Conserved quantities in human mobility

From locations to trips

Author(s):
Hong, Ye; Martin, Henry; Xin, Yanan; Bucher, Dominik; Reck, Daniel Jan; Axhausen, Kay W.; Raubal, Martin

Publication date:
2023-01

Permanent link:
https://doi.org/10.3929/ethz-b-000584618

Originally published in:
Transportation Research Part C: Emerging Technologies 146, https://doi.org/10.1016/j.trc.2022.103979

Transportation Research Part C 146 (2023) 103979

Contents lists available at ScienceDirect

Transportation Research Part C

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

Conserved quantities in human mobility: From locations to trips

Ye Hong [a,*], Henry Martin [a,c], Yanan Xin [a], Dominik Bucher [a,d,1], Daniel J. Reck [b], Kay W. Axhausen [b], Martin Raubal [a]

[a] Institute of Cartography and Geoinformation, ETH Zurich, 8093 Zurich, Switzerland
[b] Institute for Transport Planning and Systems, ETH Zurich, 8093 Zurich, Switzerland
[c] Institute of Advanced Research in Artificial Intelligence (IARAI), 1030 Vienna, Austria
[d] c.technology, 8002 Zürich, Switzerland

ARTICLE INFO

Keywords:
Human mobility
Individual activity–travel pattern
Intra-personal variability
Mobility constraints

ABSTRACT

Quantifying intra-person variability in travel choices is essential for the comprehension of activity–travel behaviour. Due to a lack of empirical studies, there is limited understanding of how an individual's travel pattern evolves over months and years. We use two high-resolution user-labelled datasets consisting of billions of GPS track points from ∼ 3800 individuals to analyse individuals' activity–travel behaviour over the long term. The general movement patterns of the considered population are characterised using mobility indicators. Despite the differences in the mobility patterns, we find that individuals from both datasets maintain a conserved quantity in the number of essential travel mode and activity location combinations over time, resulting from a balance between exploring new choice combinations and exploiting existing options. A typical individual maintains ∼ 15 mode–location combinations, of which ∼ 7 are travelled with a private vehicle every 5 weeks. The dynamics of this stability reveal that the exploration speed of locations is faster than the one for travel modes, and they can both be well modelled using a power-law fit that slows down over time. Our findings enrich the understanding of the long-term intra-person variability in activity–travel behaviour and open new possibilities for designing mobility simulation models.

1. Introduction

Transport planners have long recognised that individual travel behaviour is dynamic and varies considerably from regular personal routines when observed over an extended period. This intra-person variability exhibits needs and constraints that are not constant from day to day (Hägerstrand, 1970), and that allow for flexibility in modelling individuals' timings and activities (Axhausen, 2006; Pas and Sundar, 1995). The constraint-need interactions form the observed individuals' multidimensional activity–travel patterns over time. To understand the interaction between the different aspects of activity–travel patterns, multi-day observations of individuals' mobility traces are necessary (Schlich and Axhausen, 2003), as opposed to the cross-sectional data where individuals are asked to report their travel behaviour on a single day. Intra-person variability of activity–travel patterns is an important area of research that has many practical applications such as offering better understandings of urban life (Ahas et al., 2010), assessing policy impacts (Jones and Clarke, 1988), and optimising public transit usage (Egu and Bonnel, 2020).

* Corresponding author.
E-mail addresses: hongy@ethz.ch (Y. Hong), martinhe@ethz.ch (H. Martin), yanxin@ethz.ch (Y. Xin), dominik.bucher@gmail.com (D. Bucher), mail@daniel-reck.de (D.J. Reck), axhausen@ivt.baug.ethz.ch (K.W. Axhausen), mraubal@ethz.ch (M. Raubal).

[1] The majority of this work was done while the author was at the Institute of Cartography and Geoinformation, ETH Zurich

https://doi.org/10.1016/j.trc.2022.103979
Received 9 June 2022; Received in revised form 1 December 2022; Accepted 1 December 2022
Available online 16 December 2022

Using actively collected multi-day travel survey datasets, previous studies have gained insights into the intra-person variability, mainly focusing on a single dimension of the activity–travel patterns, such as trip frequency (Pas and Sundar, 1995), activity location (Susilo and Kitamura, 2005; Hanson and Huff, 1988), and travel mode (Heinen and Chatterjee, 2015). Comparatively few studies investigated the inter-dependency between the different behaviour dimensions and measured their variability over time, such as activity–travel–location combinations by Susilo and Axhausen (2014) and activity–location–mode sequences by Dharmowijoyo et al. (2017). All of this research provides valuable knowledge for understanding the variability in activity–travel patterns. However, their results are restricted by the employed datasets. Since multi-day datasets are costly and difficult to collect (Schlich and Axhausen, 2003; Goulet-Langlois et al., 2018), they are limited in sample size and survey duration (typically ranging from days to a few weeks). The lack of continuous observations of individuals' activity–travel patterns over months and years hinders our understanding of the intra-person variability over the long term.

With the development of information and communication technologies (ICT) (Bucher et al., 2019) and location-based services (LBS) (Huang et al., 2018), massive passively-recorded digital datasets containing human mobility traces have emerged, which provide the opportunity to study the long-term patterns of travel behaviour. Based on these "big" mobility datasets, studies have provided statistical evidence into the regularities of individual visited activity locations (Alessandretti et al., 2020; Barbosa et al., 2015; González et al., 2008). A recent seminar study proposed that the number of essential locations visited by an individual is a conserved quantity over the long term (Alessandretti et al., 2018). Nevertheless, we still lack sufficient understanding of the evolution of travel mode, the other essential component of activity–travel patterns. Given the stability of visited locations (Alessandretti et al., 2018), it is unclear how an individual's travel mode choices influence this stability and whether similar long-term stability can be observed by additionally considering travel mode information.

We aim to bridge this research gap by analysing the evolution of individuals' choices in activity location and travel mode in two large, long-term GPS datasets. We address the following research question: *Given the claimed stability of location visits, how do activity location and transport mode choices of individuals evolve in the long term?*

We address this question by statistically analysing individuals' travel mode and activity location choices over time. We provide empirical evidence that each individual's number of essential mode–location choices is a conserved quantity over the long term. This stability results from a balance between exploring new choice combinations and exploiting existing options. Our findings are derived based on high-resolution user-labelled mobility traces from participants in two longitudinal experiments: 139 participants from the Swiss Federal Railways (SBB) Green Class (GC) E-Car pilot study (GC dataset) (Martin et al., 2019), spanning over 12 months, and the research project Mobility Behaviour in Switzerland (MOBIS dataset) (Molloy et al., 2022), containing approximately 3700 persons traced for more than eight weeks. This research improves our understanding of the long-term intra-person variability in individuals' activity–travel behaviour and provides empirical support for designing new mobility simulation models.

2. Related work

2.1. Intra-person variability in activity–travel behaviour

In the travel behaviour literature, the intra-person variability analysis, which investigates the extent to which our activity–travel decisions differ from the consistent patterns over time, has sparked interest in the field for a long time (Jones and Clarke, 1988; Pas, 1987; Hanson and Huff, 1988). Over the years, a large number of empirical studies have measured the degree of this variability through indicators calculated on a single aspect of the activity–travel pattern, which can be roughly classified into trip-based methods (e.g., trip frequency (Pas and Sundar, 1995) and travel mode (Heinen and Chatterjee, 2015; Cherchi and Cirillo, 2014)), time use-related methods (Kang and Scott, 2010; Minnen et al., 2015) and activity-based methods (Schönfelder and Axhausen, 2016; Susilo and Kitamura, 2005). However, previous studies argued that the uniqueness of individuals' activity–travel patterns in each period should be understood from multiple dimensions as a form to satisfy their needs and desires that are limited by constraints (Susilo and Axhausen, 2014; Dharmowijoyo et al., 2017). Therefore, involving multiple aspects of activity–travel patterns is necessary to understand intra-person variability completely. For example, Schlich and Axhausen (2003) analysed different combinations constructed from travel mode, activity purpose, arrival time and destination location, which all show a high degree of repetition over the studied duration. Susilo and Axhausen (2014) adopt the Herfindahl–Hirschman Index (HHI) index to measure the degree of repetition of individuals' choices regarding their daily activity–travel–mode–location combinations and conclude that the variability is less correlated to travel mode choice, but more to the individuals' commitments and obligations. Dharmowijoyo et al. (2016) analysed the variability in the sequence of activity type, activity location and travel mode by applying a multidimensional sequence alignment model. They report that the variability between weekend and weekday pairs is much greater than between weekday–weekday pairs or weekend–weekend pairs. Moreover, this line of studies noted that the degree of variability differs with the employed methods and the aspect of the activity–travel pattern under consideration (Schlich and Axhausen, 2003; Raux et al., 2016).

All the studies mentioned above were conducted based on multi-day travel surveys, which require the active participation of surveyed individuals and are thus difficult to scale. Moreover, as indicated by Schlich and Axhausen (2003), high response burden and the tendency to ignore short trips are critical challenges faced by interviewed individuals when filling out a multi-day survey. Although providing valuable travel information, longitudinal travel behaviour data often cover a time range from several days up to six weeks, making it difficult to study the intra-personal variability in activity–travel behaviour over months and years. However, this knowledge is essential for understanding the evolution of travel behaviour and estimating travel demands over time. Conventional travel demand modelling approaches often implicitly assume activity–travel behaviours to be static and model them using mobility data sampled from a single day, which leads to growing model errors over time (Zhang et al., 2018). To date, we still lack a comprehensive model to describe how individual travel behaviours may have evolved over time (Chen et al., 2016).

2.2. Longitudinal human mobility studies using passively collected tracking data

Over the last decade, the availability of large-scale datasets recording human digital traces has increased in the field of human mobility and provided novel insights into its quantitative patterns (Barbosa et al., 2018; Schläpfer et al., 2021). Examples of these mobility data are geotagged posts generated from online activities (Bao et al., 2021), smart-card data collected at transit systems (Goulet-Langlois et al., 2018), call detail records (CDR) data originated from mobile phone usages (Järv et al., 2014; Yuan et al., 2012) and GPS tracking data collected through mobile devices (Hong et al., 2021). These passively collected data are generated for purposes that are not intended but can be potentially used for research. They pose little burden to the users and cover a broader population and geographic area (Chen et al., 2016; Wang et al., 2018).

Longitudinal human mobility studies applying these data have shown that an individual's movements can be described by statistical scaling laws (González et al., 2008) and have high theoretical predictability in location visits (Song et al., 2010b; Lu et al., 2013). Specifically, focusing on individual visited locations, Song et al. (2010a) reported that individuals regularly return to a few important locations over time and Alessandretti et al. (2018) extended the idea to show that the number of these locations for each individual is a conserved quantity over the long term. Moreover, studies have revealed the regularity (Stanley et al., 2018) and seasonality (Järv et al., 2014) of the individual activity spaces, which represent the observed geographical space that contains important visited locations. Another prevailing line of research aims to understand the transit usage behaviour using smart card data, where researchers measure trip rates and sequences of travel events to distinguish regular behaviour patterns over time (Egu and Bonnel, 2020; Goulet-Langlois et al., 2018) and to identify typical user groups (Liu et al., 2022).

The great opportunities for the "big" mobility data in travel behaviour studies are followed by the challenges of using these datasets. The basic data unit in passively collected datasets does not correspond to meaningful terms used in travel behaviour studies; for example, it is not straightforward to derive a location where people perform an activity directly from GPS track points or mobile phone tower data. Therefore, studies tend to identify mobility patterns (such as displacements between two inferred stops) from which predictions can be made instead of using transport-related metrics such as trip distance or time (Chen et al., 2016). Moreover, human mobility patterns inferred from these datasets often lack accurate travel-related information such as travel mode (Wang et al., 2018), which hinders their application in studying individuals' multidimensional activity–travel behaviour. The main focus of the studies on fine-grained mobility behaviours lies in understanding how individuals move from one location to another (Schläpfer et al., 2021; Alessandretti et al., 2020). To date, no study has systematically investigated the intra-person variability in the travel mode choices of individuals using passively collected mobility data.

3. Dataset

Two large-scale user-labelled GPS tracking datasets consisting of billions of GPS track points from ~3800 individuals are employed to study the long-term evolution of individual activity–travel patterns. This section provides a brief introduction to these datasets and the preprocessing methods.

3.1. Datasets

GC dataset. The *Green Class* (GC) dataset is an outcome of the SBB Green Class E-Car pilot study conducted by the Swiss Federal Railways (SBB) from November 2016 to December 2017 (Martin et al., 2019). In the study, 139 participants were offered a Mobility as a Service (MaaS) package with a fixed yearly rate, including a battery electric vehicle, a national-level season ticket (GA travelcard), as well as access to several car- and bike-sharing programs. As part of the study, the participants were asked to install a commercial application (app) on their smartphones to record their daily movement (with a median time of 13.9 s between two consecutive GPS recordings) (Bucher et al., 2020). The app pre-processed the raw GPS traces to infer *stay points* where users are stationary and *stages* of continuous movements that use a single travel mode. The participants were then requested to label the stay points with a purpose and stages with a travel mode. Although the participants were primarily selected based on their geographic location, the participation preconditions led to a bias towards the middle- and upper-class people with high mobility demand (Martin et al., 2019).

MOBIS dataset. The *Mobility Behaviour in Switzerland* (MOBIS) study aims to assess the size of the behavioural impact of mobility pricing in Switzerland (Molloy et al., 2022). Starting from Autumn 2019, ~3700 persons took part in the study and recorded their locations for eight weeks with a tracking app installed on their smartphones. Some participants continued the tracking after the MOBIS data collection was completed. The app recorded all outdoor movements and divided the GPS traces into *stay points* and *stages*. It also imputed the travel modes based on measurements such as speed and acceleration obtained from built-in smartphone sensors. The participants were encouraged to verify these imputations and add purpose labels for their recorded stay points. MOBIS focuses on people living in Switzerland who stated that they use both cars and public transport for their daily travel. Still, the sample generally matches the known socio-demographics of the Swiss population (Molloy et al., 2022), providing a strong data set for studying the mobility behaviours of the residents.

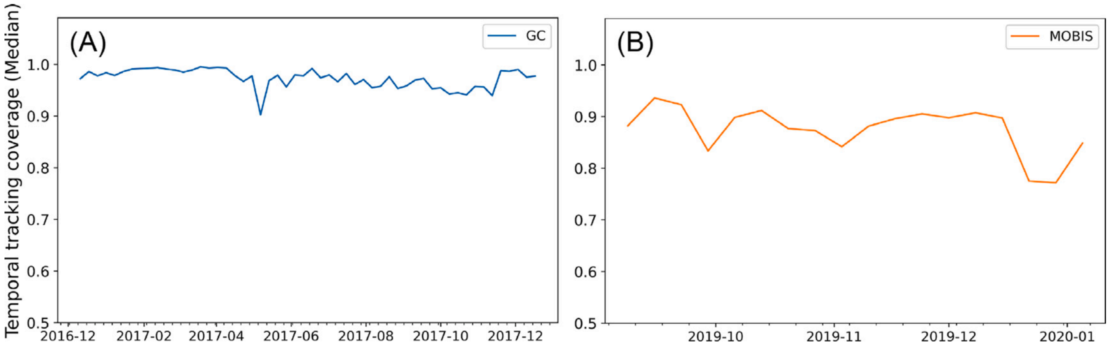

Fig. 1. The median weekly temporal tracking coverage for the (A) GC and (B) MOBIS datasets. Temporal tracking coverage is defined as the proportion of time when the user's whereabouts are recorded.

3.2. User selection and movement model generation

The quality of the recorded mobility traces typically varies across participants in app-based mobile phone tracking datasets, mainly influenced by the willingness to participate in the study and individuals' habits of using mobile phones. We pre-filter the datasets to include participants observed for a long period with high temporal tracking quality for further analysis. Specifically, we include participants observed for more than 300 days and 50 days for the GC and MOBIS datasets. Moreover, we use temporal tracking coverage, defined as the proportion of time the user's whereabouts are recorded, to evaluate the tracking quality of individuals in the temporal dimension. After this process, 115 participants in GC and 3299 participants in MOBIS remain. The median weekly temporal tracking coverage of the remaining participants shows that the quality is relatively stable and high throughout the study period for both datasets (Fig. 1).

The raw movement traces need to be processed to infer meaningful analysis units for travel behaviour analysis. In this study, the stay points and stages obtained from the tracking app need to be aggregated into *locations* and *trips*. Locations correspond to visits to the same place at different times, and trips are defined as the collection of all movement and idling between two activities (Axhausen, 2007). It is necessary to define *activities* to identify both location and trips. We regard a stay point as an activity if its duration is longer than 25 min or if it was labelled with a non-trivial purpose (we consider any available purpose as non-trivial except for *wait* or *unknown*). The process of generating locations and trips and their attached attributes are described as follows.

Locations. The stay points that are regarded as activities are aggregated to infer locations. We adopt the DBSCAN method to cluster activity stay points based on the spatial proximity of each individual to avoid generating large clusters that may cover several places (Hariharan and Toyama, 2004). We utilise the function provided in the *Trackintel* framework and used the following algorithm parameters (Hong et al., 2021): $epsilon = 50$, $num\_samples = 1$. In Appendix A.3, we describe the influence of altering these parameters on the results.

Trips. Trips are obtained by collecting all consecutive sequences of stages between two activity stay points. While numerical attributes such as length and duration can be directly obtained, the aggregation of travel modes from stages to trips has to be derived based on predetermined rules. Here, we adopt two standards for labelling the main mode of a trip (Axhausen, 2007): (A) Based on the mode with the largest share of the distance travelled and (B) based on predefined hierarchies of the assumed strength of the mode, i.e., airplane - train - tram - bus - car/motorbike - bicycle - walk. The results reported in the main text are obtained through standard A; in Appendix A.3, we show that the results are also robust for standard B. For both datasets, we prioritise the travel mode labels verified by the individual; if these are not available, we fall back to the modes imputed by the app. Finally, the main travel mode of a trip is grouped into the following three categories: non-motorised modes (i.e., walking and bicycle), private vehicle (i.e., car and motorbike) and public transport (i.e., airplane, boat, subway, train, bus and tram) (Susilo and Axhausen, 2014; Heinen and Chatterjee, 2015).

All processing steps are implemented in *Python* using the open-source *Trackintel* human movement data processing library (Martin et al., 2022). After the processing, the mobility of individual i can be described using the sequence of visited activity locations $L^i_{seq} = [loc_1, loc_2, \dots, loc_n]$ and the sequence of trips $T^i_{seq} = [trip_1, trip_2, \dots, trip_n]$, each containing n items that are ordered by their observation time. loc_k and $trip_k$ are defined as the kth element of L^i_{seq}, and T^i_{seq}, respectively. According to the movement data model, individual i conducts the kth trip to reach the kth location, i.e., L^i_{seq} and T^i_{seq}'s elements are one-to-one aligned. An activity location loc_k can be represented using arrival time t_k, geometry $l_k = (x, y)$, where x and y are spatial coordinates in a given reference system, e.g., latitude and longitude, and activity duration $d^{(l)}_k$, i.e., $loc_k = \left(l_k, t_k, d^{(l)}_k\right)$. A trip $trip_k$ contains the main travel mode e_k and the travel duration $d^{(t)}_k$, i.e., $trip_k = \left(e_k, d^{(t)}_k\right)$. Both datasets continuously record the whereabouts of each individual, resulting in a worldwide collection of trips and locations (Fig. 2A). Still, most of the recorded movements and activities are located within Switzerland (Fig. 2B).

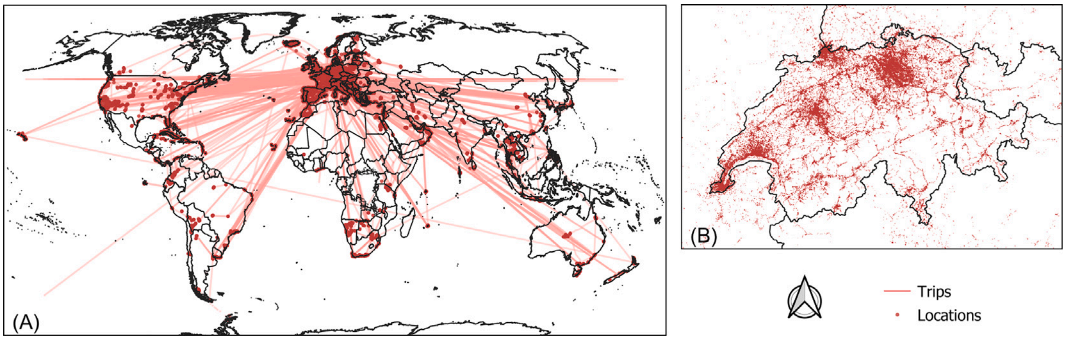

Fig. 2. Spatial distribution of the mobility for the GC and MOBIS datasets. (A) Worldwide distribution of trips and locations. (B) Enlarged plot showing the location distribution within Switzerland. Map data from the GADM Database of Global Administrative Areas, version 3.6, available at https://gadm.org/. Map projected using Robinson projection (ESRI:54030).

4. Methodology

This section provides a detailed description of the methods to study the long-term evolution of individuals' activity location and transport mode choices. We first identify the mobility behaviours of the population using mobility indicators. This assists in understanding the different mobility patterns of the two datasets. Then, the ways to quantify individuals' travel mode and activity location choices are described.

4.1. Identifying individuals' mobility behaviour

We use mobility indicators to characterise the mobility patterns of the considered individuals. Previous studies have shown that many aspects of individual mobility can be described using statistical distributions (Brockmann et al., 2006; González et al., 2008); however, the best-approximating distribution is often dataset dependent, with its parameters reflecting dataset-specific characteristics (Alessandretti et al., 2017). Here, we consider the following indicators:

- Radius of gyration (González et al., 2008). The radius of gyration Rg^i can be regarded as the characteristic distance travelled by individual i and is often used to reflect their range in the spatial dimension. It is calculated as follows:

$$Rg^i = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(l_k - l^i_{cm})^2}$$

  where l_k represents the spatial coordinates of the kth location record loc_k, and $l^i_{cm} = \frac{1}{n}\sum_{k=1}^{n} l_k$ is the centre of mass of the activity locations of individual i.
- Jump length (Brockmann et al., 2006). Jump length Δr measures the distance between consecutive displacements of human movements, i.e., $\Delta r = \|l_{k+1} - l_k\|_2$ for every $k < n$. The distribution of jump lengths followed power-laws as previously reported in Brockmann et al. (2006) and González et al. (2008).
- Location visitation frequency (González et al., 2008; Song et al., 2010a). It is widely reported that certain locations are more important than others in individuals' daily mobility, and the frequency f_k of the kth most visited location follows Zipf's law, i.e., $f_k \sim k^{-\xi}$, with $\xi \sim 1$ (González et al., 2008; Alessandretti et al., 2018). Here, we rank locations based on their recorded number of visits and show the relationship between visitation frequencies and their rank.
- Activity space. The concept of activity space represents the observed geographical space that contains locations frequently visited by an individual over a period of time (Golledge and Stimson, 1997). We use the 95% confidence ellipse to represent an individual's daily activity space, and the ellipse's area to quantify the geographical size of the space (Schönfelder and Axhausen, 2003, 2016). Practically, for each individual, we calculate the covariance matrices from daily activity locations weighted by their activity duration. The size of the 95% ellipse is then obtained from the determinant of the covariance matrix.

Additionally, we calculate basic trip and location statistics of the considered population. All the above analyses help to understand and distinguish possible differences in the general mobility patterns between the GC and MOBIS individuals.

4.2. Joint consideration of mode and location choices

We aim to quantify the evolution of individuals' travel mode and activity location choices over time. To jointly consider these two dimensions of travel behaviour, we propose the concept of *trip package* that groups individuals' movement and activity. Each unique combination of travel mode and destination location is a trip package; that is, we group the trips that share the same travel mode

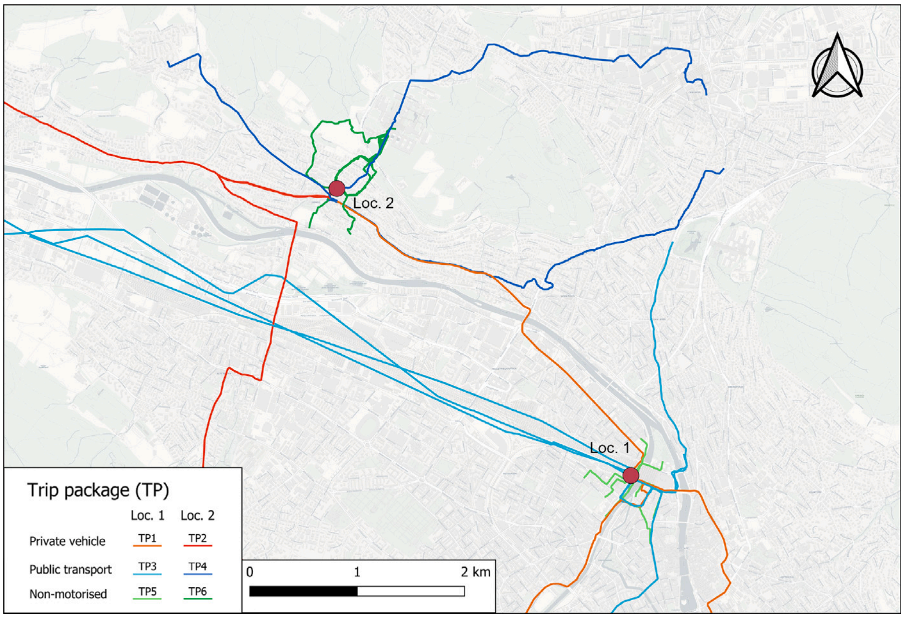

Fig. 3. The construction of trip packages. Trips are coloured according to the trip packages, defined as unique combinations of destination location and travel mode (classified into private vehicle, public transport and non-motorised). The coordinates of the locations were slightly altered to protect the subject's privacy. Map data ©OpenStreetMap contributors, ©CARTO. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and destination location as one trip package. More formally, we use p_k to represent the kth observed trip package of an individual. p_k is defined with the travel mode e_k and the destination location's geometry l_k, i.e., $p_k = (e_k, l_k)$. Therefore, the location sequence L^i_{seq} and the trip sequence T^i_{seq} of individual i can be represented using the sequence of trip packages $P^i_{seq} = [p^1_1, p^2_2, p^1_3, \dots, p^q_n]$, with n observations of q unique trip packages. We regard p^m_k and p^n_j to be the same trip package if and only if $(e_k, l_k) = (e_j, l_j)$, i.e., $p^m_k = p^n_j \iff m = n$. Additionally, we define the set of all distinct trip packages as $\mathcal{P}^i = \{p^1, p^2, \dots, p^q\}$, and therefore, $|\mathcal{P}^i| = q$.

Fig. 3 shows an example of constructing trip packages based on mobility traces from one of the authors living in Zurich, Switzerland, where six trip packages can be observed from the combination of travelling with three different travel modes to two distinct destination locations. Trip packages capture the information of *how* and *where* the individual is travelling at the same time. Therefore, considering the temporal dimension, individual mobility can be decomposed into sequences of trip packages P^i_{seq}, which allows us to study personal mobility preferences over time.

4.3. Extraction of important mode–location choices over time

A trip package encapsulates the movement behaviour for a single trip, while an individual's choice of trip packages over time captures their mobility preferences. Previous studies have shown that a small subset of all visited locations is sufficient to characterise individuals' day-to-day activity location choices (González et al., 2008; Yuan and Raubal, 2016). Following the same line of argument, we represent individuals' activity–travel behaviour using the most often observed trip packages. We propose the *behaviour set* to capture these essential trip packages. More formally, we use $P^i_{seq}(t, \delta t)$ to represent the largest sub-sequence of P^i_{seq} that is observed during a time window of δt weeks that starts at time t:

$$P^i_{seq}(t, \delta t) = [p_s, p_{s+1}, \dots, p_w]$$

where $s = \arg\min_k (\{t_k \in loc_k \in L^i_{seq} \mid t_k \geq t\})$ and $w = \arg\max_k(\{t_k \in loc_k \in L^i_{seq} \mid t_k \leq t + \delta t\})$.

Then, the behaviour set $\mathcal{B}^i(t, \delta t)$ is defined as the set of all trip packages observed at least twice and travelled for more than m minutes/week in $P^i_{seq}(t, \delta t)$:

$$\mathcal{B}^i(t, \delta t) = \left\{ p \in P^i_{seq}(t, \delta t) \mid \sum_{k=s}^{w} \mathbb{1}_{[p=p_k]} \geq 2 \land \sum_{k=s}^{w} \mathbb{1}_{[p=p_k]} \cdot d^{(t)}_k > m \cdot \delta t \right\}$$

where $\mathbb{1}_{[\cdot]}$ represent the indicator function, and $\mathbb{1}_{[C]} = 1$ if C is $True$ and $\mathbb{1}_{[C]} = 0$ otherwise. In Appendix A.1, we present an example of constructing a behaviour set from the observed trip packages in Fig. 3. The filters δt and m control the frequency and importance of trip packages contained in the behaviour set. The results presented in the section below are obtained using time window $\delta t = 5$ weeks and $m = 1$ min/week. We also tested other combinations of frequency and importance filters for defining the behaviour set, which yielded equivalent results (for sensitivity analysis on the parameter choices, see Appendix A.3). Compared to the activity set that contains an individual's most important locations over time (Alessandretti et al., 2018), the definition of the behaviour set and its construction process additionally considers trip-level information (i.e., travel mode). This ensures that the behaviour set is a more comprehensive description and a more fine-grained representation of individuals' travel behaviour.

5. Results

5.1. Mobility indicators and trip statistics

We reveal the movement characteristics of the considered population through various mobility indicators. Fig. 4 shows the distribution of radius of gyration, jump length, location visits, and the median daily activity space for the individuals in the GC and MOBIS datasets. For the radius of gyration and jump length, we fit parameter distributions to the empirical data and determine their best-fit distribution with the Akaike information criterion (AIC) and Akaike weights. We consider the log-normal distribution and the power-law (including truncated-power law) distribution as candidate distributions, as they are the most often reported distributions to approximate these two properties in the mobility literature (Zhao et al., 2015; Tang et al., 2015). The empirical distribution of radius of gyration and jump length is both best fitted using a log-normal distribution (their parameters can be found in Appendix A.2). This observation is in line with the results in Alessandretti et al. (2017), where it is reported that the jump length distribution is best described using log-normal distribution when considering GPS datasets that have high spatial and temporal resolutions, which is also the data collection method for the GC and MOBIS datasets. Moreover, the frequency of location visits and the distribution of activity spaces of both datasets are consistent with previous studies (Song et al., 2010a; Schönfelder and Axhausen, 2016; Järv et al., 2014).

While the GC and MOBIS datasets exhibit similar statistical properties for the mobility indicators, the subtle differences in their distributions reveal dataset-specific characteristics. The pattern of the frequency of the location visits is similar for the more important locations for both datasets (Fig. 4C). However, the GC participants are observed to visit more locations on average than the ones in MOBIS, as shown from the long tail of the distribution, which could partly be attributed to the more extended observation period for the GC dataset ($\sim$12 month for GC and $\sim$2 month for MOBIS). Moreover, the high probability of finding individuals with a larger radius of gyration suggests that GC participants are more active in travelling compared to the ones in the MOBIS dataset over the whole study period (Fig. 4A). This tendency could also be observed at finer time scales; we report a higher probability of travelling a longer distance to reach their next activity location (Fig. 4B) and a larger median daily activity space (Fig. 4D) for the GC individuals.

We report essential trip and activity location statistics to uncover their inter-correlated nature over the study duration. We find that, on average, individuals spent a considerable amount of their time "on the way", with the median recorded total trip duration across participants for the GC dataset reaching 2.17 h/day and the MOBIS dataset 1.47 h/day. Considering the obligatory activities such as sleep and work constraining daily movements (Hägerstrand, 1970; Schönfelder and Axhausen, 2016), this statistic emphasises the importance of trips in shaping individuals' daily schedules. Next, we investigate the relationship between activity locations and trips. We calculate two duration measures over the study period and analyse their correlation: for each location visited by an individual, we calculate (1) the total activity time spent at this location and (2) the total travel time to reach this location. The joint distribution of the two considered properties is shown in Fig. 5, where we focus on locations with longer activity durations as they are more important in individuals' mobility. We observe that the travel time increases with the location importance, with Pearson correlation coefficient $\rho = 0.71$, two-tailed $P < 0.001$ for the GC dataset and Pearson correlation coefficient $\rho = 0.75$, two-tailed $P < 0.001$ for the MOBIS dataset, implying that trip and location properties are strongly correlated. This result indicates that individuals tend to travel more to reach their essential locations, and the dimensions of activity–travel behaviour are highly correlated.

To summarise, the analysis of mobility indicators suggests that the mobility characteristics of the two considered datasets match well with previous studies. Moreover, the differences between GC and MOBIS participants reveal that GC individuals are, on average, more mobile. Last, analysing the relation between essential trip and location statistics shows that these travel behaviour dimensions are correlated and should be considered jointly when studying intra-person variability.

5.2. Stability of mode and location choices

The previous section assists in understanding the general individual mobility patterns over the entire study period yet falls short of revealing their dynamic evolution characteristics over time. By introducing the concept of trip package and behaviour set, we can disentangle the travel mode and activity location choices at any point in time and analyse their evolution by comparing them in the temporal dimension. In Fig. 6, we show that the behaviour set captures the essential part of the movement behaviour for each individual. We define the set $\mathcal{Z}^i$ as a set that contains all trip packages p that are observed in at least one behaviour set:

$$\mathcal{Z}^i = \left\{ p \mid p \in B^i(t, \delta t) \; \forall \; t \in [t_1, t_n - \delta t] \right\}$$

We then calculate the fraction of trip packages that belong to the behaviour set $r^i = \left|\mathcal{Z}^i\right| / \left|\mathcal{P}^i\right|$. The distribution of r_i over individuals reveals that only a small portion of all trip packages are part of the behaviour set (Fig. 6 light-colour histogram). Still, the frequency of observing these trip packages $f^i = \frac{1}{n} \sum_{p \in \mathcal{Z}^i} \sum_{k=1}^{n} \mathbb{1}_{[p=p_k]}$, $p_k \in P^i_{seq}$ shows that they cover the majority of trips for each individual

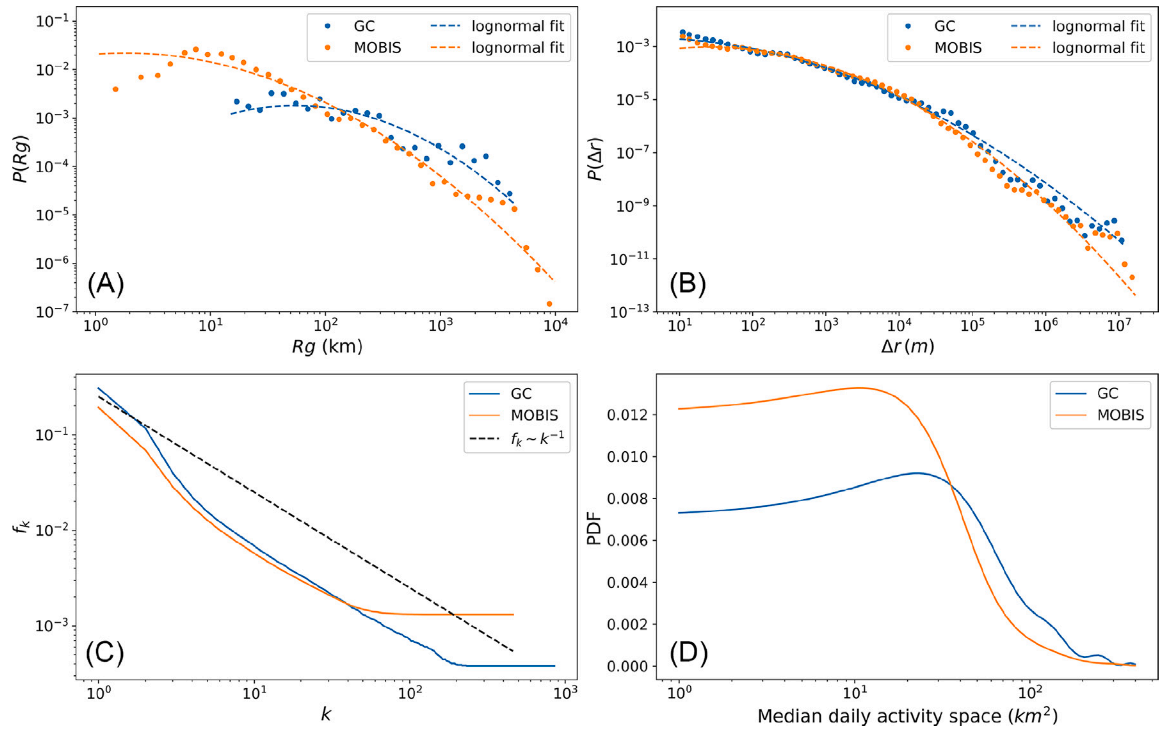

Fig. 4. The mobility indicators for the population under consideration. (A) The distribution of the radius of gyration. The dashed line represents the best fit log-normal distribution. (B) The distribution of the jump length (i.e., consecutive displacements). The dashed line represents the best fit log-normal distribution. (C) Zipf's plot showing the frequency of location visits. (D) The distribution of the median daily activity space.

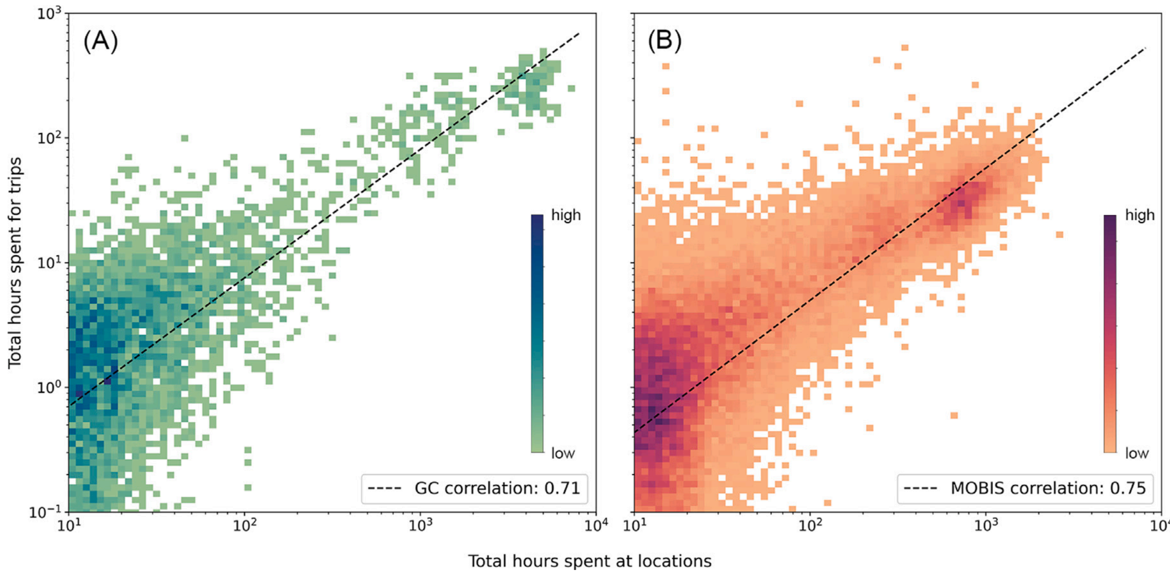

Fig. 5. Joint distribution of each location's total activity duration and total travel time (trip time) for the (A) GC and (B) MOBIS datasets. We focus on locations with a total activity duration longer than 10 h over the study period.

(Fig. 6 dark-colour histogram). In other words, the chosen criteria exclude approximately 80% of the observed trip packages yet retain the most important ones that account for more than half of daily movements.

The number of trip packages in the behaviour set $B^i(t, \delta t)$ at a given time t, denoted as the behaviour capacity $C^i(t)$, indicates the number of essential movement behaviours. In the first step, we are interested in the evolution of $C^i(t)$ that captures the number

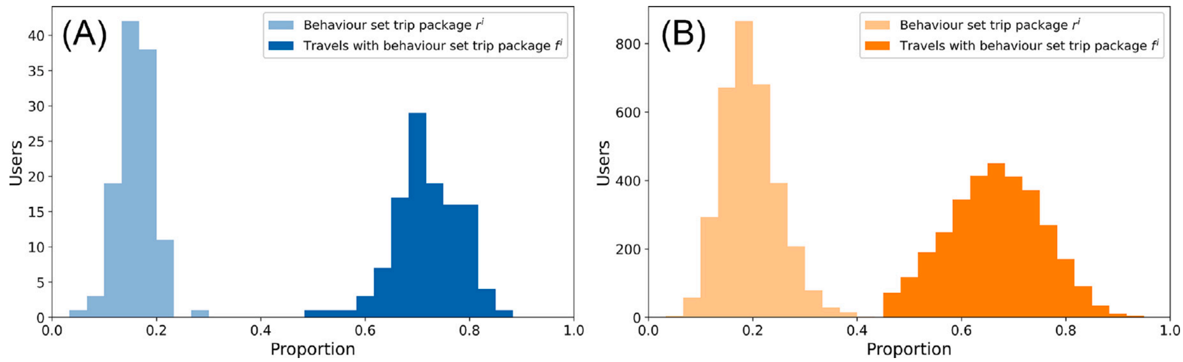

Fig. 6. The importance of the behaviour set in daily mobility of the (A) GC and (B) MOBIS datasets. Histogram of the behaviour set trip package proportion (light-colour bar) and the frequency proportion travelled with these trip packages (dark-colour bar) over individuals. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

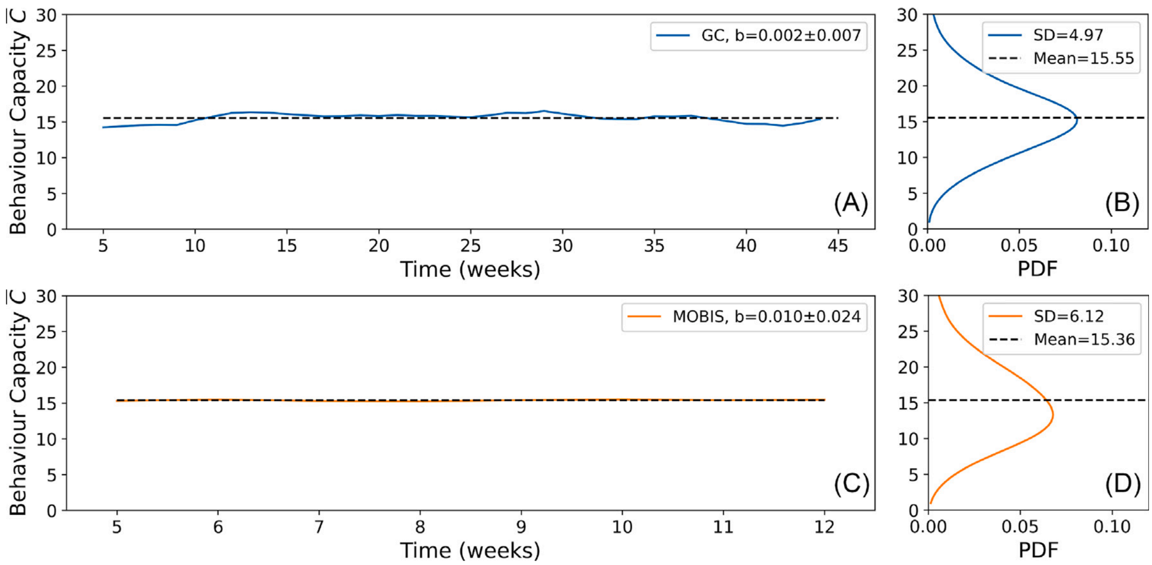

Fig. 7. Stability of individual's behaviour set. (A) and (C) The evolution of the average behaviour capacity. The regression coefficient b of a linear fit for each dataset is reported in the legend. (B) and (D) The probability density function of individuals' behaviour capacity. The standard deviation (SD) and the mean value are reported in the legend. The dashed black lines correspond to the constant mean capacity.

of distinct trip packages that individual i maintains over time. Fig. 7 shows the evolution of the average capacity across individuals $\overline{C(t)}$ for both considered datasets.

We observe that the average behaviour capacity $\overline{C(t)}$ is a conserved quantity over time. This is tested using a linear fit of the form $\overline{C(t)} = a + b \cdot t$, where b represents the slope and a is the intercept, and testing the null hypothesis $H_0 : b = 0$. We find that b is not significantly different from 0 (GC: -0.002 ± 0.007, MOBIS: -0.010 ± 0.024), and we find no evidence for rejecting the null hypothesis (GC: two-sided $P = 0.72$, MOBIS: two-sided $P = 0.68$). The individual behaviour capacity has a symmetric distribution around the sample mean (Fig. 7B and D). Additional statistical tests showing the stability of the behaviour capacity can be found in Section 6. Therefore, the number of important trip packages is stable over time on the collective population level, with a typical individual maintaining ∼15 packages every 5 week.

The collective level stability could be attributed to two distinct hypotheses: first, the behaviour capacity of each individual is stable over time (hypothesis A); and second, substantial heterogeneity can be found within the population, with some individuals increasing their behaviour capacity and others decreasing the capacity over time (hypothesis B). To distinguish between hypotheses A and B, we investigate the behaviour capacity evolution for each individual i, and define the capacity net gain $G^i(t)$ as the difference between the number of trip packages that are respectively added $A^i(t)$ and removed $R^i(t)$ at a specific time t; that is, $G^i(t) = A^i(t) - R^i(t)$. We looked at two quantities: the individual average net gain across time $\langle G^i \rangle$ and its standard deviation σ_{G^i}. Specifically, we test if an individual's average gain is smaller than the standard deviation, i.e., $\left|\langle G^i \rangle\right| < \sigma_{G^i}$, which indicates that the net gain is consistent with $\langle G^i \rangle = 0$, and suggests the behaviour capacity does not change in time for individual i. Empirical data analysis shows that this quantity holds for the majority of the individuals (GC: 100.00%, MOBIS: 88.45%). Hence, for most individuals, the average net gain over the observation period is not significantly different from zero; behaviour capacity is stable in time at the individual level, consistent with hypothesis A. Moreover, we find that the individual capacity has low variability, with the ratio between the standard deviation and the average individual capacity over time $\sigma_{C^i} / \langle C^i \rangle$ typically limited to below 22.7% for the GC dataset and 13.1% for the MOBIS dataset, demonstrating that fluctuations of the capacity are relatively small. These results indicate that the conserved quantity of behaviour capacity is observed at the population level and can also be regarded as a property inherent in individual mobility.

5.3. Sensitivity analysis of the stability

The results presented in the previous section heavily rely on the introduction of the concept trip package and behaviour set. The trip package assists in jointly considering individuals' travel mode and activity location choices, and the behaviour set helps establish a representative set of these packages to characterise movement at a given time. We are interested in whether the results still hold and how the size of the conserved quantity will possibly change when changing these prerequisites. We conduct the same analysis on the stability of the behaviour capacity when altering the definition parameters for the main travel mode, the activity location and the behaviour set. The detailed results of this sensitivity analysis can be found in Appendix A.3. At the collective level, we see no evidence of rejecting the null hypothesis $H_0 : b = 0$ at a level of $\alpha = 0.05$ for all parameter combinations, indicating that the behaviour capacity does not significantly change over time. For all participants in the GC dataset and the majority of individuals in the MOBIS dataset (>76%), the behaviour capacity is also conserved at the individual level. This analysis suggests that the stability of the behaviour capacity is independent of the definition of the behaviour set and likely to be an inherent characteristic of human movement. Moreover, we observe that the intercept a (i.e., the size of the behaviour capacity) increases and gradually saturates with the increase of the time window. This indicates that the number of essential trip packages depends on the considered time scale: a behaviour set with a smaller time window captures individuals' short-term travel behaviour, while a larger choice encapsulates travel decisions more prominent in the long term.

5.4. Evolution of travel mode choices

The individuals' behaviour set enables us to model the relationship between activity location choices and travel decision preferences. Here, as the travel modes are grouped into non-motorised, private vehicle, and public transport modes, we create behaviour sets that contain trip packages with a single type of travel mode. Therefore, the behaviour capacity $C^i(t)$ can be represented as $C^i(t) = C^i_{nm}(t) + C^i_{pv}(t) + C^i_{pt}(t)$, where $C^i_{nm}(t)$, $C^i_{pv}(t)$ and $C^i_{pt}(t)$ denote the number of trip packages (i.e., capacity) within non-motorised, private vehicle and public transport behaviour sets respectively. As a more detailed decomposition of $C^i(t)$, the evolution of $C^i_{nm}(t)$, $C^i_{pv}(t)$ and $C^i_{pt}(t)$ provides insights into individual's travel mode usages through time.

We report that the average capacity of private vehicle behaviours across individuals $\overline{C_{pv}(t)}$ is a conserved quantity over time, again shown using a linear fit of the form $\overline{C_{pv}(t)} = a + b \cdot t$, where b represents the slope and a is the intercept, and testing the null hypothesis $H_0 : b = 0$. We find no evidence for rejecting the hypothesis that $\overline{C_{pv}(t)}$ does not depend on time (GC: two-sided $P = 0.32$, MOBIS: two-sided $P = 0.98$). The analysis for individual net gain suggests that $C^i_{pv}(t)$ is stable in time for the majority of the individuals (GC: 100.00%, MOBIS: 87.97%). The intercept a stabilises at 7.26 for the GC dataset and at 7.60 for the MOBIS dataset, indicating that a typical individual maintains ~7 trip packages with private vehicle mode every 5 weeks. However, conducting the same analysis for non-motorised capacity and public transport capacity suggests that the null hypothesis is rejected at a level of $\alpha = 0.05$ for both $\overline{C_{nm}(t)}$ (GC: two-sided $P < 0.001$, MOBIS: two-sided $P = 0.02$) and $\overline{C_{pt}(t)}$ (GC: two-sided $P < 0.001$, MOBIS: two-sided $P < 0.001$). Therefore, the non-motorised capacity and the public transport capacity are not stable in time for the GC and MOBIS populations. Our analysis suggests certain mechanisms implicitly govern individuals' private vehicle behaviours, and these constraints are not found for public transport and non-motorised mode behaviours.

5.5. Exploration speed of travel behaviour dimensions

The previous sections show that the behaviour capacity is constant at both collective and individual levels. To gain insights into this invariant behaviour, we analyse the time series of the added number of trip packages $A^i(t)$ (defined in Section 5.2), which essentially captures the updating speed of the behaviour set. This section focuses on the GC dataset that contains individuals with longer tracking periods. We find that the mean added number over individuals $\overline{A(t)}$ is also constant over time. Using a linear fit in the form of $\overline{A(t)} = a + b \cdot t$, where b represents the slope and a is the intercept, we report that b is not significantly different from 0 (0.001 ± 0.002), and we find no evidence to reject the null hypothesis $H_0 : b = 0$ (two-sided $P = 0.69$). This result suggests that the behaviour set has an update rate that does not depend on the observation time. Moreover, we obtain an intercept $a = 2.13$ in the linear fit, indicating that for the behaviour set of a typical individual in the GC dataset, ~2 new trip packages can be observed per week. Hence, while the behaviour capacity is stable over the long term, the behaviour set is constantly evolving, with new trip packages continuously added to the behaviour set.

We are interested in modelling the updating speed of the behaviour set. For each individual, we distinguish between previously unobserved trip packages (exploration) and already observed behaviours (exploitation). The overall exploration, represented as the total number of trip packages $T^i(t)$ within the behaviour set observed up to time t, is well fitted using a power-law fit $T^i(t) \propto t^{\alpha_i}$, with average $\overline{\alpha} = 0.62$ (95% confidence interval (CI): $0.617 - 0.628$, two-sided $P < 0.001$). Fig. 8 shows this sublinear growth that

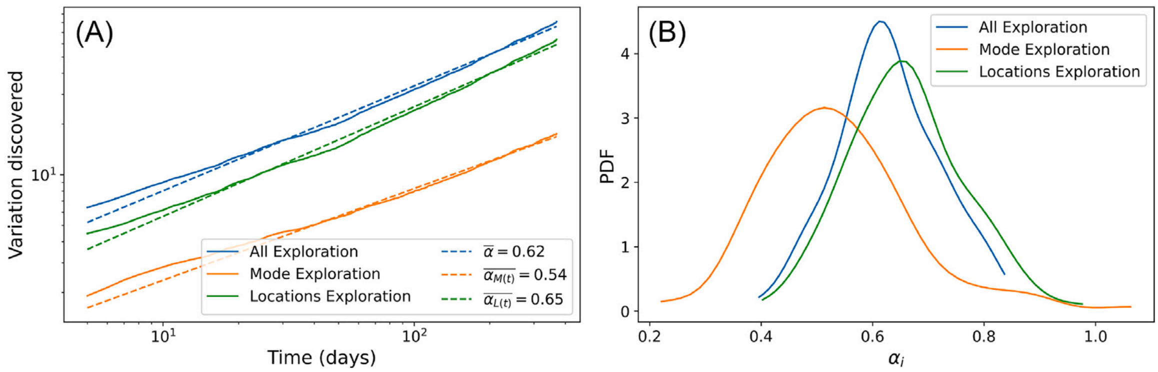

Fig. 8. The speed of exploration. (A) The total number of trip packages added to the behaviour set in time (all exploration). All exploration can be classified into explorations of a previously unobserved location (location exploration) or using a new travel mode to reach an existing location (mode exploration). The solid lines show the mean speed of exploration across participants, and the dashed lines represent a power-law fit with exponent $\overline{\alpha}$. (B) The probability density functions of the individual power-law fit exponent α_i for the different explorations.

slows down over time for the average $\overline{T(t)}$, together with the individual fit exponents a_i that distribute around the average $\overline{\alpha}$. Furthermore, we quantify the overall exploration by segmenting it into *location exploration* using the number of trip packages that arrive at a previously unobserved location ($L^i(t)$) and *mode exploration* using the ones that utilise a new travel mode to an already observed location ($M^i(t)$). Note that similar to $L^i(t)$, $M^i(t)$ has an upper limit that grows with the number of locations explored and is not restricted to the considered classes of travel mode (i.e., 3 in this study). From the empirical data, we find that both $L^i(t)$ and $M^i(t)$ can be well approximated using a power-law fit, with average $\overline{\alpha_{L(t)}} = 0.65$ (95% CI: 0.644 – 0.656, two-sided $P < 0.001$) and average $\overline{\alpha_{M(t)}} = 0.54$ (95% CI: 0.537 – 0.551, two-sided $P < 0.001$), respectively, suggesting that the exploration of locations has a faster speed than the one of travel modes (Fig. 8).

6. Discussion

This study proposes an analytical framework to consider two aspects of individual travel behaviour: travel mode and activity location. We reveal that individuals maintain a fixed size but constantly evolving behaviour set. We quantify the number of this conserved quantity with regard to the considered time scale. Moreover, we find that the behaviour set has a constant updating speed. We further analysed the exploration speed of individuals' activity location and travel mode choices by decomposing this updating process. Note that the conserved sizes of the average behaviour capacity are similar for the GC and MOBIS datasets (see Fig. 7A and C), and the distributions of the individual behaviour capacity are also similar across the two datasets (see Fig. 7B and D), despite the differences in their participants' mobility behaviour (see Section 5.1). This finding suggests that the proposed conserved quantity is robust to the intensity of an individual's mobility and can be applied as an invariant travel behaviour indicator for both active and regular populations. Moreover, these results imply that travel mode exploration is slower than activity location exploration under our definition of explorations, which suggests that individuals more often explore a new location than change the travel mode they take to reach a given location. This observation is in line with previous literature that suggests mode–location combinations are more repetitive than activity-location combinations in individuals' daily mobility (Susilo and Axhausen, 2014; Schlich and Axhausen, 2003). Here, we model the exploration speed of different travel behaviour dimensions over time. Last, our findings are based on the study by Alessandretti et al. (2018), where it is shown that the number of familiar locations an individual visits is a conserved quantity over time. Compared to only focusing on location visits, the method proposed in this study enables the consideration of additional travel behaviour aspects, such as travel mode choices. Hence, our focus is shifted towards an integrated dimension of travel behaviour; thereby, the conserved quantify of the behaviour capacity further confines individual mobility.

This data-driven study quantifies the stability of individuals' long-term travel mode and activity location choices. We believe that this stability is closely connected to the habitual behaviour (Gärling and Axhausen, 2003), or the psychological inertia effects (Gao et al., 2020) in travel choices that have been extensively discussed in previous literature (Gardner, 2009; Gao et al., 2021). Since a behaviour set contains travel behaviours that are more important to the individual, constructing a behaviour set can be understood as extracting the instances of habitual behaviours. Here, our contribution is to identify the magnitude of these habitual behaviours, empirically show their stability over time, and quantify their exploration rate in different dimensions. Moreover, it is widely accepted that individual daily mobility can be decomposed into habitual behaviours or routines (i.e., stability) and variability on the temporal dimension (Schönfelder and Axhausen, 2016; Zhong et al., 2015). From this perspective, the constructed behaviour set can be regarded as a representation of daily mobility stability. The complement set of the behaviour set (i.e., $\mathcal{P}^i \setminus \mathcal{Z}^i$), including the behaviour packages that are unimportant to the individual (i.e., observed only once or travelled less frequently), corresponds to variability. As shown in Fig. 9, the capacity of the complement set of the behaviour set shows an apparent seasonality effect, which coincides with previous reports using human activity space measures (Järv et al., 2014). Therefore, implementing the behaviour set provides a new perspective to analyse the stability and variability of individual mobility.

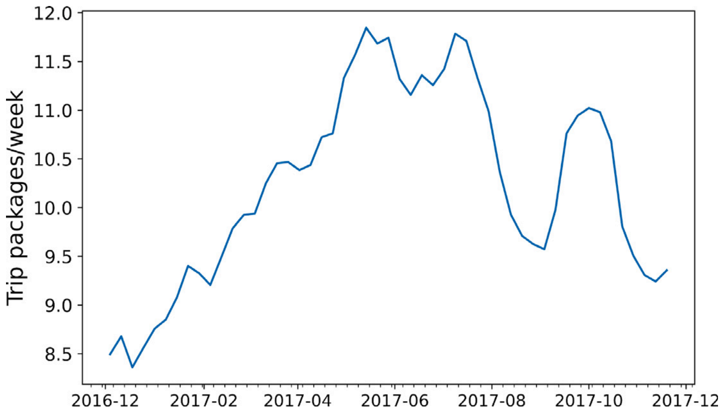

Fig. 9. The seasonality of variety-seeking behaviours. The number of trip packages that are not part of the behaviour set shows a strong seasonality effect.

Additionally, the results of this study can be applied to the formulation of alternative choice sets for choice generation problems, which is vital for many transport-related studies, such as demand prediction and infrastructure planning (Leite Mariante et al., 2018; Yao and Bekhor, 2022). Individuals face a discrete choice problem when choosing locations and travel modes for conducting the next activity. The choice set conventionally includes all alternative locations within an area and all the available travel modes for the individual, which is unrealistic if the choice set is large and likely does not conform to how people search in the real world (Chen et al., 2016). The stability of joint activity location and mode choices and their exploration speeds suggest that certain combinations are more likely than others, limiting the effective choice set of individuals. Based on this evidence, we should further explore joint models for activity location choice modelling (Zolfaghari et al., 2012; Leite Mariante et al., 2018) and travel mode choice modelling (Susilo and Axhausen, 2014) using long-term observations of individuals. The development of such models could improve the efficiency of demand generation for microscopic traffic simulation models (Horni et al., 2009; Lopez et al., 2018; Horni et al., 2016).

7. Conclusion

Understanding the long-term intra-person variability is essential for the comprehension of activity–travel behaviour, yet this line of research has always heavily relied on the employed dataset. Based on two large-scale longitudinal GPS tracking datasets involving ∼3800 individuals, this study proposes an analytical framework to jointly consider travel mode and activity location for modelling their interactions over the long term. First, we calculate various mobility indicators to characterise the datasets and show that the trip and location attributes are strongly correlated for the whole study period. Then, we propose the concept of trip package and behaviour set to capture representative mode–location choices at any given time. We found that the number of these choices is a conserved quantity over time for individuals in both considered datasets, despite their differences in general mobility patterns. A typical individual was observed to maintain ∼15 mode–location choices, of which ∼7 are travelled with a private vehicle every 5 weeks. Last, analysing the dynamics of this stability suggests that these important mode–location choices are constantly evolving. The exploration speed of locations is faster than the one for travel modes, but they can both be well modelled by a sublinear growth that slows down over time. Therefore, we provide the following answer to the proposed research question: *The activity location and travel mode choices of individuals constantly evolve, but the number of the important mode–location options maintains a dynamic balance over the long term, as a result of exploring new modes and locations at different rates.*

Despite the rapid development of travel behaviour and human mobility analysis studies (Chen et al., 2016), the understanding of the long-term intra-person variability has been limited, especially considering multiple aspects of individual activity–travel patterns. In this context, this study offers a new perspective on modelling the interactions between travel mode and activity location choices, and improves our understanding of individuals' travel decision-making process. The observed stability in the activity–travel behaviour implies that the method can be applied to distinguish the stability and variability components of individual mobility. It can also help reduce the choice options when forming long-term alternative choice sets. Moreover, the mode and location exploration rate quantification provides basic statistics for building models to describe long-term travel behaviour evolution.

It should be noted that our results are derived based on two datasets involving sample participants from Switzerland, which are specific to this developed country and cannot represent general movement patterns across all populations over the world. The activity–travel patterns of individuals depend on various location factors, such as lifestyles and transport infrastructure; therefore, the results of this study should be interpreted with care. We note, however, that the mobility indicators of our datasets generally match with previous studies. It would be interesting to compare our results with samples drawn from other parts of the world or different socio-demographic groups as a follow-up study. Moreover, considering the heterogeneous behaviour capacity across the population, future studies could aim to identify subgroups of individuals who share a similar behaviour set composition. This structural understanding of the stable behaviour set will benefit the development of fine-scaled mobility models. Lastly,

Table A.1
The construction of the behaviour set. Example trip packages' total trip duration (minutes) and the number of observations (in parenthesis) for each weekly time window are shown. Trip packages are added to the behaviour set only if they satisfy both the count and duration criteria.

$P^i_{seq}(t,\delta t)$	Weekly observation					Criteria		$\mathcal{B}^i(t,\delta t)$
	t	$t+1$	$t+2$	$t+3$	$t+4$	Count ≥ 2	Duration $> m \cdot \delta t$	
TP1	–	–	42 (1)	–	–	✗	✓	✗
TP2	14 (1)	–	–	35 (3)	13 (1)	✓	✓	✓
TP3	–	–	6 (2)	–	–	✓	✓	✓
TP4	–	–	5 (1)	–	–	✗	✗	✗
TP5	3 (2)	–	–	–	–	✓	✗	✗
TP6	10 (1)	13 (2)	6 (1)	24 (5)	7 (1)	✓	✓	✓

the proposed analytical framework enables straightforward considerations of multiple aspects of travel behaviour simultaneously. Including additional behaviour dimensions, such as route choice and start time of the trip, into the analysis for studying their interaction over time will further enhance our understanding of intra-person variability. Nevertheless, we anticipate that this study will raise attention to comprehensively consider activity–travel patterns in modelling individual mobility and open new possibilities for designing mobility simulation models.

CRediT authorship contribution statement

Ye Hong: Conceptualization, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Henry Martin:** Conceptualization, Methodology, Data curation, Writing – review & editing. **Yanan Xin:** Conceptualization, Methodology, Writing – review & editing. **Dominik Bucher:** Conceptualization, Data curation, Writing – review & editing. **Daniel J. Reck:** Conceptualization, Writing – review & editing. **Kay W. Axhausen:** Conceptualization, Data curation, Writing – review & editing. **Martin Raubal:** Conceptualization, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Raw data for the GC dataset are not publicly available due to confidentiality agreements with the participants under the European General Data Protection Regulation (GDPR). Raw data for the MOBIS dataset are not publicly available due to privacy considerations, but aggregated trip-level data are available to researchers upon request to the corresponding author.

Funding information

This work was supported by the Swiss Data Science Center (SDSC) (grant number C17-14), the Swiss Competence Center for Energy Research - Efficient Technologies and Systems for Mobility (SCCER Mobility), the ETH Zurich Foundation, Switzerland (grant number MI-01-19), the Swiss Competence Center for Energy Research (SCCER CREST - Mobility Joint Activity), the Swiss Federal Roads Office (grant number ASTRA 2017/006), and the Swiss Federal Office of Transport.

Ethical declarations

We hereby confirm that ethics approval for data collection and research in this study was received from the Institute of Cartography and Geoinformation, ETH Zurich and the Institute for Transport Planning and Systems, ETH Zurich. All participants of the study have been acknowledged with the data collection process and delivered informed consent to use data collected from them, confirming that they understand what kind of data will be collected from them and that they can withdraw from the study at any moment. We confirm that all the research methods were carried out in accordance with relevant guidelines and regulations.

Appendix

A.1. Constructing the behaviour set

We present an example to construct a behaviour set from a set of observed trip packages. Here, we consider the same trip packages as in Fig. 3. Their trip duration and the number of observations (both synthesised) starting from time t are presented in Table A.1. With parameters time window $\delta t = 5$ weeks and $m = 1$ min/week, the behaviour set $\mathcal{B}^i(t,\delta t)$ includes trip packages that simultaneously fulfil the count and duration criteria. As a result, $\mathcal{B}^i(t,\delta t)$ contains trip packages TP2, TP3, and TP6.

Table A.2
Fitting parameters of log-normal distributions.

	Jump length		Rg	
	μ	σ	μ	σ
GC	7.534	2.526	6.066	1.437
MOBIS	7.419	2.040	4.003	1.843

Table A.3
Stability for the behaviour capacity (i.e., $C^i(t)$) with altering main mode determination, location definition and time window δt. For main mode determination, *Distance* refers to determining the mode with the largest share of the distance travelled; *Hierarchy* refers to the mode with a predefined mode hierarchy. Location epsilon refers to the epsilon parameter employed in the DBSCAN algorithm to determine the locations. We show the intercept a and slope b of a linear fit in the form of $\overline{C(t)} = a + b \cdot t$, together with the p-value $p(b)$ of the null hypothesis $H_0 : b = 0$. We also report the proportion of individuals whose net gain is consistent with zero (i.e., $\left|\left\langle G^i \right\rangle\right| < \sigma_{G^i}$).

Dataset	Main mode	Location epsilon	Time window δt	Intercept a	Slope b	$p(b)$	$\left\|\left\langle G^i \right\rangle\right\| < \sigma_{G^i}$
GC	Distance	50	4	13.31	$-3.05 \cdot 10^{-3} \pm 5.89 \cdot 10^{-3}$	0.61	100%
GC	Distance	50	5	15.51	$-2.38 \cdot 10^{-3} \pm 6.61 \cdot 10^{-3}$	0.72	100%
GC	Distance	50	6	17.65	$3.31 \cdot 10^{-3} \pm 7.48 \cdot 10^{-3}$	0.66	100%
GC	Distance	50	8	21.53	$6.57 \cdot 10^{-3} \pm 9.32 \cdot 10^{-3}$	0.48	100%
GC	Distance	50	10	25.03	$1.07 \cdot 10^{-2} \pm 1.15 \cdot 10^{-2}$	0.35	100%
GC	Distance	50	15	32.19	$1.68 \cdot 10^{-2} \pm 1.79 \cdot 10^{-2}$	0.35	100%
GC	Distance	50	20	37.97	$1.47 \cdot 10^{-2} \pm 2.70 \cdot 10^{-2}$	0.59	100%
GC	Distance	50	30	46.90	$1.53 \cdot 10^{-2} \pm 6.86 \cdot 10^{-2}$	0.82	100%
GC	Distance	40	5	15.25	$-3.26 \cdot 10^{-3} \pm 6.43 \cdot 10^{-3}$	0.61	100%
GC	Distance	40	8	21.19	$-5.46 \cdot 10^{-3} \pm 9.01 \cdot 10^{-3}$	0.54	100%
GC	Distance	40	10	24.65	$-3.42 \cdot 10^{-3} \pm 1.12 \cdot 10^{-2}$	0.76	100%
GC	Distance	40	20	37.23	$-9.66 \cdot 10^{-3} \pm 2.63 \cdot 10^{-2}$	0.71	100%
GC	Distance	100	5	16.03	$3.55 \cdot 10^{-3} \pm 4.96 \cdot 10^{-3}$	0.47	100%
GC	Distance	200	5	16.11	$4.15 \cdot 10^{-3} \pm 5.23 \cdot 10^{-3}$	0.43	100%
GC	Hierarchy	40	5	15.29	$-5.14 \cdot 10^{-3} \pm 6.45 \cdot 10^{-3}$	0.43	100%
GC	Hierarchy	50	5	15.51	$2.01 \cdot 10^{-3} \pm 6.63 \cdot 10^{-3}$	0.76	100%
MOBIS	Distance	50	4	13.01	$-8.08 \cdot 10^{-3} \pm 1.70 \cdot 10^{-2}$	0.63	95%
MOBIS	Distance	50	5	15.34	$9.98 \cdot 10^{-3} \pm 2.40 \cdot 10^{-2}$	0.68	88%
MOBIS	Distance	50	6	17.54	$2.12 \cdot 10^{-2} \pm 3.37 \cdot 10^{-2}$	0.53	76%
MOBIS	Distance	40	5	14.97	$1.94 \cdot 10^{-2} \pm 2.36 \cdot 10^{-2}$	0.41	89%
MOBIS	Distance	100	5	16.19	$5.25 \cdot 10^{-3} \pm 2.49 \cdot 10^{-2}$	0.83	89%
MOBIS	Distance	200	5	16.28	$-2.61 \cdot 10^{-2} \pm 2.45 \cdot 10^{-2}$	0.29	89%
MOBIS	Hierarchy	40	5	15.00	$2.42 \cdot 10^{-2} \pm 2.36 \cdot 10^{-2}$	0.30	88%
MOBIS	Hierarchy	50	5	15.37	$1.67 \cdot 10^{-2} \pm 2.39 \cdot 10^{-2}$	0.49	89%

A.2. Fitting parameters for radius of gyration and jump length

We consider the following parameter distributions to approximate the empirical distribution of radius of gyration and jump length of the GC and MOBIS datasets:

- The log-normal distribution of a random variable x, with parameter μ and σ, with probability density function:

$$P(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right)$$

- The power-law distribution (i.e., Pareto distribution) of a random variable x, with parameter α and predefined minimum possible value of $x_{\min}$, with probability density function:

$$P(x) = \frac{\alpha x_{\min}^{\alpha}}{x^{\alpha+1}}$$

- The truncated power-law distribution of a random variable x, with parameters α, β and predefined minimum possible value of $x_{\min}$, with probability density function:

$$P(x) = (x + x_{\min})^{-\alpha} exp(-\beta x)$$

The fitting procedure is conducted using the Python package powerlaw (Alstott et al., 2014). Under the AIC criterion, the distributions of the radius of gyration and jump length are both best fitted using the log-normal distribution. The parameters of the best fit are reported in Table A.2.

A.3. Sensitivity analysis for the behaviour capacity stability

This section shows further evidence for the stability of collective and individual behaviour capacity. In Table A.3, we alter the methods to determine the main travel mode of trips, the spatial scale of locations, and the length of the time windows to construct the behaviour set. We find that the behaviour capacity is constant over time for all parameter combinations, and on both collective and individual levels.

References

Ahas, R., Aasa, A., Silm, S., Tiru, M., 2010. Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. Transp. Res. C 18 (1), 45–54. http://dx.doi.org/10.1016/j.trc.2009.04.011.

Alessandretti, L., Aslak, U., Lehmann, S., 2020. The scales of human mobility. Nature 587 (7834), 402–407. http://dx.doi.org/10.1038/s41586-020-2909-1.

Alessandretti, L., Sapiezynski, P., Lehmann, S., Baronchelli, A., 2017. Multi-scale spatio-temporal analysis of human mobility. PLoS One 12 (2), e0171686. http://dx.doi.org/10.1371/journal.pone.0171686.

Alessandretti, L., Sapiezynski, P., Sekara, V., Lehmann, S., Baronchelli, A., 2018. Evidence for a conserved quantity in human mobility. Nat. Hum. Behav. 2 (7), 485–491. http://dx.doi.org/10.1038/s41562-018-0364-x.

Alstott, J., Bullmore, E., Plenz, D., 2014. Powerlaw: A python package for analysis of heavy-tailed distributions. PLoS One 9 (1), e85777. http://dx.doi.org/10.1371/journal.pone.0085777.

Axhausen, K.W., 2006. Moving through nets: An introduction. In: Moving Through Nets: the Physical and Social Dimensions of Travel. Elsevier, Oxford, pp. 1–7. http://dx.doi.org/10.3929/ethz-a-005226264.

Axhausen, K.W., 2007. Definition of movement and activity for transport modelling: Contribution to the handbooks in transport: transport modelling. Handb. Transp. Model.: Transp. Model. 1, 329–343. http://dx.doi.org/10.3929/ethz-a-005278091.

Bao, Y., Huang, Z., Li, L., Wang, Y., Liu, Y., 2021. A BiLSTM-CNN model for predicting users' next locations based on geotagged social media. Int. J. Geogr. Inf. Sci. 35 (4), 639–660. http://dx.doi.org/10.1080/13658816.2020.1808896.

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

Barbosa, H., de Lima-Neto, F.B., Evsukoff, A., Menezes, R., 2015. The effect of recency to human mobility. EPJ Data Sci. 4 (1), 21. http://dx.doi.org/10.1140/epjds/s13688-015-0059-8.

Brockmann, D., Hufnagel, L., Geisel, T., 2006. The scaling laws of human travel. Nature 439 (7075), 462–465. http://dx.doi.org/10.1038/nature04292.

Bucher, D., Mangili, F., Cellina, F., Bonesana, C., Jonietz, D., Raubal, M., 2019. From location tracking to personalized eco-feedback: A framework for geographic information collection, processing and visualization to promote sustainable mobility behaviors. Travel Behav. Soc. 14, 43–56. http://dx.doi.org/10.1016/j.tbs.2018.09.005.

Bucher, D., Martin, H., Hamper, J., Jaleh, A., Becker, H., Zhao, P., Raubal, M., 2020. Exploring factors that influence individuals' choice between internal combustion engine cars and electric vehicles. AGILE: GISci. Ser. 1, 1–23. http://dx.doi.org/10.5194/agile-giss-1-2-2020.

Chen, C., Ma, J., Susilo, Y., Liu, Y., Wang, M., 2016. The promises of big data and small data for travel behavior (aka human mobility) analysis. Transp. Res. C 68, 285–299. http://dx.doi.org/10.1016/j.trc.2016.04.005.

Cherchi, E., Cirillo, C., 2014. Understanding variability, habit and the effect of long period activity plan in modal choices: a day to day, week to week analysis on panel data. Transportation 41 (6), 1245–1262. http://dx.doi.org/10.1007/s11116-014-9549-y.

Dharmowijoyo, D.B.E., Susilo, Y.O., Karlström, A., 2016. Day-to-day variability in travellers' activity-travel patterns in the Jakarta metropolitan area. Transportation 43 (4), 601–621. http://dx.doi.org/10.1007/s11116-015-9591-4.

Dharmowijoyo, D.B., Susilo, Y.O., Karlström, A., 2017. Analysing the complexity of day-to-day individual activity-travel patterns using a multidimensional sequence alignment model: A case study in the Bandung Metropolitan Area, Indonesia. J. Transp. Geogr. 64, 1–12. http://dx.doi.org/10.1016/j.jtrangeo.2017.08.001.

Egu, O., Bonnel, P., 2020. Investigating day-to-day variability of transit usage on a multimonth scale with smart card data. A case study in Lyon. Travel Behav. Soc. 19, 112–123. http://dx.doi.org/10.1016/j.tbs.2019.12.003.

Gao, K., Shao, M., Axhausen, K.W., Sun, L., Tu, H., Wang, Y., 2021. Inertia effects of past behavior in commuting modal shift behavior: interactions, variations and implications for demand estimation. Transportation 49 (4), 1063–1097. http://dx.doi.org/10.1007/s11116-021-10203-6.

Gao, K., Yang, Y., Sun, L., Qu, X., 2020. Revealing psychological inertia in mode shift behavior and its quantitative influences on commuting trips. Transp. Res. F 71, 272–287. http://dx.doi.org/10.1016/j.trf.2020.04.006.

Gardner, B., 2009. Modelling motivation and habit in stable travel mode contexts. Transp. Res. F 12 (1), 68–76. http://dx.doi.org/10.1016/j.trf.2008.08.001.

Gärling, T., Axhausen, K.W., 2003. Introduction: Habitual travel choice. Transportation 30 (1), 1–11. http://dx.doi.org/10.1023/A:1021230223001.

Golledge, R.G., Stimson, R.J., 1997. Spatial Behavior: A Geographic Perspective. Guilford Press.

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

Goulet-Langlois, G., Koutsopoulos, H.N., Zhao, Z., Zhao, J., 2018. Measuring regularity of individual travel patterns. IEEE Trans. Intell. Transp. Syst. 19 (5), 1583–1592. http://dx.doi.org/10.1109/TITS.2017.2728704.

Hägerstrand, T., 1970. What about people in regional science? Pap. Reg. Sci. Assoc. 24 (1), 6–21. http://dx.doi.org/10.1007/BF01936872.

Hanson, S., Huff, O.J., 1988. Systematic variability in repetitious travel. Transportation 15 (1), 111–135. http://dx.doi.org/10.1007/BF00167983.

Hariharan, R., Toyama, K., 2004. Project lachesis: Parsing and modeling location histories. In: Geographic Information Science. Springer, pp. 106–124. http://dx.doi.org/10.1007/978-3-540-30231-5_8.

Heinen, E., Chatterjee, K., 2015. The same mode again? An exploration of mode choice variability in great britain using the national travel survey. Transp. Res. A 78, 266–282. http://dx.doi.org/10.1016/j.tra.2015.05.015.

Hong, Y., Xin, Y., Martin, H., Bucher, D., Raubal, M., 2021. A clustering-based framework for individual travel behaviour change detection. In: 11th International Conference on Geographic Information Science (GIScience 2021) - Part II, Vol. 208. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, p. 4. http://dx.doi.org/10.4230/LIPIcs.GIScience.2021.II.4.

Horni, A., Nagel, K., Axhausen, K.W. (Eds.), 2016. The Multi-Agent Transport Simulation MATSim. Ubiquity Press, http://dx.doi.org/10.5334/baw.

Horni, A., Scott, D.M., Balmer, M., Axhausen, K.W., 2009. Location choice modeling for shopping and leisure activities with MATSim: Combining microsimulation and time geography. Transp. Res. Rec. 2135 (1), 87–95. http://dx.doi.org/10.3141/2135-11.

Huang, H., Gartner, G., Krisp, J.M., Raubal, M., Van de Weghe, N., 2018. Location based services: ongoing evolution and research agenda. J. Locat. Based Serv. 12 (2), 63–93. http://dx.doi.org/10.1080/17489725.2018.1508763.

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

Jones, P., Clarke, M., 1988. The significance and measurement of variability in travel behaviour. Transportation 15 (1), 65–87. http://dx.doi.org/10.1007/BF00167981.

Kang, H., Scott, D.M., 2010. Exploring day-to-day variability in time use for household members. Transp. Res. A 44 (8), 609–619. http://dx.doi.org/10.1016/j.tra.2010.04.002.

Leite Mariante, G., Ma, T.-Y., Van Acker, V., 2018. Modeling discretionary activity location choice using detour factors and sampling of alternatives for mixed logit models. J. Transp. Geogr. 72, 151–165. http://dx.doi.org/10.1016/j.jtrangeo.2018.09.003.

Liu, S., Yamamoto, T., Yao, E., Nakamura, T., 2022. Exploring travel pattern variability of public transport users through smart card data: Role of gender and age. IEEE Trans. Intell. Transp. Syst. 23 (5), 4247–4256. http://dx.doi.org/10.1109/TITS.2020.3043021.

Lopez, P.A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., Wiessner, E., 2018. Microscopic traffic simulation using SUMO. In: 2018 21st International Conference on Intelligent Transportation Systems (ITSC). pp. 2575–2582. http://dx.doi.org/10.1109/ITSC.2018.8569938.

Lu, X., Wetter, E., Bharti, N., Tatem, A.J., Bengtsson, L., 2013. Approaching the limit of predictability in human mobility. Sci. Rep. 3 (1), 2923. http://dx.doi.org/10.1038/srep02923.

Martin, H., Becker, H., Bucher, D., Jonietz, D., Raubal, M., Axhausen, K.W., 2019. Begleitstudie SBB green class - abschlussbericht. Arbeitsberichte Verk. Raumplanung 1439, http://dx.doi.org/10.3929/ethz-b-000353337.

Martin, H., Hong, Y., Wiedemann, N., Bucher, D., Raubal, M., 2022. Trackintel: An open-source python library for human mobility analysis. URL https://arxiv.org/abs/2206.03593.

Minnen, J., Glorieux, I., van Tienoven, T.P., 2015. Transportation habits: Evidence from time diary data. Transp. Res. A 76, 25–37. http://dx.doi.org/10.1016/j.tra.2014.12.013.

Molloy, J., Castro, A., Götschi, T., Schoeman, B., Tchervenkov, C., Tomic, U., Hintermann, B., Axhausen, K.W., 2022. The MOBIS dataset: a large GPS dataset of mobility behaviour in Switzerland. Transportation 1–25. http://dx.doi.org/10.1007/s11116-022-10299-4.

Pas, E.I., 1987. Intrapersonal variability and model goodness-of-fit. Transp. Res. A 21 (6), 431–438. http://dx.doi.org/10.1016/0191-2607(87)90032-X.

Pas, E.I., Sundar, S., 1995. Intrapersonal variability in daily urban travel behavior: Some additional evidence. Transportation 22 (2), 135–150. http://dx.doi.org/10.1007/BF01099436.

Raux, C., Ma, T.-Y., Cornelis, E., 2016. Variability in daily activity-travel patterns: the case of a one-week travel diary. Eur. Transp. Res. Rev. 8 (4), 1–14. http://dx.doi.org/10.1007/s12544-016-0213-9.

Schläpfer, M., Dong, L., O'Keeffe, K., Santi, P., Szell, M., Salat, H., Anklesaria, S., Vazifeh, M., Ratti, C., West, G.B., 2021. The universal visitation law of human mobility. Nature 593 (7860), 522–527. http://dx.doi.org/10.1038/s41586-021-03480-9.

Schlich, R., Axhausen, K.W., 2003. Habitual travel behaviour: Evidence from a six-week travel diary. Transportation 30 (1), 13–36. http://dx.doi.org/10.1023/A:1021230507071.

Schönfelder, S., Axhausen, K.W., 2003. Activity spaces: measures of social exclusion? Transp. Policy 10 (4), 273–286. http://dx.doi.org/10.1016/j.tranpol.2003.07.002.

Schönfelder, S., Axhausen, K.W., 2016. Urban Rhythms and Travel Behaviour: Spatial and Temporal Phenomena of Daily Travel. Routledge, http://dx.doi.org/10.4324/9781315548715.

Song, C., Koren, T., Wang, P., Barabási, A.-L., 2010a. Modelling the scaling properties of human mobility. Nat. Phys. 6 (10), 818–823. http://dx.doi.org/10.1038/nphys1760.

Song, C., Qu, Z., Blumm, N., Barabasi, A.-L., 2010b. Limits of predictability in human mobility. Science 327 (5968), 1018–1021. http://dx.doi.org/10.1126/science.1177170.

Stanley, K., Yoo, E.-H., Paul, T., Bell, S., 2018. How many days are enough?: capturing routine human mobility. Int. J. Geogr. Inf. Sci. 32 (7), 1485–1504. http://dx.doi.org/10.1080/13658816.2018.1434888.

Susilo, Y.O., Axhausen, K.W., 2014. Repetitions in individual daily activity-travel-location patterns: a study using the Herfindahl-Hirschman index. Transportation 41 (5), 995–1011. http://dx.doi.org/10.1007/s11116-014-9519-4.

Susilo, Y.O., Kitamura, R., 2005. Analysis of day-to-day variability in an individual's action space: Exploration of 6-week mobidrive travel diary data. Transp. Res. Rec. 1902 (1), 124–133. http://dx.doi.org/10.1177/0361198105190200115.

Tang, J., Liu, F., Wang, Y., Wang, H., 2015. Uncovering urban human mobility from large scale taxi GPS data. Physica A 438, 140–153. http://dx.doi.org/10.1016/j.physa.2015.06.032.

Wang, Z., He, S.Y., Leung, Y., 2018. Applying mobile phone data to travel behaviour research: A literature review. Travel Behav. Soc. 11, 141–155. http://dx.doi.org/10.1016/j.tbs.2017.02.005.

Yao, R., Bekhor, S., 2022. A variational autoencoder approach for choice set generation and implicit perception of alternatives in choice modeling. Transp. Res. B 158, 273–294. http://dx.doi.org/10.1016/j.trb.2022.02.015.

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 (8), 1594–1621. http://dx.doi.org/10.1080/13658816.2016.1143555.

Yuan, Y., Raubal, M., Liu, Y., 2012. Correlating mobile phone usage and travel behavior - a case study of Harbin, China. Comput. Environ. Urban Syst. 36 (2), 118–130. http://dx.doi.org/10.1016/j.compenvurbsys.2011.07.003.

Zhang, A., Kang, J.E., Axhausen, K., Kwon, C., 2018. Multi-day activity-travel pattern sampling based on single-day data. Transp. Res. C 89, 96–112. http://dx.doi.org/10.1016/j.trc.2018.01.024.

Zhao, K., Musolesi, M., Hui, P., Rao, W., Tarkoma, S., 2015. Explaining the power-law distribution of human mobility through transportation modality decomposition. Sci. Rep. 5 (1), 9136. http://dx.doi.org/10.1038/srep09136.

Zhong, C., Manley, E., Müller Arisona, S., Batty, M., Schmitt, G., 2015. Measuring variability of mobility patterns from multiday smart-card data. J. Comput. Sci. 9, 125–130. http://dx.doi.org/10.1016/j.jocs.2015.04.021.

Zolfaghari, A., Sivakumar, A., Polak, J.W., 2012. Choice set pruning in residential location choice modelling: a comparison of sampling and choice set generation approaches in greater London. Transp. Plan. Technol. 35 (1), 87–106. http://dx.doi.org/10.1080/03081060.2012.635420.