ORIGINAL ARTICLE

Understanding the impact of temporal scale on human movement analytics

Rongxiang Su[1] [ID] · **Somayeh Dodge[1]** [ID] · **Konstadinos G. Goulias[1]**

Abstract

Movement is manifested through a series of patterns at multiple spatial and temporal scales. Movement data today are becoming available at increasingly fine-grained temporal granularity. These observations often represent multiple behavioral modes and complex patterns along the movement path. However, the relationships between the observation scale of movement data and the analysis scales at which movement patterns are captured remain understudied. This article aims at investigating the role of temporal scale in movement data analytics. It takes up an important question of "how do decisions surrounding the scale of movement data and analyses impact our inferences about movement patterns?" Through a set of computational experiments in the context of human movement, we take a systematic look at the impact of varying temporal scales on common movement analytics techniques including trajectory analytics to calculate movement parameters (e.g., speed, path tortuosity), estimation of individual space usage, and interactions analysis to detect potential contacts between multiple mobile individuals.

Keywords Human movement · Temporal scale · Trajectory analytics · Activity space · Interaction analysis · GPS tracking

1 Introduction

Human movement data obtained from Global Positioning System (GPS) tracking or other location-aware technologies (LATs) in the form of trajectories (i.e., time-ordered sequences of locations) are used for a variety of purposes including understanding human mobility behavior, activity patterns, and social interactions (for an early assessment of this potential see Goulias and Janelle 2005). These data can also be used to complement person daily diary surveys to verify reporting of activities and travel (Wolf et al. 2014a, b; Shen and Stopher 2014; Klous et al. 2017; Su et al.

✉ Rongxiang Su
   rongxiangsu@ucsb.edu

[1] Department of Geography, University of California Santa Barbara, Santa Barbara, CA, USA

2020, 2021), as a stand-alone tracing of the movement of people to replace expensive travel surveys (Procter et al. 2018; Marra et al. 2019), and understanding the choices and movement decisions of people requiring travel details that cannot be collected in other ways (Hood et al. 2011; Ciscal-Terry et al. 2016; Zimmermann et al. 2017). Movement observations often represent multiple behavioral modes along the movement path and complex patterns across different spatial and temporal scales. When these observations are represented by trajectories, the influence of analysis scale is unavoidable and may affect the outcomes of machine learning and knowledge discovery tasks. In this article, *temporal scale* is defined as the granularity of movement observations (i.e., the average time interval between two consecutive tracking points). Other terms for temporal scale include temporal resolution, temporal granularity, sampling frequency, sampling rate, and sampling interval, as used interchangeably in this article. The choice of granularity can influence the results of analysis performed on movement data (Laube and Purves 2011; Moreira et al. 2010).

Many analytical approaches to understand movement patterns rely on computing first order movement properties from movement trajectories, such as daily travel distance, speed (e.g., point-to-point speed along the movement trajectory and/or average travel speed between origin and destination of an uninterrupted movement), acceleration, turning angle, and path tortuosity or sinuosity (a metric of the amount of variability in movement direction and the shape of a trajectory) (Laube et al. 2007; Dodge et al. 2008). The first order movement parameters form the fundamental components of more complex movement patterns, and hence they are often incorporated as building blocks of machine learning techniques and methods that are required to represent, quantify, contextualize, and analyze movement trajectories to better understand complex human movement patterns such as human interaction in space and time. A primary task prior to movement pattern analysis is trajectory characterization and segmentation based on the first order movement parameters and their derivatives (Dodge et al. 2009). For example, for human interaction analysis, it is essential to first recognize what transport mode people are taking to better identify and contextualize critical interactions. Speed and path tortuosity are two fundamental features that can be easily extracted from raw tracking data to assist trajectory characterization and transport mode identification tasks (Buchin et al. 2011; Schüssler and Axhausen 2008). Hence, it is crucial to understand the impact of temporal scale on computation of these primitive movement parameters prior to more sophisticated human movement analytics. If the fundamental movement parameters needed to infer higher level parameters yield the same values at different scales, one can select the scale which satisfies minimum energy consumption. That is, since movement speed is used to infer transport mode, the sampling scales that result in the same ranges of speed values can also lead to the detection of the same travel modes. In contrast, if the selection of a higher scale yields different values, it may lead to erroneous inferences about the route selected, travel mode, or the estimated time of arrival. In fact, a coarse temporal scale can yield location and/or speed estimates with high error which can compromise many of the parameters controlling the functionality of important real-world applications (e.g., mapping congestion in cities, optimal logistics distribution development) and result in disastrous

occurrences and accidents. In addition to exploring the impact of temporal scale on calculating the fundamental movement parameters, previous studies also discuss the impact on more sophisticated human mobility applications such as identifying stops where major activities take place, computing movement entropy which measures the randomness of daily visited locations, and estimating radius of gyration which captures the spatial dispersion of locations where individual activities happened (Ranjan et al. 2012; Zhao et al. 2018, 2019). When it comes to the application of identifying human interaction patterns, it is important to take into account the individuals' activity spaces and accessible areas along the movement paths for potential overlaps (Hägerstrand 1970; Miller 2005; Dodge et al. 2021). It is therefore important to understand the impact of temporal scale on estimation of individual space usage and identifying the interaction between two or more moving individuals. These studies highlight the importance of analytical scale in movement analytics. However, the relationships between the observation scale of movement data and the analysis scales at which movement patterns are captured remain understudied.

This article aims at investigating the role of temporal scale in human movement analytics. With this, we take up an important question of "how do decisions surrounding the scale of movement data and analyses impact our inferences about movement patterns?". Compared to existing studies that are more focused on animal movement in the Euclidean space (Laube and Purves 2011), we investigate the impact of varying temporal scales on understanding human movement which occurs in a more confined network space (i.e., parking lots, roadways, walkways, bikeways, and public transportation stops, stations, and routes). Through a set of computational experiments in the context of human movement, we take a systematic look at the impact of varying temporal scales on common computational movement analytics techniques including calculating first order movement parameters (e.g., speed, path tortuosity), estimation of individual space usage, and interactions analysis among mobile individuals. The outcomes of this study are important in our decisions about temporal scale in movement data collection, mobility analytics, as well as simulation of movement for real-world applications.

2 Background

2.1 Movement analytics and temporal scale

With the advancement of LATs, movement data today are becoming available at increasingly higher volumes and variety at multiple temporal granularity collected through diverse modes such as GPS data recorders, call detail records (CDR) of smart phones, radio-frequency identification (RFID) tags, Wi-Fi and Bluetooth sensors, and georeferenced social media applications (Batty 2012). Beside contributing to research on understanding of human mobility behavior, other applications of tracking human movement include, but not limited to mapping congestion levels in cities (Kan et al. 2018a; Stipancic et al. 2019), verifying design characteristics of built transportation infrastructure components (Deng et al. 2018), obtaining real-time provision of information about transit services such as expected arrival time of

vehicles at boarding stops and stations (Shalaby and Farhan 2004; Brakewood and Watkins 2019), monitoring near real-time information for car travels (Martínez-Díaz and Soriguera 2021), planning optimal logistics distribution development (Žunić et al. 2020), estimation of greenhouse gas (GHG) emissions and travel (Kan et al. 2018b; Neves and Brand 2019; Sui et al. 2019), and modeling disease spread such as COVID-19 (Fang et al., 2020; Kraemer et al. 2020; Lai et al. 2020; Tian et al. 2020; Su and Goulias 2021). Different applications using LAT data necessitate different degrees of precision and accuracy (Fillekes et al. 2019; Marra et al. 2019; Schneider et al. 2016; Wolf et al. 2014b).

Using a suitable temporal scale is critical in movement analytics. For example, using movement data of cows, Laube and Purves (2011) show that the median and variance in speed drop with a coarser sampling rate, as the actual path traveled is systematically underestimated. In addition, median values of turning angle and path sinuosity might result in higher values in finer temporal windows compared to coarser data. This knowledge is essential in customizing tracking to optimize the trade-off between tracking costs (i.e., energy and time consumption, battery life), sampling interval, and knowledge inference from movement tracking data (Marra et al. 2019). It can also contribute to characterizing the temporal dimension and advancing the representation of time in movement context (Aigner et al. 2007). For example, for travel surveys in transportation studies that aim to collect data on all modes of transport, practitioners have settled to use a scale between 3 and 5 s per fix (Wolf et al. 2014a). For the purpose of traffic surveillance, a common practice is to apply a large-scale floating car system (i.e., GPS tracking of vehicles) to track vehicles' locations at one fix per minute to reduce the cost of data transmission (Chen et al. 2014). This type of data can also support the research of human's spatial and social behavior and their interaction with each other and the environment through data mining and knowledge discovery (see the review in Lu and Liu (2012)). The sampling rate for some passive mode detection algorithms may be 30–60 s per fix and considered to be sufficient when the data collection period is long (Burkhard et al. 2020). For individual activity space estimation, in general, a lower resolution data may be sufficient. Zhao et al. (2019) argue that using a human tracking data set collected at two-hour sampling interval can depict fairly accurately the spatial extent of individual's locations of major activities in a day. Smolak et al. (2021) find that an entropy measure to quantify the predictability of individual mobility can decrease with increasing sampling interval of human tracking data up to a resolution of one day. This paper adds to these studies by providing a systematic investigation of the impact of varying temporal scales on common computational movement analytics.

2.2 Activity space analysis

In addition to computing the first order movement parameters, individual activity space approximation is another important subject in travel behavior and time-geography domains (Golledge and Stimson 1997; Lee et al. 2016; Miller 1991). Human activity space usually is measured by a set of visited locations. The definitions of individual activity space or space utilization vary greatly in terms of the

spatial and temporal resolutions. For a large-scale study such as daily or weekly, individual activity space can be characterized by its shape, size, and spatial distribution of activity locations. Minimum convex polygon (Worton 1987) is widely used as the simplest estimator of home range of animals in movement ecology as well as in human activity space approximation (Buliung and Kanaroglou 2006a, b; Hirsch et al. 2014; Lee et al. 2016). It is defined as a polygon that contains all tracking locations of a moving entity and has no internal angle exceeding 180 degrees. However, it is also well known that minimum convex polygon has many drawbacks and tends to overestimate the activity space as the measure only considers outermost points and does not consider internal space usage and constraints (Worton 1987). One way to compensate for this shortage is to take into account the spatial distribution of the tracking points. The indicator of radius of gyration incorporates the spatial distribution of visited locations and has been widely used to approximate the spatial dispersion of human's activity space (González et al. 2008; Song et al. 2010; Yuan and Raubal 2016; Xu et al. 2018). Besides these two measures characterizing the shape around the activity locations, kernel density estimators (KDE) is another way to approximate activity space by transforming the point pattern into a continuous probability density surface of visit (Silverman 1986; Worton 1987). Specifically, animal's space utilization or the home range usually can be approximated by the 95% cumulative intensity contours of probability density surfaces (Downs and Horner 2012; Selkirk and Bishop 2002; Worton 1987). This approach has also been applied to approximating the size of human activity space (Kwan 2000; Schönfelder and Axhausen 2003). However, it does not consider the temporal structure of tracking data and the results are highly sensitive to bandwidth selection (Byrne et al. 2014; Hemson et al. 2005; Horne and Garton 2006). Bandwidth in this case is used to define the extent of the kernel distribution. In movement ecology, the Brownian bridge movement model (BBMM) has been recognized as a more favorable approach to traditional KDE to estimate animal's home range as the BBMM incorporates both the transitions and the amount of time between consecutive locations of an individual (Bullard 1991; Horne et al. 2007). Similar to KDE, the 95% BBMM can be used to delineate the standard home range size of animals. In this study, we investigate the temporal scale impact on three common measures of human activity space including minimum convex polygon, radius of gyration and KDE.

For a fine-scale approximation of individual activity space during movement (i.e., individual space utilization between successive tracking points), time-geographic approaches are more appropriate than other aggregate measurements. In time-geography, the activity space of a moving entity can be estimated by a *space–time prism* (Hägerstrand 1970). As shown in Fig. 1, the space–time prism is shaped by a pair of origin and destination locations, a time budget, and the maximum speed capacity for the travel mode. The projection of this prism on a two-dimensional space is called *Potential Path Area (PPA)* (Burns 1979; Lenntorp 1976). The PPA ellipse captures the accessible locations that a moving individual can visit under the given space–time constraints. For mathematical definitions of PPA, the reader can refer to Miller (2005).
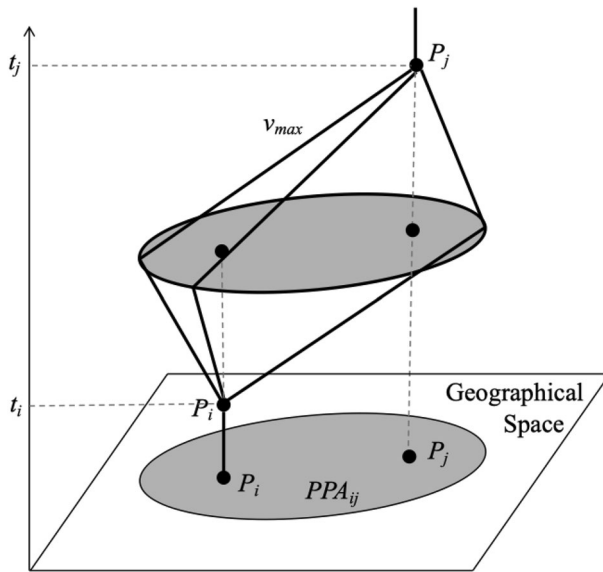
Fig. 1 Illustration of space–time prism and potential path area (modified from Miller 2005)

2.3 Movement interaction analysis

Recent advances in tracking technologies enable the development of fined-grained movement data. However, it remains a challenge to establish the appropriate scale for interaction analysis of collective human movement in space and time. Interaction analysis is a critical task in movement pattern analysis (Long et al. 2015). Identifying spatiotemporal patterns of interaction among moving individuals can contribute to understanding traffic dynamics, human social relationships, urban land use dynamics, virus transmissions through human contact, to name a few. Interaction can take two different forms based on when it occurs: concurrent (or direct/synchronous) interaction and delayed (or indirect/asynchronous) interaction. Existing methods quantifying interactions mostly rely on the spatial proximity between two individuals and many of them require user defined spatial and temporal thresholds (see comprehensive reviews by Long et al. (2014); Miller (2015); Joo et al. (2018)). However, these proximity-based approaches are limited when interacting individuals are not following a similar path simultaneously due to signal loss or imperfect tracking, or when the interactions are delayed (e.g., two persons visit the same grocery store at different time). In contrast, time-geography or PPA-based approaches are found to provide a more robust framework to identify both concurrent and delayed interactions between individuals (Long et al. 2015; Hoover et al. 2020; Dodge et al. 2021). This is mainly because PPA itself incorporates the uncertainty of positioning and gaps in movement data by considering the locations accessible to moving individuals between tracking points, while the proximity-based approaches lack this capacity (Dodge et al. 2021). Choices on observation scale might lead to significantly different outcomes in identified patterns of interaction. This can be crucial in

applications such as public health where underestimation or overestimation of critical contacts might have serious implications (Cencetti et al. 2021). It is therefore important to study the relationships between the observation scale of movement data and the analysis scales at which movement patterns are captured. In this article, we discuss the role of temporal scale in human interaction analytics.

3 Data

The data used in this study comes from the 2012–2013 California Household Travel Survey (CHTS) which includes a single-day travel diary and three days of GPS tracking data containing the same single-day during which a travel diary was collected from a subset of the total recruited respondents to CHTS (NuStats 2013). In the travel diary, respondents report their trip start time and end time, origin and destination coordinates (in longitude and latitude format) and corresponding location types (e.g., home, work, school, other) for every trip in the assigned diary day. The GPS component of the CHTS tracked individual's locations with a speed greater than one mile per hour every three seconds (i.e., only records movement but not stops) using the GlobalSat GPS Data Loggers that can be easily carried in a pocket, bag, or purse. Each GPS record tracks a unique ID of each respondent, current location in longitude and latitude format, and local time. In this study, we use the subset that had a perfect match between the travel diary and the GPS tracking component. In this subset, the transport mode of each individual trip is labeled manually through a software named Trip Identification and Analysis System (TIAS) which uses speed profiles from the GPS tracking data collected every 3 s as input (NuStats 2013). This subset of data contains 3622 individual respondents from 1683 households reporting 55,147 trips in total that happened from September 2012 to January 2013. Trip is the one-way movement from an origin to a destination. The average trip length is 9.94 km with a standard deviation of 41.48 km. The average trip duration is 17.71 min with a standard deviation of 94.94 min. Figure 2 shows the data
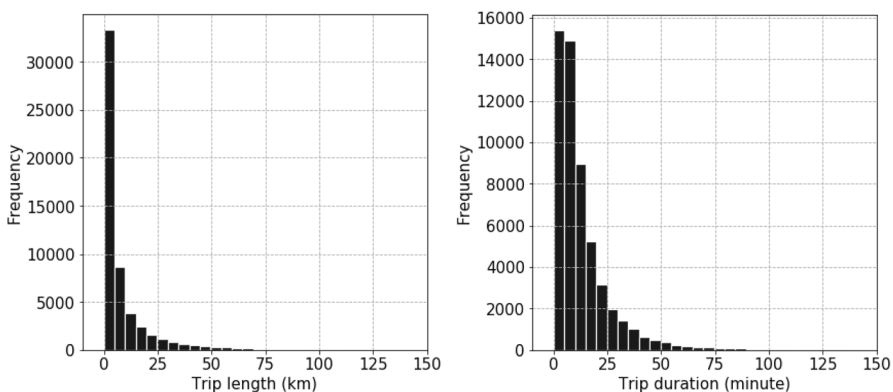

Fig. 2 The distributions of trip length and trip duration

distributions of trip length and trip duration. In our sample, 60.69% of the trips are shorter than 5 km and 55.07% of the trips last less than 10 min.

It is essential to distinguish different modes of transport in the computation of speed, path tortuosity, and space utilization. To avoid the influence of mixed mode trips, in the experiments below, only the trips that were made using a single mode according to the travel diary are used. The total number of single-mode trips are 39,381 trips (71.41% of the whole data set). According to CHTS, the transport modes are labeled and categorized into five major classes including *auto* (69.40%), *transit bus* (1.56%), *rail transport* (1.75%), *walk* (23.97%), and *bike* (2.55%), which account for 99.22% of the total single-mode trips in our sample. The rest of the 0.78% single-mode trips for other non-conventional modes such as streetcar/cable car/trolley, wheelchair/mobility scooter, other non-motorized modes, or unknown are omitted for this experiment. Specifically, the auto category includes mostly private vehicles, taxis, and hired cars. The transit bus category comprises of buses, shuttles, and other private mass transportation. The rail transport category includes metro, subway, and train. It is worth noting that even though metro/subway/train can be counted as a transit system, we distinguish them from buses and shuttles because the movement path of metro uses a straighter network, and metro usually travels at a constantly faster speed than buses and shuttles.

4 Computational experiments and the results

In human movement analytics, a primary task is to compute movement parameters such as speed, acceleration, direction, path tortuosity to characterize human movement in space and time (Schüssler and Axhausen 2008). In this section we first use several computational experiments to demonstrate how movement parameters can be influenced by varying temporal scales of the GPS tracking data. This is followed by an investigation of the impact of temporal scale on estimation of individual space usage computed using different approaches (e.g., minimum convex polygon, 95% Kernel Density Estimation home range, radius of gyration, and potential path area). The last focus of this section is to examine the scale impact on the analysis of human interactions between mobile individuals using a time-geographic approach.

4.1 Movement parameters

Two basic movement parameters, speed and path tortuosity, are used to demonstrate the impact of varying sampling intervals on computing movement parameters. Given the original sampling interval is 3 s per fix, the raw data are manually downsampled to 9 s, 15 s, 30 s, 60 s (1 min), 180 s (3 min), 300 s (5 min), 600 s (10 min), 900 s (15 min), and 1800 s (30 min) by skipping multiple trajectory points (see the example in Fig. 3). Granted that it is less likely to use 10 min or above sampling intervals for human movement analysis, these coarse sampling intervals are retained in our experiments for observing more holistically the scale impact on movement analytics. In addition, in many applications, high-resolution GPS tracking data are
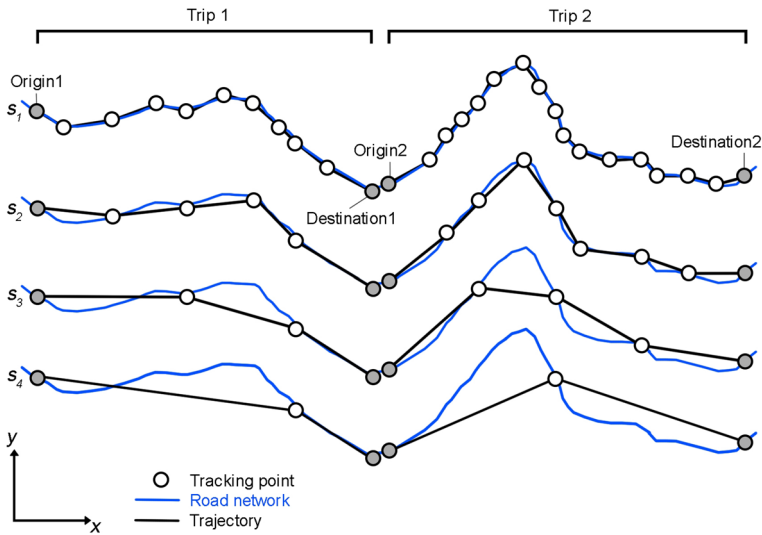
Fig. 3 Deriving trajectories at variable temporal scales. In this sketch, s_1, s_2, s_3, s_4 denote four different temporal scales of the individual tracking data. The origin location of the second trip is not exactly the same as the destination of the first trip because the GPS logger being used in this study will turn off tracking when people move slower than 1 mph (i.e., when they participate in an activity at Destination 1)

not available and tracking is done via other stationary sensors at times when moving individuals interact with or pass by these sensors. These forms of commonly used movement data including call detail records, smart card data, geotagged social media check-in data, etc., are obtained in various resolutions, often coarser than 30 min or 1 h (Barbosa et al. 2018). Therefore, it is important to investigate the impact of such coarse temporal scales on movement analytics. Considering that the origin and destination of individual trips are places where activities happened, the original origin and destination locations of each trip are retained in the process of down-sampling (see the example in Fig. 3). It is important to note that when the sampling interval is increased to 1800 s, only origin and destination locations are left for 93.06% of individual trips as the durations of most trips are shorter than 1800 s. As shown in Fig. 4, the number of GPS fixes substantially decreases after down-sampling.

An illustration of the trajectory of the same individual at different sampling intervals is presented in Fig. 5. The original trajectory collected at 3 s (shown in dark red) represent the continuous movement of the individual as a smooth trajectory, while those down-sampled to 9 s, 15 s, and beyond exhibit a 'jagged' geometry. The finer resolution tracking data can delineate the shape of the road network (also see the trajectory at the scale s_1 in Fig. 3). As shown in the two inset figures, the finer scale data captures turnings at intersections and the moving along a curved road well, while the coarser trajectories might overlook these details. To tackle this issue, a common practice is to snap individual's locations to road segments by map matching (Hashemi and Karimi 2014; Quddus et al. 2007). In this way, it becomes possible to compute network distance that incorporates the shape of road network
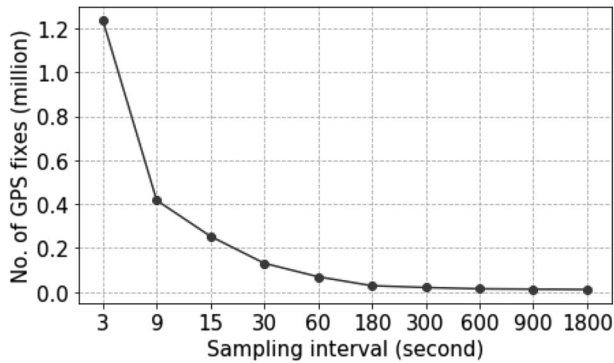
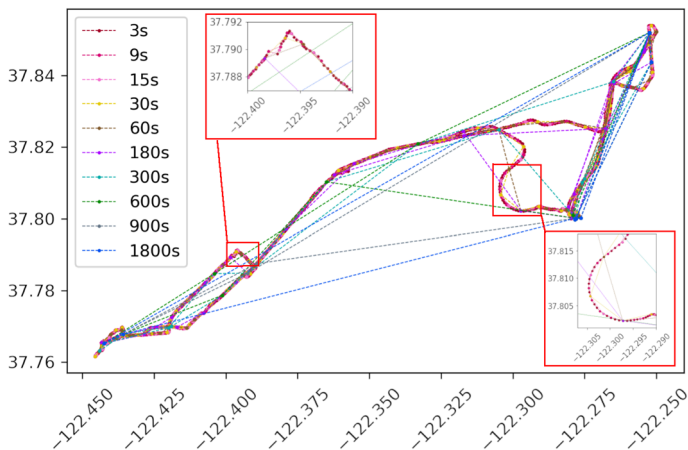**Fig. 4** Number of GPS fixes at different sampling intervals

Fig. 5 Illustration of an individual trajectory at different temporal scales

instead of taking Euclidean distance between points of the raw trajectory. However, when adding map matching to preprocess trajectories, one should be careful about the temporal scale. As sampling interval increases, the actual movement path can be distorted (see Figs. 3 and 5), which may cause erroneous results in the map matching process and computation of trajectory shape and movement distance. In our experiments, we do not apply map matching as the original data come in a very high resolution and represent the shape of the network with high precision.

4.1.1 Speed

Two types of speed are used here that are the point-to-point speed along the trajectory and the average speed of an individual trip. A trajectory of a trip consisting of n tracking points can be denoted as

$T = \{(x_0, y_0, t_0), (x_1, y_1, t_1), \ldots, (x_i, y_i, t_i), \ldots, (x_n, y_n, t_n)\}$, where (x_i, y_i) represents the geographic coordinate and t_i denotes the recorded timestamp. The point-to-point speed $v_{i,i+1}$ can be calculated as follows using a pair of consecutive fixes (x_i, y_i) and (x_{i+1}, y_{i+1}).

$$v_{i,i+1} = \frac{\sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}{t_{i+1} - t_i} \quad with\ i \in [0, n-1] \tag{1}$$

The average speed of an individual trip $\bar{v}$ can be defined as the total distance traveled between the origin (x_0, y_0, t_0) and destination (x_n, y_n, t_n) divided by the elapsed time as shown below.

$$\bar{v} = \frac{\sum_{i=0}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}}{t_n - t_0} \tag{2}$$

The goal of this experiment is twofold: (1) to reveal the impact of sampling intervals on computing both the point-to-point speed and the average speed of individual trip, and (2) to understand how that impact differs for different transport modes.

Point-to-point speed along the path Figure 6 illustrates the impact of varying temporal scales on overall point-to-point speeds, disaggregated by different transport modes. The boxplots present the variation of point-to-point speed in our sample at different sampling intervals. Each boxplot displays a median value, a box enclosing the twenty-fifth to seventy-fifth percentiles, and an upper whisker and a lower whisker representing the maximum and minimum values which are less than a distance of 1.5 times the interquartile range from the upper quartile and the lower quartile, respectively (Tukey 1977). In addition, the mean value is also displayed using a triangle mark. Outliers (values beyond the whiskers) are excluded to improve readability. The results visualized in Fig. 6 suggest that as the sampling interval increases, the median and mean of point-to-point speed tend to decrease, which is consistent with existing research (Laube et al. 2007; Laube and Purves 2011). This is reasonable because the actual point-to-point distance traveled is underestimated as the sampling interval increases. As the example in Fig. 5 shows, when comparing the original trajectory collected at 3 s with the trajectory down-sampled to 30 s or above, it is obvious that while the travel time remains the same between the same pair of tracking points at different sampling intervals, the distance tends to shrink due to the aggregation, cutting network corners, and elimination of network details, which results in reducing point-to-point speed. Except for the walk mode which presents a stable variance in speed across different sampling rates, the variance in point-to-point speeds of other modes decreases as the temporal scale increases. In general, except for the transit bus mode, the mean and median values of point-to-point speed are very close. Presumably, transit buses usually make more stops than other modes which causes more low values of point-to-point speed compared to other modes.

Considering that the distribution of speed values is positively skewed, a nonparametric test, Kruskal–Wallis test (Kruskal and Wallis 1952) is applied to examine whether there is significant difference among the mean values of speed at the ten
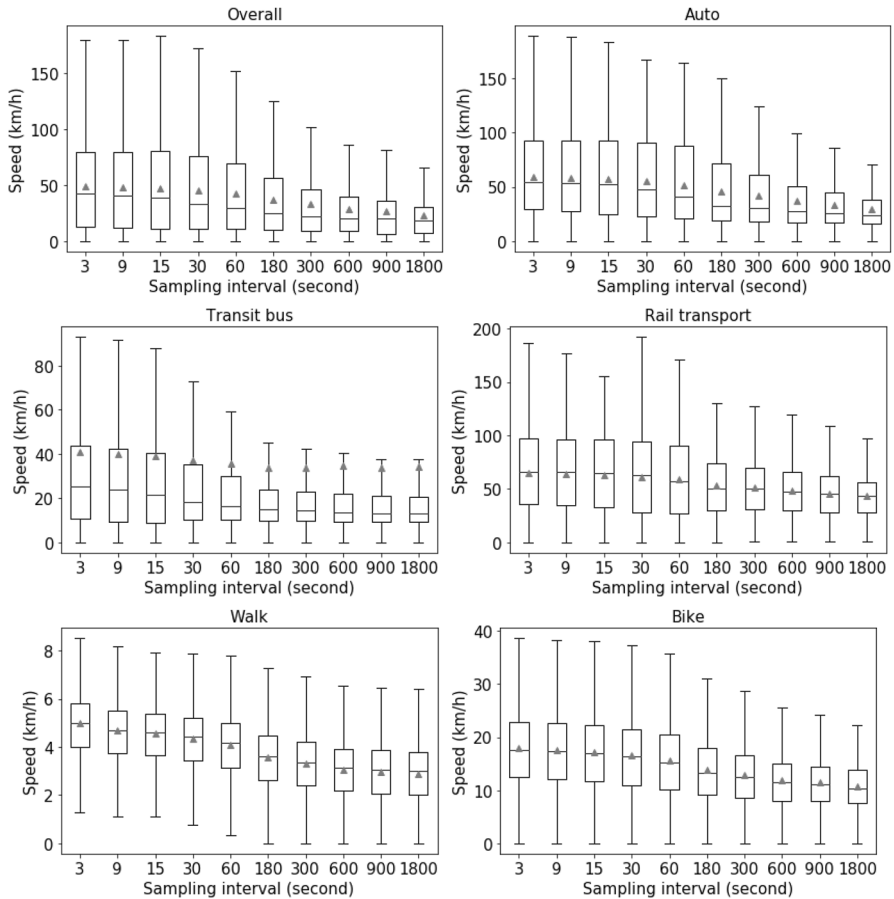
Fig. 6 Box plots of point-to-point speed values by different transport modes at different temporal scales

different sampling intervals. The results indicate that there is significant difference ($p < 0.01$) among the ten groups of speed values at different sampling intervals both for the overall sample and the subsets by the five categories of transport modes. Next, a Dunn's test (Dunn 1964) is applied to determine which specific means are significantly different from the others. The results of Dunn's test on overall point-to-point speed indicate that all pairs of sampling intervals (i.e., speed of the same mode computed at two different temporal scales) are significantly different from each other (all pairwise $p < 0.05$). Similarly, the subsets of auto and walk modes also presents the same relationship (all pairwise $p < 0.05$). The mean point-to-point speed of auto mode drops significantly from 59.12 to 29.84 km/h when sampling interval increases from 3 to 1800 s (30 min). For walking, it drops significantly from 5 to 2.88 km/h. For bike mode, only the groups of 600 s and 900 s are found not significantly different ($p = 0.07$). The average speed of bike trips drops significantly from 17.94 to 10.74 km/h as the sampling rate becomes coarser. That is to say, increasing

sampling intervals (even just from 3 to 9 s) will underestimate significantly the point-to-point speed when people travel by auto, walk, and bike. However, if we take a close look into the drops of mean speed of auto mode, for instance, the difference of the mean speeds between 3 and 15 s is only 1.93 km/h which should not be considered as a large difference in the transportation context even though it is statistically significant. In terms of transit bus and rail transport modes, the significance test results suggest that the mean speed of many groups of different sampling intervals are not significantly different. Regarding transit bus, we find no significant difference between 180 and 300 s ($p = 0.18$), 300 s and 600 s ($p = 0.35$), 300 s and 900 s ($p = 0.09$), 600 s and 900 s ($p = 0.46$), 600 s and 1800 s ($p = 0.23$), 900 s and 1800 s ($p = 0.62$). For rail transport, only the groups of 600 s and 900 s ($p = 0.08$), 900 s and 1800 s ($p = 0.06$) are found not significant. The average speed drops from 64.76 to 59 km/h when the sampling interval increases from 3 to 60 s. Given that these two mean speed values are significantly different, the absolute difference is only 5.76 km/h which should not be considered as a large difference because rail transport usually travels at a constant speed. We will further discuss these findings in Sect. 5 to elaborate on their practical implications.

Average trip speed Figure 7 illustrates the impact of varying sampling intervals of GPS tracking data on the overall average speed of individual trip and disaggregated by different transport modes. Similar to Fig. 6, the results suggest that as the sampling interval increases, the overall median and mean speed values decrease. However, the data distribution of average speed of individual trip along with the varying temporal scales is more stable as compared to the point-to-point speed shown in Fig. 6. Especially for rail transport mode, the variance is very stable and even the mean and median do not change much as the sampling interval increases from 3 to 1800 s. This is reasonable because rail transport usually travels at a relatively constant speed and runs in a straighter path (designated tracks) compared to the other modes. Hence, the average speed of individual trip by rail transport is not very sensitive to the temporal scale. The results also suggest that at the higher the sampling intervals, the average speed of individual trips by auto, transit bus, walk, and bike are more underestimated. This is reasonable because as the sampling interval increases, the many details of movement paths are neglected, which systematically underestimates the actual trip distance traveled and thus results in a lower average speed of individual trips (see the example of Fig. 5). Similar to point-to-point speed for the transit mode in Fig. 6, the mean values of average trip speed for transit bus are substantially greater than the median values across different sampling rates.

The Kruskal–Wallis test results indicate significant difference among different sampling intervals for the whole sample ($p < 0.01$) as well as for the five subsets grouped by different transport modes (all p-values are below 0.01). We next apply Dunn's test to identify which specific means are significantly different from the others at various time scales. Considering the whole data set, the Dunn's test results indicate that the ten groups of different sampling intervals are significantly different from each other (all p-values are below 0.01). The Dunn's test applied to the subsets of auto and walk modes suggests that the mean values of speed of individual trips at the ten sampling intervals are significantly different from each other (all p-values
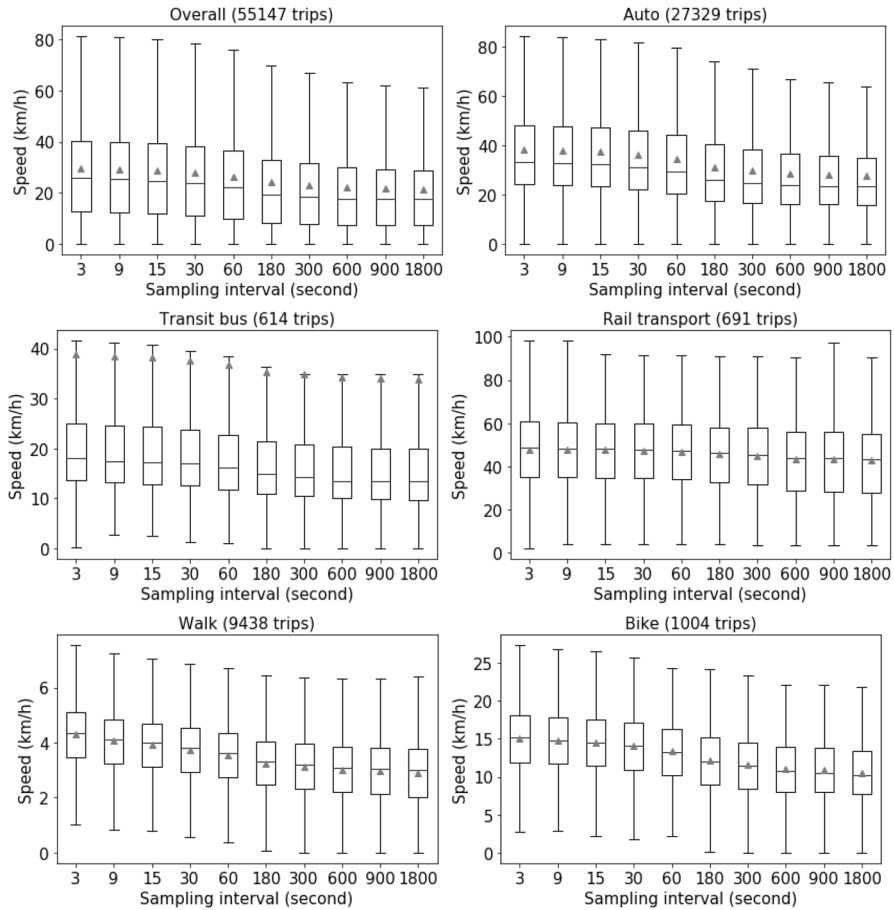
Fig. 7 Box plots of the average speed values for individual trips by different transport modes at different temporal scales

are below 0.01). These observations are consistent with the case of point-to-point speed by auto and walk modes. However, for transit bus mode, there is no significant difference ($p > 0.05$) between the groups of 3 s and 9 s, 9 s and 30 s, 30 s and 60 s, 300 s and 600 s, and 600 s and 1800 s. For rail transport, we find no significant difference ($p > 0.05$) between the group of 3 s and the other three groups of 9 s, 30 s, and 60 s, respectively. In addition, there is no significant difference ($p > 0.05$) between the groups of 9 s and 30 s as well as 60 s, and between the groups of 30 s and 60 s, 300 s and 600 s, 600 s and 1800 s. In terms of bike mode, we only find no significant difference between the groups of 3 s and 9 s and the other pairs of groups are significantly different from each other at 0.05 significance level.

4.1.2 Path tortuosity

The second movement parameter of interest is path tortuosity. Various measures have been developed to estimate tortuosity of the movement path of a moving object using trajectory data. Examples include but not limited to fractal dimension (Bovet and Benhamou 1988; Turchin 1998; Falconer 2004), sinuosity (Bovet and Benhamou 1988; Benhamou 2004), straightness index (Batschelet 1981), and angular variance of turning angles (Estevez and Christman 2006). Considering similarities across these measures, here we only consider *straightness index (SI)* to illustrate the influence of varying temporal scales on computing path tortuosity. SI of a trip (i.e., global path tortuosity) is defined as the ratio of the distance between origin and destination to the actual total distance traveled along the movement path as shown formally in Eq. (3).

$$SI = \frac{\sqrt{(x_n - x_0)^2 + (y_n - y_0)^2}}{\sum_{i=0}^{n-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}} \quad with\ SI \in [0, 1] \tag{3}$$

where (x_0, y_0) and (x_n, y_n) are the coordinates of the origin (first point) and destination (last point) of a trip, respectively. The SI value is 0 when the moving object returns to the origin location of a trip and is 1 when the movement path of a trip is completely straight.

SI can also be computed over a sliding window k. This is used as the local measure of path tortuosity of a portion of an individual trip as defined below (Dodge et al. 2009).

$$SI'(p, k) = \frac{\sqrt{(x_{p+k} - x_{p-k})^2 + (y_{p+k} - y_{p-k})^2}}{\sum_{i=p-k}^{p+k-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}} \quad with \quad SI'(p, k) \in [0, 1] \tag{4}$$

where $SI'(p, k)$ is the local SI of the p th point of an individual trajectory given a sliding window with a width k. We consider a sliding window of 5 points in the following experiment. The narrower the width of the sliding window k, the noisier are the results. For coarsely sampled data such as 900 s and 1800 s, it is very likely that many individual trips contain fewer than 5 tracking points. In such case, the local measure of path tortuosity will be invalid.

Global path tortuosity Figure 8 illustrates the influence of varying temporal scales on the values of global SI. It shows that as the sampling interval increases, the global SI is more overestimated and approaching 1 which represents a complete straight movement path. This is reasonable because increasing sampling interval neglects the details and complexity of the movement path, which can result in a straighter path compared to the actual path delineating by fine-grained data (see the example of Fig. 5). It is worth to note that when the sampling interval increases to 1800 s, only the origin and destination will be left for the majority of individual trips (93.06%). Therefore, the movement path is completely straight ($SI = 1$) at the sampling interval of 1800 s for the majority of trips. On the other hand, the mean value is consistently
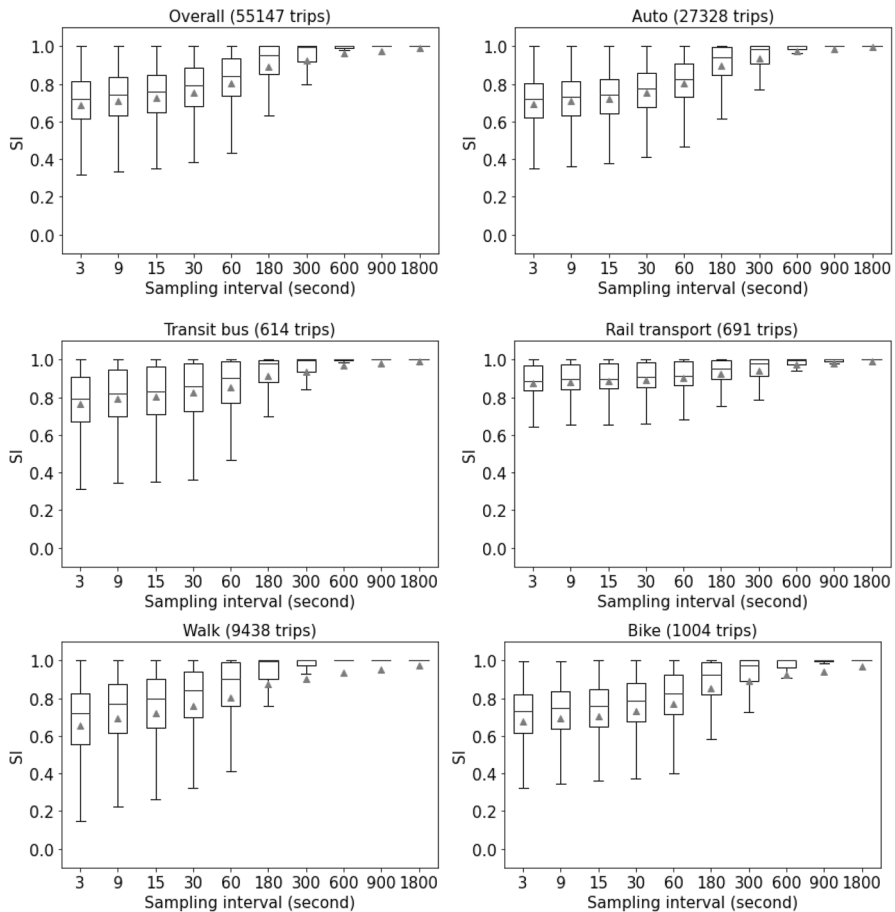
Fig. 8 Box plots of the global straightness index of individual trips by different transport modes at different temporal scales

below the median for every mode across different temporal scales. The skewness is negative for every group of the data indicating a negatively skewed distribution.

For relatively high-speed movement such as auto and transit bus modes, the results suggest that the movement path becomes straighter quickly along with the increasing sampling intervals. This is reasonable because a vehicle moving on road networks usually needs to make turns to reach the destination, and increasing sampling intervals will be very likely to ignore these turns which can result in a straighter movement path as compared to the actual traveled path. The distribution of the global *SI* values does not vary much as the sampling interval increases from 3 to 60 s for rail transport compared to the other modes. This observation indicates that the path tortuosity of rail transport is less impacted by varying scales, which is expected because railways run on fixed and relatively straight tracks with a constant speed as we observed above. For lower speed movement such as walking and biking,

we find the movement path becomes straighter quickly along with the increasing sampling intervals. This is reasonable because the trip length of biking/walking usually is shorter than vehicle-based trips and pedestrians and bicyclists are less constrained by road networks compared to vehicles (i.e., their movement is more flexible in terms of direction). Increasing sampling intervals is very likely to ignore wandering and turning for pedestrians and bicyclists. These findings are important for map matching applications which is the process of matching GPS tracking points to a real road network (segments). At intersections, if turns are missed due to a coarser temporal scale then the map matching can result in more erroneous matches. This can alter both movement speed and travel path.

The Kruskal–Wallis test results show significant differences in the mean values of the global *SI* at different sampling intervals (all *p*-values are below 0.01). The Dunn's test applied to the subsets of auto and railway transport modes suggests that the mean values of the global *SI* are only not significantly different between the groups of 3 s and 9 s, 9 s and 15 s, 15 s and 30 s, 30 s and 60 s ($p > 0.05$). For transit bus, only the groups of 9 s and 15 s are found not significant ($p > 0.05$). For bike mode, no significant difference ($p > 0.05$) is found between the groups of 3 s and 9 s, 9 s and 15 s. For walk mode, all groups are found significantly different from each other ($p < 0.05$).

Local path tortuosity Figure 9 illustrates the influence of varying temporal scales on the values of local SI over a sliding window of five tracking points. Individual trips that contain fewer than five tracking points are excluded in the results here. Generally, the values of local *SI* decrease as the sampling interval increases. This is as expected because when the temporal resolution is high, the movement over the sliding window of five tracking points only lasts for a few seconds (e.g., 12 s at 3 s temporal scale) and thus the path tends to be straight and smooth. However, at higher sampling intervals, the sliding window captures movement over a longer period of time, which can result in less straight movement path. It is important to note that the local SI is very similar to the global SI when the sampling interval reaches 300 s and above since the duration of 86.57% of trips is shorter than 25 min.

Regarding different travel modes, the results of auto and rail transport show that the local *SI* stays at a very high level close to one (meaning that the path is almost completely straight) at various sampling rates. This observation suggests that the movement by auto and rail transport over a segment of individual trips tends to be straight. This reflects the shape of roadway segments and especially railways that are very straight. Also, for coarser tracks, only the origin and the destination, and in some cases, a few more tracking points, remain and therefore, the local and global SI become the same. However, it is surprising that even when the sampling rate increases to 60 s per fix and above for auto trips, the local path tortuosity is still near 1. Presumably, the majority of drivers follow the shortest path or a straight highway when heading to the destination. In respect of rail transport, the local *SI* deviates slightly from 1 when sampling interval is increased to 30 s and above. This is because rail transport movement can be straight over a sliding window of five tracking points even when the sampling interval is large. The local path tortuosity of transit bus, walking, and biking is affected by temporal scale substantially. Transit buses
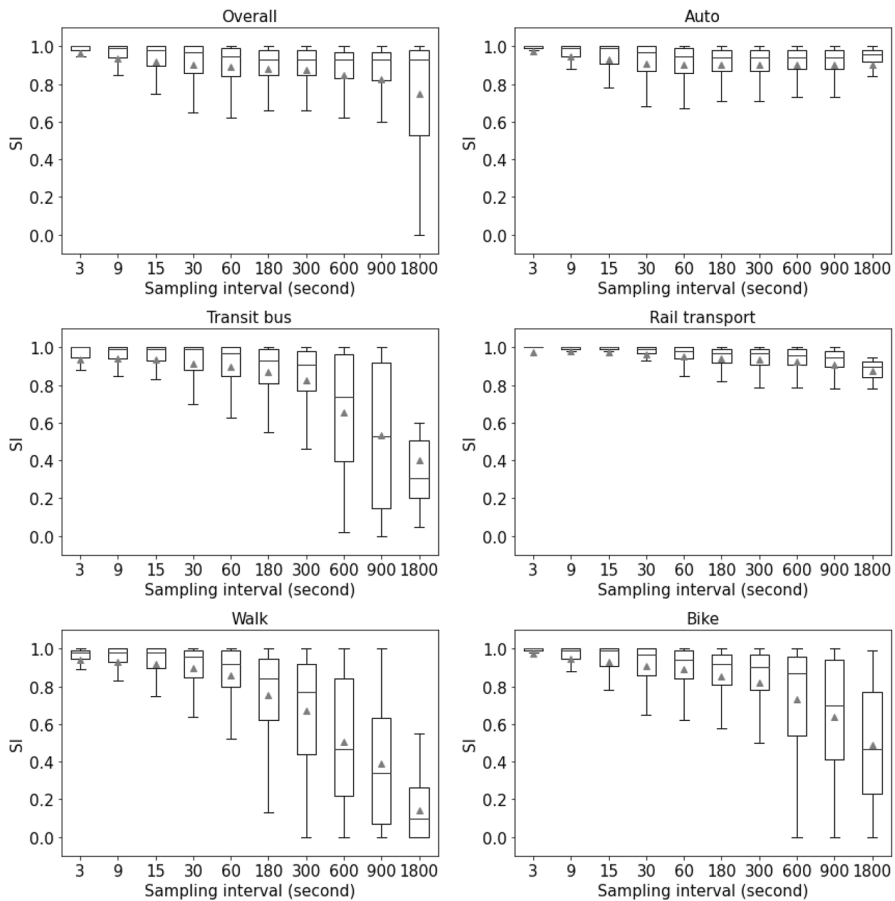
Fig. 9 Box plots of the local straightness index by different transport modes at different temporal scales

move slowly compared to autos and rail vehicles and are more likely to make stops during the movement. Therefore, the local SI can capture less straight movement over a sliding window of five tracking points as the sampling interval increases. Walking and biking are usually in lower speed and follow less constrained paths. The segment trajectory by walking and biking that consists of five tracking points can be less straight with coarser sampling.

The Kruskal–Wallis test results show significant differences in the mean values of the global *SI* at different sampling intervals (all *p*-values are below 0.01). The Dunn's test results suggest that there is significant difference ($p < 0.05$) in the mean values of the local SI among the groups of 3 s, 9 s, 15 s, 30 s, 60 s, and 180 s for the overall sample as well as the other subsets of each mode.

4.2 Space utilization

As discussed in the introduction, activity space estimation is a fundamental task prior to human interaction analysis. We consider activity space or space utilization as one of the key indicators for travel behavior that could potentially be impacted by different temporal scales. The objective of this experiment is to use three approaches to estimate human's space utilization during movement and to assess the impact of the temporal scale on activity space estimation. We first use three measures namely *minimum convex polygon* (MCP) (also called a convex hull), *95% Kernel Density Estimation (KDE) home ran*ge, and *radius of gyration* to measure individual activity space at the daily level (see the review in Sect. 2). These three indicators can capture the spatial range and spatial dispersion of human's daily activities. In addition, we apply a time-geography-based approach to estimate the accessible area at a finer scale along the movement path based on consecutive GPS tracking points. The radius of gyration is formulated as follows.

$$R_g = \sqrt{\frac{\sum_{i=0}^{n}[(x_i - \bar{x})^2 + (y_i - \bar{y})^2]}{n}} \tag{5}$$

where $(\bar{x}, \bar{y})$ represents the center of mass of the trajectory, specifically, $\bar{x} = \sum_{i=0}^{n} x_i/n, \bar{y} = \sum_{i=0}^{n} y_i/n$. The greater the value of R_g, the larger the activity space of an individual.

Figure 10 illustrates the difference between the MCP and the radius of gyration. If only a minimum convex polygon is applied, the activity space of person #1 and person #2 will be exactly the same, and it cannot capture the difference in the spatial distributions of their tracking points. However, person #1 has a larger radius
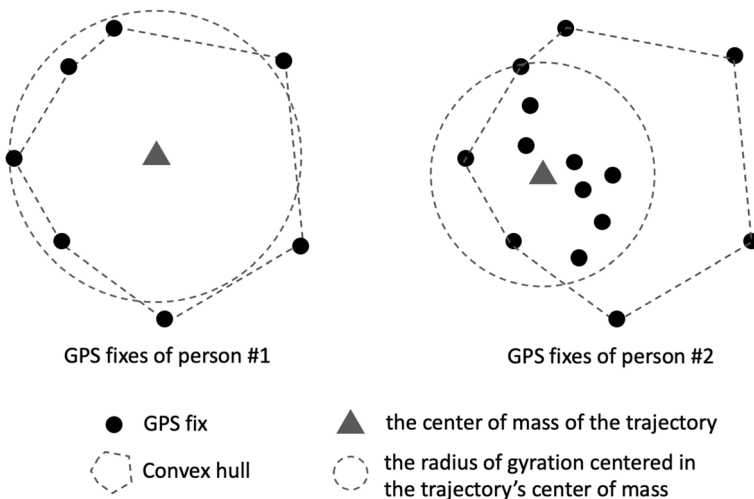

GPS fixes of person #1   GPS fixes of person #2

● GPS fix   ▲ the center of mass of the trajectory

⬡ Convex hull   ⬡ the radius of gyration centered in the trajectory's center of mass

Fig. 10 Illustration of minimum convex polygon and radius of gyration

of gyration than person #2, which indicates that person #1 has a more dispersed activity space while person #2's activities are more concentrated. Here, we use both measures to delineate the individual daily activity space. Figure 11(a) shows the example of the MCP and the radius of gyration for a moving individual using a real track from CHTS. The area of the MCP is 0.042 km^2 and the radius of gyration is 6.79 km.

To approximate individual space utilization using KDE, the bivariate normal kernel function is applied:

(a)

(b)

Fig. 11 Illustration of the minimum convex polygon and the radius of gyration (**a**) and the 95% Kernel Density Estimation home range (**b**) of an individual daily trajectory collected at 3 s per fix. In figure (**a**), the trajectory is shown in black. The green triangle is the center of mass of the trajectory. The red polygon depicts the minimum convex polygon, and the blue circle delineates the radius of gyration. In figure (**b**), the bold line in black represents the same trajectory

$$K(\text{x}) = \frac{1}{2\pi}\exp\left(-\frac{1}{2}\text{x}^t\text{x}\right) \tag{6}$$

where $K(\text{x})$ is the bivariate normal kernel function and x is a vector containing the coordinates of a point on the plane. Subsequently, the kernel density estimation of the utilization distribution at a given point x of the plane can be obtained by:

$$\hat{f}(\text{x}) = \frac{1}{nh^2}\sum_{i=1}^{n}K\left\{\frac{1}{h}(\text{x} - \mathbf{X_i})\right\} \tag{7}$$

where h is a smoothing parameter (i.e., the bandwidth), n is the number of relocations (i.e., data points), and $\mathbf{X_i}$ is the ith relocation of the sample. We employ one of the most common approaches, the reference bandwidth (Silverman, 1986), as defined below:

$$h = \frac{1}{2}(\sigma_x + \sigma_y)n^{-1/6} \tag{8}$$

where σ_x and σ_y are the standard deviations of the x and y coordinates of the relocations, respectively.

The result of KDE is a continuous density surface reflecting different probabilities of visiting in each location. The area of KDE can be measured by 100% contour of the density surface (i.e., include all nonzero activity density area) or any other specified level (e.g., 95% contour) in order to exclude areas that may consist of infrequent activities. The derived area is considered as the individual activity space. Figure 11(b) shows the example of the 95% KDE area for the same moving individual in Fig. 11(a). For implementation of KDE, readers may refer to the documentation of the R package "adehabitatHR" developed by Calenge (2006).

To estimate individual daily activity space, the minimum convex polygon, the radius of gyration, and the 95% KDE home range are computed using the individual daily records as the unit of analysis. As mentioned in Sect. 3, the GPS component of CHTS was collected over a three-day period. In total we have 9349 individual daily tracking records. It is important to note that the sample size for the experiment here is lower than the original sample size (9349) due to three reasons: (1) we exclude individuals who travel outside California; (2) the convex hull approach is invalid when an individual has only two tracking points in their daily diary (i.e., at least three points are required to construct a valid convex hull); and (3) the KDE approach requires more than five tracking points for an individual's daily movement. Hence, it is invalid to use the convex hull approach for people who only made one outward trip on the survey day (e.g., left home and did not come back) when the sampling interval is 1800 s. Similarly, the KDE approach may also be invalid when none to very few tracking points between the origin and destination were left as the sampling rate becomes coarser.

Figure 12 shows the impact of varying temporal scales on computing the minimum convex polygon, the radius of gyration, and the area of the 95% KDE. The results in Fig. 12(a) suggest that the area of minimum convex polygon remains very stable across different sampling rates, presumably because we retain all origin and
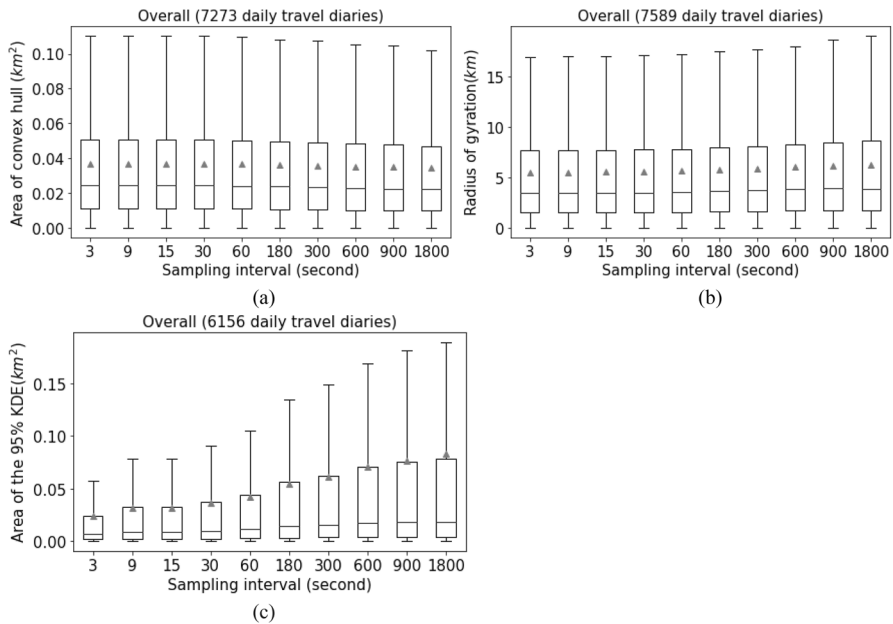
Fig. 12 Box plots of the area of minimum convex polygon (**a**) and the radius of gyration (**b**) and the area of the 95% KDE (**c**) of individual daily movement at different temporal scales

designation locations in the process of down-sampling. Given these are major locations where individual daily activities happened, it is reasonable that the estimated minimum convex polygons for the same person under different sampling intervals remain similar. The Dunn's test results indicate no significant difference among the groups of 3 s, 9 s, 15 s, 30 s, 60 s, and 180 s ($p > 0.05$). In terms of the radius of gyration, no significant difference ($p > 0.05$) among the groups of 3 s, 9 s, 15 s, 30 s, and 60 s is found. In contrast to the stable trend found with the minimum convex polygon, there is a slightly increasing trend of the radius of gyration when the temporal scale becomes coarser (Fig. 12(b)). Specifically, the mean value of radius of gyration increases from 5.50 km (at 3 s) to 6.26 km (at 1800 s). Also, the median value increases from 3.46 to 3.85 km. This observation indicates that at the higher sampling intervals, the spatial dispersion of human's activity space is overestimated. In terms of the area of the 95% KDE, Fig. 12(c) shows that when the temporal scale becomes coarser, the area of individual activity space estimated by KDE increases rapidly. Presumably, individual's activity locations become more sparsely distributed in space when the temporal scale becomes coarser, which results in larger 95% KDE surface. The Dunn's test results indicate only no significant difference among the groups of 600 s, 900 s, and 1800 s ($p > 0.05$). In addition to the above analysis on individual daily activity space, we also experimented on three-day aggregation of individual activity space. However, the results are similar to daily aggregation of individual activity space and hence are not presented here.

The previous analysis examined the impact of varying sampling intervals on approximating human activity space at an aggregate daily level (i.e., one activity

space per day per individual). Next, we apply a time-geography approach to estimate the activity space (accessible areas) during movement at a finer scale (i.e., between consecutive tracking points per trip). Figure 13 illustrates an example of the PPAs for a moving individual that was tracked at 60 s per fix. A PPA is shaped by a pair of consecutive tracking points, a time budget (i.e., sampling interval in this case), and the maximum speed capacity during a specific time interval which is estimated by a floating average of speed over an exponential kernel. For mathematical definitions and implementation of PPA, the reader can refer to Miller (2005) and Dodge et al. (2021).

Figure 14 illustrates the impact of varying temporal scale on computing the area of PPA. It shows that overall, using a coarser sampling rate the data generates bigger PPAs in spite of different transport modes people use. This is mainly because a larger sampling interval allows more time budget (equals to the sampling interval between two consecutive fixes) which is a critical parameter to build a PPA. In addition, the magnitude of the area of PPA varies largely among different modes as the maximum speed capacity for each mode is substantially different. Specifically, people using auto, transit bus, and rail transport modes are more likely to have larger size of PPA. The rail transport mode has the largest size of PPA because it usually runs on designated tracks with a constant speed and less likely to be impacted by traffic. However, this does not necessarily mean that the individuals traveling by rail have access to all locations captured in the PPAs due to the restricted nature of rail transport, but in general they can cover longer distances given a larger time budget. One might need to refine the PPA approach that accounts for the rail networks or resort to different approaches to approximate the accessible locations of mobile individuals if they travel by rail (e.g., develop station-based and schedule constrained PPAs). People using walk and bike modes usually generate a smaller size of PPA
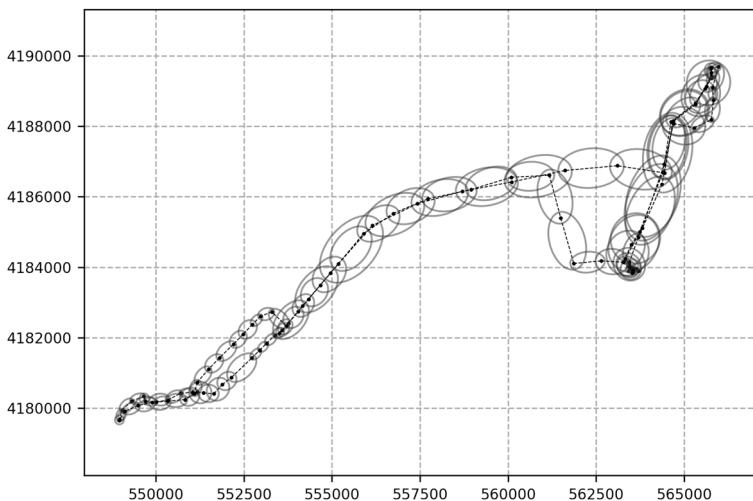

Fig. 13 Illustration of PPAs along the movement path of the same individual in Fig. 11 collected at 60 s per fix
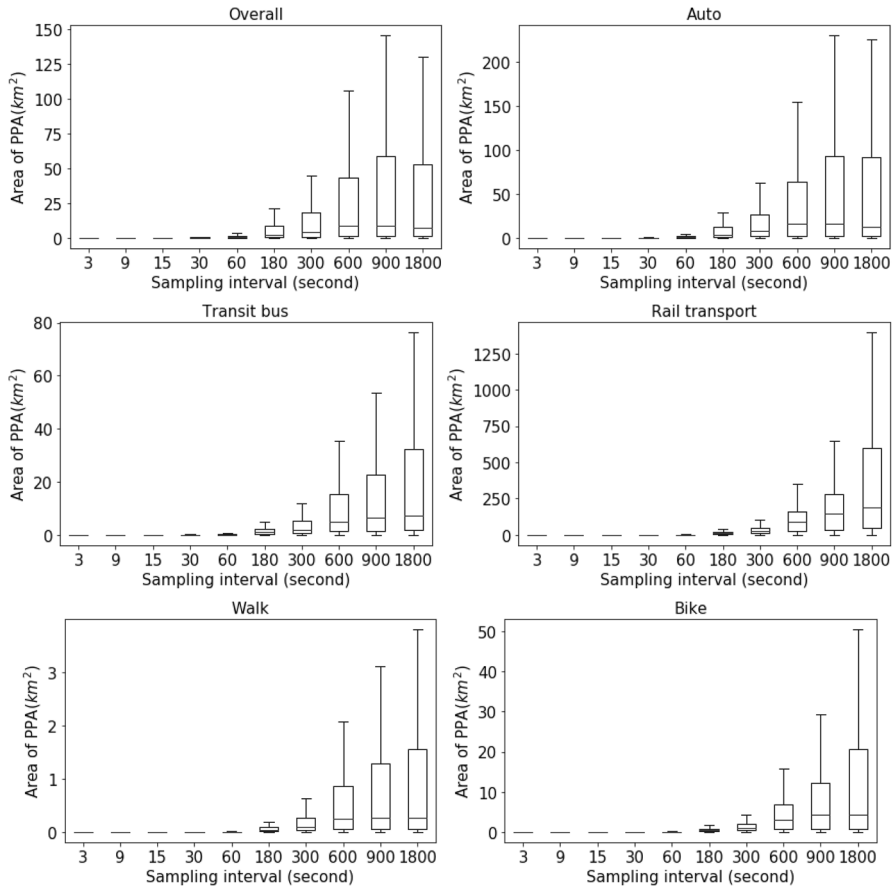
Fig. 14 Box plots of the PPA areas of consecutive GPS fixes by different transport modes at different temporal scales

because the maximum speed by walk and bike is relatively lower compared to the other modes. The Dunn's test results of auto indicate that the mean values of the area of PPA are significantly different ($p < 0.05$) from each other when the sampling interval is within the range of 3 s and 600 s. As the sampling rate becomes coarser (600 s, 900 s, 1800 s), no significant difference is found ($p > 0.05$). For transit bus, railway transport, and bike modes, the mean values are significantly different ($p < 0.05$) from each other when the sampling interval is within the range of 3 s and 180 s. As the sampling interval increases from 180 s, most of the groups are found not significantly different ($p > 0.05$). In terms of the walk mode, we find each group is significantly different ($p < 0.05$) from each other when the sampling interval is within the range of 3 s and 600 s. In general, as the temporal scale becomes very high, no significant difference in the mean values of the area of PPA can be identified. This is because many trips would not last longer than 900 s (15 min). Therefore, only the origin and destination will be retained for these short trips in the

process of down-sampling. Eventually, the estimated PPAs of these short trips will be the same at very high temporal scales when the sampling interval exceeds the trip duration.

4.3 Human interaction analysis

The goal of this experiment is to understand the impact of temporal scale on identifying potential concurrent and delayed interactions between moving individuals. Two different approaches are applied to identify potential interactions: (1) a proximity-based approach which uses a spatial buffer and a temporal window to identify space–time contacts between mobile individuals, and (2) a time-geographic-based approach named ORTEGA, proposed in Dodge et al. (2021), which uses the PPA to find potential concurrent and delayed contacts between mobile individuals. It is important to note that the identified contacts by both approaches can only be considered as potential interactions, and individuals may or may not socially or physically interact when they come into close contacts spatially and temporally. Guided by the Dodge et al. (2021) study, the proximity-based approach is implemented by intersecting spatial buffers of a 100-m distance threshold centered on synchronous GPS tracking points of two individuals. To relax the restriction of requiring synchronous fixes when identifying concurrent interactions, a time window of 5 min is allowed when determining if two spatial buffers are intersected. For the PPA-based approach, two individuals are considered to have a potential interaction if their PPAs are intersected. Even though the PPA-based approach does not require a predefined distance buffer or a set time window we allow the same 5-min time lag for intersection in this experiment to make it comparable with the proximity-based approach. That is, if the time intervals of the two intersecting PPAs of the two individuals are overlapped synchronously or have a 5-min time lag, we identify it as a potential interaction (i.e., concurrent interactions). If the PPAs of the two individuals are intersected but the time lag is longer than 5 min, we consider this as a delayed interaction. The frequency of interactions among a group of people is then quantified by the count of pairs of individuals that have interacted with each other during the survey day for both the proximity-based and PPA-based approaches. The outcomes of these two approaches at different temporal scales are then compared as described below.

Considering that people living far away are less likely to interact, instead of taking the whole sample of CHTS, we only use the portion of the data that is collected in Santa Clara county, where we have the largest number of respondents as a case study. This portion of the data contains GPS tracking of 850 persons from 380 households that were collected from February 3, 2012 to January 31, 2013. Given the information of household composition, we separate the detected interactions by interactions that happened among persons in the same household and interactions that happened with persons that are not from the same household to incorporate the social relationship factors in human interaction analysis. Individuals within the same household tend to travel together and interact more, while individuals from different households might have occasional social interactions and various random contacts.

Figure 15 illustrates the impact of temporal scale on identifying concurrent interactions using the PPA-based approach versus the proximity-based approach between individuals of the same household and different households. In general, both approaches are sensitive to the temporal scale. The proximity-based approach can identify a greater number of contacts within household interactions when the temporal resolution is 1 min as opposed to coarser sampling rates beyond 1 min. Specifically, the total number of within household interactions identified is 367 at 1-min interval, whereas this number drops to 361 at 5-min interval, and to 359 at 10-min interval and remain unchanged at 20-min and 30-min intervals. Likewise, the proximity-based approach can identify 73 outside household interactions at 1-min interval while substantially fewer number of outside household interactions are identified when sampling interval increases (i.e., 6, 2, 2, 2 outside household interactions are identified at 5 min, 10 min, 20 min, 30 min interval, respectively). On the contrary, the PPA-based approach can identify substantially more outside household concurrent interactions when coarse temporal scales are used. This is mainly because using larger sampling intervals results in bigger PPAs which are more likely to intersect with each other. But the total number of within household interactions remains stable as the sampling interval increases. People living in the same household are more likely to travel together during a day (e.g., parents escort children to school) (Lee and Goulias 2018), which increases the chance of physical contacts with each other. Even when the sampling interval is increased, these people are still very likely to be tracked simultaneously. Hence, the PPA-based approach and the proximity-based can identify very closely the number of within household interactions and the numbers remain stable along with the increasing sampling intervals. Unlike people living in the same household, people from different households are less likely to travel together and their physical interactions are often unintentional (e.g., two strangers pass by each other on the street) and asynchronous (e.g., two persons visit the same grocery store at different time). The proximity-based approach is less effective in this situation because it requires a synchronous and predefined distance buffer threshold
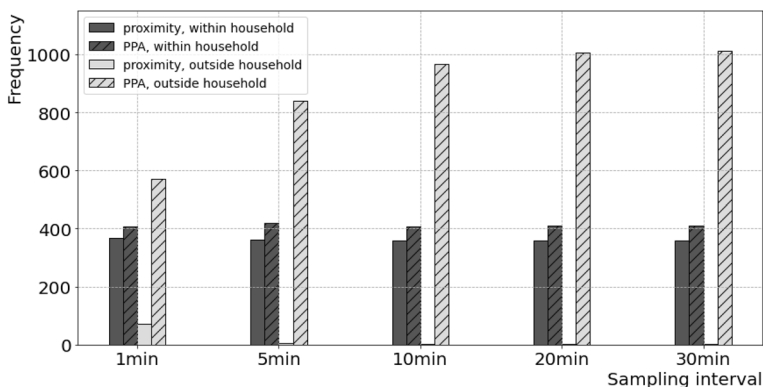

Fig. 15 Illustration of the impact of temporal scale on identifying concurrent interactions (allowing 5 min time lag) using the PPA-based approach versus the proximity-based approach (100 m buffer) between individuals of the same household and different households

which is set by the user (such as 100 m in this case) to determine if two moving entities come close to each other. However, the PPA-based approach does not have such constraint as it is based on the actual locations accessible to people, their time budget and speed capacities. All of these parameters can be extracted from the given data without a need for a user defined threshold. Therefore, we observe more outside household concurrent interactions identified by the PPA-based approach as opposed to the proximity-based approach. As shown in Dodge et al. (2021), the outcomes of the proximity-based approach is also sensitive to the buffer size. Specifically, they find that by applying 500 m distance buffer, the proximity-based approach can identify almost the same number of concurrent intrahousehold interactions but interhousehold interactions are still underestimated. Employing a distance buffer larger than 500 m may not be reasonable as the potential concurrent interactions are defined as close contact between moving individuals in this study and individuals stay away from each other more than 500 m should not be considered as potential concurrent interaction.

We next focus on the impact of temporal scale on identifying delayed interactions using the PPA-based approach. A delayed interaction is context-specific and a meaningful temporal delay can vary in different applications. For example, in the context of COVID-19 transmission, interactions with a delay of up to 30 min may be considered critical or risky contacts as the droplets may stay in the air for a considerable amount of time. As shown in Fig. 16, allowing a longer time lag can identify more delayed interactions within households as well as outside households. This is reasonable because allowing a longer time lag increases the chance of intersection between two PPAs. However, the increase in the amount of identified delayed interactions outside households is substantially higher than the interactions within households. This is mainly because people living in the same household are more likely to interact synchronously. Therefore, allowing longer time lag does not help much with identifying delayed interactions within households. Nevertheless, people from different households are more likely to have asynchronously interactions, resulting in more delayed interactions when allowing longer time lags. In addition, using different sampling intervals of the data seems to have no significant impact on identifying delayed interactions between individuals of the same household, whereas more delayed interactions between individuals of different households can be identified if a coarser sampling rate of data are used. This is similar to our previous findings regarding concurrent interactions. Presumably, people living in the same household are more likely to travel together and increasing sampling intervals results in larger PPAs but does not impact very much the chances of intersecting between PPAs. On the other hand, larger PPAs of individuals from different households as a result of a coarser time scale increases the probability of intersecting PPAs.

5 Discussion

The presented work investigates the question: "how do decisions surrounding the scale of movement data and analyses impact our inferences about movement patterns?". The outcomes reveal some practical implications of the impact of temporal
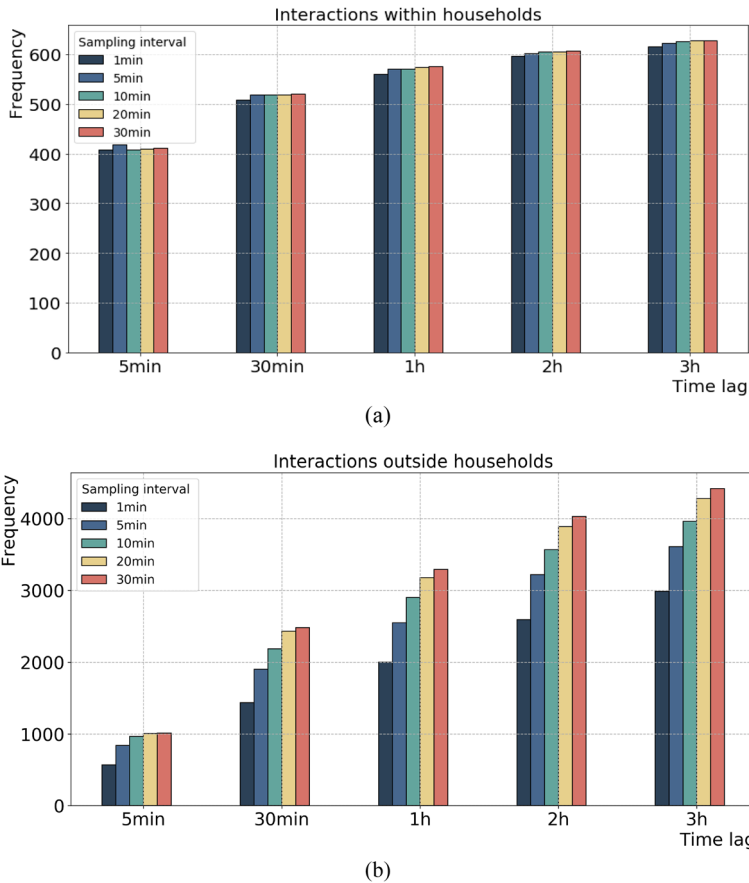
Fig. 16 Illustration of the impact of temporal scale on identifying delayed interactions using PPA-based approach between individuals of the same household (**a**) and different households (**b**)

scale on movement analysis results. Here we discuss main findings in the light of our research question.

The outcomes suggest that increasing sampling intervals (even just from 3 to 9 s) can significantly underestimate the point-to-point speed when people travel by auto, walk, and bike. However, the difference of the mean speed of auto mode between 3 and 15 s is only 1.93 km/h which should not be considered as a large difference in the transportation context. On the other hand, when the sampling interval increases to 60 s, the mean speed decreases by 7.29 km/h (4.53 MPH). This is very important for some practical applications such as mapping motorist safety by identifying road segments where motorists frequently exceed the speed limit or a safe speed based on the historical tracking data of individuals. Using coarse tracking data (sampling interval equals to or greater than 60 s) will very likely miss many speed limit violations. In addition, the variation in computed speed for walk and bike modes across varying temporal scales is evident. This is important as selecting appropriate

temporal resolution data plays a critical role in mapping bicycling or pedestrian exposure and safety risk (Ferster et al. 2021). For high-speed movement such as auto and transit modes and low-speed movement such as walking and biking, we find the global movement path becomes straighter quickly along with coarser sampling. But for rail transport, the global path tortuosity is less impacted by the temporal scale. Regarding the local path tortuosity, auto and rail transport are found less impacted by the temporal scale while the segment of individual trips by transit, walking, or biking becomes less straight with coarser sampling. The results of the present study may suggest that analysis using coarser tracking data such as call detail records and geotagged social media check-in data, or aggregate mobility indices (e.g., average distance traveled from home, vehicle miles traveled, average trip speed), derived from tracking data of cellphone users, can be impacted by varying temporal scales. These indices have become ubiquitous in studying the impact of the COVID-19 pandemic in urban areas (Oliver et al. 2020).

The impact of different temporal scales is exacerbated when different modes such as walking, bicycling, ride hailing, and public transport modes are combined. This is the situation with integration of multiple modes into one platform as in Mobility as a Service (MaaS) (Goodall et al. 2017). If different applications and users employ different scales on their respective devices, the expected arrival time communicated to a user about each vehicle will be wrong and the impact of this can span from unreliable level of service to security and safety issues. Ideally, we want to identify one temporal scale that works well with the users of transportation services and provides the most reliable arrival times of any service. Our findings here indicate that for coordinating multiple modes, the 3 s recording is the best approach. However, we cannot tell if recording location more often will yield any additional benefits for MaaS users due to data limitation.

Our research findings suggest that the impact of temporal scale on activity space estimation highly depends on the method being used. In general, we find that the MCP and the radius of gyration are not affected by temporal scale significantly as long as the origin and destination locations (i.e., major activity locations) are recorded and retained in the movement data. Using a sampling rate of up to one min per fix should be sufficient for approximating activity space using these two measures. However, the results of KDE show that the estimated activity space is substantially affected by temporal scale. Using a coarse sampling data scheme may overestimate the activity space of individual by KDE approaches. This implies a need to develop new multi-scale approaches that can work across scales for quantifying activity space (Miller et al. 2019). For example, we may need to use a different scale to identify potential destination locations to build choice sets (Yoon et al. 2012) requiring detailed pinpointing of business establishments versus developing built environment indicators of accessibility to test their correlation with activity and travel behavior in a day (Chen et al. 2011; Yoon and Goulias 2010) that do not require pinpointing of exact opportunity locations.

Our analysis focusing on human interaction analysis suggests that both the proximity-based approach and PPA-based approach are sensitive to the temporal scale of the input data, specifically for concurrent interactions. This can be important when interaction analysis is used for public health applications such as contact tracing and

identifying critical contacts. In particular, the concurrent interactions between individuals identified by the proximity-based approach can substantially be underestimated at coarser sampling intervals. In contrast, the PPA-based approach can identify substantially more concurrent interactions outside households when a coarser temporal scale is used. But the number of within household interactions remains stable as the sampling interval increases. In addition, using data of different sampling intervals seems to have no significant impact on identifying delayed interactions between individuals of the same household, whereas more delayed interactions between individuals of different households can be identified with a coarser temporal scale. The outcomes suggest that movement data of fine-grained temporal granularity such as one min per fix is required for a more reliable interaction analysis (see the interpretation in Sect. 4.3). Coarser data might overestimate the number of contacts. It is important to note that the approaches used in this study for human interaction analysis (both PPA-based and proximity-based methods) are limited to identifying encounters or co-location as potential cases for interaction between individuals and therefore does not indicate actual occurrence of social interaction. Nevertheless, an underlying requirement for a physical social interaction is encounter. A promising future work of human interaction analysis is to take into account contextual factors (e.g., movement mode, behavior, geographic, and environmental variables) in order to better distinguish and contextualize co-location behavior and actual interaction between moving individuals.

The size of a PPA is particularly important for activity-based travel demand forecasting and simulation that include a destination choice component (Bhat et al. 2013). In these models the daily activity-travel pattern of a person is simulated minute-by-minute and includes predictions of destination locations for activity participation (e.g., eat meal at a restaurant). Destinations selected by a person are from a PPA created from the available time windows identified in the daily schedule simulation. Sets of destinations (called the destination choice set and is the number of alternatives considered for an activity location) from which an individual selects one, are enumerated from a PPA (Yoon et al. 2012). Using the choice set for this time window the simulation uses discrete choice models (i.e., models of the probability of selecting a destination, (Train 2009)). These models are estimated before the simulation using observed destinations in a survey and their identification as suitable alternatives is based on PPAs too. If the destination of a choice set is underestimated or overestimated, as Thill (1992) shows, the parameter estimates and the computation of choice probabilities are biased and therefore the simulated choices can be heavily biased. When destination choices are the outcome of joint scheduling of activities by multiple persons this bias may worsen. The findings here suggest we need to perform experiments computing choice sets from different types of PPAs and time scales and then estimate discrete choice models to identify the best fitting models to observed behavior.

6 Conclusion

With this research, we discuss the impact of temporal scale on human movement analytics focusing on computing movement parameters including speed and path tortuosity, estimation of individual space usage, and human interaction analysis.

We find both the point-to-point speed and the average speed of an individual trip drop with coarser sampling. Varying temporal scale also impact the outcomes of local and global straightness of the movement path and as a result changes the overall trajectory shape and lengths, which can impact other derivative measures obtained from these fundamental movement parameters. In terms of activity space, the area of minimum convex polygon remains stable across different sampling rates. In contrast, there is a steadily increasing trend of the radius of gyration with a coarser sampling rate. These observations indicate that the higher the sampling intervals, the more overestimated the spatial dispersion of human's activity space will be. The 95% KDE of individual activity space also presents an increasing trend when sampling rate becomes coarser. In addition, overall, using a coarser data generates bigger PPAs in spite of different transport modes people used. The magnitude of the area of PPA varies largely among different modes as the maximum speed capacity for each mode is substantially different. As the size of the PPA increases with coarser data, the PPA-approaches can result in an overestimation of individual encounters when used for interaction analysis.

This research, however, has also some limitations. First, in the investigation of scale impact on movement speed, we only considered the factor of different transportation infrastructure by separating trips by modes but did not take into account the shape of the underlying street network. Although the high-resolution tracking data used in this study to a great extent correctly captures the shape of the network. Future research on this topic can consider the factor of various network shapes (e.g., distinguish data in a grid vs a cul-de-sac urban design). Second, even though we were able to distinguish the scale impact on different modes by annotating information on the transport mode used for each individual trip based on the travel survey, a more holistic investigation of scale impact on mode detection tasks using GPS tracking data is needed. Further analysis should consider inferencing the transportation mode for segments of a trip (e.g., implementing mode detection algorithms) and then study the parameters we explored here. This will provide more precise human interaction analysis (e.g., to know if an interaction happened for a portion of a trip during which individuals used the same mode and vehicle).

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethics approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

References

Aigner W, Miksch S, Müller W, Schumann H, Tominski C (2007) Visualizing time-oriented data—A systematic view. Comput Graph 31:401–409. https://doi.org/10.1016/j.cag.2007.01.030

Barbosa H, Barthelemy M, Ghoshal G, James CR, Lenormand M, Louail T, Menezes R, Ramasco JJ, Simini F, Tomasini M (2018) Human mobility: Models and applications. Phys Rep 734:1–74. https://doi.org/10.1016/j.physrep.2018.01.001

Batschelet E (1981) Circular statistics in biology. Taylor & Francis, London

Batty M (2012) Smart Cities, Big Data. Environ Plann B Plann Des 39:191–193. https://doi.org/10.1068/b3902ed

Benhamou S (2004) How to reliably estimate the tortuosity of an animal's path: straightness, sinuosity, or fractal dimension? J Theor Biol 229:209–220

Bhat CR, Goulias KG, Pendyala RM, Paleti R, Sidharthan R, Schmitt L, Hu H-H (2013) A household-level activity pattern generation model with an application for Southern California. Transportation 40:1063–1086. https://doi.org/10.1007/s11116-013-9452-y

Bovet P, Benhamou S (1988) Spatial analysis of animals' movements using a correlated random walk model. J Theor Biol 131:419–433

Brakewood C, Watkins K (2019) A literature review of the passenger benefits of real-time transit information. Transp Rev 39:327–356

Buchin M, Driemel A, Van Kreveld M, Sacristán V (2011) Segmenting trajectories: A framework and algorithms using spatiotemporal criteria. J Spatial Infor Sci 31(3):33–63

Buliung RN, Kanaroglou PS (2006a) A GIS toolkit for exploring geographies of household activity/travel behavior. J Transp Geogr 14:35–51

Buliung RN, Kanaroglou PS (2006b) Urban form and household activity-travel behavior. Growth Chang 37:172–199

Bullard F (1991) Estimating the home range of an animal: a Brownian bridge approach. University of North Carolina, Chapel Hill

Burkhard O, Becker H, Weibel R, Axhausen KW (2020) On the requirements on spatial accuracy and sampling rate for transport mode detection in view of a shift to passive signalling data. Transportat Res Part c Emerg Tech 114:99–117

Burns LD (1979) Transportation, temporal and spatial components of accessibility. Lexington Books, Lexington

Byrne ME, Clint McCoy J, Hinton JW, Chamberlain MJ, Collier BA (2014) Using dynamic brownian bridge movement modelling to measure temporal patterns of habitat selection. J Anim Ecol 83:1234–1243. https://doi.org/10.1111/1365-2656.12205

Calenge C (2006) The package adehabitat for the R software: tool for the analysis of space and habitat use by animals. Ecol Model 197:1035

Cencetti G, Santin G, Longa A, Pigani E, Barrat A, Cattuto C, Lehmann S, Salathe M, Lepri B (2021) Digital proximity tracing on empirical contact networks for pandemic control. Nat Commun 12:1–12

Chen Y, Ravulaparthy S, Deutsch K, Dalal P, Yoon SY, Lei T, Goulias KG, Pendyala RM, Bhat CR, Hu H-H (2011) Development of indicators of opportunity-based accessibility. Transp Res Rec 2255:58–68. https://doi.org/10.3141/2255-07

Chen BY, Yuan H, Li Q, Lam WHK, Shaw S-L, Yan K (2014) Map-matching algorithm for large-scale low-frequency floating car data. Int J Geogr Inf Sci 28:22–38. https://doi.org/10.1080/13658816.2013.816427

Ciscal-Terry W, Dell'Amico M, Hadjidimitriou NS, Iori M (2016) An analysis of drivers route choice behaviour using GPS data and optimal alternatives. J Transp Geogr 51:119–129

Deng M, Huang J, Zhang Y, Liu H, Tang L, Tang J, Yang X (2018) Generating urban road intersection models from low-frequency GPS trajectory data. Int J Geogr Inf Sci 32:2337–2361

Dodge S, Weibel R, Lautenschütz A-K (2008) Towards a taxonomy of movement patterns. Inf vis 7:240–252. https://doi.org/10.1057/PALGRAVE.IVS.9500182

Dodge S, Weibel R, Forootan E (2009) Revealing the physics of movement: COMPARING the similarity of movement characteristics of different types of moving objects. Comput Environ Urban Syst 33:419–434. https://doi.org/10.1016/j.compenvurbsys.2009.07.008

Dodge S, Su R, Johnson J, Simcharoen A, Goulias K, Smith JLD, Ahearn SC (2021) ORTEGA: an object-oriented time-geographic analytical approach to trace space-time contact patterns in movement data. Comput Environ Urban Syst 88:101630. https://doi.org/10.1016/j.compenvurbsys.2021.101630

Downs JA, Horner MW (2012) Analysing infrequently sampled animal tracking data by incorporating generalized movement trajectories with kernel density estimation. Comput Environ Urban Syst 36:302–310. https://doi.org/10.1016/j.compenvurbsys.2011.12.001

Dunn OJ (1964) Multiple comparisons using rank sums. Technometrics 6:241–252

Estevez I, Christman MC (2006) Analysis of the movement and use of space of animals in confinement: The effect of sampling effort. Appl Anim Behav Sci 97:221–240

Falconer K (2004) Fractal geometry: mathematical foundations and applications. Wiley, New Jersey

Fang H, Wang L, Yang Y (2020) Human mobility restrictions and the spread of the novel coronavirus (2019-ncov) in china. J Public Econ 191:104272

Ferster C, Nelson T, Laberee K, Winters M (2021) Mapping bicycling exposure and safety risk using Strava Metro. Appl Geogr 127:102388. https://doi.org/10.1016/j.apgeog.2021.102388

Fillekes MP, Giannouli E, Kim E-K, Zijlstra W, Weibel R (2019) Towards a comprehensive set of GPS-based indicators reflecting the multidimensional nature of daily mobility for applications in health and aging research. Int J Health Geogr 18:1–20

Golledge RG, Stimson RJ (1997) Spatial behavior: a geographic perspective. The Guilford Press, New York

González MC, Hidalgo CA, Barabási A-L (2008) Understanding individual human mobility patterns. Nature 453:779–782. https://doi.org/10.1038/nature06958

Goodall W, Dovey T, Bornstein J, Bonthron B (2017) The rise of mobility as a service. Deloitte Rev 20:112–129

Goulias K, Janelle D (2005) GPS Tracking and time geography: applications for activity modeling and microsimulation, position papers. FHWA-sponsored peer exchange and CSISS specialist meeting

Hägerstrand T (1970) What about people in regional science?. In: Papers of the regional science association

Hashemi M, Karimi HA (2014) A critical review of real-time map-matching algorithms: current issues and future directions. Comput Environ Urban Syst 48:153–165. https://doi.org/10.1016/j.compenvurbsys.2014.07.009

Hemson G, Johnson P, South A, Kenward R, Ripley R, Macdonald D (2005) Are kernels the mustard? Data from global positioning system (GPS) collars suggests problems for kernel home-range analyses with least-squares cross-validation: Kernel home-range analyses and LSCV problems. J Anim Ecol 74:455–463. https://doi.org/10.1111/j.1365-2656.2005.00944.x

Hirsch JA, Winters M, Clarke P, McKay H (2014) Generating GPS activity spaces that shed light upon the mobility habits of older adults: a descriptive analysis. Int J Health Geogr 13:1–14

Hood J, Sall E, Charlton B (2011) A GPS-based bicycle route choice model for San Francisco, California. Transportat Lett 3:63–75

Hoover BA, Miller JA, Long J (2020) Mapping areas of asynchronous-temporal interaction in animal-telemetry data. Trans GIS 24:573–586. https://doi.org/10.1111/tgis.12622

Horne JS, Garton EO (2006) Likelihood cross-validation versus least squares cross-validation for choosing the smoothing parameter in kernel home-range analysis. J Wildl Manag 70:641–648. https://doi.org/10.2193/0022-541X(2006)70[641:LCVLSC]2.0.CO;2

Horne JS, Garton EO, Krone SM, Lewis JS (2007) Analyzing animal movements using Brownian bridges. Ecology 88:2354–2363. https://doi.org/10.1890/06-0957.1

Joo R, Etienne M-P, Bez N, Mahévas S (2018) Metrics for describing dyadic movement: a review. Mov Ecol 6:26. https://doi.org/10.1186/s40462-018-0144-2

Kan Z, Tang L, Kwan M-P, Ren C, Liu D, Li Q (2018a) Traffic congestion analysis at the turn level using Taxis' GPS trajectory data. Comput Environ Urban Syst. https://doi.org/10.1016/j.compenvurbsys.2018.11.007

Kan Z, Tang L, Kwan M-P, Ren C, Liu D, Pei T, Liu Y, Deng M, Li Q (2018b) Fine-grained analysis on fuel-consumption and emission from vehicles trace. J Clean Prod 203:340–352. https://doi.org/10.1016/j.jclepro.2018.08.222

Klous G, Smit LA, Borlée F, Coutinho RA, Kretzschmar ME, Heederik DJ, Huss A (2017) Mobility assessment of a rural population in the Netherlands using GPS measurements. Int J Health Geogr 16:1–13

Kraemer MU, Yang C-H, Gutierrez B, Wu C-H, Klein B, Pigott DM, Du Plessis L, Faria NR, Li R, Hanage WP et al (2020) The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 368:493–497

Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. J Am Stat Assoc 47:583–621

Kwan M-P (2000) Interactive geovisualization of activity-travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. Transportat Res Part c: Emerg Tech 8:185–203. https://doi.org/10.1016/S0968-090X(00)00017-6

Lai S, Ruktanonchai NW, Zhou L, Prosper O, Luo W, Floyd JR, Wesolowski A, Santillana M, Zhang C, Du X, Yu H, Tatem AJ (2020) Effect of non-pharmaceutical interventions to contain COVID-19 in China. Nature 585:410–413. https://doi.org/10.1038/s41586-020-2293-x

Laube P, Purves RS (2011) How fast is a cow? Cross-Scale Analysis of Movement Data: Cross-Scale Trajectory Analysis. Trans GIS 15:401–418. https://doi.org/10.1111/j.1467-9671.2011.01256.x

Laube P, Dennis T, Forer P, Walker M (2007) Movement beyond the snapshot–dynamic analysis of geospatial lifelines. Comput Environ Urban Syst 31:481–501

Lee JH, Goulias KG (2018) Companionship and time investment in social fields at different life cycle stages: Implications for activity and travel modeling and simulation. Res Transp Econ 68:18–28

Lee JH, Davis AW, Yoon SY, Goulias KG (2016) Activity space estimation with longitudinal observations of social media data. Transportation 43:955–977. https://doi.org/10.1007/s11116-016-9719-1

Lenntorp B (1976) Paths in space-time environments: a time-geographic sudy of movement possibilities of individuals. Lund Studies in Geography Number 44, Royal University of Lund, Sweden

Long JA, Nelson TA, Webb SL, Gee KL (2014) A critical examination of indices of dynamic interaction for wildlife telemetry studies. J Anim Ecol 83:1216–1233

Long JA, Webb SL, Nelson TA, Gee KL (2015) Mapping areas of spatial-temporal overlap from wildlife tracking data. Mov Ecol 3:1–14

Lu Y, Liu Y (2012) Pervasive location acquisition technologies: Opportunities and challenges for geospatial studies. Comput Environ Urban Syst 36:105–108. https://doi.org/10.1016/j.compenvurbsys.2012.02.002

Marra AD, Becker H, Axhausen KW, Corman F (2019) Developing a passive GPS tracking system to study long-term travel behavior. Transportat Res Part c: Emerg Tech 104:348–368

Martínez-Díaz M, Soriguera F (2021) Short-term prediction of freeway travel times by fusing input-output vehicle counts and GPS tracking data. Transportat Lett 13(3):193–200

Miller HJ (1991) Modelling accessibility using space-time prism concepts within geographical information systems. Int J Geogr Inf Syst 5:287–301. https://doi.org/10.1080/02693799108927856

Miller HJ (2005) A measurement theory for time geography. Geogr Anal 37:17–45. https://doi.org/10.1111/j.1538-4632.2005.00575.x

Miller JA (2015) Towards a better understanding of dynamic interaction metrics for wildlife: a null model approach. Trans GIS 19:342–361

Miller HJ, Dodge S, Miller J, Bohrer G (2019) Towards an integrated science of movement: converging research on animal movement ecology and human mobility science. Int J Geogr Inf Sci 33:855–876. https://doi.org/10.1080/13658816.2018.1564317

Moreira A, Santos MY, Wachowicz M, Orellana D (2010) The impact of data quality in the context of pedestrian movement analysis, In: Geospatial Thinking. Springer, Berlin

Neves A, Brand C (2019) Assessing the potential for carbon emissions savings from replacing short car trips with walking and cycling using a mixed GPS-travel diary approach. Transportat Res Part a: Policy Practice 123:130–146

NuStats (2013) 2010–2012 California household travel survey final report. California department of transportation

Oliver N, Lepri B, Sterly H, Lambiotte R, Deletaille S, De Nadai M, Letouzé E, Salah AA, Benjamins R, Cattuto C, Colizza V, de Cordes N, Fraiberger SP, Koebe T, Lehmann S, Murillo J, Pentland A, Pham PN, Pivetta F, Saramäki J, Scarpino SV, Tizzoni M, Verhulst S, Vinck P (2020) Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. Sci Adv. https://doi.org/10.1126/sciadv.abc0764

Procter DS, Page AS, Cooper AR, Nightingale CM, Ram B, Rudnicka AR, Whincup PH, Clary C, Lewis D, Cummins S et al (2018) An open-source tool to identify active travel from hip-worn accelerometer, GPS and GIS data. Int J Behav Nutr Phys Act 15:1–10

Quddus MA, Ochieng WY, Noland RB (2007) Current map-matching algorithms for transport applications: State-of-the art and future research directions. Transportat Res Part c: Emerg Tech 15:312–328. https://doi.org/10.1016/j.trc.2007.05.002

Ranjan G, Zang H, Zhang Z-L, Bolot J (2012) Are call detail records biased for sampling human mobility? ACM SIGMOBILE Mobile Comput Communicat Rev 16:33–44

Schneider C, Zutz S, Rehrl K, Brunauer R, Gröchenig S (2016) Evaluating GPS sampling rates for pedestrian assistant systems. J Locat Based Serv 10:212–239. https://doi.org/10.1080/17489725.2016.1259509

Schönfelder S, Axhausen KW (2003) Activity spaces: measures of social exclusion? Transp Policy 10:273–286. https://doi.org/10.1016/j.tranpol.2003.07.002

Schüssler N, Axhausen KW (2008) Identifying trips and activities and their characteristics from GPS raw data without further information. Arbeitsberichte Verkehrs-und Raumplanung. https://doi.org/10.3929/ethz-a-005589980

Selkirk SW, Bishop ID (2002) Improving and extending home range and habitat analysis by integration with a geographic information system. Trans GIS 6:151–159. https://doi.org/10.1111/1467-9671.00102

Shalaby A, Farhan A (2004) Prediction model of bus arrival and departure times using AVL and APC data. J Public Transp 7:3

Shen L, Stopher PR (2014) Review of GPS travel survey and GPS data-processing methods. Transp Rev 34:316–334

Silverman BW (1986) Density estimation for statistics and data analysis. Chapman & Hall, London

Smolak K, Siła-Nowicka K, Delvenne J-C, Wierzbiński M, Rohm W (2021) The impact of human mobility data scales and processing on movement predictability. Sci Rep 11:15177. https://doi.org/10.1038/s41598-021-94102-x

Song C, Qu Z, Blumm N, Barabasi A-L (2010) Limits of predictability in human mobility. Science 327:1018–1021. https://doi.org/10.1126/science.1177170

Stipancic J, Miranda-Moreno L, Labbe A, Saunier N (2019) Measuring and visualizing space–time congestion patterns in an urban road network using large-scale smartphone-collected GPS data. Transportat Lett 11:391–401

Su R, Goulias K (2021) Evolution of the Chinese spring festival travel network during the COVID-19 early outbreak. Transportat Lett. https://doi.org/10.1080/19427867.2021.1896065

Su R, McBride EC, Goulias KG (2020) Pattern recognition of daily activity patterns using human mobility motifs and sequence analysis. Transportat Res Part c: Emerg Tech 120:102796. https://doi.org/10.1016/j.trc.2020.102796

Su R, McBride EC, Goulias KG (2021) Unveiling daily activity pattern differences between telecommuters and commuters using human mobility motifs and sequence analysis. Transportat Res Part a: Policy Practice 147:106–132. https://doi.org/10.1016/j.tra.2021.03.002

Sui Y, Zhang H, Song X, Shao F, Yu X, Shibasaki R, Sun R, Yuan M, Wang C, Li S, Li Y (2019) GPS data in urban online ride-hailing: a comparative analysis on fuel consumption and emissions. J Clean Prod 227:495–505. https://doi.org/10.1016/j.jclepro.2019.04.159

Thill J-C (1992) Choice set formation for destination choice modelling. Prog Hum Geogr 16:361–382

Tian H, Liu Y, Li Y, Wu C-H, Chen B, Kraemer MUG, Li B, Cai J, Xu B, Yang Q, Wang B, Yang P, Cui Y, Song Y, Zheng P, Wang Q, Bjornstad ON, Yang R, Grenfell BT, Pybus OG, Dye C (2020) An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. Science 368:638–642. https://doi.org/10.1126/science.abb6105

Train KE (2009) Discrete choice methods with simulation. Cambridge University Press, Cambridge

Tukey JW (1977) Exploratory Data Analysis. Reading, MA, Addison-Wesley

Turchin P (1998) Quantitative analysis of movement: measuring and modeling population redistribution in animals and plants. Sinauer Associates, MA

Wolf J, Bachman W, Oliveira M, Auld J, Mohammadian AK, Vovsha P, Zmud J (2014a) Applying GPS data to understand travel behavior, volume II: guidelines

Wolf J, Bachman W, Oliveira M, Auld J, Mohammadian AK, Vovsha P, Zmud J (2014b) Applying GPS data to understand travel behavior, Volume I: Background, methods, and tests

Worton B (1987) A review of models of home range for animal movement. Ecol Model 38:277–298

Xu Y, Belyi A, Bojic I, Ratti C (2018) Human mobility and socioeconomic status: analysis of Singapore and Boston. Comput Environ Urban Syst. https://doi.org/10.1016/j.compenvurbsys.2018.04.001

Yoon S, Goulias K (2010) Impact of time-space prism accessibility on time use behavior and its propagation through intra-household interaction. Transportat Lett 2:245–260

Yoon SY, Deutsch K, Chen Y, Goulias KG (2012) Feasibility of using time–space prism to represent available opportunities and choice sets for destination choice models in the context of dynamic urban environments. Transportation 39:807–823

Yuan Y, Raubal M (2016) Analyzing the distribution of human activity space from mobile phone usage: an individual and urban-oriented study. Int J Geogr Inf Sci 30:1594–1621. https://doi.org/10.1080/13658816.2016.1143555

Zhao Z, Yin L, Shaw S-L, Fang Z, Yang X, Zhang F (2018) Identifying stops from mobile phone location data by introducing uncertain segments. Trans GIS 22:958–974

Zhao Z, Shaw S-L, Yin L, Fang Z, Yang X, Zhang F, Wu S (2019) The effect of temporal sampling intervals on typical human mobility indicators obtained from mobile phone location data. Int J Geogr Inf Sci 33:1471–1495. https://doi.org/10.1080/13658816.2019.1584805

Zimmermann M, Mai T, Frejinger E (2017) Bike route choice modeling using GPS data without choice sets of paths. Transportat Res Part c: Emerg Tech 75:183–196

Žunić E, Delalić S, Donko D (2020) Adaptive multi-phase approach for solving the realistic vehicle routing problems in logistics with innovative comparison method for evaluation based on real GPS data. Transportat Lett. https://doi.org/10.1080/19427867.2020.1824311