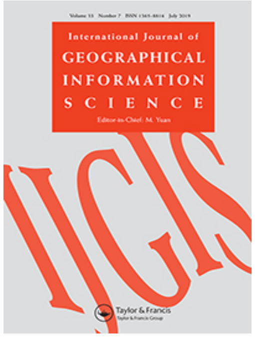

Taylor & Francis
Taylor & Francis Group

International Journal of Geographical Information Science

ISSN: 1365-8816 (Print) 1362-3087 (Online) Journal homepage: https://www.tandfonline.com/loi/tgis20

The effect of temporal sampling intervals on typical human mobility indicators obtained from mobile phone location data

Zhiyuan Zhao, Shih-Lung Shaw, Ling Yin, Zhixiang Fang, Xiping Yang, Fan Zhang & Sheng Wu

To cite this article: Zhiyuan Zhao, Shih-Lung Shaw, Ling Yin, Zhixiang Fang, Xiping Yang, Fan Zhang & Sheng Wu (2019) The effect of temporal sampling intervals on typical human mobility indicators obtained from mobile phone location data, International Journal of Geographical Information Science, 33:7, 1471-1495, DOI: 10.1080/13658816.2019.1584805

To link to this article: https://doi.org/10.1080/13658816.2019.1584805

INTERNATIONAL JOURNAL OF GEOGRAPHICAL INFORMATION SCIENCE
2019, VOL. 33, NO. 7, 1471–1495
https://doi.org/10.1080/13658816.2019.1584805

RESEARCH ARTICLE

The effect of temporal sampling intervals on typical human mobility indicators obtained from mobile phone location data

Zhiyuan Zhao[a,b], Shih-Lung Shaw[c], Ling Yin[d], Zhixiang Fang[a,e], Xiping Yang[f], Fan Zhang[d] and Sheng Wu[b,g]

[a]State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, PR China; [b]Spatial Information Research Center of Fujian Province, Fuzhou University, Fuzhou, PR China; [c]Department of Geography, University of Tennessee, Knoxville, TN, USA; [d]Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, PR, China; [e]Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan, PR China; [f]School of Geography and Tourism, Shaanxi Normal University, Xi'an, PR China; [g]Fujian Collaborative Innovation Center for Big Data Applications in Governments, Fuzhou, PR China

ABSTRACT
Mobile phone location data have been extensively used to understand human mobility patterns through the employment of mobility indicators. The temporal sampling interval (TSI), which is measured by the temporal interval between consecutive records, determines how well such data can describe human activities and influence the values of human mobility indicators. However, systematic investigations of how the TSI affects human mobility indicators remain scarce, and characterizing those relationships is a fundamental research question for many related studies. This study uses a mobile phone location dataset containing 19,370 intensively sampled individual trajectories (TSI < 5 minutes) to systematically assess the impacts of the TSI on four typical mobility indicators that describe human mobility patterns from different aspects, which are movement entropy, radius of gyration, eccentricity, and daily travel frequency. We find that different TSIs have complex impacts on the values of different mobility indicators. Specifically, (1) coarser TSIs tend to underestimate the values of the four selected indicators with different degrees; (2) the degrees of underestimation vary significantly among users for eccentricity and daily travel frequency but exhibit high inter-user consistency for radius of gyration and movement entropy. The above findings can help better understand the variations among human mobility studies.

ARTICLE HISTORY
Received 4 May 2017
Accepted 7 February 2019

KEYWORDS
Human mobility; temporal sampling intervals; mobile phone location data; modifiable temporal unit problem(MTUP)

1. Introduction

During the last decade, massive mobile phone location data have contributed to human mobility studies in urban planning, epidemic control, and traffic analysis as well as other areas (Ratti *et al*. 2006, Ahas *et al*. 2007, Wang *et al*. 2012, Blondel *et al*. 2015, Mao *et al*. 2016). Mobility indicators such as movement entropy (Song *et al*. 2010) and travel frequency (Çolak *et al*. 2015) play key roles in these studies. However, because they are mainly collected for billing and service purposes rather than for research purposes,

CONTACT Shih-Lung Shaw sshaw@utk.edu; Ling Yin yinling@siat.ac.cn
Supplemental data for this article can be accessed here.

mobile phone location datasets have some quality concerns (e.g. spatiotemporal resolution, sampling bias) when they are used in human mobility research. Currently, it remains unclear how these quality concerns in mobile phone location datasets affect human mobility studies (Goodchild 2013, Liu *et al*. 2016). This study focuses on the temporal resolution issue. We note that 'human mobility' is a broad concept. Human mobility patterns contain short-term patterns such as daily travel frequency (e.g. Çolak *et al*. 2015) as well as long-term patterns such as monthly variability in human activity spaces (e.g. Järv *et al*. 2014). In this study, we focus on the short-term patterns and select four typical indicators (movement entropy, radius of gyration, eccentricity and daily travel frequency) to measure human mobility patterns from different perspectives. We then systematically investigate how the indicator values obtained from mobile phone location data change with different temporal resolutions.

To investigate the effect of temporal resolution in mobile phone data on human mobility indicators, this study introduced the concept of temporal sampling interval (TSI), which is defined as the temporal interval between pairs of consecutive records, to characterize the temporal resolution of mobile phone location data. The TSI refers to the inter-event time in call detail record (CDR) data and the temporal interval between records corresponding to other types of signals such as periodically updated signals.

Impacts caused by different TSIs on human mobility indicators could refer to the temporal aggregation effect in the modifiable temporal unit problem (MTUP) (Cheng and Adepeju 2014), which is similar to the scaling effect in the well-known modifiable areal unit problem (MAUP) (Openshaw 1984). The temporal aggregation effect indicates that different temporal resolutions may result in varying outcomes (Cheng and Adepeju 2014). Previous studies have shown that the use of CDRs may introduce biases into human mobility research because of their sparse temporal resolutions (e.g. Ranjan *et al*. 2012, Zhao *et al*. 2016). However, a series of designed TSIs is needed to systematically investigate and better understand the temporal aggregation effect. This design, on the one hand, is in agreement with previous studies that investigate the temporal aggregation effect (Cheng and Adepeju 2014, Liu *et al*. 2017). On the other hand, it can provide practical insights in collecting datasets regarding mobile phone location data with customized TSIs according to particular research purposes, which is becoming possible (e.g. Ratti *et al*. 2006).To systematically investigate the temporal aggregation effect in human mobility studies using mobile phone location data, this study attempts to (1) reveal and quantify the impact of TSIs on typical mobility indicators of different types; (2) based on the sensitivity of mobility indicators on TSI, offer insights on how to choose proper temporal resolutions or choose suitable mobility indicators to answer particular mobility-related issues.

The rest of this paper is organized as follows. Section 2 reviews several related studies. Sections 3 and 4 describe the dataset and the selected indicators used in this study, respectively. Section 5 investigates the changes in the value of the four selected indicators caused by the use of different TSIs from both individual and aggregate perspectives. Finally, we discuss the results and draw conclusions in Sections 6 and 7, respectively.

2. Related works

2.1. *Measuring human mobility patterns using mobile phone location data*

Compared to traditional datasets such as travel diary data or GPS-assistant survey data, large amounts of mobile phone location data exhibit advantages of large sample sizes and low-cost data collection processes (Yue *et al*. 2014, Birenboim and Shoval 2016) but fall short of the spatiotemporal resolution and richness of semantic information of users and their activities (Widhalm *et al*. 2015, Diao *et al*. 2016). Given the fluctuations in people's activity locations, which range from daily to monthly (Järv *et al*. 2014, Widhalm *et al*. 2015), coarse temporal resolutions may result in some activity locations being ignored. In general, mobile phone location datasets offer a useful source for understanding people's daily lives by extracting socioeconomic characteristics with the help of various mobility indicators (Liu *et al*. 2015, Steenbruggen *et al*. 2015). However, most of these studies were conducted using one or several mobile phone location datasets with particular TSIs. In the era of big data, few cross-validations have been carried out in these studies (Lazer *et al*. 2014), and the extent to which we can trust the conclusions of human mobility studies remains an open question (Miller and Goodchild 2015). Whether related datasets are suitable for answering particular questions or what kinds of impacts the quality of the datasets may introduce need further investigation before we use them to draw any useful knowledge (Shaw *et al*. 2016).

Various indicators have been developed to measure different dimensions of human mobility patterns based on mobile phone location data. For instance, movement entropy has been used to reflect the randomness of people's daily activity locations and evaluates the predictability of the next activity location (Eagle and Pentland 2006, Song *et al*. 2010). In addition, approximate standard ellipses and their eccentricities can describe the shape of people's visited locations (Yuan *et al*. 2012, Ahas *et al*. 2015). Regarding the extent of daily activity locations, both radius of gyration and the diameter of all the activity locations can provide appropriate measures for corresponding purposes (González *et al*. 2008, Xu *et al*. 2016a). For travel strength, daily travel frequency and commuting distance are commonly used indicators for investigating the difference in people's mobility patterns and the degrees of home-work balance between cities (Isaacman *et al*. 2010, Xu *et al*. 2016a).

2.2. *Temporal sampling intervals in mobile phone location data*

Currently, there are two main types of mobile phone location data collected using cell identification positioning techniques (Ludden *et al*. 2012) in existing studies: (1) mobile phone user-based data, which contain records from a user generating a communication event (e.g. a phone call or Internet access) through a mobile network and (2) cell-tower-based data, which mainly contain records such as handovers between cell towers and periodic update records generated by mobile network operators (Yue *et al*. 2014).

The temporal resolutions of both types of mobile phone location data are often sparse and differ from each other (e.g. Chen *et al*. 2016). CDRs form a mobile phone user-based type of dataset that is collected passively for billing purpose when users engage in communication activities, such as making phone calls or sending text messages (Wang *et al*. 2010). The TSIs of CDRs depend on the temporal patterns of communication events, which show burst patterns over time (Barabási 2005), implying that the temporal resolutions of CDRs are uneven and sparse. Conversely, datasets that contain records generated by mobile network operators by periodically detecting the location of each mobile phone represent a cell-tower-based type of dataset (Calabrese *et al*. 2014). Even for network-based datasets, no standard exists that states how frequently the location of each mobile phone should be updated. Hence, each mobile phone location dataset may have its own distinct TSI. For instance, the average TSI of the dataset used in Schneider *et al*. (2013) was approximately 30 minutes, whereas the TSI of the dataset used in Yue *et al*. (2017) was 1 hour. In Ahas *et al*. (2015), the average daily sampling frequency of the CDR datasets for Paris and Harbin were 8.13 times and 5.5 times, respectively.

The TSI can reflect how finely the dataset describes the daily activities of corresponding users in the temporal dimension. As Figure 1 illustrates, brief activities may be neglected when the TSIs of mobile phone location data are coarse (e.g. the activity at location C in Figure 1). In such cases, the values of mobility indicators may change (e.g. the daily activity locations and travel frequency change between trajectory II and trajectory III in Figure 1). Mobile phone location datasets have specific TSIs (Chen *et al*. 2016); thus, identification of the potential impacts of TSIs on mobility indicator values requires further investigation.

Researchers have noted that different patterns can be observed for different temporal resolutions (Çöltekin *et al*. 2011). Similar to the MAUP, which involves the scaling and

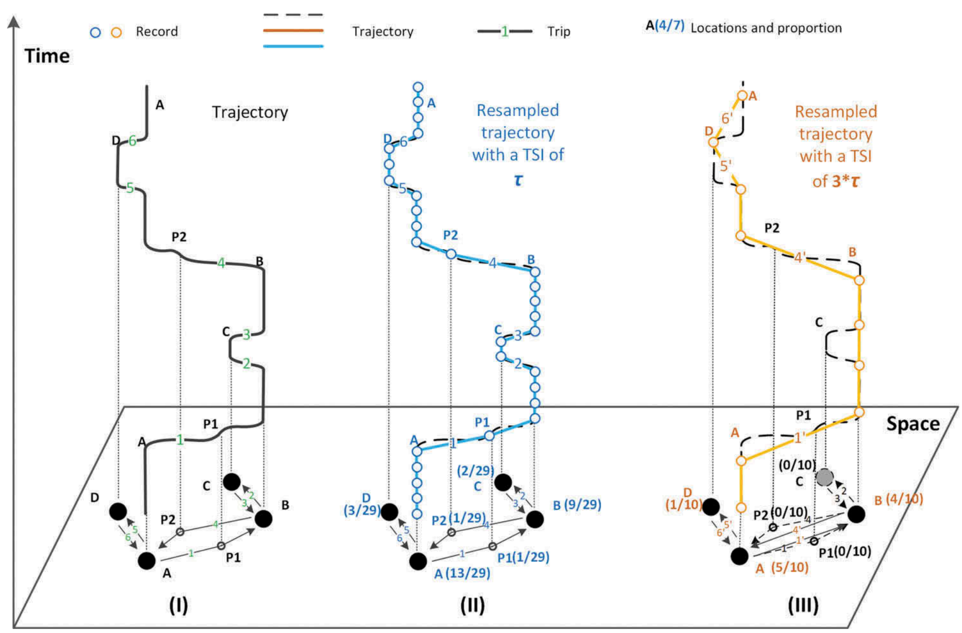

Figure 1. A trajectory example of how different temporal sampling intervals (TSIs) result in different human mobility indicator values (e.g. activity locations and travel frequency). Trajectory (I) is the origin one, while trajectory (II) and trajectory (III) are two resampled trajectories with different TSIs. The TSI of trajectory (III) is three times the TSI of trajectory (II).

zoning effects of data in spatial dimensions, Cheng and Adepeju (2014) proposed the MTUP, which consists of three temporal effects: aggregation, segmentation and boundary. Among them, the temporal aggregation effect, which corresponds to the scaling effect in MAUP, indicates that the results of a space-time clustering method are related to the temporal resolution of the dataset. In a study of movement, Dodge (2015) noted that a better interpretation should take the temporal granularity of trajectories into account. This issue is also prevalent in human mobility studies using mobile phone location data because the TSIs of those datasets are coarse and varied, but it has not been systematically investigated.

2.3. *The impacts of the temporal sampling issue on human mobility studies*

In human mobility studies based on mobile phone location datasets, there are several temporal sampling issues that should be considered, for example, call frequency, and sampling rates. For instance, Ranjan *et al.* (2012) examined the potential biases when using only voice call-based data compared to an unfiltered dataset that also contained text message-based data and the records corresponding to Internet access activities. They found that voice call-based data provided a better estimate of the radius of gyration of the users than movement entropy. Zhao *et al.* (2016) investigated the potential biases of CDRs extracted from the dataset that also contains other types of records such as handovers and periodic update records. They found that the low sampling rates of CDRs could cause significant underestimates of the values of indicators such as movement entropy and daily travel distance; however, the data provided a fairly good basis for estimating the radius of gyration. By collecting the CDRs and corresponding GPS tracking data for 84 users over eight months, Hoteit *et al.* (2016) found that CDRs have few impacts on the extraction of important locations such as homes and workplaces, but they affect the values of the radius of gyration.

We note that most of these studies have focused on biases in CDRs, which represent a specific kind of mobile phone location data. Recently, the impacts caused by TSIs for other mobile phone datasets have received increasing attentions. Lu *et al.* (2017) used a network-based dataset to demonstrate how the representativeness of mobile phone on the temporal dimension characterized by different TSIs affected the reported human mobility patterns. Zhao *et al.* (2018) argued that coarse temporal resolutions decrease the effectiveness of identifying stops based on a filtered mobile phone location dataset containing 329 trajectories with a TSI of five minutes and they proposed a new method to improve the result. Cuttone *et al.* (2018) explored two types of predictability in human mobility and discussed how the temporal resolution affect the accuracies of the predictability using a trajectory dataset collected by smartphones with a TSI of 15 minutes. These studies caution researchers to be aware that different TSIs can lead to varying human mobility results. However, a systematic investigation supported by a series of designed TSIs is needed. It is helpful to better understand the temporal aggregation effect from the theoretical perspective and gain more practical insight into collecting customized datasets targeted at particular research purposes or choosing proper research questions based on given datasets.

3. Dataset

3.1. *Data source*

This study employed a mobile phone location dataset that was collected by a mobile phone company in Shenzhen, China. The dataset contained records collected on a workday in 2013 from more than 1.4 million subscribers. When a user connected to the Internet, made or received a phone call or text message through the mobile network, the activity time and the location of the servicing cell tower were recorded (Table 1). Even during the night, given a smartphone, some apps (e.g. social media apps such as WeChat) continuously communicated with their servers in the background through the mobile Internet and generated corresponding records as long as it is on. This dataset differed from the conventional CDR datasets that contained only phone call or text message activities, which often lacked records during the night since people mainly communicate with each other during the daytime. The average TSI of this dataset was approximately 30 minutes.

Table 1. Sample records of the mobile phone location data.

In addition, the dataset contained data from approximately 3,400 cell towers; the average distance among the 5 nearest neighbors of each cell tower was approximately

Table 1. Sample records of the mobile phone location data.

UserID	TimeStamp	Longitude	Latitude
460XXXXXX9251	2013–10-XXTXX: XX: XX.000Z	114.XXXX	22.XXXX
460XXXXXX2565	2013–10-XXT XX: XX: XX.000Z	114.XXXX	22.XXXX
460XXXXXX1646	2013–10-XXT XX: XX: XX.000Z	114.XXXX	22.XXXX
460XXXXXX3757	2013–10-XXT XX: XX: XX.000Z	114.XXXX	22.XXXX

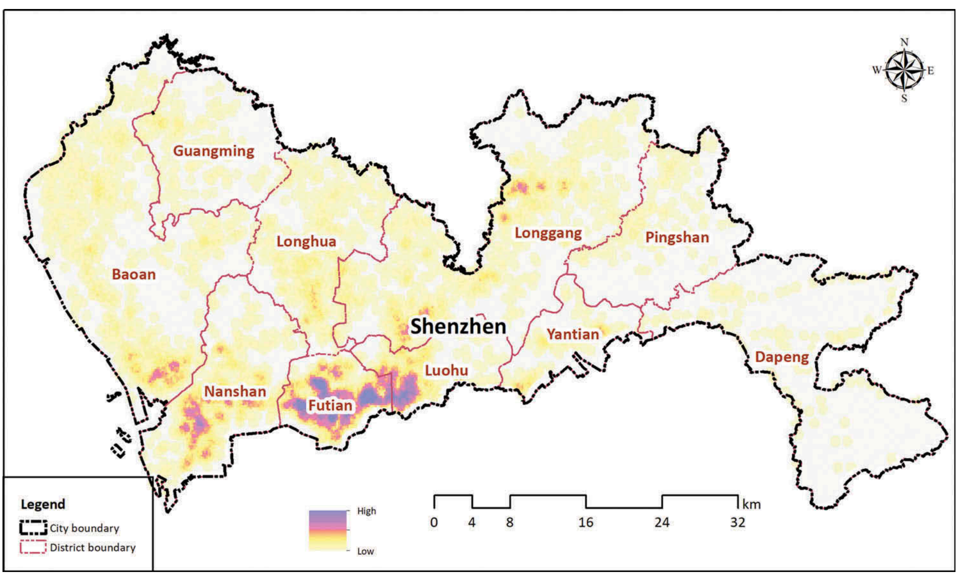

Figure 2. Spatial density of the cell towers in the study area.

550 meters (see Figure 2). To maintain the privacy of the users, all of the phone numbers in the dataset were anonymized.

3.2. *Resampling datasets with TSIs*

To investigate the impacts of TSIs, it is necessary to compare the values of each mobility indicator calculated from datasets with different TSIs. Therefore, we need a benchmark dataset and a series of TSIs. In this study, we first filtered a subset of users from the raw dataset with intensively sampled trajectories as the benchmark dataset and denoted the dataset *D*0. Every trajectory in *D*0 is the same as the corresponding user's trajectory in the raw dataset. We then applied a downscaling method to resample the dataset with different TSIs. Next, four typical, commonly used mobility indicators were selected. By exploring the changes in their values across datasets with different TSIs, this study revealed how the TSIs of mobile phone location data affect related mobility studies. The flowchart of the research design is shown in Figure 3.

3.2.1. *The characteristics of the benchmark dataset (D0)*

To generate *D*0, we applied three rules to guarantee a high sampling frequency. (1) The total number of records in one day must exceed 288. This value ensures an average TSI of less than 5 minutes, which is shorter than most major daily activities. (2) The elapsed time between the first and the last records should exceed 16 hours to ensure relatively complete coverage of daily activities. (3) The maximum time interval between consecutive records should be smaller than one hour to guarantee that the records are temporally well distributed. We obtained 19,370 intensively sampled trajectories in total based on the above three rules. Some noise might exist in the raw trajectories in *D*0. For example, the 'ping-pong' phenomenon, which was represented in the data as quick moves between neighboring locations, is a typical type of noise (Iovan *et al*. 2013). We followed the method used in the work of Horn *et al*. (2014) to mitigate the impacts of that noise by removing abnormal records using a velocity threshold of 80 km/h (which is the speed limit of the urban traffic systems in many cities in China). A record will be determined to be abnormal if both the speeds of the trip between the record and its previous one and the trip between the record and its following one are greater than 80 km/h. As a result, the average TSIs of more than 75% of the trajectories in D0

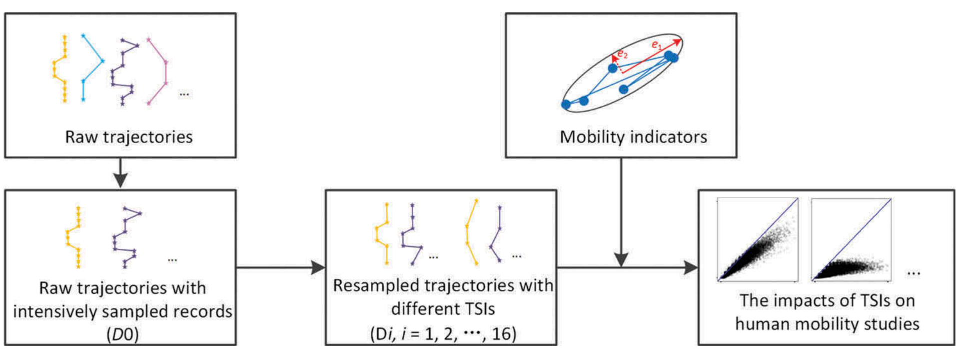

Figure 3. The research design in this study.

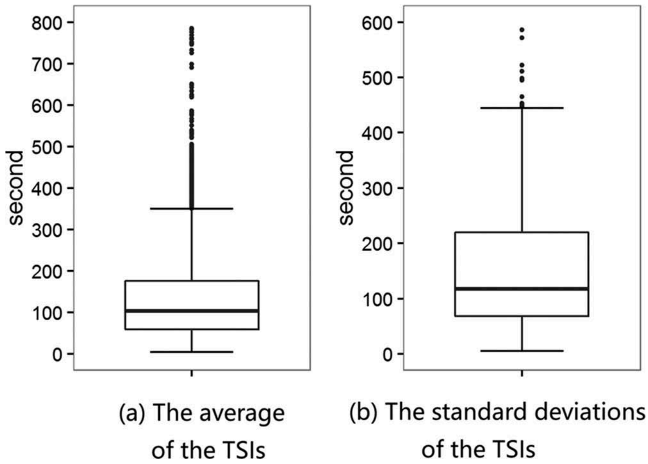

Figure 4. Boxplots of the statistical information for the TSIs for the trajectories in D0.

were smaller than 200 seconds (Figure 4), and the standard deviation of the TSIs of almost 75% of the trajectories in D0 were also smaller than 200 seconds, which indicates that the trajectories in D0 were both intensively sampled and relatively evenly distributed in time. The specific distributions of the average TSIs and the standard deviation of the TSIs could be found in Figure S12 in the supplementary file.

3.2.2. *The selection of TSIs*

To systematically examine the effects of the TSIs, we need to introduce a series of datasets with different TSIs. As far as we knew, the finest average TSI of the large amounts of mobile phone location data used in previous studies was approximately 30 minutes (Schneider *et al*. 2013, Widhalm *et al*. 2015). Considering that fact, we adopted 15 minutes, or half the previous finest average TSI, as the first TSI of the series and used 15 minutes as the increment interval for the resampled datasets. Considering that the D0 used in this study covered only one day, if a TSI was too large, the number of records for each user in the resampled dataset was too small to appropriately reflect the characteristics of the activities of those users. Hence, we fixed the upper end of the TSI range at four hours, which was far longer than the TSI values used in many related studies (Ratti *et al*. 2006, Xu *et al*. 2016a). Thus, we obtained 16 resampling TSIs that started at 15 minutes and increased in 15-minute steps to 4 hours (240 minutes). Here, D*i* (where *i* = 1, 2, 3… 16) represents the datasets resampled from D0 using each TSI in the series. The TSI of each D*i* in this study was computed using the following equation. For instance, the TSI of the resampled dataset D2 was 30 = 15 * 2 minutes.

$$TSI_{Di} = 15 \times i \ (minute), \ i = 1, 2, 3, \ldots, 16 \tag{1}$$

3.2.3. *Resampling trajectory datasets (Di) with TSIs*

For the resampling operation, we needed a start time t_0 and a specific TSI τ_i (see Figure 5). Given a t_0, if one or more records fell within a resampled time window defined by τ, we chose the record closest to the center of the time window (Figure 5). Otherwise,

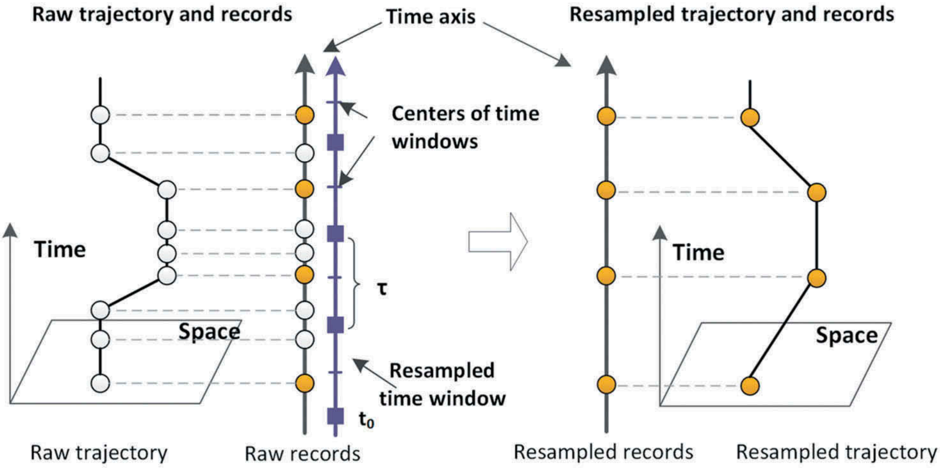

Figure 5. The progress of the resampling method.

there was no record in the time window and the resampled trajectory. For each specific τ, we used 50 different randomly selected t_0 values to generate corresponding resampled trajectories and calculated the average value of each mobility indicator to reduce the potential influence of the particular start times. The probability distributions and the cumulative probability functions of the average TSIs for users in Di are shown in Figures S13 and S14 in the supplementary file.

4. Typical mobility indicators and the measurement of the effects of TSIs

4.1. *Typical mobility indicators*

We selected four commonly used mobility indicators that describe human mobility patterns from different perspectives. Specifically, we used movement entropy and the radius of gyration to measure spatial randomness and extent, respectively, eccentricity to depict the shape of daily activity locations, and daily travel frequency to reflect the strength of people's daily movements.

4.1.1. *Movement entropy*

In the field of information theory, entropy refers to the randomness and general uncertainty of target items (Shannon 1948). Researchers have introduced movement entropy, which is based on this concept, to measure the randomness of daily activity locations of an individual (Eagle and Pentland 2006, Qin *et al.* 2012). The following equation shows the general calculation:

$$H = -\sum_{n=1}^{N} p_n \times log_2 p_n \quad (2)$$

where H is the entropy value, N is the total number of unique items, and p_n is the proportion of each item (see Figure 1 for examples). The movement entropy can be calculated after defining p_n according to the purpose of the study (Song *et al*. 2010). In this study, N is the total number of activity locations, and p_n is the percentage of records at location n.

4.1.2. *Radius of gyration*

Radius of gyration (ROG) refers to the distribution of a component of an object around an axis (Goldstein *et al*. 2013). Considering that a trajectory consists of a series of location records, researchers have applied this concept to measure the spatial extent of the activity of people by taking each record to represent a mass point (González *et al*. 2008). ROG is normally calculated using Equation (3):

$$R_g = \sqrt{\frac{1}{N} \times \sum_{n=1}^{N} DIS\left(Pt_n, Pt_{,mean}\right)^2} \tag{3}$$

where R_g is the ROG value, N is the total number of daily records, Pt_n is the location of the n^{th} record, Pt_{mean} is the mean center point of all of the recorded locations and DIS is the distance function. Larger values of R_g generally denote larger spatial extents.

4.1.3. *Eccentricity*

Eccentricity measures the degree of deviation of an ellipse from a circle (Simmons 1996). From a morphological perspective, the shape of an individual's activity locations on a two-dimensional surface can be approximated as an ellipse. By considering each record as a mass point, the two derived principal moments of inertia correspond to the semi-axes of the ellipse (Yuan *et al*. 2012). The eccentricity of people's daily activity locations can be calculated using the following equation:

$$\varepsilon = \sqrt{1 - (e_2/e_1)^2} \tag{4}$$

where ε is the eccentricity value, and e_1 and e_2 correspond to the principal moments of inertia (González *et al*. 2008, Yuan *et al*. 2012). Possible values of ε range between zero and one. A larger value of ε indicates a larger difference between the lengths of e_1 and e_2 (i.e. the ellipse is narrower). When the lengths of e_1 and e_2 are the same, ε will be zero (i.e. the ellipse is a circle). If a person remains at one location all day long, her/his eccentricity is zero because e_1 and e_2 are equal.

4.1.4. *Daily travel frequency*

Daily travel frequency (DTF) is a commonly used indicator in both travel demand analyses (Martin *et al*. 1998) and human mobility studies (Carrion *et al*. 2014). Typically, a distance threshold such as 300–500 meters is applied to eliminate short movements from consideration as trips in traditional travel survey data. To that end, given a trajectory, this study first decomposed it into stays and moves (Equation (5)). The stops and moves of trajectories (SMoT, see Spaccapietra *et al*. 2008) model was applied in this study to achieve the 'stays-and-moves' identification purpose. The SMoT model has been widely applied for detecting stops using both GPS tracking data (Zheng *et al*. 2009) and mobile phone location data (Widhalm *et al*. 2015, Tu *et al*. 2017). Each move corresponds to a trip. DTF can then be calculated using Equation (6):

$$Trajectory \rightarrow [\ldots, stay_p, move_q, stay_{p+1}, \ldots] \tag{5}$$

$$tFreq = COUNT(move_q) \tag{6}$$

where *tFreq* is the value of the DTF, $stay_p$ is the p^{th} stay, $move_q$ is the q^{th} move, and *COUNT* is the counting function. This equation implies that every move segment has an equal effect on *tFreq*.

This study adopted the approach used by Jiang *et al*. (2013) to extract stays from mobile phone location data using two parameters, a distance threshold and a stay duration threshold. In this study, considering that the spatial resolution of the cell towers was approximately 550 meters, we set the distance threshold to 500 meters. For the stay duration threshold, because the average TSI in the D0 dataset was as small as 5 minutes, we choose 10 minutes as the duration threshold.

4.2. *Measurement of the effects of TSIs on mobility indicators*

To better estimate the underestimations, we refer to Zhao *et al*. (2016) and use the parameters of a linear regression model to quantify the effects of varying TSIs by fitting the values of each mobility indicator among the datasets with different TSIs:

$$V_i = \beta_{i,j} \times V_j \tag{7}$$

where V_i and V_j are the values of each user's mobility indicators in D*i* and D*j*, respectively, $i,j = (0,1,2,\ldots,16)$, $i > j$, and $\beta_{i,j}$ is the slope of the regression model between V_i and V_j.

With this linear regression model, firstly, we use $\beta_{i,j}$ to reflect the relationship between the values contained in datasets with different TSIs at an aggregate level. $\beta_{i,j} < 1$ implies that the value of a particular indicator decreases and a coarser TSI results in underestimations. Otherwise, $\beta_{i,j} > 1$ implies that a coarser TSI results in overestimations. Here, we use $(1 - \beta_{i,j})$*100% to measure the impacts caused by coarser TSIs. Both underestimation and overestimation in this study are relative concepts that indicate the trend of the changes between two different TSIs. The corresponding results would be more informative if we had a ground truth dataset. Secondly, we further adopt the coefficient of determination of the linear regression model (R^2) to assess how well the slopes of the linear regression model explains the changes in the mobility indicators among users. An R^2 close to one suggests that the related change exhibits high consistency among users and that the corresponding models are able to adequately explain the degrees of change among users.

5. Analysis and results

In this section, we first investigate how each user's mobility indicators change with TSIs to learn the basic change patterns for each indicator. Then, to better understand the trends and extract more practical insights, we quantify the aggregate impacts of TSIs and compare the differences of the impacts between the four selected indicators.

To compare the changes in the values of the mobility indicators at different TSIs, we assess both (1) the changes between the raw trajectories in D0 and the resampled trajectories in each dataset D*i* to learn the degree of impact that each TSI can cause and (2) the marginal changes between pairs of datasets with adjacent TSIs (e.g. 15-minute and 30-minute) to learn how many impacts each range of TSI contributes to the total impacts.

5.1. *Exploratory analysis of the effects of TSIs on mobility indicators*

To analyze how the values of the selected mobility indicators change with the TSIs, we first use scatter plots to explore the change patterns from an individual level. In the plots, the horizontal axis represents the mobility values of the same user in the dataset with a finer TSI, whereas every point in the plots corresponds to the specific mobility indicators of a specific individual with different TSIs. To better describe the characteristics of the plots, we employ the Spearman correlation coefficient (*r*) to measure the consistency of the changes among users (see Tables S1 and S2 in the *Supplementary file*) and the max value (*maxV.*) to reflect the changes in the value ranges. The content of this paper would be too long if all the plots of D0 vs. Di and Di vs. Di+1 were presented (16 plots for each group); therefore, we selected 8 plots for each group for each mobility indicator (Figures 6, 7, 9 and 11). All other plots are listed in the *Supplementary file* (Figures S1–S8).

5.1.1 *Movement entropy*

Regarding movement entropy, we first find that the values decrease with high inter-user consistency ($r > 0.96$) of degrees as TSI increased (see Figure 6(a)). Second, the marginal changes caused by 15-minute increases of TSIs tend to be minor (see Figure 6(b)), which implies that a 15-minute difference has few impacts on each user's movement entropy.

Movement entropy reflects the randomness of daily activity locations. Its value will decrease when some activity locations are neglected with increasing TSIs. However, activities with short duration times (e.g. eating lunch outside or shopping after work) are more easily neglected, and movement entropy is weighted by the duration times of daily activities. We refer to the concept of k-ROG (Pappalardo *et al*. 2015, see Section 5.1.2 for more detailed information) and introduce k-En to test the influences of activity duration times on the movement entropy. The k-En represents the movement entropy derived from the top K most frequent activity locations. As Figures S17 and S18 in the supplementary file indicate, the 2-En (meaning the movement entropy derived from the top two most frequent activity locations) in the resampled datasets contribute the most significant part for the movement entropy. Specifically, more than 25% of users' 2-En contribute more than 80% of the benchmark values derived from D0 and more than 40% of users' 2-En contribute more than 80% of the movement entropy derived from the corresponding resampled dataset. Therefore, these neglected activities with short duration times have limited impacts on this indicator.

5.1.2. *Radius of gyration*

The values of the ROG exhibit both high robustness and high consistency of change degrees ($r > 0.977$) among users as the TSI increases (see Figure 7). Moreover, the maximum ROG value is also very stable and changes from only 31.83 km to 30.84 km as the TSI increases in D1 to D15. One of the key reasons is that the value of ROG is weighted by the

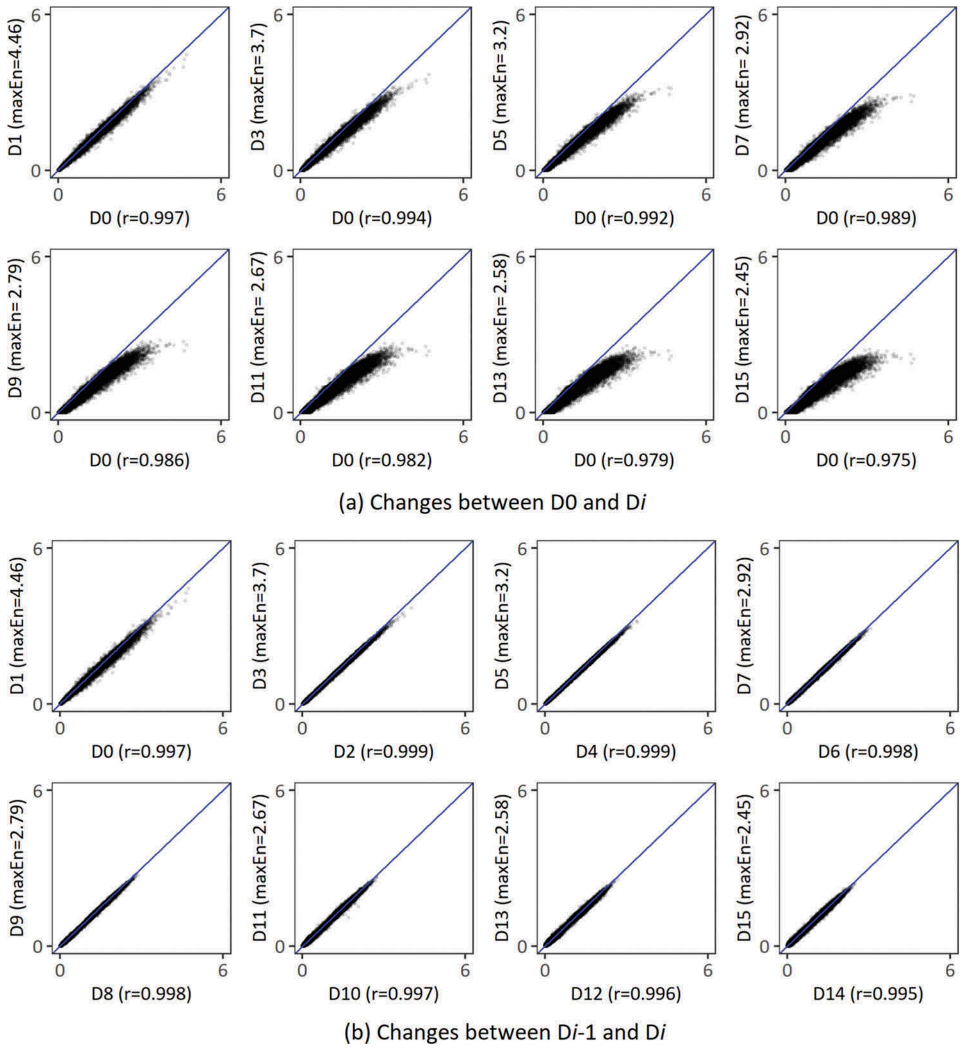

Figure 6. Changes in movement entropy between datasets with different temporal sampling intervals.

duration time of activity locations and activities with short duration times are more easily neglected, which is similar to the situation of movement entropy. Pappalardo *et al.* (2015) introduced a concept of k-ROG to present the radius of gyration derived from the top K most frequent activity locations. This concept was used to test the contributions of k-ROG to the conventional ROG derived from all the records. We found that more than 50% of users' 2-ROG (meaning the ROG value derived from the top two most frequent activity locations) derived from the resampled datasets contribute 80% more of the benchmark values derived from corresponding users in D0. With regard to the values of ROG derived from the resampled datasets, more than 70% of users' 2-ROG contribute 80% more accounts. (Figures S19 and S20). These results indicate that the locations of the two major activities determine the main part of the ROG of a user.

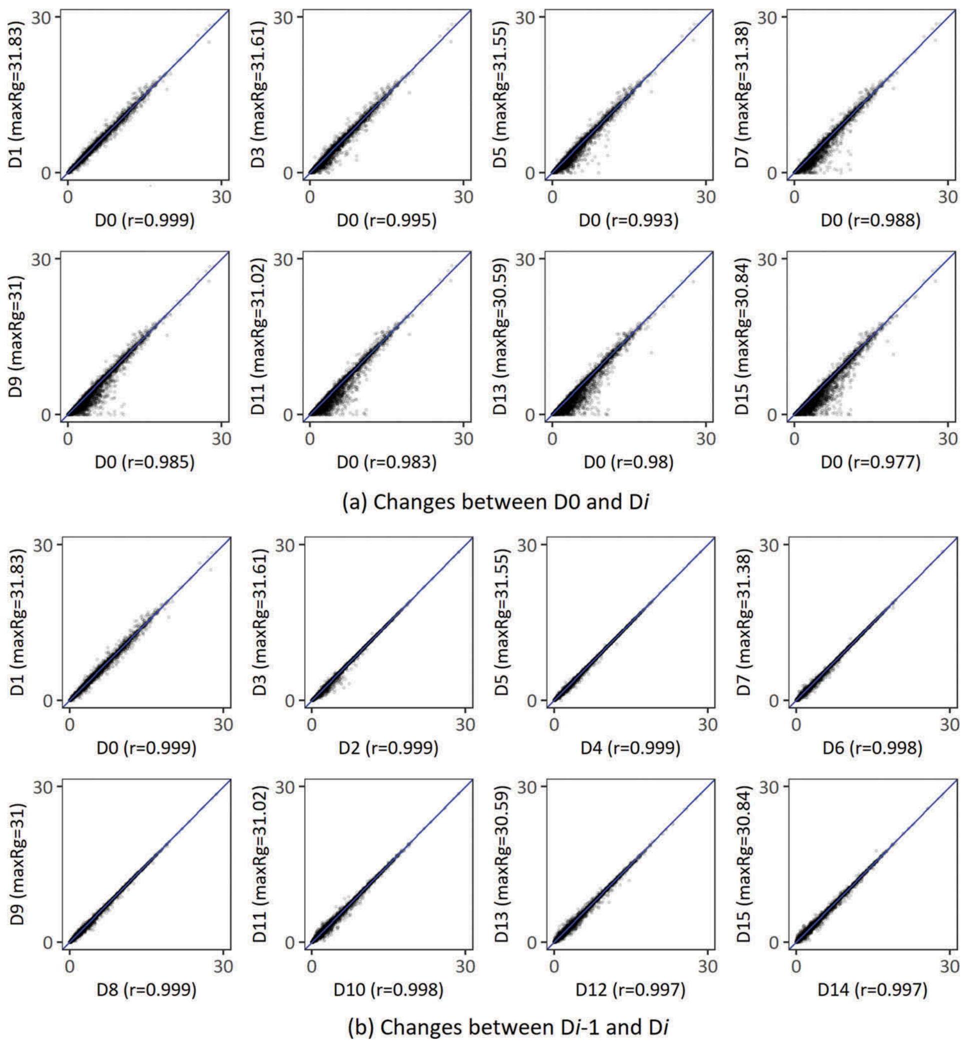

Figure 7. Changes in the radius of gyration between datasets with different temporal sampling intervals (Units: km).

In addition to the duration of daily activities, ROG is further related to the distances between activity locations and the center locations. When some activity locations are neglected, activity locations far away from the center location will offset the influences caused by the activity location located close to the center. As Figure S9 illustrates, the neglect of location B near the center location results in an increase in ROG from 0.991 to 1.091, while neglecting location D far from the center location results in a decrease in ROG from 0.991 to 0.9. If both locations are neglected, the change in ROG is even smaller (e.g. a change from 0.991 to 1.0) than if either location is neglected. During this process, however, the more activity locations that are neglected, the greater the movement entropy decreases, regardless of location (Figure S9). This offset effect makes ROG exhibit higher stability with varying TSIs than movement entropy.

The above results suggest that datasets with very coarse TSIs (e.g. two hours) can provide good estimations of the spatial extent of the locations of the major activities (e.g. sleep and work) for each user's daily life measured by the ROG.

We also note that for users with small spatial extents of daily activity locations (the points close to the origin in Figure 7(a)), the degree of changes in ROG values and the variations in the changes among users are larger than those for other users. This situation can be explained by the concept of space-time constraints from the time-geography framework (Hägerstraand 1970). When a user has a small spatial extent of daily activity locations, space-time constraints imposed by travel distances in the physical world become less critical. She/he can arrange daily activities more flexibly, and frequent short-duration activities are feasible. Alternatively, long-distance travel requires considerable time and limits the ability of people to perform other activities. These people tend to have fewer, longer-duration activities. As the TSI increases, short-duration activities can easily be ignored, resulting in a greater change in ROG values for users who have a small spatial extent of daily activity locations.

5.1.3. *Eccentricity*

For eccentricity, both the changes between D0 and D*i* and the marginal changes exhibit lower consistencies among users than movement entropy and radius of gyration. Figure 8(a) indicates that there are more points close to the axis $\varepsilon_{D0} = 1$ (meaning the locations of the user's daily activities are located along a line) and the axis $\varepsilon_{Di} = 0$ (meaning the user stays at one location all day long or the locations of her/his activities are evenly located in a circle) as the TSI increases. It indicates that many users' eccentricities decreased from values close to a value smaller than one to one or zero directly, which implies decreasing patterns with increasing TSIs.

Eccentricity is weighted by both the duration times of daily activities and the spatial distribution of activity locations. Its value may change from the maximum to the minimum as TSI increases, especially for users with few outdoor short-duration activities (e.g. morning jogging and after-dinner recreation are popular for elderly, retired people in Shenzhen, China). Both the short-duration outdoor activities (Figure S10(a)) and the 'fake moves' (Zhao *et al*. 2018) caused by undetected ping-pong phenomena (Figure S10(b)) for these users are easily neglected. As a result, the daily activity locations derived from the dataset may change from many to two or one, which results in significant changes in eccentricities (Figure S10). However, the short-duration outdoor activities and the 'fake moves' for these users have limited impacts on the values of movement entropy and radius of gyration (Figure S10). Therefore, the eccentricity exhibits an obviously lower inter-user consistency in change degree than the above two indicators.

Although the scatter plots do not show clear linear correlations (Figure 8(a)), the Spearman correlation coefficients remain significant and relatively high (e.g. D0 vs. D15; $r > 0.7$). One key reason underlying this phenomenon is the skewed distribution of the eccentricity values of the users because the eccentricities of many users were close to one, given that the top two major places accounted for the majority of most people's daily lives (Bagrow and Koren 2009) and these two locations form the shape of a 'line'.

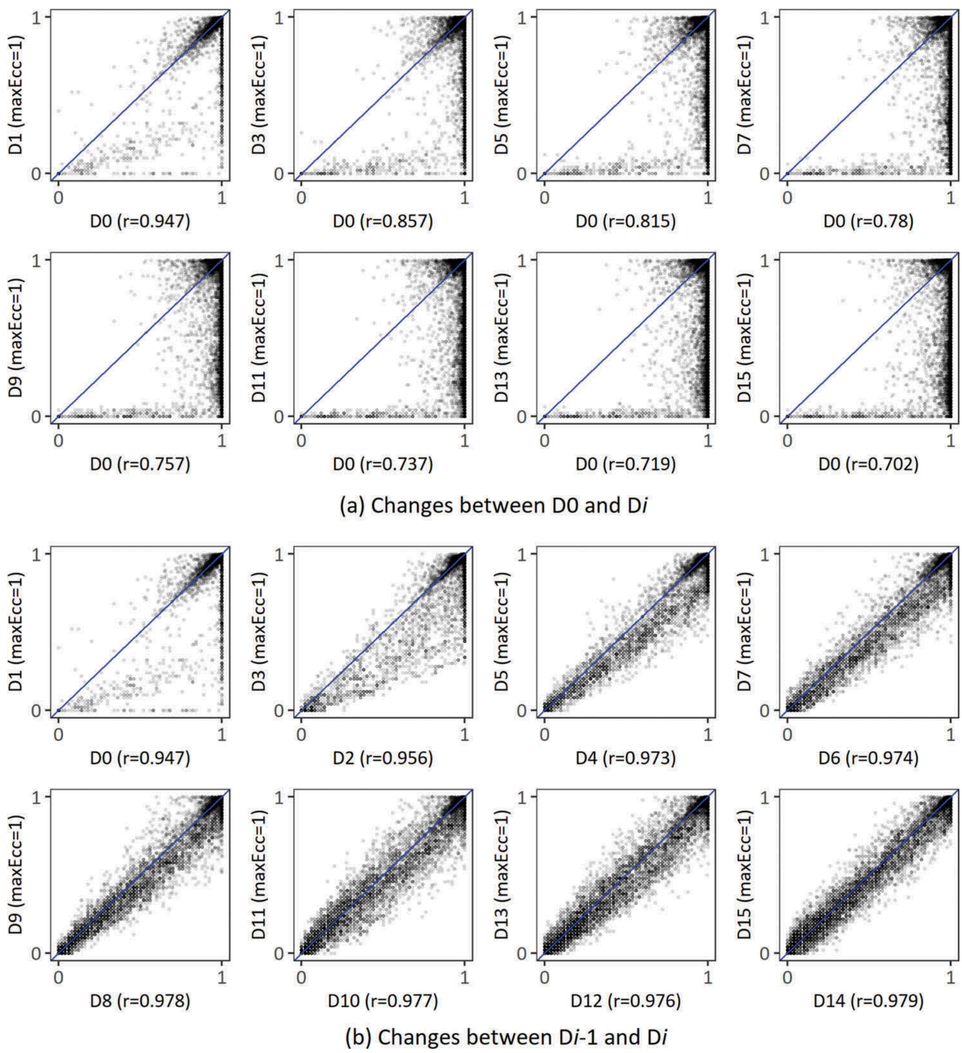

Figure 8. Changes in eccentricity between datasets with different temporal sampling intervals.

5.1.4. *Daily travel frequency*

From Figure 9, we first find that a certain coarse TSI (e.g. two hours for D8) leads to a significant decrease in DTF compared to the results derived from D0. The degree of the decrease is clearly larger than the above three indicators. In addition, the DTF of people who travel more frequently is more highly affected by the changing TSIs (Figure 9(a)). Second, the degrees of marginal change decrease with increasing TSI. Moreover, when the TSI was smaller than one hour (D4), the marginal change tended to be larger (Figure 9(b)). Third, the DTF values of many users decrease to approxmately two as the TSI increases (Figure S16), no matter how large they grew from D0, especially when the TSIs exceeded two hours (i > 8). In other words, the rates of decrease vary among users. This situation results in a relative low Spearman coefficient for the changes between D0

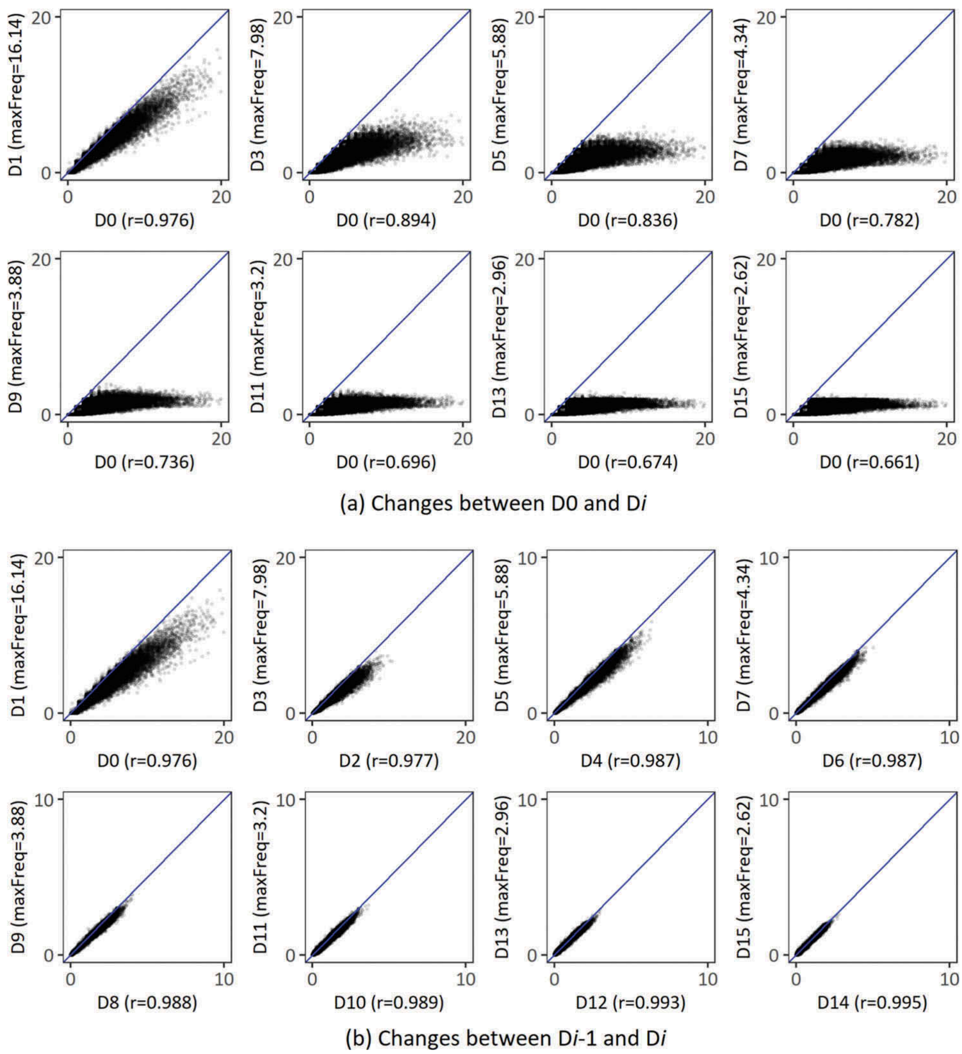

Figure 9. Changes in daily travel frequency between datasets with different temporal sampling intervals.

and D*i* among users for DTF. In fact, those two trips probably corresponded to the two commuting trips that were relatively robust to changes in the TSI.

Unlike the above three indicators, a short-duration daily activity contributed equally to the value of DTF as a long-duration daily activity. These short-duration activities account for a significant part of people's daily lives (Golledge and Stimson 1997). However, they are easily neglected as the TSI increases, which results in a significant decrease in DTF.

5.2. *Quantitative analysis of the effects of TSIs on mobility indicators*

Figure 10 shows the quantitative results of the effects of TSIs on the four selected mobility indicators measured by the linear regression model for the changes between

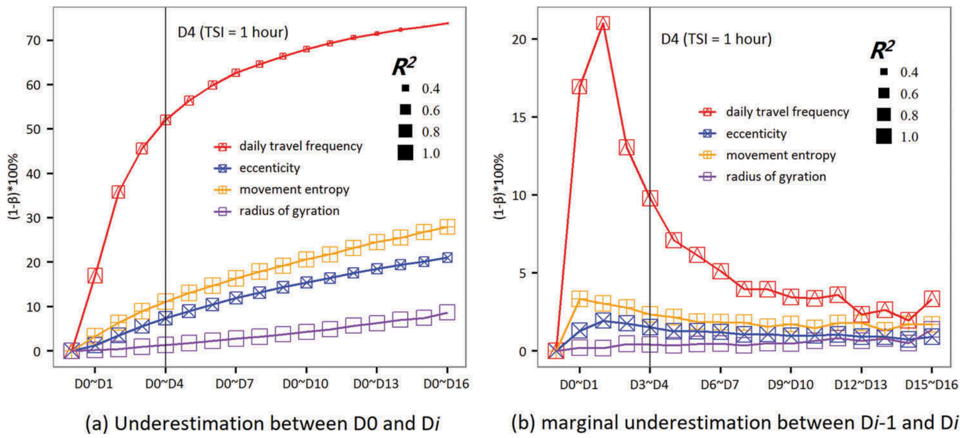

Figure 10. The impacts of the temporal sampling interval (TSI) on the values of different mobility indicators, as measured by (1-β)*100%. The size of each point corresponds to the R^2 of the corresponding linear regression model. Smaller R^2 values indicate greater variations in the degree of underestimation among users for the different mobility indicators.

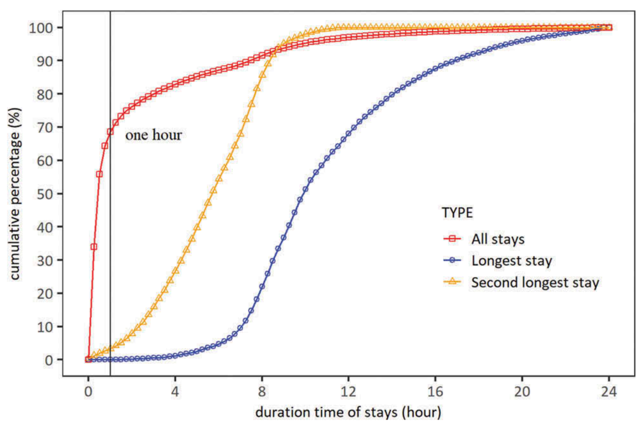

Figure 11. Duration times of the stays in the 'stays-and-moves' identification results derived from D0 based on the SMoT.

D0 and D*i* (Figure 10a) and marginal changes between D*i*-1 and D*i* (Figure 10b), where *i* = 1, 2, 3,…, 16. The corresponding regression parameter values are listed in Tables S3 and S4 in the *Supplementary file*.

We find that coarser TSIs tend to underestimate the values of the selected mobility indicators and the underestimations vary across indicators. Specifically, Figure 10(a) indicates that the underestimations of daily travel frequency compared to D0 exceed those of the other three indicators. The radius of gyration is the most stable indicator. Even when the TSI is as large as four hours, the underestimation is less than 10% compared to D0. Movement entropy experiences a similar but slightly smaller underestimation than that for eccentricity. In addition, for TSIs smaller than one hour (i < 4), the marginal underestimations caused by 15-minute increases of TSI tend to be higher, especially for daily travel frequency (Figure 10b). For instance, a 15-minute increase in

TSI leads to more than 10% (even 21% for D1 vs. D2) underestimation in the DTF when the TSI is smaller than one hour, while it mostly causes less than 5% underestimation in the DTF otherwise.

The sizes of the markers in Figure 10 represent the R^2 of corresponding models. In terms of the underestimation compared to D0, the consistency of daily travel frequency among the users decreases rapidly with increasing TSIs, whereas those for radius of gyration and movement entropy are relatively stable. It implies that a certain coarse TSI (e.g. two hours) may lead to very different impacts across users. Regarding movement entropy and eccentricity, their values experience similar degrees of underestimation at an aggregate level for a certain coarse TSI. However, the corresponding underestimations exhibit different degrees of consistency among users (Figure 10a), especially for relatively coarse TSIs (e.g. four hours). It suggests that researchers should be aware of the heterogeneity of the impacts on the derived mobility patterns caused by coarse TSIs among populations. In addition, Figure 10(b) illustrates that the marginal underestimations exhibit high degrees of consistency among users. It implies that a 15-minute increase in TSI causes minor inter-user consistency for the four selected mobility indicators.

6. Discussion

First, the degrees of the impact of TSIs on different mobility indicators are related to what activity features these indicators measure and how these features are quantified. Daily travel frequency is related to each stay in the 'stay-and-move' results regardless of its duration time. As Figure 11 indicates, the duration time of approximately 70% of stays derived from D0 based on the SMoT (see Section 4.4) is shorter than one hour. As the result of Carrion *et al*. (2014) implied, mobile phone location data with fine temporal resolution tended to capture more short-duration daily activities (e.g. meal/eating break or recreation) than traditional survey data. These stays with short duration times are easily neglected with increasing TSIs, and the daily travel frequency decreases directly. However, the other three indicators are related to the location and the duration time of stays. Stays with longer duration times are more easily sampled with changing TSIs, and they usually involve more records in corresponding trajectories (e.g. the activities at locations A and B in Figure 1). As Figure 11 indicates, more than 80% of the longest stays last longer than 8 hours and more than 70% of the second longest stays last longer than 4 hours. The duration-time-weighted indicators are more determined by the activities corresponding to these two stays mathematically than other activities (Kang *et al*. 2012). Therefore, a given TSI leads to a higher underestimation of daily travel frequency than the other three indicators. For example, most of the stays in D0 that last less than one hour cannot be identified as stays if the TSI becomes one hour; thus, 70% of stays can lead to more than five times higher underestimation of the daily travel frequency than the other three indicators (see Figure 10).

Second, the impact of TSIs on different indicators are complex at the individual level. Although coarser TSIs tend to underestimate the four selected indicators, the phenomenon in which some points are located above the diagonal lines in Figures 6–8 indicates that the values of the movement entropy, the radius of gyration and the eccentricity may increase with the TSI for certain users. In addition, the plots in Figure 9(a) indicate that for people who travel frequently, their daily travel frequencies are underestimated more. With regard to eccentricity, if the number of activity locations of a user decreases to two dues to the neglected activities with increasing TSIs (e.g. trajectories in Figure S10 (a)), its value will increase directly to one (the points at the axis of ε_{Di} = 1 in Figure 8), which is the maximum value. When the TSI increases, the radius of gyration will increase with an increase in the proportion of the locations that are far from the center of daily activities (e.g. location D in Figures S9 and S11), while the movement entropy will increase if all the locations are retained and the proportions of the locations with long-duration activities decreases (e.g. location A in Figure S11). As for daily travel frequency, people who have more daily trips tend to have more activities with short duration times, which are more easily neglected for a given TSI. Understanding the influences of the TSI among people can provide helpful insights to estimate the potential biases in policy-making based on mobile phone location data. For instance, short-distance trips are important for non-motor travel planning such as public bicycle planning (Xu *et al*. 2016b). These trips often involve short-duration activities, which, however, are easier to be neglected if the TSI is relatively large (e.g. one hour). As a result, the travel demand of these short-duration activities will be underestimated, especially for the people who travel more frequently.

Third, the selection of TSI is highly relevant to the research question. On one hand, when the research question is related to the spatial extent of activity locations with long duration [e.g. using the radius of gyration to study regional poverty (Blumenstock *et al*. 2015)], coarse TSIs (e.g. four hours) would be acceptable. When the research question is about the daily travel frequency such as using the daily travel frequency to evaluate urban travel demands (Çolak *et al*. 2015), a small TSI (e.g. smaller than 0.5 hours) should be used. On the other hand, given a mobile phone location dataset, if its TSI is sufficiently large (e.g. two hours or larger), indicators such as daily travel frequency and eccentricity are not suitable to describe the corresponding characteristics of the involved users. However, it can potentially provide a relatively good estimation of the radius of gyration and movement entropy for the descriptions of the corresponding characteristics of users' activity patterns. This issue refers to the temporal aggregation effect in the MTUP, which reminds researchers to choose the proper temporal resolutions when answering specific questions.

Fourth, a better interpretation of the relationship between TSI-related features and mobility indicators must account for the impacts of TSI. For example, Yuan *et al*. (2012) found that people who have higher call frequencies have higher movement entropy. This result can be partially explained by the effects of TSIs. People who have higher call frequencies, indicating more frequent location records and smaller TSIs, receive smaller underestimations from the TSIs and tend to have a higher movement entropy (see Figure 10). This issue is related to the temporal aggregation effect in MTUP. In fact, the MTUP is a fundamental question in space-time analysis, similar to the MAUP. Researchers need to be aware of this factor when they interpret corresponding outcomes to avoid potential misinterpretation of related conclusions.

Last, the dataset adopted in this study partly limited certain promising research directions. For instance, this study only reveals how the TSI, ranging from 15 minutes to 4 hours, affects the outcomes of the daily mobility patterns because the dataset in this study only covered one workday. However, human mobility patterns exhibit both regularity and fluctuations across different time scales (e.g. González *et al*. 2008, Järv *et al*. 2014, 2017). Investigating how the outcomes change with the TSIs in deriving activity patterns over different time scales (e.g. seasonal activities such as tourism) is another basic research topic. In addition, considering that the time spans of most mobile phone location datasets in previous publications vary significantly, answering 'how long is long enough' for a specific research purpose is worthy of further investigation. This issue further involves another practical topic that how to balance data quantity and data quality in answering a particular question. These meaningful topics require corresponding dataset support that the dataset in this study could not provide.

7. Conclusions

This study quantifies and reveals the complex impacts of TSIs of mobile phone location datasets on four typical human mobility indicators with a series of designed TSIs that increase at a fixed interval of 15-minutes from 15 minutes to 4 hours. We find that the impacts of TSIs are strongly related to the mobility feature that each indicator measures and how an indicator quantifies such feature mathematically. Specifically, (1) movement entropy receives a certain amount of underestimation from increasing TSIs compared with the value derived from the intensively sampled trajectory (D0, TSI < 5 minutes). The degrees of underestimation have a high consistency among users. (2) The radius of gyration is a very stable indicator when the TSI changes. It implies that even a mobile phone dataset with a relatively coarse TSI (e.g. two hours) can adequately estimate users' radius of gyration. (3) Eccentricity receives close underestimation with movement entropy from the aggregated level. However, the degrees of the impacts for this indicator vary significantly across users, especially for users with few short-duration daily activities. (4) Daily travel frequency derived from mobile phone location data can receive more than five times higher underestimation from increasing TSIs than the other three indicators. The underestimations are mainly derived from the increase in the TSI when the TSI is smaller than one hour. In addition, the degrees of underestimation also change significantly among users and users who travel more frequently receive higher underestimation for their daily travel frequencies. (5) Except for daily travel frequency, the other three indicators may increase with TSI for certain particular users.

The above findings demonstrate that the temporal aggregation effect of MTUP (Cheng and Adepeju 2014) have different influences on the results derived based on different mobility indicators. Being aware of the MTUP is useful to better understand the conclusions derived from current mobile phone location datasets such as the relationship between users' call frequencies and movement entropies (Yuan *et al*. 2012), or the travel demand of a public bicycle system (Xu *et al*. 2016b). In addition, related findings can also provide insightful suggestions in designing appropriate research questions based on given datasets and choosing proper datasets for particular research purposes.

In the future, we plan to examine the impact of the TSI on human mobility indicators across populations with typical activity patterns; for example, people who stay in a local area all day and people who travel a lot during the day. Such work will provide more informative implications for group-based human mobility studies. In addition, how the time span of mobile phone location datasets (the time span of the dataset in this study is only one day) and temporal resolutions jointly affect the outcomes of human mobility patterns derived from mobile phone location data is another meaningful topic. Investigations on this topic can provide more insightful suggestions for choosing appropriate datasets to answer particular questions.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This study was jointly supported by the National Natural Science Foundation of China [41231171, 41771441 and 41371420, 41801373]; Basic Research Project of Shenzhen City [JCYJ20170307164104491]; the Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Land and Resources [KF-2018-03-006] and the Alvin and Sally Beaman Professorship and the Arts and Sciences Excellence Professorship of the University of Tennessee.

Notes on contributors

Zhiyuan Zhao received his Bachelor degree in GIS, M.Sc. and Ph.D. degrees in Cartography and GIS from Wuhan University, China. His research interests include GIS for transportation and spatio-temporal data mining.

Shih-Lung Shaw is Alvin and Sally Beaman Professor and Arts and Sciences Excellence Professor of Geography at the University of Tennessee, Knoxville. His research interests cover space-time GIS, GIS for transportation, space-time analytics of human dynamics, transportation geography, and time geography.

Ling Yin received her Bachelor degree in Geography and Master degree in GIS from Nanjing University, China, and obtained her Ph.D. degree in GIS from the University of Tennessee at Knoxville. Her research interests include GIS for transportation, spatiotemporal data mining and time geography.

Zhixiang Fang received his M.Sc. and Ph.D. degrees from Wuhan University in 2002 and 2005, respectively. He is currently a Professor at the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His research interests include spatiotemporal modeling of human behavior, space-time GIS for transport, and intelligent navigation.

Xiping Yang received his Bachelor degree in surveying and mapping engineering from Chengdu University of Technology, China, and obtained his M.S. and Ph.D. degrees in GIS from Wuhan University. His research interests include spatiotemporal trajectory big data mining, urban human mobility, and travel behavior analysis.

Fan Zhang received his Ph.D. degree from Huazhong University of Science and Technology, China. He is currently a Professor at the Shenzhen Institutes of Advanced Technology, Chinese Academy Sciences. His research interests include big data applications in smart cities and cloud computing security.

Sheng Wu received his Ph.D. degree in cartography and GIS from Information Engineering University, China. He is currently a Professor at the Spatial Information Research Center of Fujian Province, Fuzhou University, China. His research interests include analysis and visualization of big data in government, applications in smart cities, and emergency management.

References

Ahas, R., *et al*., 2007. Seasonal tourism spaces in Estonia: case study with mobile positioning data. *Tourism Management*, 28 (3), 898–910. doi:10.1016/j.tourman.2006.05.010

Ahas, R., *et al*., 2015. Everyday space–time geographies: using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. *International Journal of Geographical Information Science*, 29 (11), 2017–2039. doi:10.1080/13658816.2015.1063151

Bagrow, J.P. and Koren, T., 2009. Investigating bimodal clustering in human mobility. *In*: *2009 International conference on computational science and engineering. Presented at the 2009 international conference on computational science and engineering*. Vancouver, BC, Canada, 944–947.

Barabási, A.-L., 2005. The origin of bursts and heavy tails in human dynamics. *Nature*, 435 (7039), 207–211. doi:10.1038/nature03459

Birenboim, A. and Shoval, N., 2016. Mobility research in the age of the smartphone. *Annals of the American Association of Geographers*, 106 (2), 283–291.

Blondel, V.D., Decuyper, A., and Krings, G., 2015. A survey of results on mobile phone datasets analysis. *EPJ Data Science*, 4 (1), 10. doi:10.1140/epjds/s13688-015-0046-0

Blumenstock, J., Cadamuro, G., and On, R., 2015. Predicting poverty and wealth from mobile phone metadata. *Science*, 350 (6264), 1073–1076. doi:10.1126/science.aad1815

Calabrese, F., Ferrari, L., and Blondel, V.D., 2014. Urban sensing using mobile phone network data: a survey of research. *ACM Computing Surveys*, 47 (2), 25:1–25: 20. doi:10.1145/2655691

Carrion, C., *et al*., 2014. Evaluating FMS: a preliminary comparison with a traditional travel survey. *In*: *Presented at the transportation research board 93rd annual meeting*. Washington, DC USA, No. 14-5541.

Chen, C., *et al*., 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation Research Part C: Emerging Technologies*, 68, 285–299. doi:10.1016/j.trc.2016.04.005

Cheng, T. and Adepeju, M., 2014. Modifiable Temporal Unit Problem (MTUP) and its effect on space-time cluster detection. *PloS one*, 9 (6), e100465. doi:10.1371/journal.pone.0100465

Çolak, S., *et al*., 2015. Analyzing cell phone location data for urban travel: current methods, limitations and opportunities. *Transportation Research Record: Journal of the Transportation Research Board*, 2526, 126–135. doi:10.3141/2526-14

Çöltekin, A., *et al*., 2011. Modifiable temporal unit problem. *In*: *Presented at the ISPRS/ICA workshop "Persistent problems in geographic visualization"(ICC2011)*. Paris, France.

Cuttone, A., Lehmann, S., and González, M.C., 2018. Understanding predictability and exploration in human mobility. *EPJ Data Science*, 7 (1), 2. doi:10.1140/epjds/s13688-017-0129-1

Diao, M., *et al*., 2016. Inferring individual daily activities from mobile phone traces: a Boston example. *Environment and Planning B: Planning and Design*, 43 (5), 0265813515600896. doi:10.1177/0265813515600896

Dodge, S., 2015. From observation to prediction: the trajectory of movement research in GIScience. *In*: H. Onsrud and W. Kuhn, eds. *Advancing geographic information science: the past and next twenty years*. Needham, MA: GSDI Association Press, 123–136.

Eagle, N. and Pentland, A., (Sandy), 2006. Reality mining: sensing complex social systems. *Personal Ubiquitous Comput*, 10 (4), 255–268. doi:10.1007/s00779-005-0046-3

Goldstein, H., Safko, J.L., and Poole, C.P., 2013. *Classical mechanics*. Pearson New International ed of 3rd Revised edition ed. London, UK: Pearson.

Golledge, R.G. and Stimson, R.J., 1997. *Spatial behavior: a geographic perspective*. New York, NY, USA: The Guilford Press.

González, M.C., Hidalgo, C.A., and Barabási, A.-L., 2008. Understanding individual human mobility patterns. *Nature*, 453 (7196), 779–782. doi:10.1038/nature06958

Goodchild, M.F., 2013. The quality of big (geo)data. *Dialogues in Human Geography*, 3 (3), 280–284. doi:10.1177/2043820613513392

Hägerstrand, T., 1970. What about people in regional science? *Papers in Regional Science*, 24 (1), 7–24. doi:10.1111/j.1435-5597.1970.tb01464.x

Horn, C., *et al.*, 2014. Detecting outliers in cell phone data: correcting trajectories to improve traffic modeling. *Transportation Research Record: Journal of the Transportation Research Board*, 2405, 49–56. doi:10.3141/2405-07

Hoteit, S., *et al.*, 2016. Filling the gaps: on the completion of sparse call detail records for mobility analysis. *In*: *Proceedings of the Eleventh ACM Workshop on Challenged Networks*. New York, NY, USA: ACM, 45–50.

Iovan, C., *et al.*, 2013. Moving and calling: mobile phone data quality measurements and spatio-temporal uncertainty in human mobility studies. *In*: D. Vandenbroucke, B. Bucher, and J. Crompvoets, eds. *Geographic information science at the heart of Europe*. Springer International Publishing, Basel, Switzerland, 247–265.

Isaacman, S., *et al.*, 2010. A tale of two cities. *In*: *Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications*. New York, NY, USA: ACM, 19–24.

Järv, O., Ahas, R., and 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. doi:10.1016/j.trc.2013.11.003

Järv, O., Tenkanen, H., and Toivonen, T., 2017. Enhancing spatial accuracy of mobile phone data using multi-temporal dasymetric interpolation. *International Journal of Geographical Information Science*, 31 (8), 1630–1651. doi:10.1080/13658816.2017.1287369

Jiang, S., *et al.*, 2013. A review of urban computing for mobile phone traces: current methods, challenges and opportunities. *In*: *Proceedings of the 2Nd ACM SIGKDD International Workshop on Urban Computing*. New York, NY, USA: ACM, 2:1–2:9.

Kang, C., *et al.*, 2012. Intra-urban human mobility patterns: an urban morphology perspective. *Physica A: Statistical Mechanics and Its Applications*, 391 (4), 1702–1717. doi:10.1016/j.physa.2011.11.005

Lazer, D., *et al.*, 2014. The parable of google flu: traps in big data analysis. *Science*, 343 (6176), 1203–1205. doi:10.1126/science.1248506

Liu, J., *et al.*, 2016. Rethinking big data: a review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, 115, 134–142. doi:10.1016/j.isprsjprs.2015.11.006

Liu, X., *et al.*, 2017. The impact of MTUP to explore online trajectories for human mobility studies. *In*: *Presented at the ACM SIGSPATIAL Workshop on Prediction of Human Mobility (PredictGIS)*. California, USA, 1–9.

Liu, Y., *et al.*, 2015. Social sensing: a new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105 (3), 512–530. doi:10.1080/00045608.2015.1018773

Lu, S., *et al.*, 2017. Understanding the representativeness of mobile phone location data in characterizing human mobility indicators. *ISPRS International Journal of Geo-Information*, 6 (1), 7. doi:10.3390/ijgi6010007

Ludden, B., *et al.*, 2012. *Report on implementation issues related toaccess to location information by emergencyservices (E112) in the European Union*. EU: Coordination Group on Access to Location Information for Emergency Services, No. Final report.

Mao, L., *et al.*, 2016. Mapping intra-urban transmission risk of dengue fever with big hourly cellphone data. *Acta tropica*, 162, 188–195. doi:10.1016/j.actatropica.2016.06.029

Martin, W.A., *et al.*, 1998. *Travel estimation techniques for urban planning*. Washington, DC, USA: National Academy Press.

Miller, H.J. and Goodchild, M.F., 2015. Data-driven geography. *GeoJournal*, 80 (4), 449–461. doi:10.1007/s10708-014-9602-6

Openshaw, S., 1984. *The modifiable areal unit problem*. Norwich, UK: Geo Books.

Pappalardo, L., *et al.*, 2015. Returners and explorers dichotomy in human mobility. *Nature Communications*, 6 (8166), 1–8. doi:10.1038/ncomms9166
Qin, S.-M., *et al.*, 2012. Patterns, entropy, and predictability of human mobility and life. *PloS one*, 7 (12), e51353. doi:10.1371/journal.pone.0051353
Ranjan, G., *et al.*, 2012. Are call detail records biased for sampling human mobility? *SIGMOBILE Mobile Computing and Communications Review*, 16 (3), 33–44. doi:10.1145/2412096
Ratti, C., *et al.*, 2006. Mobile landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design*, 33 (5), 727–748. doi:10.1068/b32047
Schneider, C.M., *et al.*, 2013. Unravelling daily human mobility motifs. *Journal of the Royal Society Interface*, 10 (84), 20130246. doi:10.1098/rsif.2013.0246
Shannon, C.E., 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27 (3), 379–423. doi:10.1002/bltj.1948.27.issue-3
Shaw, S.-L., Tsou, M.-H., and Ye, X., 2016. Editorial: human dynamics in the mobile and big data era. *International Journal of Geographical Information Science*, 30 (9), 1687–1693. doi:10.1080/13658816.2016.1164317
Simmons, G., 1996. *Calculus with analytic geometry*. 2nd ed. New York, NY, USA: McGraw-Hill Education.
Song, C., *et al.*, 2010. Limits of predictability in human mobility. *Science*, 327 (5968), 1018–1021. doi:10.1126/science.1177170
Spaccapietra, S., *et al.*, 2008. A conceptual view on trajectories. *Data & Knowledge Engineering*, 65 (1), 126–146. doi:10.1016/j.datak.2007.10.008
Steenbruggen, J., Tranos, E., and Nijkamp, P., 2015. Data from mobile phone operators: a tool for smarter cities? *Telecommunications Policy*, 39 (3–4), 335–346. doi:10.1016/j.telpol.2014.04.001
Tu, W., *et al.*, 2017. Coupling mobile phone and social media data: a new approach to understanding urban functions and diurnal patterns. *International Journal of Geographical Information Science*, 31 (12), 2331–2358. doi:10.1080/13658816.2017.1356464
Wang, H., *et al.*, 2010. Transportation mode inference from anonymized and aggregated mobile phone call detail records. *In*: *13th International IEEE Conference on Intelligent Transportation Systems. Presented at the 13th International IEEE Conference on Intelligent Transportation Systems*, 318–323.
Wang, P., *et al.*, 2012. Understanding road usage patterns in urban areas. *Scientific Reports*, 2, 1001. doi:10.1038/srep00386
Widhalm, P., *et al.*, 2015. Discovering urban activity patterns in cell phone data. *Transportation*, 42 (4), 597–623. doi:10.1007/s11116-015-9598-x
Xu, Y., *et al.*, 2016b. Estimating potential demand of bicycle trips from mobile phone data—an anchor-point based approach. *ISPRS International Journal of Geo-Information*, 5 (8), 131. doi:10.3390/ijgi5080131
Xu, Y., *et al.*, 2016a. Another tale of two cities: understanding human activity space using actively tracked cellphone location data. *Annals of the American Association of Geographers*, 106 (2), 489–502.
Yuan, Y., Raubal, M., and Liu, Y., 2012. Correlating mobile phone usage and travel behavior – A case study of Harbin, China. *Computers, Environment and Urban Systems*, 36 (2), 118–130. doi:10.1016/j.compenvurbsys.2011.07.003
Yue, Y., *et al.*, 2014. Zooming into individuals to understand the collective: a review of trajectory-based travel behaviour studies. *Travel Behaviour and Society*, 1 (2), 69–78. doi:10.1016/j.tbs.2013.12.002
Yue, Y., *et al.*, 2017. Measurements of POI-based mixed use and their relationships with neighbourhood vibrancy. *International Journal of Geographical Information Science*, 31 (4), 658–675. doi:10.1080/13658816.2016.1220561
Zhao, Z., *et al.*, 2016. Understanding the bias of call detail records in human mobility research. *International Journal of Geographical Information Science*, 30 (9), 1738–1762. doi:10.1080/13658816.2015.1137298
Zhao, Z., *et al.*, 2018. Identifying stops from mobile phone location data by introducing uncertain segments. *Transactions in GIS*, 22 (4), 958–974. doi:10.1111/tgis.v22.4
Zheng, Y., *et al.*, 2009. Mining interesting locations and travel sequences from GPS Trajectories. *In*: *Proceedings of the 18th International Conference on World Wide Web*. New York, NY, USA: ACM, 791–800.