specific case study and how to derive home-work activity patterns from it; Section 4 develops the case study of home-work activity patterns in Wuhan, China; Section 5 discusses the results of the empirical analysis; and Section 6 draws the conclusions of our study and points to future work in this line of research.

2. Literature review

In recent years, many studies have been undertaken to understand transit users' travel habits from smart cards data, which contain detail records of every instance when people board or alight the public transport system. The temporal variation of swiping in or out helps to compile the ridership statistics (Tu et al., 2018), replicate dynamic travel demand and simulate operational conditions of the transit system (Liu et al., 2018), predict individual mobility (Zhao et al., 2018), and so on. The spatial interaction between stations across the urban space reveals the dynamic flows that shape the internal structure of the urban region (Roth et al., 2011), public activity centers (Cats et al., 2015), and urban functionality at the station level (Zhou et al., 2017). The study of variability in temporal regularities between cities (Zhong et al., 2016), as well as shifts in temporal habits across years (Briand et al., 2017) also provide valuable insights into how human travel patterns change in systematic ways across different spatial and temporal scales.

Commuting behavior in public transportation systems is both "repetitious" and "variable" (Hanson and Huff, 1988). It has observed that variability in the use of public transport happens between individuals (inter-personal variations), but also within individuals over time (intra-personal variations). On the one hand, it is believed that inter-personal variations in travel patterns are the external expression of the intrinsic socio-demographic characteristics of the rider. Whereas the smart card data record passengers' activity over time, they enable the partitioning or grouping of passengers according to their temporal similarity in patterns of check-in and check-out (Faroqi et al., 2018; El Mahrsi et al., 2017), or to focus on certain types of extreme transit riders based on specific rules (Long et al., 2016). Moreover, there is a rising interest in studying the relationship between the transit behavior of passengers and their socio-economic attributes. For instance, El Mahrsi et al. (2017) discovered clusters of passengers who have similar temporal boarding times and studied how socio-economic attributes can affect travel patterns based on their local residences. Along this line, Liu and Cheng (2018) extracted transit patterns by using a text mining technology and enhanced the interpretation jointly with open geodemographics derived from census data. Goulet-Langlois et al. (2016) studied passenger heterogeneity by identifying clusters from activity sequences spanning multiple weeks.

On the other hand, intra-personal activity behavior is driven by associated needs and desires of individuals, and governed by a set of constraints, which are typically examined with time-space prisms from a time geographic perspective (Hägerstrand, 1970; Kitamura and Yamamoto, 2006; Zhang and Thill, 2017). The time-consuming process of creating time-space prism makes it hard to handle a big dataset with millions of individuals. Research has concentrated on the non-spatial outcomes of travel behavior, like the frequency of boarding in SCD (Deschaintres et al., 2019). Other efforts have been made to analyze the human activity space variability by using mobile phone data (Järv et al., 2014; Xu et al., 2016). For the purpose of investigating variability in commuting behaviors, Huang et al. (2018) tracked the home and work location changes in four mobility groups with the help of a 7-year metro SCD. With these rather limited research developments, it is clear that further study on the variability of individual commuting behavior from day to day, or week to week with big geospatial data is needed. A possible reason for the dearth of research in this area is that only locations are recorded in these GPS tracking or smart card tracking datasets, rather than a full suite of activity attributes (e.g., activity type, duration, etc.), which curtails the ability to identify variability from the activity perspective.

With most SCD datasets, trip purposes and personal information are unavailable. Therefore, researchers resort to various approaches to infer

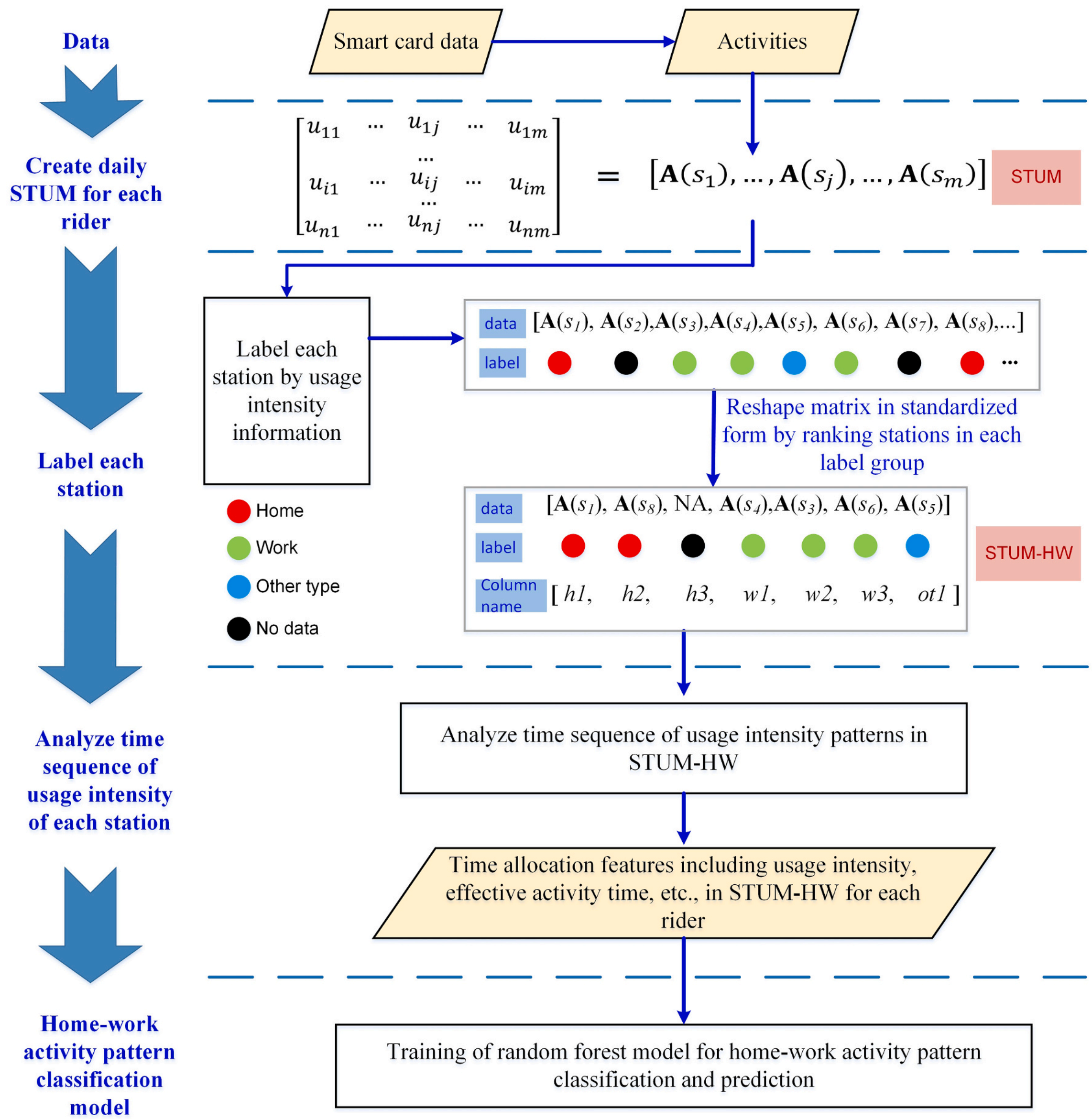


Fig. 1. Flowchart of the complete methodology.

home and work activities from limited data at the first step of their analysis. Ma et al. (2017) adopted several criteria including travel days, departure time, number of stops visited, and home/work route frequencies, by using a rough set clustering method to differentiate commuters from non-commuters. Other studies have identified the home or workplace station from SCD by some predefined rule-based method. For example, Long and Thill (2015) started with one-day data to identify the daily home stop to be the departure bus stop of the first trip, and then identified the final home location as the largest spatial cluster of daily home stops over one week. Similarly, Zhou et al. (2014) identified the workplace and home places based on activity duration rules that the duration should be more than 6 h. Liu and Cheng (2018) identified the residential station to be the most frequently visited first boarding station of frequent passengers. However, as Wang et al. (2017) pointed out, a drawback of the rule-based method is that it is hard to correctly handle commuters who have multiple home/workplace stations or have flexible or irregular work hours. Moreover, thresholds may be hard to decide upon without the support of travel survey data, considering the divergence of travel behavior between cities. A uniform rule that does not consider personal variations is likely to overlook diverse and flexible individual patterns.

Several other studies combine spatio-temporal movement patterns with surrounding land use or Points of Interest (POIs), aiming at inferring trip purposes from SCD with a probabilistic rule-based model or a decision tree (Alsger et al., 2018; Lee and Hickman, 2014). However, multi-home, multi-employment or work shift patterns are not fully discussed in those studies. Hence, the potential of SCD for understanding commuting patterns effectively and efficiently at the personal level is still challenging and has not yet to be fully exploited.

3. Methodology

Fig. 1 shows the methodological framework for home-work activity pattern classification. The logical flowchart consists of five steps. In the data processing step, activities are generated from the SCD after data cleaning. For our purposes, an activity entails the traveler's physical presence over the time interval between two consecutive trips, on condition that the alighting station of the previous trip is the same as the boarding station of the next trip, following in this respect the approaches in Chakirov and Erath (2012) and Zhou et al. (2017). Then, we create a daily STUM for each rider from all their activities. In the third step, each station in STUM is labeled by considering the usage intensity information in specific hours of the day. The STUM is then reshaped to a standardized matrix dubbed the STUM-HW so the matrix profile of each rider has the same structure. Next, we analyze the time sequence of usage intensity patterns for each station in the STUM to get the time allocation features including total usage intensity, effective activity time, and so on. In the fifth step, the exact features of each STUM-HW are input as a vector to train a random forest model for the purpose of classifying and predicting home-work activity patterns for all riders.

3.1. Space time usage matrix

The usage of a place by a traveler can be measured by how frequently and how long he/she conducts activities at that place. For each traveler, we create a profile that documents the spatio-temporal usage patterns of each metro station. We incorporate all workday activities in our various spatio-temporal measurements.

We derive a traveler's profile by constructing their STUM. The STUM is designed as a matrix (eq. (1)) with each column standing for a spatial unit, namely a station s and its surrounding in our research, $s \in (s_j, j = 1, \ldots, m)$, where m is the number of metro stations, and each row is a temporal unit t, $t \in (t_i, i = 1, \ldots, n)$, where n is the number of time slots of uniform span Δt in a day. Accordingly, the value u_{ij} in the matrix $\mathbf{U}_c$ speaks for the usage intensity in the vicinity of station s_j at time slot t_i by transit rider c.

$$\mathbf{U}_c = \begin{bmatrix} u_{11} & \ldots & u_{1j} & \ldots & u_{1m} \\ & & \ldots & & \\ u_{i1} & \ldots & u_{ij} & \ldots & u_{im} \\ & & \ldots & & \\ u_{n1} & \ldots & u_{nj} & \ldots & u_{nm} \end{bmatrix} = (u_{ij}) \in \mathbb{R}^{n\times m} \tag{1}$$

To specify matrix $\mathbf{U}_c$, we take a strategy of temporal decomposition of recorded activities that works as follows. If an activity $a = (ts, te, s)$ is detected for a traveler at station s with a start time ts and an end time te, then it can be decomposed into a binary vector of fixed length, as shown in eq. (2):

$$\mathbf{a}(s) = [x_1, \ldots, x_t, \ldots, x_n] \tag{2}$$

with $x_t = \begin{cases} 1, ts \le t \le te \\ 0, otherwise \end{cases}$, if $te \ge ts$ and $x_t = \begin{cases} 1, t1 \le t \le ts, te \le t \le tn \\ 0, otherwise \end{cases}$, if $te < ts$.

A value of 1 means the traveler is using or staying around s_1 during that time slot. An example is shown in Fig. 2. In the decomposition, the condition $te < ts$ happens when the activity starts after noon and ends before noon of the following day. Overwhelmingly, activities meeting this condition would involve staying home. Occasionally, however, the activity could be work, which would be the case of someone working on a night shift. However, because our framework is designed under the assumption that no travel behavior survey data are available to supplement SCD, it is appropriate to state that staying home is the activity associated with condition $te < ts$.

Through the decomposition process, the activity at any station s_j is represented as a usage sequence vector of time. The activity duration is obtained by simple algebraic aggregation as $\sum_{t=1}^{n} x_t \Delta t$. The decomposition $\mathbf{a}(s)$ in eq. (2) preserves the relevant information on the activity while offering great flexible for mathematical operations. Since all the activities of a transit rider over a certain study period are represented by a vector of the same length, activities that occur in the vicinity of the same station over this period can be summed up to derive this rider's usage intensity of this station. In eq. (3), $\mathbf{A}(s_j)$ is introduced to represent the accumulated usage sequences of the same person around station s_j.

$$\mathbf{A}(s_j) = \sum_{s=s_j} \boldsymbol{a}(s) \tag{3}$$

Hence, the STUM $\mathbf{U}_c$ comprises the stack of $\mathbf{A}(s_j)$ at all m stations, given as eq. (4):

$$\mathbf{U}_c = \begin{bmatrix} u_{11} & \ldots & u_{1j} & \ldots & u_{1m} \\ & & \ldots & & \\ u_{i1} & \ldots & u_{ij} & \ldots & u_{im} \\ & & \ldots & & \\ u_{n1} & \ldots & u_{nj} & \ldots & u_{nm} \end{bmatrix} = [\mathbf{A}(s_1), \ldots, \mathbf{A}(s_j), \ldots, \mathbf{A}(s_m)] \tag{4}$$

To make it possible to compare the STUMs across different travelers, we standardize $\mathbf{U}_c$ to $\mathbf{U}_c'$ using eq. (5):

$$\mathbf{U}_c' = \left(u_{ij}'\right) \in \mathbb{R}^{n\times m},$$

$$with\ u_{ij}' = \frac{u_{ij}}{\sum\limits_{\substack{1\le i\le n \\ 1\le j\le m}} u_{ij}} \tag{5}$$

Each normalized value u_{ij}' in $\mathbf{U}_c'$ denotes the relative usage intensity of a person, which shall be in the [0,1] range. A higher u_{ij}' indicates that the cardholder stays more frequently around station s_j in the time slot ti.

Certain portions of the entire space-time domain may be of interest. Let TP = $[t{:}\ t']$ represent a time span from t to t', while S represents a set of stations. Then $\mathbf{U}'[\mathrm{TP}, \mathrm{S}]$ is a subset of $\mathbf{U}_c'$ that only depicts the usage intensity in the vicinity of targeted stations during the specified time range. If we wish to know the aggregate usage intensity $ui(\mathbf{U}'[\mathrm{TP}, \mathrm{S}])$ around the set of stations S during TP, it is straightforward to add up the values in $\mathbf{U}_c'$ that satisfy $s_j \in S$ and TP = $[t{:}\ t']$, as shown in eq. (6):

$$ui(\mathbf{U}'[\mathrm{TP}, S]) = \sum_{\substack{t\le i\le t' \\ s_j\in S}} u_{ij}'. \tag{6}$$

Based on activities derived from complete records of smart card swipes, the STUM is computed from the spatio-temporal decomposition of activities and their accumulation with consistent dimensionality $n*m$. Thanks to this framework, repetitive or regular activities like staying home or working would show higher usage intensity than occasional and discretionary activities, like shopping or recreation. Furthermore, when one type of activity occurs in the vicinity of multiple stations (for example people having multiple workplaces), it is very convenient to study the entirety of typical activities with eq. (6). The distinctive

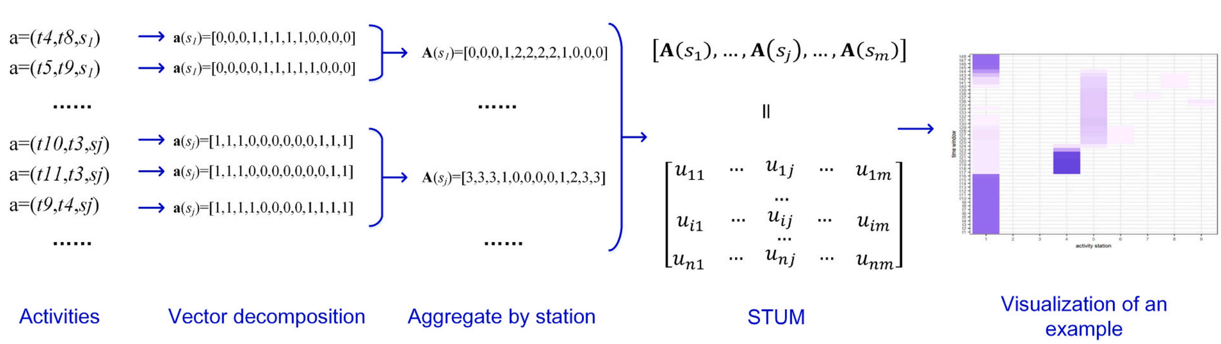
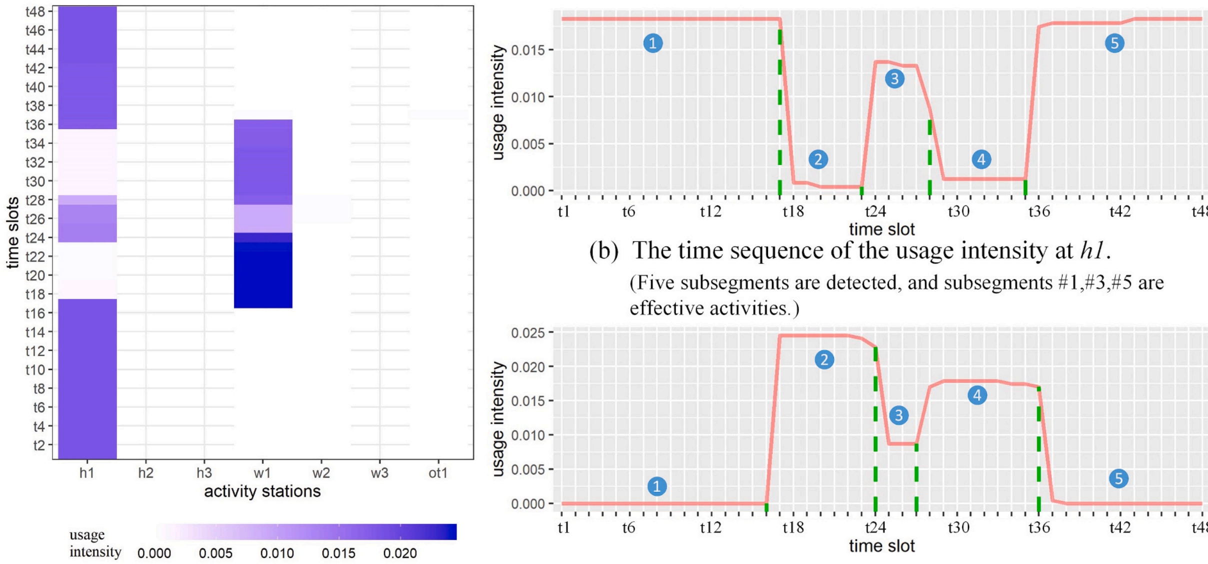


Fig. 2. Illustrative example of STUM when $n = 12$.



(a) The STUM-HW of a traveler.

(b) The time sequence of the usage intensity at *h1*.
(Five subsegments are detected, and subsegments #1,#3,#5 are effective activities.)

(c) The time sequence of the usage intensity at *w1*.
(Five subsegments are detected, and subsegments #2,#3,#4 are effective activities.)

Fig. 3. Example of time sequence segmentation at *h1* and *w1*.

advantage of STUM over a simple check-in and check-out frequency profile is to embed information on the start time, end time, activity duration, and the frequency of activities.

The temporal unit used in the STUM ought to be thoughtfully selected because it controls how and at what level of temporal granularity activities can be decomposed or aggregated. If we choose 0.5 h as time unit, then each activity will be decomposed by 30-min segment using eq. (2), which will allow to extract activities from the SCD that would pass unnoticed with a longer time slot. On the other hand, if we choose a week as a time unit, then all activities in a week will be aggregated using eq. (3) and fine-grain detail of the individual's space-time activity pattern will be smoothed out. In the following study, we first divide a day into 48 time windows (i.e., $n = 48$) numbered from t1 to t48, with each time slot standing for a $\Delta t = 0.5h$ increment. So, we create a daily individual STUM $\mathbf{U}_c'$ with a dimension of 48*m to help identify the stations associated with home, work and other types of activities. The daily STUM can be scaled up to a coarser STUM, like a weekly STUM, where the temporal scale is broadened by choosing a week as a time unit. This approach will be used in Section 5.2 to study the longer term (week to week) intra-personal change in multiple home locations that features on the space-time activity pattern.

3.2. Home-work activity characteristics

3.2.1. Labeling home and work stations in STUM

In this section, we construct daily STUMs with $n = 48$ to label potential home and work locations on each rider's time allocation pattern. A daily STUM is consistent with the day-to-day repetitiveness of many elements of travel behavior. It is customary that the home activity happens at night while the work activity happens during the daytime and our labeling will be consistent with these tendencies. The following timeframes are used: daytime is from 9 am to 5 pm, nighttime from 9 pm to 6 am, while the remaining hours are treated as "other times". Hence, we use the following terminology:

DT = [t19: t34] for daytime,
NT = [t43: t48, t1: t12] for nighttime,
OT = [t13: t18, t35: t42] for "other times",
ALL = [t1: t48] for all time slots.

By summing up the usage frequencies in each time slot via eq. (6), for each card holder we can calculate $ui(\mathbf{U}'[\text{DT}, s_j])$, $ui(\mathbf{U}'[\text{NT}, s_j])$, and $ui(\mathbf{U}'[\text{OT}, s_j])$ at each station s_j. Stations with zero usage intensity in all three timeframes are labeled 'NA' to indicate they are not used by this card holder. Then, we label each of the other stations s_j as a potential location for home, work, or other type, according to the highest frequencies among $ui(\mathbf{U}'[\text{DT}, s_j])$, $ui(\mathbf{U}'[\text{NT}, s_j])$, and $ui(\mathbf{U}'[\text{OT}, s_j])$. We are aware that activities other than work, like shopping or recreation, may also happen during DT. In order to learn the most probable activities from frequency and duration features, we will study the time sequence of ui at each station by introducing several indices in a subsequent stage of the process. It should be clarified that the three timeframes are not used as strict rules to label any single activity, but to help identify the dominant activities around each station.

Once stations have been assigned a label (i.e., home, work, or other type), we select the top-ranking stations in each label group by $ui(\mathbf{U}'[\text{ALL}, s_j])$. For the purpose of this study, we keep the top 3 stations of the home group to be *(h1,h2,h3)*, the top 3 stations of work group to be *(w1, w2, w3)*, and the top station of the other type group to be *(ot1)*. In other words, the STUM is truncated and resized to seven columns with column names *h1, h2, h3, w1, w2, w3, ot1*, as illustrated in Fig. 3(a). Here we choose the top 3 stations, considering that the fraction of people with more than 3 homes or workplaces would be quite low according to surveys in Cheng et al. (2020), and that the data show that the usage intensity at the fourth station is quite low, if not null. Each person is treated independently to identify his/her stational functions in STUM. The resulting subset matrix that focuses mainly on potential home and work stations is named the STUM-HW matrix. By repeating these steps for each traveler, a full set of individual STUM-HW matrices is generated.

3.2.2. Usage intensity characteristics

So far, we have identified a list of possible home and workplace stations and arranged them in order of the usage intensity in each rider's STUM-HW. The start time, end time as well as the duration of the

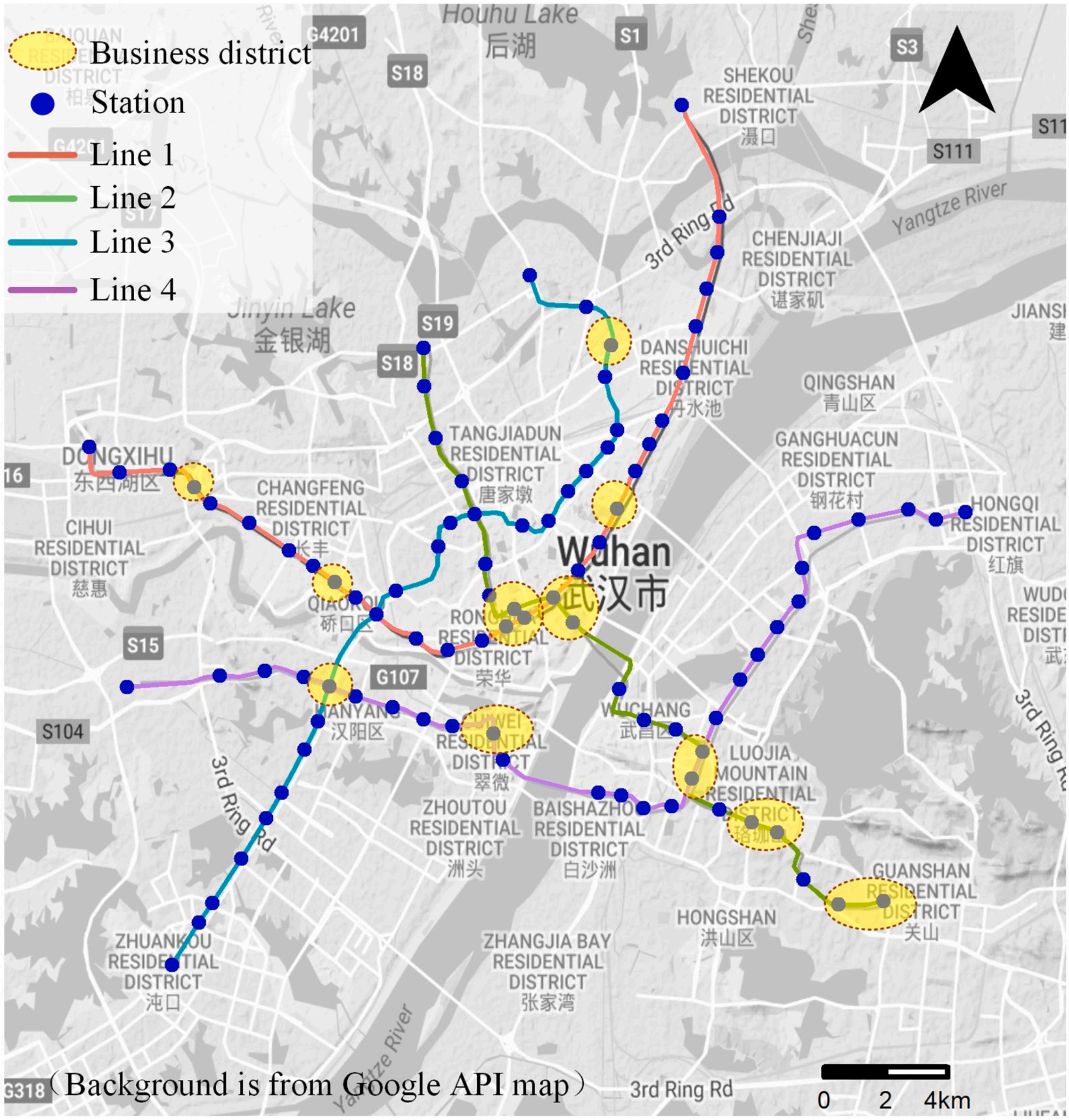


Fig. 4. Map of Wuhan's metro system in 2016.

activity at each station also matter to study how people use the station. Thus, it is important to study the time sequence of the usage intensity around each station, which is a vector of $\mathbf{U}'[\text{ALL}, s_j]$ in each column in STUM-HW. For example, the STUM-HW in Fig. 3(a) shows a traveler's activity pattern at $h1$ and $w1$; the corresponding time sequences of the usage intensity depicted in Fig. 3(b) and (c), respectively.

Since $\mathbf{U}'[\text{ALL}, s_j]$ is calculated from the accumulation of all decomposed activities, it is necessary to segment the time series based on usage intensity and find the discrete time slots that delineate significant activities (green dashed line in Fig. 3). The significant activities are associated with sub-segments of high usage intensity. We use a simplified version of the changing point detection algorithm proposed in Lavielle (2005) to cut the usage intensity sequence into effective sub-segments, with parameters *lmin* (the minimum length of a sub-segment) and α (the lagged difference). The process is as follows:

Step 1: Find all changing points k ($1 \leq k \leq n$) in the original series that satisfy $u'_{k,\,j} - u'_{k-2,\,j} >= \alpha$, cut the original time series into $k + 1$ sub-segments, and calculate the mean intensity value *x.mu* and segment length *x.sl* for each sub-segment x.

Step 2: if $x.sl < lmin$, merge the current segment x to the prior one or to the post segment, whichever has the smaller difference to *x.mu*. Loop until all segment lengths are longer than *lmin*.

After detecting all sub-segments of stational time series $\mathbf{U}'[\text{ALL}, s_j]$, we filter those effective activities considering the *x.mu* account for the maximum *ui* value max(*x.mu*) of all sub-segments. If $x.mu / \max(x.mu) \geq \beta$, then sub-segment x is considered an effective activity. As an example, there are 3 effective activities, marked 1, 3, and 5, for case featured in Fig. 3(b). The total effective activity duration can be calculated from all effective sub-segments as:

$$atl\left(\mathbf{U}'\left[\text{ALL}, s_j\right]\right) = \sum x.sl \times \Delta t, \text{if } x \text{ is effective.} \tag{7}$$

In addition, we use the usage intensity as a weight to calculate the weighted mean activity time:

$$wat\left(\mathbf{U}'\left[\text{ALL}, s_j\right]\right) = \sum \frac{x.mu^{*}x.sl}{max(x.mu)} \times \Delta t, \text{if } x \text{ is effective.} \tag{8}$$

This is particularly meaningful for the work station series, as this

Table 1
Description of the metro SCD.

Month	Transactions	Reconstructed trips	Activities	Number of cardID
2016.08 (23 workdays)	64,905,143	31,549,032	12,982,471	2,738,291
2016.09 (21 workdays)	63,522,933	30,863,748	13,454,683	2,791,085
2016.10 (18 workdays)	53,907,173	26,245,030	11,560,056	2,505,399
TOTAL	182,335,249	88,657,810	37,997,210	5,059,822

approach permits us to get the work start time $work_{start}$ as the start time of the first effective segments and the work end time $work_{end}$ as the end time of the last effective segments.

So far, the proposed analytical framework has led to propose several indices to depict *ui* characteristics at each station. This information is then used to classify the home-work patterns into different groups and study the work time flexibility in the next step.

3.3. Home-work activity patterns classification

After depicting the space-time behavioral patterns embedded in the columns of the STUM-HW using *ui*, *atl*, and *wat* (eqs. (6)–(8)), we study the interpersonal variability in home-work activity patterns that metro users exhibit. To this end, we classify card holders by a number of measures of their space-time behavioral patterns. We choose the random forest (RF) algorithm for this purpose because it is known to reduce the chance of overfitting to training data. An RF consists of many decision trees and assures the diversification and low correlation across trees by taking a bagging strategy as well as by using feature randomness when building each individual tree. The important parameters in a random forest are the number of trees to grow (*ntree*) and the number of variables randomly sampled as candidates at each split (*mtry*). Increasing the *mtry*, or *ntree* parameters will improve the performance of the model, while also increasing the computational time.

In general, we will have a STUM-HW of 48*7 dimensions for each card holder. For every column (station s_j) in the matrix, we will calculate the stational usage intensity $ui(\mathbf{U}'[\mathrm{ALL}, s_j])$ and the effective activity duration $atl(\mathbf{U}'[\mathrm{ALL}, s_j])$. Plus, we get the work start time $work_{start}$, and

Table 2
Classification results of the RF model.

Classes	Description	Training dataset	Testing dataset	Forecasting results
C1	1 home station & 1 work station	101	49	119,252
C2	1 home station & multiple work stations	49	22	20,325
C3	Multiple home stations & 1 work station	25	12	9865
C4	Multiple home stations & multiple work stations	29	10	3800
C5	Home station & other type activity station	24	21	5051
C6	Outliers	12	6	1511
Total	–	240	120	159,804

the work end time $work_{end}$ from aggregated time series of work locations $\mathbf{U}'[\mathrm{ALL}, s_j \in (w1, w2, w3)]$ as well as the weighted mean activity time *wat*. Thus, for each card holder, we have 17 features as an input to feed the RF to classify home-work activities patterns.

4. Case study

4.1. Dataset

In this study, we use metro SCD over three months (2016.08–2016.10, including 62 workdays) collected in Wuhan, China. There were 4 metro lines operating in Wuhan in 2016, and they included a total of 96 stations, as shown in Fig. 4. Wuhan's metro system was designed so that stations are spaced in the urban fabric with a catchment area of 500–600 m, in the majority of cases. Some stations, like line endpoints, may have a wider catchment area of 1000 m or more.

The original SCD record tracks each traveler's cardID, stationID where a transaction was made, the type of transaction (check-in or check-out), and the time of the transaction. In Table 1, we provide tallies on the SCD data during workdays (Monday through Friday, excluding holidays). During the 3-month period, there are about 89 million trips reconstructed from the two successive check-in (boarding) and check-out (alighting) records by same cardID after data cleaning. Activities are then identified between two consecutive rides following the

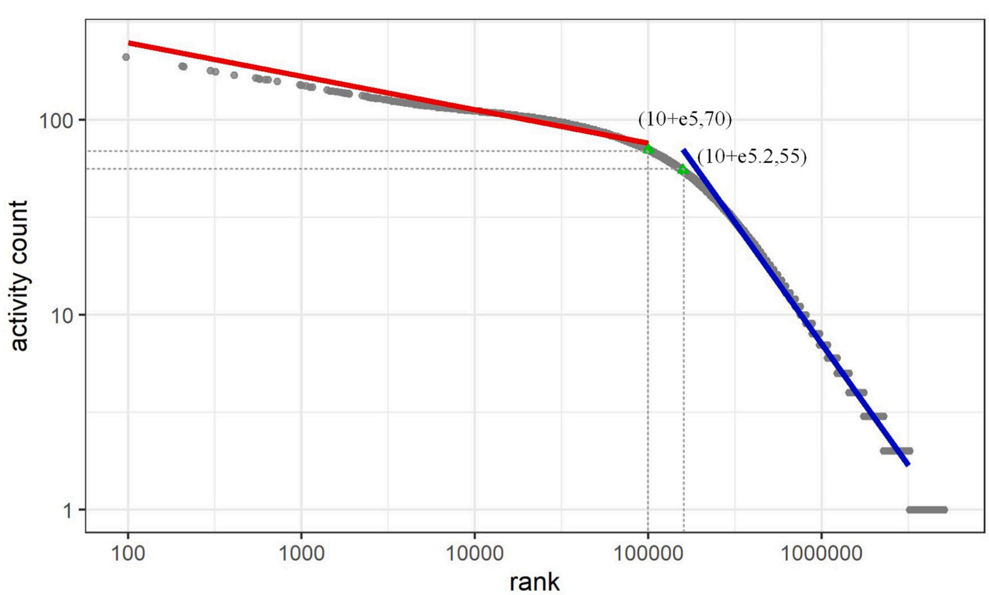


Fig. 5. The log-log plot of the activity count of transit riders.

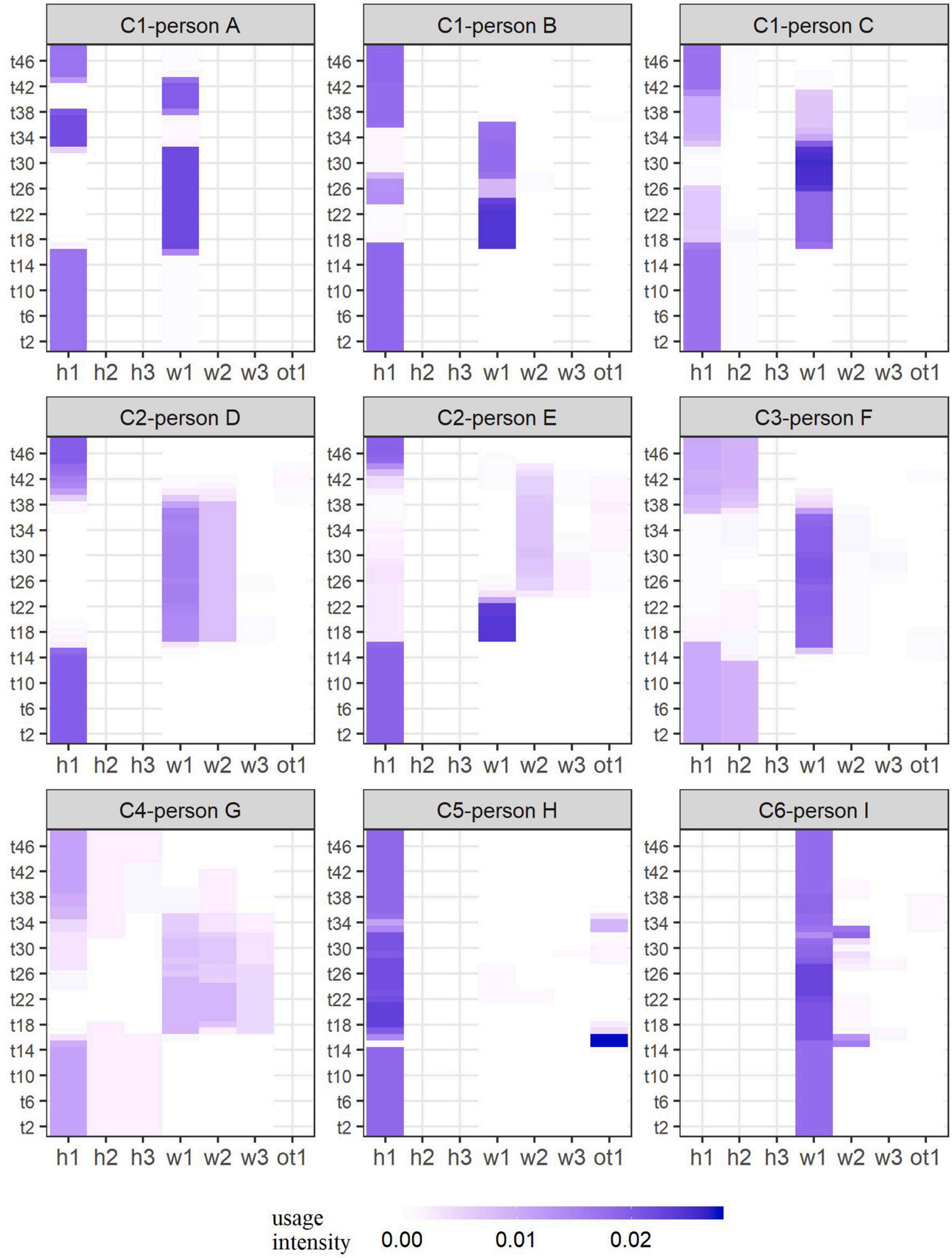


Fig. 6. STUM examples in various classes.

approach in Chakirov and Erath (2012) and Zhou et al. (2017). We generate nearly 38 million activities performed by 5 million unique card holders. As a first step, we filter out infrequent travelers to focus on travelers whose commuting trips can be retrieved. The log-log plot of weekday activity counts of each person (Fig. 5) shows that the empirical distribution of activity counts is well fitted by two truncated power law distributions. The inflection of the plot is between [55, 70]; accordingly, we choose 55 as the minimum number of weekday activities that should be recorded in the dataset for a transit rider to be regarded a frequent traveler. By this definition, we have about 12.8 million activities by 160,164 frequent travelers on workdays. That is to say, we selected 3.2% of travelers to be frequent commuters, accounting for 33.8% of total recorded activities. These activities are used to create individual STUMs in the following study.

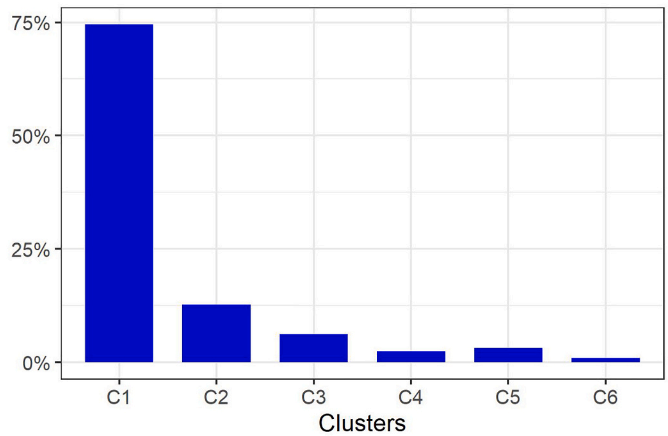


Fig. 7. RF class frequencies.

4.2. Home-work activity pattern classification

Following a sensibility analysis, we select $\alpha = 0.002$ and $lmin = 3$ as parameters in the changing point detection algorithm to cut the time sequence of usage intensity into sub-segments. If α is too high, the changing point at a gentle slope would not be detected, which would affect people who have flexible home/work departure or arrival times. $lmin = 3$ makes sure that the actual activity lasts at least 1.5 h; a high $lmin$ would mistakenly merge short sub-segments. The *atl* value in eq. (7) would be affected more than the *wat* in eq. (8) by these parameters, because the latter takes the usage intensity as weight when calculating the activity time length. Also, we set $\beta = 1/3$ to filter effective sub-segments; this constrains the *ui* of an effective activity to be at least one-third of the highest *ui*.

We trained the RF classification model over a range of class numbers and settled for a solution with six classes as it is readily interpretable in terms of mobility behaviors. The statistical description of each class in the RF model is provided in Table 2. The training set consists of a random sample of 240 riders and the testing set contains 120 riders; the remaining 159,804 travelers (160,164–240–120) are used for classification forecasting. In the RF model, the feature of the usage intensity of station $h3$ and $w3$ contributes to class prediction in the order of 2.8% and 8.8%, respectively, and the effective activity time length *atl* in eq. (7) of station $h3$ and $w3$ contributes to the model by 0.7% and 3.5%, respectively. This evidence suggests that omitting the third potential station of home and work may lead to missing on some non-trivial information on activities. The unweighted Kappa statistic measures agreement for categorical data, which is calculated from the confusion matrix on the testing data with the *caret* package in R (Kuhn, 2008). The value of 0.8057 is strong indication that the model is good for forecasting classification.

Representative cases of STUM-HWs in each class are displayed in Fig. 6. People belonging to C1 have ordinary mobility patterns that encompass a single home location and a single work location. They constitute the vast majority of the whole dataset (74.5%). However, this does not mean they all lead the same life. For example, persons A, B, C featured in Fig. 6 show different working schedules. People like C1-A and C1-B have rather fixed working times, but they differ by the times of return home for lunch or dinner. Person B is found to work from 9 am to 6 pm, while person A works extra time from t38 to t43, after dinner. Deeper differences can be detected in the activity behavior of card holder C1-C. This individual's STUM-HW shows a much higher usage intensity during t26 ~ t33 than in other time slots in *w1*. When tracing back to the original SCD records, we find that this card holder has two regular work times on different days of the week. One schedule is 8:30 am ~ 4:30 pm, while the other is 1:00 pm ~ 9:00 pm, which would be consistent with working on different shifts at the same location. These two work schedules overlap during t26 ~ t33, hence the higher usage intensity during this timeframe. Intra-personal variability in activity schedules of this sort is further discussed in Section 5.3.

In Fig. 6, person D and person E both belong to class C2 whose main property is that mobility patterns include multiple work locations. Person D tends to have fixed work times at both places of work. In contrast, person E goes to one workplace in the morning and travel to another workplace in the afternoon. As the survey in Shen et al. (2013) showed, people with alternative workplaces may be business people who have branch companies, or need to go out to meet business partners, or may routinely work out of an office, like salespersons. Moreover, person F in C3 has multiple home locations, while person G in C4 has both multiple home and work locations. Person H in C5 does not show any typical work locations, possibly as an unemployed or retired individual, but he/she spends short times at *ot* type stations with high regularity. Person I in C6 would be a typical outlier, whose home-work patterns defy interpretation.

As the results of the RF model in Table 2 and Fig. 7 indicate,

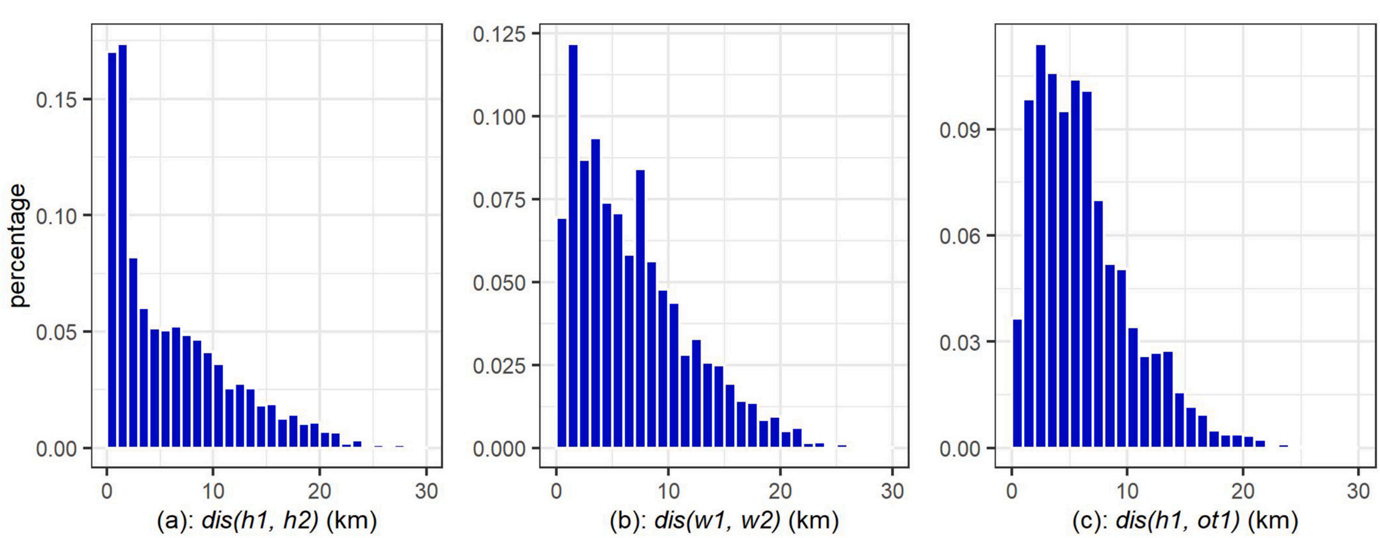


Fig. 8. Histograms of distance between (a) multiple home stations; (b) multiple work stations; (c) home station to *ot* type station (bin width = 1 km).

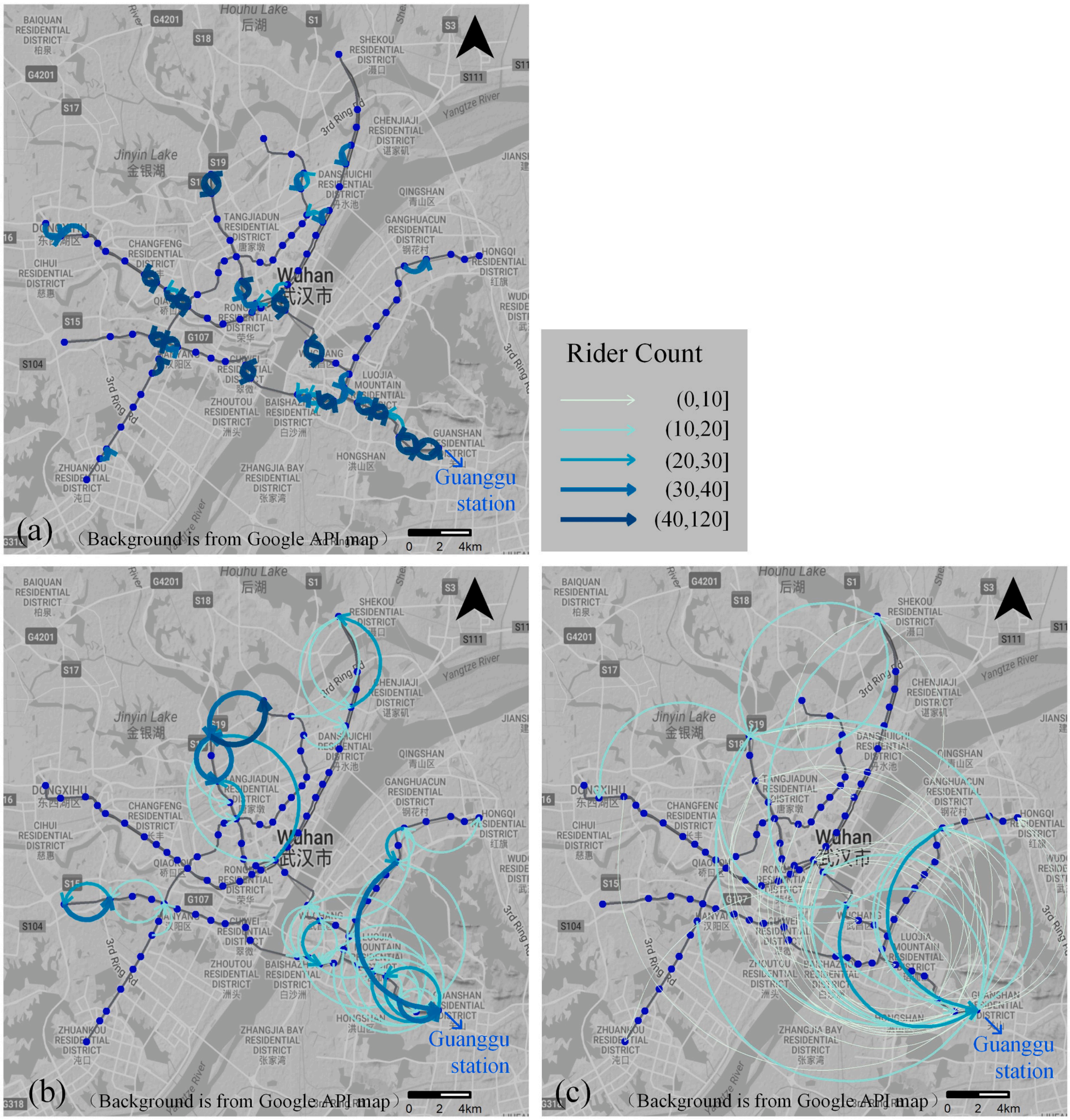


Fig. 9. Geographic interactions (brown and red arrows) between *h1* and *h2* for travelers with multiple home stations when (a) *dis(h1, h2)* ≤*2 km;* (b) *2 km* < *dis(h1, h2)* ≤*10 km;* (c) *dis(h1, h2)* > *10 km.* (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

commuters with a single home and a single work location (class C1) account for 74.5% of all card holders. In addition, 15.1% of riders (C2 + C4) exhibit multiple workplaces, and 8.6% of riders (C3 + C4) exhibit multiple home locations. A recent household survey showed that more than 15% of Wuhan households own multiple homes in the same city (Huang et al., 2020, p. 6 in Fig. 1). Our own result is lower than this rate because we just focus on metro transit riders, who include a higher percentage of people from lower and middle socioeconomic classes, whose real estate ownership is more limited. Another reason that may explain this difference is that here our instances with multiple homes stand for homes that are in fact in use as dwellings, instead of property ownership in the study by Huang et al. (2020).

Finally, 3.2% of riders (C5) show home to other activity patterns; they stay home most of the time but have regular and brief activities away from home. About a quarter of the population have a sophisticated home-work activity pattern that is beyond the ordinary 1 home-1 workplace pattern. Furthermore, even people who fit the latter case, have quite differentiated work timeframe patterns.

From the above results, it is striking that, with the help of the STUM construct and the proposed analytic framework, we are able to draw such a clear and detailed picture of how people allocate time at a fine temporal granularity across home and work activities. Meanwhile, questions arise as to how individuals maintain their specific kind of home-work activity patterns. For example, how do people who have multiple home locations arrange switches between these home locations? Do they keep hopping between multiple residences across the study period? How many people have flexible work schedules? We are going to discuss these spatial (home locations) and temporal (work times) aspects derived from the analysis of the STUM in the next section.

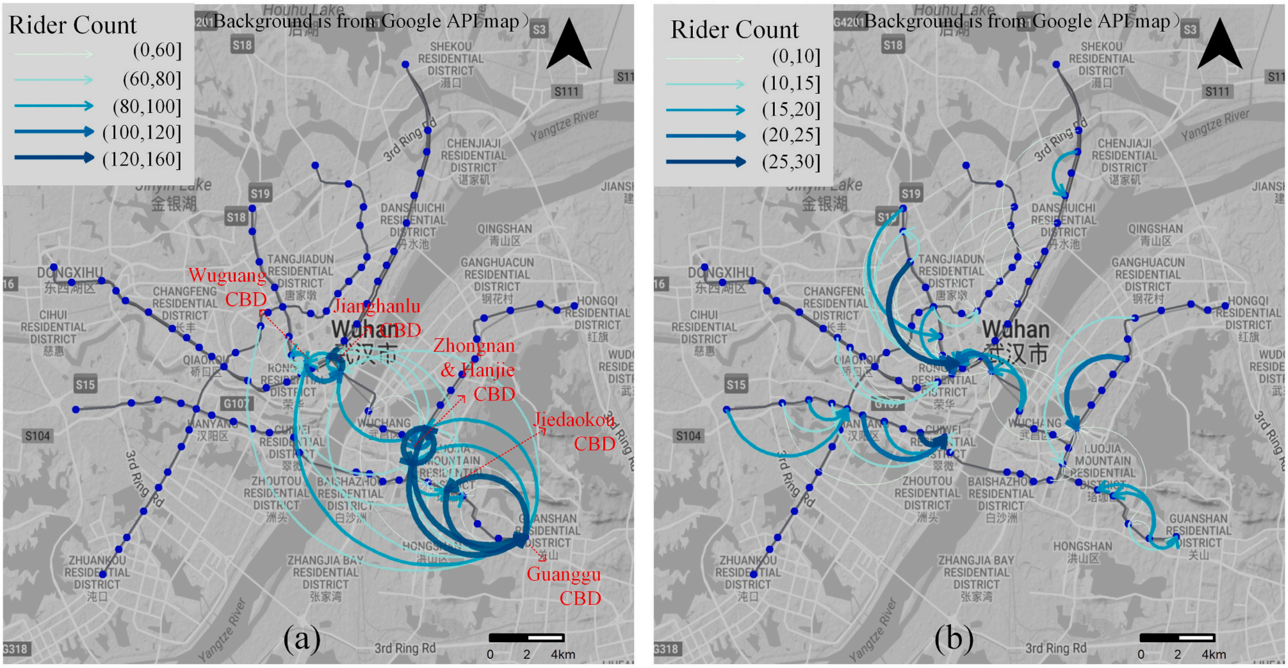


Fig. 10. The 50 strongest geographic interactions between (a) *w1* to *w2*; (b) *h1* to *ot1*.

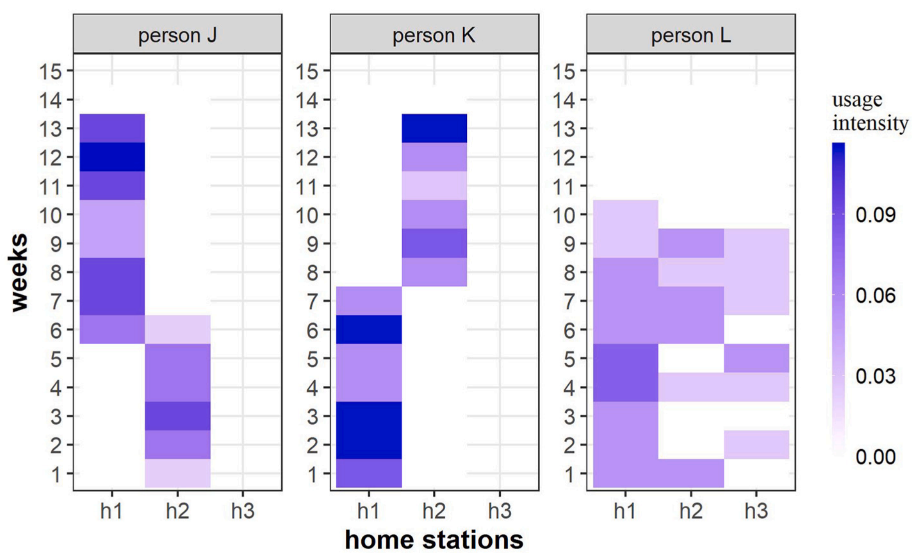


Fig. 11. Three typical weekly STUMs of travelers with multiple home locations.

5. Analysis and discussion

5.1. Geographical distribution of activity locations

For people who maintain multiple home stations (classes C3 and C4), we first examine how far these home stations are located from each other. Fig. 8(a) is a histogram of the distance between the two most frequent home stations, i.e., *dis(h1,h2)*, for people in C3 or C4. In this group, 34.4% of distances between home stations *h2* and *h1* are under 2 km, which is approximately the average distance between two neighborhood metro stations in Wuhan. This suggests that these people may in fact live some distance away between the two stations and rely on either of these stations in their commuting travel dependent of other circumstances of their travel behaviors, such as the conduct of some other activities in proximity of one or the station. As shown in Fig. 9(a), a dozen pairs of stations are concerned by these switching behaviors between substitute stations, and most of these pairs involve one station that is in the vicinity of a large business district (see also Fig. 4).

A large share of card holders (43.2%) have home stations 2 to 10 km apart (Fig. 8(a)). A few pairs of substitute stations can still be found, like north of the city center and in the southeastern section of the city (Fig. 9 (b)). Given the greater geographic separation between *h1* and *h2* home stations, it is quite probable that many card holders in this group in fact have multiple home bases. In the context of living experiences in a major Chinese city, it is not unusual for a worker to own, lease, or occupy more than one housing unit as a facet of a personal investment strategy, or as a co-housing arrangement with close family, or a transitional stage before permanent relocation (Huang and Yi, 2011). This is further discussed in Section 5.2. These trends carry over to card holders with home stations more than 10 km apart (22.4%). This is particularly prominent in the

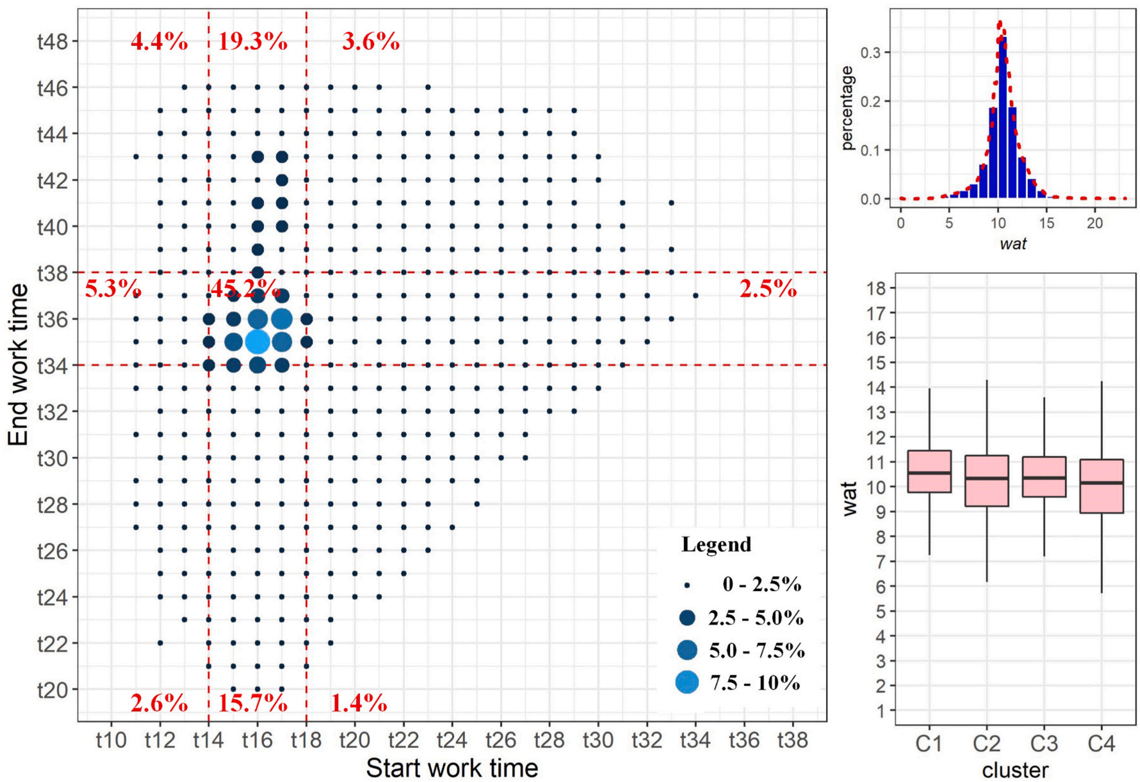


Fig. 12. Distribution of work time of commuters.

southeastern section of the city around the Guanggu station where many high-tech companies have attracted young people and where the housing market has been very vibrant (Fig. 9(c)).

For card holders with multiple work stations (class C2 or C4), we find a rather different distance distribution as for multiple home stations (Fig. 8(b)). The histogram of *dis(w1, w2)* (Fig. 8(b)) displays a distinctive crater shape and there are several secondary peaks between 2 and 10 km. This implies that, unlike the situation where many people rely on two neighboring stations for home, people do not usually rely on two nearby stations to reach their place of work. Fig. 10(a) visualizes the 50 strongest geographical interactions between *w1* and *w2*. These are true instances of riders working at more than one location inside the city limits. The *w1* to *w2* interactions mainly happen between the five commercial and central business districts (CBD) of Wuhan.

Furthermore, we explore the geographic interactions between *h1* and *ot1* for people in class C5. On average, these people stay at *ot1* locations 1 h around 7:30 am to 8:30 am, a short while just before the typical start of the workday. The geographic interactions from *h1* to *ot1* in Fig. 10(b) show a clear tendency for traveling from peripheral neighborhoods of the city. In line with this, we find that 77.7% of riders in class C5 travel to the *ot1* location from home within 9 km (Fig. 8(c)). This class is very likely formed of parents taking their time just before rush hour to take their children to school, after which they return home or hang out at other ancillary places nearby. This kind of pick-up/drop-off activity always shows a high degree of spatial repetition (Buliung et al., 2008), which can thus be detected with the STUM.

5.2. Variabilities in multiple home activities

To further investigate how people switch between multiple home stations, we focus on riders in classes C3 and C4. The time unit is adjusted to be 1 week (5 workdays) and we construct a weekly STUM for each traveler whose home-work pattern includes multiple homes locations. Accordingly, we set $n = 14$ and $m = 3$, indicating 14 weeks and 3 home stations in the weekly STUM. Whenever people have a home activity at week i, $u_{ij} = 1$ is set in eq. (4).

Fig. 11 depicts three typical weekly STUMs. They enable us to assemble evidence on whether people share multiple homes within the same timeframe or transition between homes over the timeframe. Fitting the later situation, person J stays at *h2* during the first 5 to 6 weeks, and then moves on to the *h1* station. Similarly, person K stays at *h1* and then moves to *h2* at the 8th week. In contrast, person L hops between three different home locations during the study period. By varying the temporal unit in the STUM, it is quite clear we witness a change in home location at the weekly scale. Furthermore, we calculate the cosine similarity between each traveler's weekly STUM and the matrix of these three typical people to generate a rough classification. As a result, we find that among the 13,741 people having multiple home stations, 40.6% share person J's pattern, 25.3% are similar to person K, while the remaining 33.9% share multiple home stations throughout the study period, like person L in Fig. 11. Thus, it can be estimated that about two thirds of people show multiple home locations on their activity pattern inferred from the SCD due to relocation, but still quite a few people keep hopping between their multiple home stations and multiple actual home locations for an extended period of time.

5.3. Work time flexibility

Finally, we study the plasticity of the work schedule of commuters by focusing specifically on typical working individuals who fit in classes C1, C2, C3 or C4. We calculate the weighted mean work time $wat(\mathbf{U}[\text{ALL}, s_j \in (w1, w2, w3)])$ with eq. (8). Work time closely follows a normal distribution (Fig. 12). Most people work 10–11 h per weekday. Note that this work time may include lunch time if people do not travel on the metro during the interval. The boxplot shows that the average *wat* does not vary much across clusters, but the *wat* statistic of people with a single workplace (C1 and C3) varies less widely than among people with multiple workplaces. Fig. 12 also displays the start work time $work_{start}$

and end work time $work_{end}$ distribution. 12.3% of regular commuters start working before 7 am (t14), while 7.5% of commuters start after 9 am (t18). On the other hand, 19.7% end before 5 pm (t34), while 27.3% of commuters end working after 7 pm (t38). Only 45.2% of people follow the common routine that starts between 7–9 am and ends between 5–7 pm. This leaves about 55% of metro riders having a different work time than the standard work timeframe. Hence, it can be seen from the daily activity records in SCD that people are behaviorally quite diverse and exhibit very flexible work schedules.

6. Conclusions

It is a commonly accepted view that people have one home and one workplace, and that they go to work at 7–9 am in the morning, while tracking back home at 5–7 pm in the afternoon. But does everybody really follow that standard routine? At closer inspection of where, when, and how long people stay at a place, we can conclude that individuals exhibit quite contrasted home-work activity patterns. Some people are parents who only travel ahead of rush hours to drop off children at school; some are busy traveling during work hours between multiple sites; some are hopping between different home bases; and some are juggling two work shifts. What a harried and diverse city life!

We introduced the innovative concept of space time usage matrix to measure the accumulation of activities hidden in transit smart card data. With the help of the STUM and an analytical framework involving spatio-temporal decompositions and aggregations, usage intensity along the space-time gradient can be leveraged to identify multiple home or work locations as well as the plasticity of people's time allocation across the home and work continuum. The case study in personal mobility in Wuhan shows that the STUM is effective at depicting the spatio-temporal regularities and variabilities in home, work, and other activities. Our results show that about 25% of the population have a sophisticated home-work activity pattern other than the ordinary 1 home-1 workplace pattern. Furthermore, even for this latter group, quite differentiated work timeframe patterns exist. Results show that only 45.2% of people follow the common routine that starts around 7–9 am and ends around 5–7 pm in Wuhan. Although we use metro SCD in this study, the STUM construct can work with a variety of datasets containing bus records, mobile phone data, and any other datasets that record personal travel sequences. The methodology and results will contribute to more refined analysis in transport policy, urban planning and social studies.

It is worth pointing out that some bias may exist, since our study focuses on frequent commuters using the metro transit system, and additional studies would be needed to generalize to the broader resident population. The analysis of variabilities for non-recurrent and occasional users should be important to gain valuable insights into the whole picture of mobility and activity of city residents. Survey data would be instrumental to support the validation of results derived from SCD. Moreover, in combination with travel behavior survey data, the method could be enhanced to better handle the case of people who stay home during daytime and work at nighttime.

This work can be enhanced in several ways. Firstly, while the concept of the STUM and the associate analytical framework are flexible enough to be applied at various temporal granularities, at present this is possible only when the analyses are done separately. We believe that the integration of cross-scalar analyses would allow to better conceptualize repetition of diverse longitudinal patterns, as well as variability in repetitive behaviors. Second, weekend activities would gain to be incorporated in the STUM framework to fill the gap in inter-personal and intra-personal variability of space-time activities left unanswered by the present study. Finally, since big geo-spatial data always lack the social-demographic background of individuals, travel survey data would supplement the socio-spatial grounding of our results, but help offer insights in the relationships between individual socio-demographic profiles and activity behaviors.

Acknowledgements

The authors wish to thank the journal editor and anonymous reviewers who critically commented on the research reported herein and were instrumental in enhancing the quality of the manuscript. This work was supported by the National Natural Science Foundation of China (No. 42001399), the Fundamental Research Funds for the Central Universities (CCNU20ZT020, CCNU19TD001, CCNU19TD002), and the Open Fund of Key Laboratory of Geospatial Big Data Mining and Application, Hunan Province (No. 2019-01).

References

Alsger, A., Tavassoli, A., Mesbah, M., Ferreira, L., Hickman, M., 2018. Public transport trip purpose inference using smart card fare data. Transport. Res. Part C: Emerging Technol. 87, 123–137. https://doi.org/10.1016/j.trc.2017.12.016.

Briand, A.S., Côme, E., Trépanier, M., Oukhellou, L., 2017. Analyzing year-to-year changes in public transport passenger behaviour using smart card data. Transport. Res. Part C: Emerging Technol. 79, 274–289. https://doi.org/10.1016/j.trc.2017.03.021.

Buliung, R., Roorda, M.J., Remmel, T.K., 2008. Exploring spatial variety in patterns of activity-travel behaviour: initial results from the Toronto Travel-Activity Panel Survey (TTAPS). Transportation 35, 697–722. https://doi.org/10.1007/s11116-008-9178-4.

Cats, O., Wang, Q., Zhao, Y., 2015. Identification and classification of public transport activity centres in Stockholm using passenger flows data. J. Transp. Geogr. 48, 10–22. https://doi.org/10.1016/j.jtrangeo.2015.08.005.

Chakirov, A., Erath, A., 2012. Activity identification and primary location modelling based on smart card payment data for public transport activity identification and primary location modelling based on smart card payment data for public transport. In: The 13th International Conference on Travel Behaviour Research(IATBR). Toronto, Canada, July 15–20, pp. 1–22.

Cheng, Z., Prakash, K., Smyth, R., Wang, H., 2020. Housing wealth and happiness in Urban China. Cities 96, 102470. https://doi.org/10.1016/j.cities.2019.102470.

Deschaintres, É., Morency, C., Trépanier, M., 2019. Analyzing transit user behavior with 51 weeks of smart card data. Transp. Res. Rec. 2673, 33–45. https://doi.org/10.1177/0361198119834917.

El Mahrsi, M.K., Côme, E., Oukhellou, L., Verleysen, M., 2017. Clustering smart card data for urban mobility analysis. IEEE Trans. Intelligent Trans. Sys. 18, 712–728.

Faroqi, H., Mesbah, M., Kim, J., Tavassoli, A., 2018. A model for measuring activity similarity between public transit passengers using smart card data. Travel Behav. Soc. 13, 11–25. https://doi.org/10.1016/j.tbs.2018.05.004.

Gao, Q.L., Li, Q.Q., Yue, Y., Zhuang, Y., Chen, Z.P., Kong, H., 2018. Exploring changes in the spatial distribution of the low-to-moderate income group using transit smart card data. Comput. Environ. Urban. Syst. 72, 68–77. https://doi.org/10.1016/j.compenvurbsys.2018.02.006.

Goulet-Langlois, G., Koutsopoulos, H.N., Zhao, J., 2016. Inferring patterns in the multi-week activity sequences of public transport users. Transport. Res. Part C: Emerging Technol. 64, 1–16. https://doi.org/10.1016/j.trc.2015.12.012.

Hägerstrand, T., 1970. What about people in Regional Science? Papers Reg. Sci. Assoc. 24, 6–21. https://doi.org/10.1007/BF01936872.

Hanson, S., Huff, O.J., 1988. Systematic variability in repetitious travel. Transportation 15, 111–135. https://doi.org/10.1007/BF00167983.

Huang, J., Levinson, D., Wang, J., Zhou, J., Wang, Z., 2018. Tracking job and housing dynamics with smartcard data. Proc. Natl. Acad. Sci. 115, 12710–12715. https://doi.org/10.1073/pnas.1815928115.

Huang, Y., Yi, C., 2011. Second home ownership in transitional Urban China. Hous. Stud. 26, 423–447. https://doi.org/10.1080/02673037.2011.542100.

Huang, Y., Yi, D., Clark, W.A.V., 2020. Multiple home ownership in Chinese cities: an institutional and cultural perspective. Cities 97, 102518. https://doi.org/10.1016/j.cities.2019.102518.

Järv, O., Ahas, R., Witlox, F., 2014. Understanding monthly variability in human activity spaces: a twelve-month study using mobile phone call detail records. Trans. Res. Part C, Emerg. Technol. 38, 122–135. https://doi.org/10.1016/j.trc.2013.11.003.

Kitamura, R., Yamamoto, T., 2006. How routine is a routine? An analysis of the day-to-day variability in prism vertex location. Transp. Res. A Policy Pract. 40, 259–279. https://doi.org/10.1016/j.tra.2005.07.002.

Kuhn, M., 2008. Building predictive models in R using the caret Package. J. Stat. Softw. 28, 1–26. http://www.jstatsoft.org/article/view/v028i05/v28i05.pdf.

Kwan, M.P., 2007. Mobile communications, social networks, and urban travel: hypertext as a new metaphor for conceptualizing spatial interaction. Prof. Geogr. 59, 434–446. https://doi.org/10.1111/j.1467-9272.2007.00633.x.

Lavielle, M., 2005. Using penalized contrasts for the change-point problem. Signal Process. 85, 1501–1510. https://doi.org/10.1016/j.sigpro.2005.01.012.

Lee, S.G., Hickman, M., 2014. Trip purpose inference using automated fare collection data. Public Transport 6, 1–20. https://doi.org/10.1007/s12469-013-0077-5.

Liu, X., Zhou, Y., Rau, A., 2018. Smart card data-centric replication of the multi-modal public transport system in Singapore. J. Transp. Geogr. 76, 254–264. https://doi.org/10.1016/j.jtrangeo.2018.02.004.

Liu, Y., Cheng, T., 2018. Understanding public transit patterns with open geodemographics to facilitate public transport planning. Transportmetrica A: Trans. Sci. 16, 76–103. https://doi.org/10.1080/23249935.2018.1493549.

Long, Y., Thill, J., 2015. Combining smart card data and household travel survey to analyze jobs – housing relationships in Beijing. Comput. Environ. Urban. Syst. 53, 19–35. https://doi.org/10.1016/j.compenvurbsys.2015.02.005.

Long, Y., Liu, X., Zhou, J., Chai, Y., 2016. Early birds, night owls, and tireless/recurring itinerants: an exploratory analysis of extreme transit behaviors in Beijing, China. Habitat Int. 57, 223–232. https://doi.org/10.1016/j.habitatint.2016.08.004.

Ma, X., Liu, C., Wen, H., Wang, Y., Wu, Y.J., 2017. Understanding commuting patterns using transit smart card data. J. Transp. Geogr. 58, 135–145. https://doi.org/10.1016/j.jtrangeo.2016.12.001.

Roth, C., Kang, S.M., Batty, M., Barthe, M., 2011. Structure of urban movements: polycentric activity and entangled hierarchical flows. PLoS One 6, 2–9. https://doi.org/10.1371/journal.pone.0015923.

Shen, Y., Kwan, M., Chai, Y., 2013. Investigating commuting flexibility with GPS data and 3D geovisualization: a case study of Beijing, China. J. Transp. Geogr. 32, 1–11. https://doi.org/10.1016/j.jtrangeo.2013.07.007.

Tu, W., Cao, R., Yue, Y., Zhou, B., Li, Qiuping, Li, Qingquan, 2018. Spatial variations in urban public ridership derived from GPS trajectories and smart card data. J. Transp. Geogr. 69, 45–57. https://doi.org/10.1016/j.jtrangeo.2018.04.013.

Wang, Y., Correia, G.H.D.A., de Romph, E., Timmermans, H.J.P., 2017. Using metro smart card data to model location choice of after-work activities: an application to Shanghai. J. Transp. Geogr. 63, 40–47. https://doi.org/10.1016/j.jtrangeo.2017.06.010.

Xu, Y., Shaw, S., Zhao, Z., Yin, L., Lu, F., Chen, J., Xu, Y., Shaw, S., Zhao, Z., Yin, L., Lu, F., Chen, J., Fang, Z., 2016. Another tale of two cities: understanding human activity space using actively tracked cellphone location data. Ann. Assoc. Am. Geogr. 106, 489–502. https://doi.org/10.1080/00045608.2015.1120147.

Zhang, W., Thill, J.-C., 2017. Detecting and visualizing cohesive activity-travel patterns: a network analysis approach. Comput. Environ. Urban. Syst. 66, 117–129. https://doi.org/10.1016/j.compenvurbsys.2017.08.004.

Zhao, Z., Koutsopoulos, H.N., Zhao, J., 2018. Individual mobility prediction using transit smart card data. Trans. Res. Part C, Emerg. Technol. 89, 19–34. https://doi.org/10.1016/j.trc.2018.01.022.

Zhong, C., Batty, M., Manley, E., Wang, J., Wang, Z., Chen, F., Schmitt, G., 2016. Variability in regularity: mining temporal mobility patterns in London, Singapore and Beijing using smart-card data. PLoS One 11, 1–17. https://doi.org/10.1371/journal.pone.0149222.

Zhou, J., Murphy, E., Long, Y., 2014. Commuting efficiency in the Beijing metropolitan area: an exploration combining smartcard and travel survey data. J. Transp. Geogr. 41, 175–183. https://doi.org/10.1016/j.jtrangeo.2014.09.006.

Zhou, Y., Fang, Z., Zhan, Q., Huang, Y., Fu, X., 2017. Inferring social functions available in the metro station area from passengers' staying activities in smart card data. ISPRS Int. J. Geo Inf. 6, 394–411. https://doi.org/10.3390/ijgi6120394.