MAKING LARGE-SCALE SEMI-PASSIVE GPS TRAVEL DIARIES VALUABLE: A QUALITY ENHANCEMENT METHOD

Victoria Dahmen, Corresponding Author
Chair of Traffic Engineering and Control
Technical University of Munich
Arcisstr. 21, 80333, Munich, Germany
Email: v.dahmen@tum.de

Santiago Álvarez-Ossorio Martinez, Corresponding Author
Chair of Traffic Engineering and Control
Technical University of Munich
Arcisstr. 21, 80333, Munich, Germany
Email: santiago.alvarez@tum.de

Allister Loder
Chair of Traffic Engineering and Control
Technical University of Munich
Arcisstr. 21, 80333, Munich, Germany
Email: allister.loder@tum.de

Klaus Bogenberger
Chair of Traffic Engineering and Control
Technical University of Munich
Arcisstr. 21, 80333, Munich, Germany
Email: klaus.bogenberger@tum.de

Word Count: 6400 words + 2 table(s) $\times$ 250 + 600 words for references = 7500 words

Submission Date: July 31st 2023

Paper accepted for presentation at the 103rd Transportation Research Board Annual Meeting, Washington D.C., January 2024

ABSTRACT

The last decade has seen a growing interest in *semi-passive travel diaries*. These diaries are characterized, in contrast to *fully-passive* ones, by the active validation and correction by the participants of automatically-generated trips. Albeit promising and with important benefits in terms of cost, scalability, and trip-recall quality, these diaries still face challenges resulting from data collection errors and imperfect validation by users. In an aim to become an integral part of Household Travel Surveys, it is essential to develop a method for enhancing the quality of these diaries, increasing their reliability, correctness, and usability in further mobility analyses, however, such methodology has yet to be discussed in the literature. In long-duration studies one can prioritize quality over quantity, due to the sheer amount of data, to yield a highly meaningful sample.

In this paper, we present a data quality enhancement method for large-scale long-duration semi-passive travel diaries that targets erroneous records (noise, or from poor validation), enriches the data (e.g., trip and tour detection) and adds supplementary information. We demonstrate its benefits when applied to a one-year study with over a thousand participants. Furthermore, we share our experience working with this unique data and provide insights about the participants' behavior in validation and app interaction that could be of interest for the design of future studies. The output of the proposed method is a meaningful design agnostic dataset; hence facilitating further mobility data analyses. We further recommend that future studies promote active correction and validation by the user.

Keywords: semi-passive travel diaries, data processing, travel behavior, tracking data

INTRODUCTION

For decades, transportation researchers have sought to understand and measure individuals' travel behavior. For this purpose, they have traditionally relied –and still mostly do– on *active solicitation* data (i.e., where the subjects of the study self-report their activities and *trips* by means of a questionnaire or interview). The methods and tools employed to collect these data, also known as *memory-based travel diaries*, have evolved, shifting from paper surveys and in-person interviews to Computer-Aided Telephone/Personal Interviews (*CATI/CAPI)* and Computer-Assisted Self Interviews (*CASI*) (*1, 2*). These advancements enabled faster, cheaper, and more accurate data collection, but still involved very high costs and effort from the surveying agency or institution. In the mid 80s research began to explore the collection of passive tracking data (i.e., without the direct intervention of the subjects, who are just asked to carry a GPS-logger or install it in their vehicles) (*3*). Nevertheless, the inflection point in the use of passive tracking data was the popularization of smartphone devices equipped with GPS antennas. Not only does this enable more extensive and easily-scalable studies at lower costs, as data can be generated without special equipment, but the additional motion sensors can provide valuable information for detecting movement patterns.

The raw data obtained from passive tracking devices/apps typically consists of a sequence of coordinates and matching timestamps, collected at (ir)regular intervals depending on the specific smartphone and operating system (to conserve battery, the GPS sensor is usually software triggered), privacy set-up, and even battery-saving mode, but they lack any contextual information. This is, whether an instantaneous observation (a pair of XY coordinates and a timestamp) corresponds to a static activity –and its purpose– or to a movement – and the employed mode.

In theory, as we will discuss in the literature review, it is possible to detect whether an individual is moving, and predict which transport mode is used, or if the user is static, and impute the purpose of the *stay*, thus generating *fully-automated* –also known as *fully-passive*– travel diaries. However, in practice, the complexity and heterogeneity of human travel patterns (*4*), and GPS noise often lead to erroneous results (e.g., the segmentation of one single *stay* into multiple, disconnected shorter *stays*). In most cases –and due to the lack of ground truth to assess their quality– these diaries may be inadequate as input for further mobility behavior analysis or modeling.

As a result of these limitations, an alternative approach seeking to combine the benefits of passively-generated travel diaries and traditional *CASI* surveys is gaining relevance among researchers and practitioners: *semi-automated* –or *semi-passive*– travel diaries (*2*). Concisely, this solution consists of 1) recording passively the movement of individuals using the smartphone's GPS, 2) automatically generating draft travel diaries (whose complexity depends on the specific algorithms implemented), and 3) asking the participants to review, correct, and validate the draft using an app or online platform. Thus, in comparison to traditional travel surveys, the workload for the participant is significantly reduced, short *trips* can be successfully recorded (mitigating recall errors), *trip* duration and lengths are accurately retrieved, and precise *stay* locations and *trip* routes can be collected (*2, 5, 6*).

Multiple pilot studies with dedicated apps have been conducted in different countries during the last decade (*7–11*), which have mostly focused on discussing aspects such as the app design, recruitment process, accuracy of *trip* and mode-choice detection, and comparing the overall results with existing travel surveys. However, to the best of our knowledge, the literature providing insights on how to enhance the quality of the data obtained from these apps and identify valuable observations –particularly for long-duration studies– is scarce (perhaps because most studies employed proprietary software). We believe this deserves to be studied, as *semi-automated* travel diaries, albeit promising and of better quality than *fully-automated* travel diaries, still face specific challenges derived from the imperfect validation by the users (*2*) and errors in the data collection. Additionally, for the sheer amounts of data recorded in large-scale studies it is not feasible to manually correct the recordings. For this reason, it is important to perform data quality enhancement on these data to improve the correctness and usability of the travel diaries for further mobility analyses. This also entails removing noisy and irreparable data, as due to the scale of such long-duration studies, quality is preferred over quantity.

In this paper, we share our experience working with a long-duration, large-scale *semi-passive travel diary* dataset, detail a data quality enhancement method, and present insights for others dealing with similar data. The data is obtained in the context of the *Mobilität.Leben* project (*12*). With a total of 1,192 participants tracked over 13 months, this study faced unprecedented challenges due to its large size and duration (comparable studies to date rarely exceed two months, as we will see in the literature review). Importantly, this paper does not intend to provide an overall discussion of the project (design, recruitment, analysis of the mobility behavior, etc.). Rather, this paper contributes with a method for enhancing the quality of such long-duration *semi-passive* travel surveys; this an essential step that improves the suitability and relevance of the data source for further analyses. Nevertheless, we also make recommendations and provide learnings about the participants' behavior that could be of interest for the design of future studies (e.g., the elapsed time until participants validate their trips, the amount of users who remain active *validators* during the project, and the amount who abandon the project).

The paper is structured as follows. We first provide a brief introduction to the automatic generation of travel diaries and the *Mobilität.Leben* study. Then we present our methodological framework and discuss the results of its application to our dataset. After discussing the improvements of the diaries, we finally provide recommendations and insights on implications for future studies.

BACKGROUND

(*Semi-*)Automated generation of travel diaries

For decades, the automatic generation of GPS-based travel diaries has been a popular field of research in the transportation and geoinformatics fields (*2, 3*). This is a complex process involving a multitude of steps, which have been widely discussed in literature. In this section, we provide a concise review of the topic and introduce relevant fundamental concepts. The interested reader is referred to the cited references.

The process begins by recording the participant's location using a GPS receiver. This location is intrinsically noisy, particularly in dense urban areas due to the *canyon effect*, so filtering outliers and smoothing are necessary. Nowadays most studies rely on the private smartphones of the participants, which introduces a critical trade-off: battery consumption vs. tracking accuracy. This can be partly addressed by employing the device's accelerometer to avoid reading the GPS position when the device is static (*13*). Once a trajectory (a sequence of coordinates and their timestamp) is recorded, it is segmented into –static– *stays* and –dynamic– *moves* using heuristic rules or data-driven methods (*14*). Then, the travel mode of a *move* can be detected based on the speed, acceleration, transport network, distance between observations, etc. (*15*). Likewise, but with poorer accuracy, the *stay* purpose can be imputed employing attributes such as the land-use information, duration, and time of day (*16*).

Studies involving semi-passive mobility tracking apps rarely exceed the duration of two months (*7–10, 17*), while Molloy et al. (11) conducted the initially 8-week MOBIS study with 3,680 users, but many continued to use the app for more than a year. Similarly, the on-going Lake Geneva Sustainability Panel, conducted by EPFL, will also track approximately 2,500 participants for three weeks (*18*). A key learning from comparative studies employing both passive tracking apps and traditional survey methods is that short trips are underreported in the latter (*9*) and that there is a high diversity between phones (*19*). While the aforementioned studies compare various experimental set-ups and recruiting methods (*8*), or app design (*2, 7*), the processing and enhancement of the data and its errors are rarely discussed.

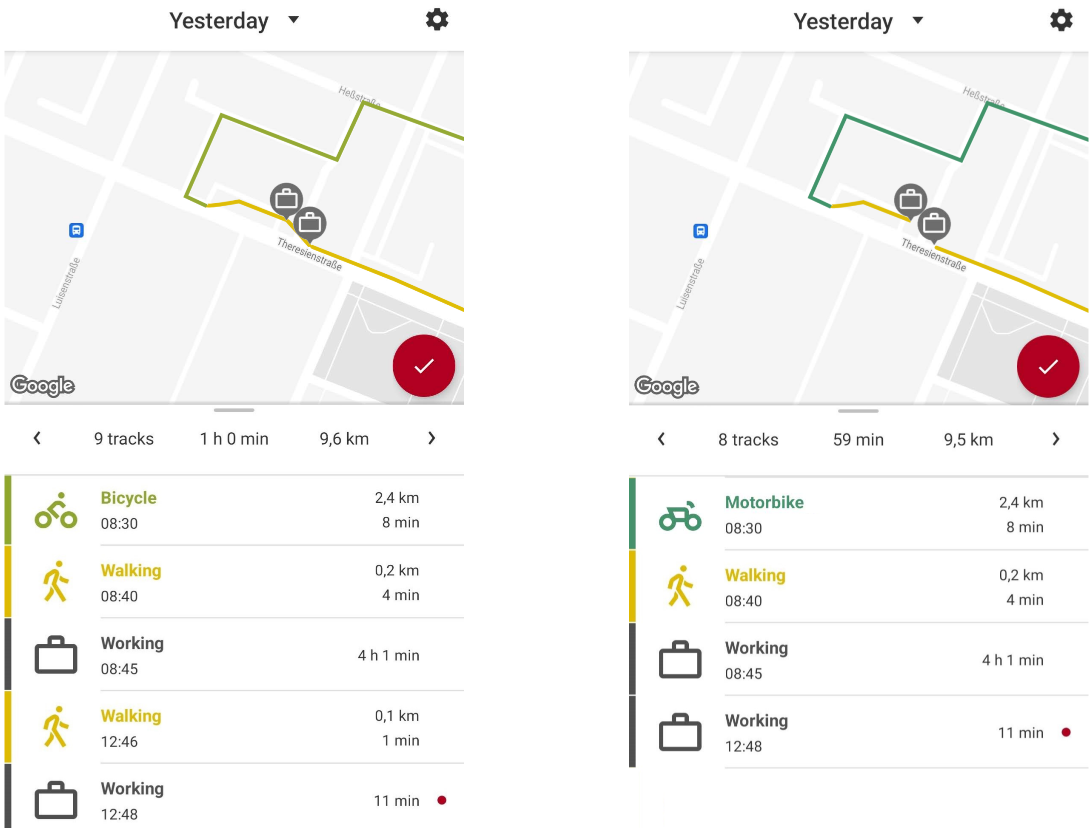
Widely-acknowledged public libraries for the analysis of spatio-temporal tracking data are available in different programming languages (*20, 21*). However, these libraries use the raw tracking data as input and do not assume the availability of user-validated information (i.e., *semi-passive travel diaries*). In practice, most research agencies do not have the expertise nor the resources to conduct the whole process, from app development and data collection to travel behavior analysis. Therefore, we expect that many will employ proprietary software to generate the *semi-passive travel diaries*. Thus, we propose a processing method that builds upon *user-validated travel diaries*, hence addressing the gap in literature, and demonstrate its benefits when applied to a unique long-duration dataset.

The *Mobilität.Leben* project

In spring 2022, the German parliament passed an amendment allowing the use of local Public Transport (PT) for a fee of 9 euros per month between June and August. The so-called 9-Euro-Ticket was valid throughout Germany with the exception of long-distance rail services. In this unprecedented context, the *Mobilität.Leben* project was initiated to study the impacts on travel behavior and evaluate the effectiveness of transport policy instruments (*12*). Initially conceived to last until early Autumn 2022, the study was extended into 2023 when the successor ticket –the 49-Euro *Deutschlandticket*, starting in May 2023– was announced (additional participants were recruited to compensate for those who abandoned after the first phase). In total, the data collection lasted for 13 months, and in this paper we report on the currently-available first 12 months.

The study included a multi-wave survey with 2,569 participants (collecting mobility tool ownership, socio-economic, attitudinal, and travel behavior information). Besides, a subset of 1,192 respondents –most of them living in the Munich region– installed a GPS-based tracking app in their smartphones (available for *Android* and *iOS*), which recorded their movements and *stays* and generated a *fully-passive travel diary*. Individuals responding to all surveys and recording data for more than a week received a monetary incentive. In the app, participants could visualize their diaries and –partially– edit them. It was possible to modify the automatically-detected transport mode, merge consecutive *tracks*, select the purpose of *stays*, or remove incorrect *tracks/stays*. If a participant did not open the app in five days, they received a daily pop-up notification. The app learns the purpose of previously annotated locations, otherwise the default *unknown* purpose is assigned. In Figure 1, we illustrate a travel diary before and after modifying the transport mode and removing an erroneously detected walk. The *Mobilität.Leben* app was developed by *Motiontag* and is similar to those used in other research projects such as Molloy et al. (11). Importantly, the way participants were recruited does not ensure a fully representative sample of the region's population.

Hereafter, the following nomenclature will be employed when discussing the components of the travel diaries (as illustrated in Figure 2). An *activity* is a generic term to refer to any obser-



(a) *Fully-passive travel diary* (b) *Semi-passive travel diary* after user's correction

FIGURE 1: *Mobilität.Leben*'s app *track* validation interface in Android

vation in the raw data (i.e., *track* or *stay*). A *track* (popularly called *tripleg* or stage) is a movement of a user by a single mode of transport. A *stay* corresponds to a static *activity* with a given purpose. A *trip* is a set of *tracks* and wait-*stays* between two consecutive non-wait-*stays*. A *tour* is a set of consecutive *trips* that begin and end with a home *stay*. *Tracks* are associated to one of 20 possible modes (including different PT, private, sharing, and active modes). 15 possible purposes are allowed for *stays* (work, home, errands, leisure, etc). Each *trip* has a main-mode (that used for the longest distance, as employed in (*22*)).

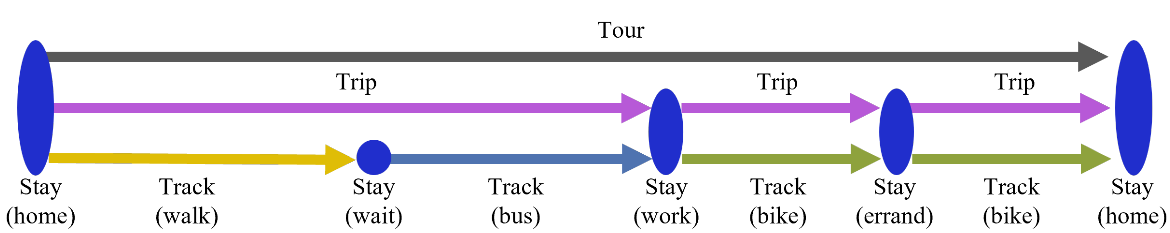


FIGURE 2: Schematic illustration of the terminology used in this work.

METHODOLOGICAL FRAMEWORK

In the following section, our data quality enhancement method for *semi-passive travel diaries* generated by the *Mobilität.Leben* app will be introduced. Our overall objective is to detect and either correct or remove errors in the data (often resulting from poor user validation) and enrich the dataset by integrating relevant external data sources. This results in a dataset that is smaller in size, yet qualitatively superior and richer in information, hence increasing its value for mobility analyses. The design agnostic output facilitates easy-to-use and custom data selection at a range of levels: stage (*track*), *trip* or *tour*-based (*1*).

In Figure 3, we provide an overview of the *Mobilität.Leben* project data processing approach, from the sensor data collection to the final output data. The first component, the data collection and *trip* diary generation, spans from the raw trajectory acquisition to the generation of *semi-passive travel diaries*, as discussed in the Background section. The second component, the data quality enhancement of the *semi-passive travel diaries*, is the focus of this paper and will be explained in detail in the coming paragraphs. Finally, the third component, includes the possible applications of the resulting data in future studies.

Our quality enhancement method consists of three stages: *cleansing and processing*, *data enrichment*, and *integration of external data sources and assessment of tracking-quality*. Each of these stages integrates multiple steps, whose purpose and basic functioning will be described in the corresponding paragraphs. Some of these steps are based on previous studies and consider specific thresholds (e.g., the maximum allowed speed of a bike *track* to deem it valid). For the sake of brevity and ease of reading, we summarize all relevant thresholds values with their source and explanation in Table 1.

The ***cleansing and processing*** stage seeks to perform basic sanity checks on the *semi-passive travel diaries* provided by the *Mobilität.Leben* app, detect anomalous observations, and correct/remove them. In a study of small size and short duration, or in a large one with enormous resources, it would be possible to hire human "reviewers" to analyze the diaries of each user and correct potential errors. However, this approach becomes untenable when hundreds –or thousands– of users are monitored for long periods of time, requiring an automated method.

Thus, we begin by removing *tracks* whose average speed is over a transport-mode-specific threshold. These could result from the erroneous transport mode assignment or from tracking failures, both being observed in the data. Then, *tracks* with excessively short/long duration are also detected and removed. Short *tracks* are often present in our dataset when participants move within buildings (e.g., at work) and long *tracks* (in relation to the traveled distance and the employed mode) happen –seldom– when the app fails to detect a *stay* and considers an individual as moving although he/she is in the same location for several hours/days. Removing such short *tracks* often leads to unconnected, consecutive *stay* locations (i.e., two *stays* with the same purpose, almost at the same location, but with a short temporal gap between them). We address this by detecting and merging these consecutive *stays*. If a short *stay* without annotated purpose is observed immediately (in space and time) before a PT *track*, the main purpose of this *stay* is imputed as waiting (importantly, the app is sensitive to small movements and it can detect very short walks; e.g., from a supermarket to the bus stop in front of it). The benefit of wait imputation is that we can detect more "real" *trips* (i.e., from origin to destination, without fictitious intermediate stops). For a similar reason, if an abnormally short *stay* has no annotated purpose, we remove it from the diary. This can lead to the risk of eliminating some real, short *stays*, but given the lack of cooperation from the user, we prioritize *trip* completeness. To compensate for the possible tracking gaps created in the

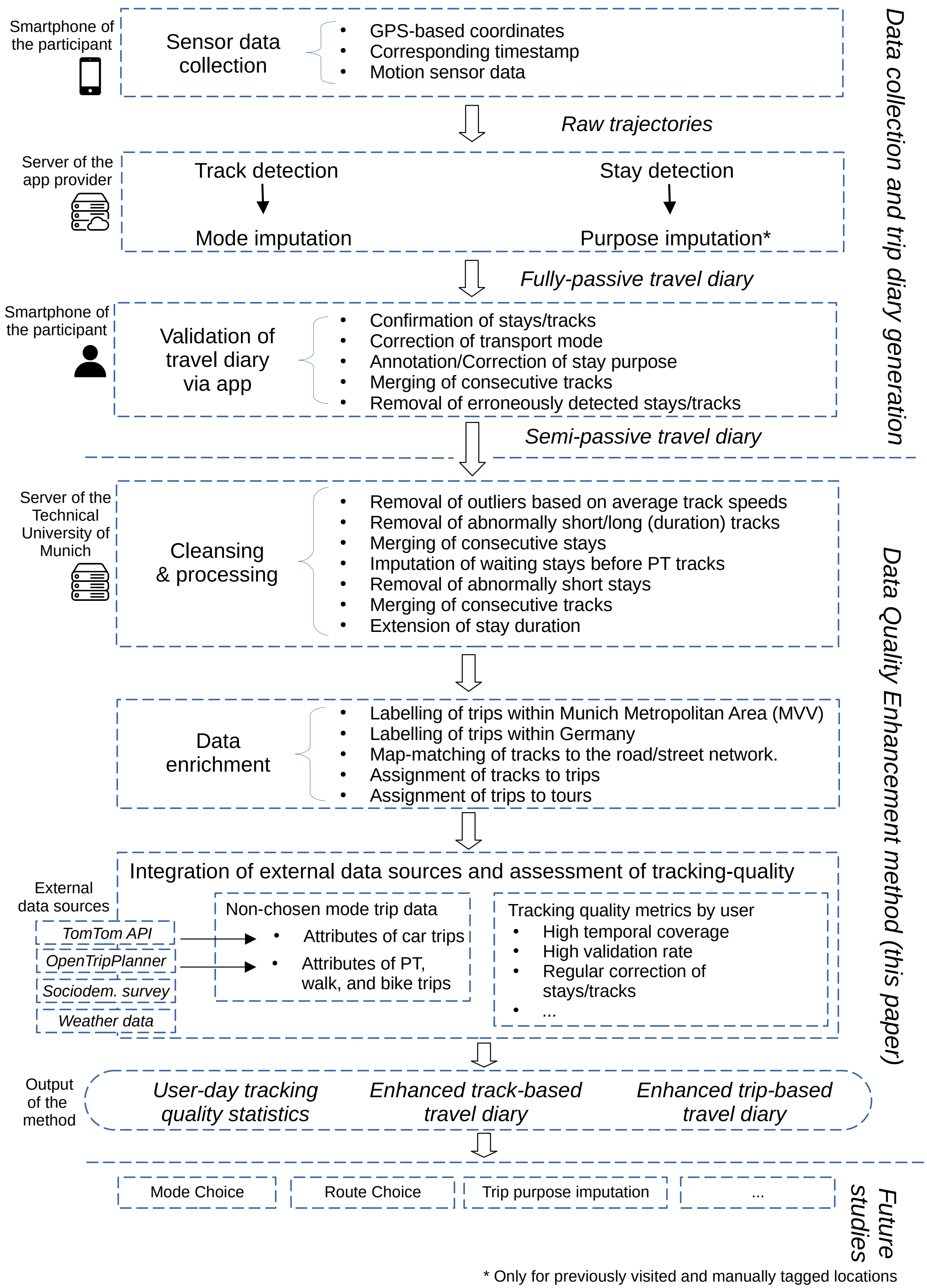


FIGURE 3: Overview of the proposed methodological framework

TABLE 1: Threshold parameters employed in our data quality enhancement method

Threshold parameter	**Values and Reference/Justification**
Cleansing and processing	
Max. average speed per *track* to consider it an outlier	Mode-dependent (99th percentile in the dataset) For instance Bike 28 km/h; Tram: 39 km/h
Abnormally short *tracks* between *stays* with same purpose	Straight distance O/D <100m and *Track* duration <3 min Similar approach and thresholds as (*20, 23, 24*)
Abnormally long or slow tracks	Mode-dependent minimum average *track* speed and/or maximum *track* duration (both 99th percentile)
Max. allowed gap to merge two stays without intermediate tracks	Straight distance <50 m to account for GPS noise (*23*) and Temporal gap <4 min (to account for abnormally short *tracks*)
Parameters for waiting imputation of *stays* with unknown purpose	Immediately before PT *track* and max. duration = 5 min. Based on the duration of *stays* with annotated wait purpose.
Abnormally short *stays*	Min. duration of stay = 5 min (similar to (*24*))
Merging of consecutive *tracks*	Same transport mode, max. 50 m gap between *tracks* (*23*), and max. 4 min gap between *tracks* (shorter than in (*20*))
Extension of *stay* duration until the beginning/end of the following/previous *track*	Max. 100 m gap from O/D of *track* and stay (to account for *cold start* issues (*24*)) and max. 72 h duration of the *stay* (to consider multi-day *stays*).
Data enrichment	
Max allowed gap between *activities* in a *trip*	Max. 5 min temporal gap and 75 m spatial gap between *tracks* or*tracks* and wait *stays* within a *trip* (to account for *cold start* issues (*24*))
Max allowed gap between the *trips* in a tour	Max. spatial gap between consecutive *trips* = 200 m Max. *tour* duration 24 h (we focus on typical days)
Output generation	
User tracking-quality evaluation metrics	Temporal coverage per user per day In-app validation of travel diaries In-app correction of *activities* during last/previous week Active status (1+ recorded *activities* on a day) Mobile status (1+ recorded *tracks* on a day)

previous step, we proceed by merging consecutive *tracks* with the same transport mode if the end of the first *track* is very close in space-time to the beginning of the second *track*. In the last step of this stage, we address the lack of GPS tracking when the participant is static (e.g., when the phone is turned-off or underground). In this case, if a *stay* is detected and later a *track* starts in that same location, the *stay* is extended to match the beginning of the *track*.

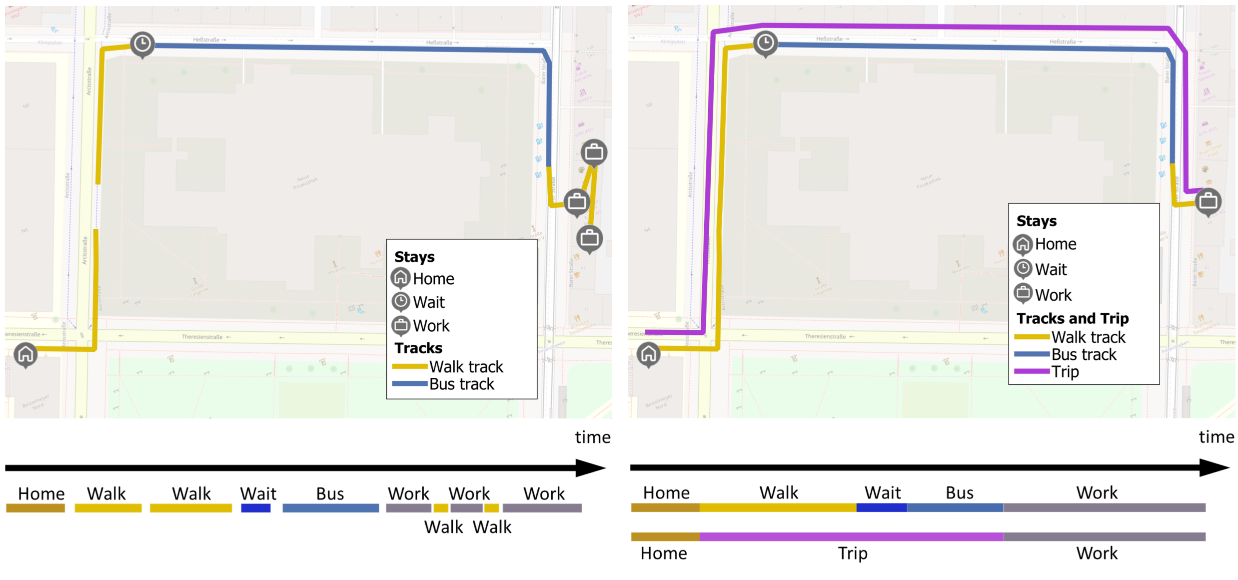
In the ***data enrichment*** stage, we seek to derive additional relevant attributes from the cleaned and processed *semi-passive travel diaries*, without –or with minimal– additional external data sources. In particular, our focus is on detecting *trips* and *tours* from the *track* data, since they are commonly used for travel behavior analyses.

In a first step, we annotate whether a *track* is (partially) within the boundaries of the Munich public transportation network (MVV) and the German national borders. This aids in the easy selection of relevant data when assessing mobility behavior. Subsequently, we implement a popular open-source map-matching tool (*25*) and match the trajectories to a topologically simplified version of the OpenStreetMap network (*26*). This enhances significantly the value of the travel diaries, as *track* trajectories are no longer just spatial points, but they can be associated with a specific sequence of links in the transport network, enabling, for example, the detailed study of the participants' routes. To reduce runtime, a maximum lattice width of 30 was only used if the average observation distance after the first iteration exceeded a threshold value (0.0005). To conclude with this stage, the travel diary of each user is analyzed chronologically to detect complete *trips* and *tours* (as defined in the Background section). In the first case, our approach considers as a *trip* all *tracks* and wait *stays* between two observed *stays*, as long as certain maximum spatial and temporal gaps are respected between consecutive *tracks* (or *tracks* and wait *stays*). We opted for this approach since, for longer temporal/spatial gaps, we cannot guarantee that the participant is not undertaking unobserved *activities*; and because due to the large dataset available in the study, we prioritize a better quality of the *trips* rather than quantity. A similar approach is adapted to detect *tours*, but with slightly looser thresholds (the more *tracks* are involved, the higher the chances of exceeding the thresholds). The exact threshold values for each step were decided based on a combination of literature and the exploratory analysis of validated or deleted tracks, as applicable, but it is important to note that they are highly influenced by the tracking app and the preceding steps. *Activities* not assigned to a *trip* or *tour*, are not discarded, as they remain valuable for *activity*-based analyses.

The results of the *data enrichment* stage are *enriched semi-passive travel diaries*, with better quality and additional attributes than the *semi-passive diaries* generated by the *Mobilität.Leben app*. Figure 4 illustrates a multitude of the steps implemented in the preceding two stages. In the unprocessed trajectory the walk segment is split in two due to a lost GPS signal and there are several short walk segments between work *stays* inside a building. Additionally, there are temporal gaps between various consecutive activities. These three issues are fixed using the presented framework: the two successive walk segments are merged into one, the duration of the activities is extended to maximize the temporal coverage, the work *stays* are consolidated into one, and the overall *trip* (pink line) is generated.

Finally, in the ***integration of external data sources and assessment of tracking-quality*** we extend and process the *enriched semi-passive travel diaries* to create four *modules* that can be directly employed for specific further mobility analyses, thus reducing the workload that researchers must devote to preparing the data. We introduce these in the following:

1. A module to automatically generate the *trip* characteristics (length, duration and route) for non-chosen travel modes. This information is necessary, for example, for studies dealing with mode choice based on revealed preference data (*27*). Car *trip* data is queried from *TomTom Routing API* (*28*) and the remaining modes are generated offline in a server using *OpenTripPlanner API* with Munich's transport network and real *GTFS* for the studied period.
2. The derivation of relevant *user-day tracking-quality statistics*, which can be used to identify various user groups and assess the completeness and reliability of the travel diary of each user. These include the following: temporal daily coverage (% hours tracked in a day), distance by mode, if the user is *active* (any *activities*), and *mobile* (any tracks). Two further metrics are computed that reflect the involvement of the user in the study: *validating* (share of passively-generated *activities* accepted in the past/next week), or *correcting* (if a user has merged/deleted/modified any *activities* in the past/next week).



(a) Unprocessed trajectory

(b) Processed trajectory

FIGURE 4: Synthetic *semi-passive* vs. *enriched semi-passive* detected trip

3. To facilitate *trip*-based analyses, we also offer a consolidated form of *trip*-related information and derive additional relevant attributes, in addition to the enhanced track-based travel diaries. These include: the location and purpose of origin and destination *stays*; the total distance/duration by mode, for multimodal *trips*; and the main-mode, that with the longest distance, as in (*22*). With this consolidated information it is possible to identify *round-trips* –those starting and finishing in the same location and without detected intermediate *stays*; frequently for leisure/sports– and, if necessary, exclude them from further analyses.
4. A module to integrate socio-demographic data from the survey, as well as historical hourly/daily weather conditions in Munich from the German weather service.

APPLICATION TO THE *MOBILITÄT.LEBEN* DATASET

In the following section we will present the results of applying the proposed framework to the data recorded in the *Mobilität.Leben* app. In this paper, we use data from from June 1st 2022 to May 31st 2023. This section will follow the structure of the methodological framework.

Cleansing and processing

The original data comprises 1,648,867 *tracks* and 1,261,117 *stays*. At each processing step we tracked the number of changes made relative to the previous step, to be able to observe the effect of these. Based on the maximum average speed threshold, 1.2% of *tracks* were classified as outliers. Abnormally short *tracks* and *stays* made up 3.8% and 10.2% of *tracks* and *stays*, respectively, and were subsequently removed. Meanwhile, 1.5% of *tracks* and 7.0% of *stays* were successfully merged. Lastly, 3.3% of the *stays* with *unknown* purpose –which accounted for 40.9% of the total *stays*– were imputed as *wait* and the duration of 1.5% of the *stays* was extended.

In this stage the number of *tracks* and *stays* was reduced by 8.6% and 13.9%, resulting in 1,507,059 *tracks* and 1,086,058 *stays*. Importantly, while this drop seems large, it also reflects merged *activities*, explaining why the observed total duration for all users decreased by just 2.5%. The large size of the dataset justifies the removal of erroneous activities to improve overall data quality, yet from hereon no further data will be discarded. The average *track* length increased slightly by 1.6% (to 10.0 km), as a result of removing abnormally short *tracks* and merging consecutive ones. Across all travel modes, the average walking *track* length increased most at 3.8% (to 690 m). Similarly, the duration of *tracks* increased: 5.2% overall, 8.8% for walk, and 3.3% for PT *tracks*. The average duration of *stays* with a *work* purpose increased by about half an hour (by 12.5%).

Data enrichment

The assignment of *tracks* to *trips* and *trips* to *tours* is a valuable step in the data enrichment stage. Overall, 92.6% of the *tracks* were assigned to a *trip*. *Tours* can only be detected if the home of a user is known ("only" 64.7% of users annotated it); for these users the *tour* detection rate –i.e., the share of *trips* assigned to *tours*– is 53.1%. Importantly, *round-trips* (approximately 7% of the total) were, by definition, not assigned to tours. If the threshold values are increased to 500 m, the detection rate rises to 61.5%. Overall, around half of the participants –with annotated home– have a tour-detection rate above 60%, while the upper and lower 10% reach around 80% and 30%, respectively.

Integrating supplementary information & design agnostic output

This stage of the data quality enhancement method focuses on adding value to it by integrating and deriving supplementary information, rather than further altering or removing *activities*. It is key to the agnostic user design and ensures that a wide range of information is easily accessible and usable for further analyses. Across all detected *trips*, the average length is 14.5 km and 26.0 minutes, compared to 10.0 km and 15.5 minutes for *tracks*. The average number of *trips* per user per day is 4.6, where on average 1.6 *tracks* are assigned to each *trip*. Typically, every third PT trip includes wait *stays*, where the median total wait duration per trip is 6.9 minutes (importantly, we observed that short PT transfers are occasionally not detected by the app). Regarding *tours*, a user makes on average 1.3 *tours* per day (each with 3.1 *trips*), and 27.7 km and 68 minutes per *tour*.

When considering the enhanced data, the following is observed. For the *active* users, the average temporal coverage is 89.6% (21.4 hrs/day), while 66.1% of user-days are fully recorded (100%) and 91% of user's days have at least 12 hours of *activities*. On average users participated in the study for 209 days. Furthermore, 67.8% of users perform correcting behavior (mode change, *track*/*stay* deletion, *track* merging) on a bi-weekly basis. 79.6% of users *validate* all of their *activities* within 30 days.

DISCUSSION

In this section we will first address the improvement of the travel diaries as a result of our quality enhancement method and compare the results to a regional travel survey. Then we will discuss the possible use-cases of the output data, along with some insights that assist in the design of similar studies.

Quality improvement

Both the *cleansing and processing* and the consolidation of *tracks* into *trips* impact the travel diaries in a multitude of ways. In Table 2, we compare the results of the *Mobilität in Deutschland*

(*MiD*) travel survey from 2017 for Munich (*22*), with the corresponding subset of *tracks* or *trips* for various stages of our methodology. It is of great importance to note that the *Mobilität.Leben*'s sample is not fully representative of Munich's population, yet with this comparison we aim to show that the proposed data quality enhancement method leads to values that are more similar to those of the large-scale representative travel survey collected using conventional methods. For instance, considering the average number of *trips*/*tracks* per day, we observe that our *trip* detection leads to a value much closer to *MiD* than for the raw or processed *tracks*. This also applies to the duration and distance traveled by a user per day. Regarding the distance traveled by a user per *trip* or *track* we observe that the values get more realistic, the further the processing progresses. Nonetheless, the remaining differences to *MiD* can be explained because participants of traditional travel surveys tend to underestimate their number of *trips* per day and misestimate *trip* duration and length (*9*).

When comparing the aggregate results in terms of the modal split by frequency (and distance), as shown in Figure 5, it becomes evident that both the enhanced *tracks* and *trips* have an improved modal split compared to the raw data. The walk mode share decreases from 5.6% to 4.1% (41% drop) after the *trip* detection, as frequently walking is not the main mode of a *trip* but only the access mode. Regarding the modal split by frequency, the share of bike and car *trips* grows compared to PT and walking *trips*, as the latter are more likely to be multi-leg *trips*.

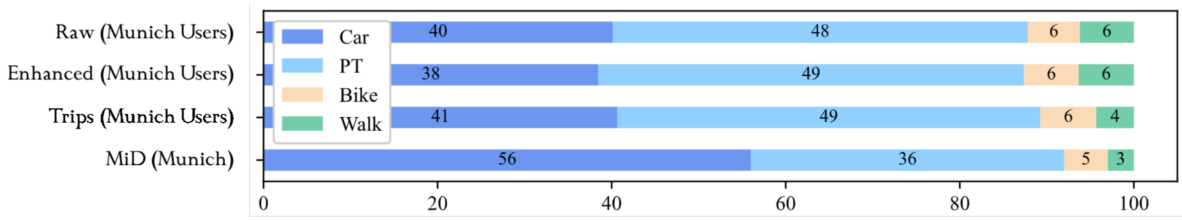
TABLE 2: Comparison pre-/post-enhancement results for users living in the Munich area from September 1st 2022 till November 1st 2022 and *MiD* 2017 travel survey

	Raw *tracks*	Enhanced *tracks*	Detected *trips*	*MiD* Munich
No. (*trips* or *tracks*)	150,693	138,128	80,604	-
(*trips* or *tracks*)/user/day	7.8	7.2	4.4	3.2
hr/user/day	1.8	1.7	1.8	1.8
km/user/day	49.6	46.3	44.2	42
km/user/(*trip* or *track*)	7.2	7.3	11.3	12.5
hr/user/(*trip* or *track*)	0.2	0.3	0.4	0.3
Mean daily temp. coverage	90.8	87.7	85.7	-

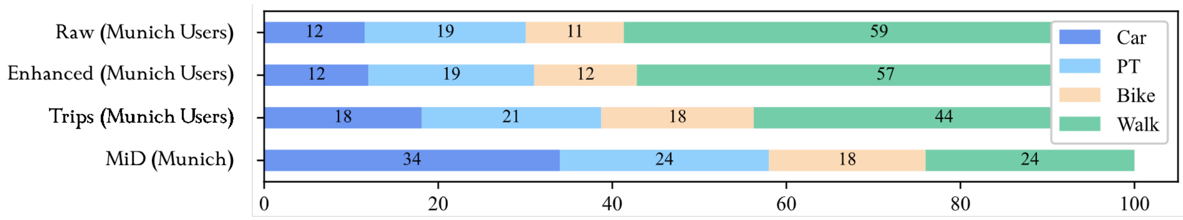
Relevance, use-cases, and insights

Having shown the qualitative improvement in the travel diaries, we now move on to presenting the impact and use that the enriched data can have – i.e., to highlight it's potential. In addition to looking at survey participation and user involvement, we will suggest exemplary use-cases.

The users of any given day were grouped into five categories, based on their level of involvement: 1) users that abandoned the experiment, 2) users that are still involved but not *active* on that day, 3) users that have at least one *activity* on that day, 4) users that recorded more than 80% of that day, 5) users that recorded more than 80% of that day and additionally *corrected* an activity in the app the week before/after that day. The evolution of these behavioral groups throughout the experiment are shown in Figure 6. The upper bound of the curve indicates the cumulative number of participants since the start of the study, which is steady throughout most of the study until a sharp increase is observed at the start of the *Deutschlandticket*. As shown in the figure, in the first four months participants abandoned the study at a steady rate of around 4% per month; then, after the first wave, this rate increased to close to 10%. Subsequently, until the beginning of



(a) Modal split by distance



(b) Modal split by frequency

FIGURE 5: Modal split by distance and frequency, respectively, annotated by percentage share - City of Munich, September-November 2022.

the third wave, the abandonment rate returned to values close to 4% (even though the number of actively correcting users with high temporal coverage –at the bottom– stayed stable). Interestingly, as shown by the fluctuations in the mobile users curve, a large share of participants (approximately 20%) did not record movements on Sundays compared to other weekdays.

The in-app *validation* of a *trip* was assessed in depth and it was found that 80% of all *activities* are *validated* within 4 days. In Figure 7, alongside with the values for all users, two groups were compared: the 100 users with the highest and the lowest average daily temporal coverage and *correction rate* (if users were editing *activities* in the week before/after a day). The latter group, which performed fewer corrections and had many gaps in their diaries, has a notable delay in the *validation* of *activities*, with 12.0% of *tracks* not *validated* within 30 days. The opposite is observed for the more involved user group. This shows that people who reliably *correct* erroneous *activities* tend to *validate* them sooner. This is intuitive, as it is easier to recollect recent *activities* better (and thus correct them).

IMPLICATIONS AND RECOMMENDATIONS FOR FUTURE STUDIES

In the following we will share our learnings and insights to aid in the design of future studies employing *semi-passive travel diaries*.

Firstly, researchers can benefit from the involvement of users with the app (i.e., when users edit and correct *activities*) to identify the most frequent flaws in the draft travel diaries, and design, accordingly, methods to address them automatically if users neglect validation and correction. This is not possible with *fully-passive travel diaries*, as, by definition, users cannot modify the generated travel diaries.

Secondly, the high *trip*-detection rate shows that it is possible –after thorough processing– to successfully derive *trips* from *semi-passive travel diaries*, even when many participants have low commitment (*correcting* and *validating* activities in the app). In the case of home-based *tours*, the results are quite different. Yet, even for participants with recorded homes, it is often impossible to recover tours based on the maximum allowed spatial gaps, as a complex *tour* can contain many

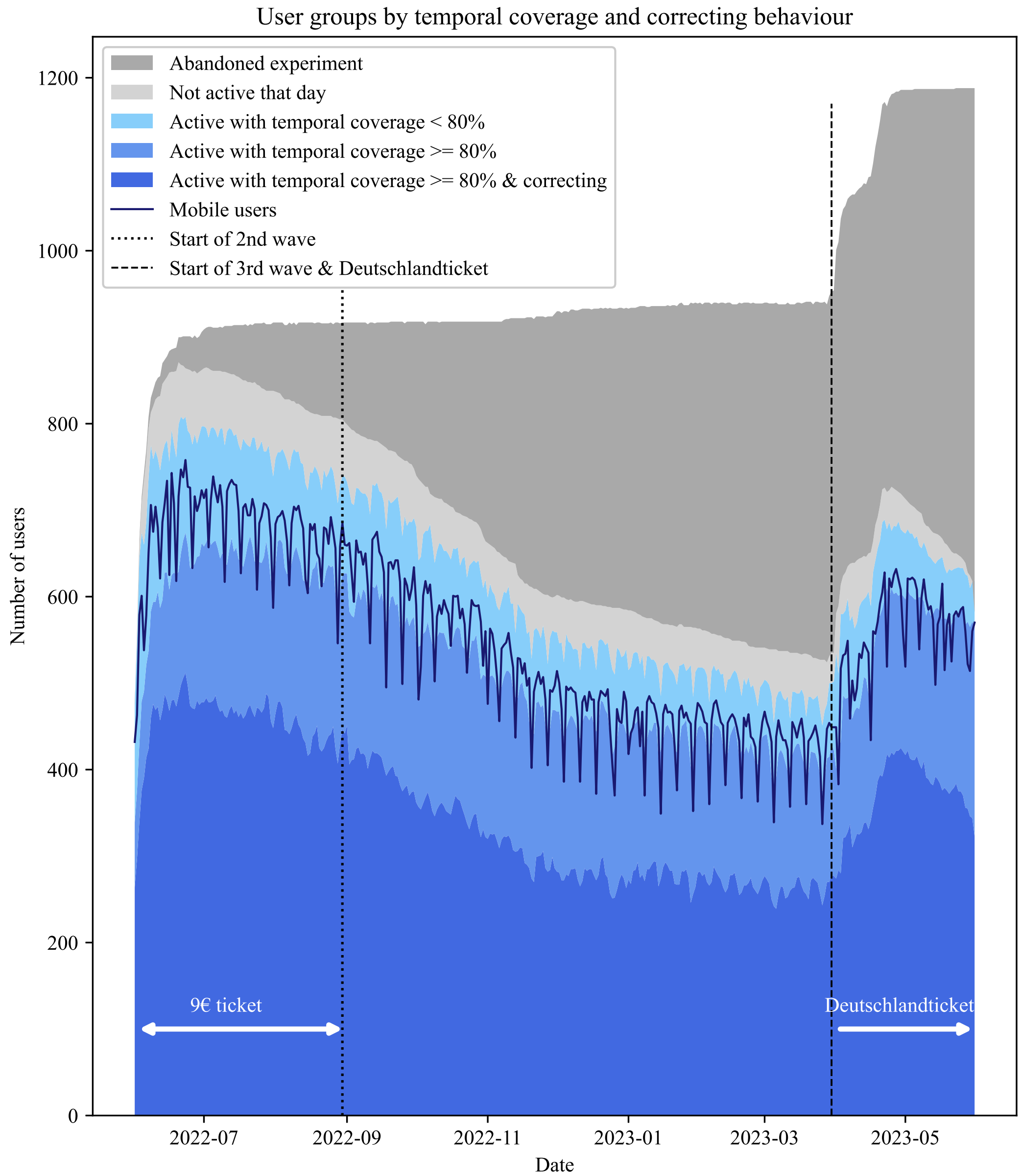


FIGURE 6: Evolution of user participation throughout the study

trips making it more likely to exceed the thresholds once (or to have some un/mis-recorded *tracks*). It is also noted that for users with higher correction rate (top 100 users) 58.1% of the trips can be assigned to a *tour*, compared to 40.9% for the least *correcting* users. Importantly, if a researcher is interested in recovering a higher number of *tours*, it would be possible to relax the current thresholds, at the expense of accepting tours with poorer quality. In our case, and due to the huge size of the dataset, we prioritized quality over quantity. If home locations are not imputed, *tours* cannot be derived for a large number of participants. Therefore, we strongly recommend to incentive users to validate their home location within the first days of a study (with pop-up notifications, emails, or similar).

We also observe that *corrections* are performed less frequently than *validations*: for 68.4%

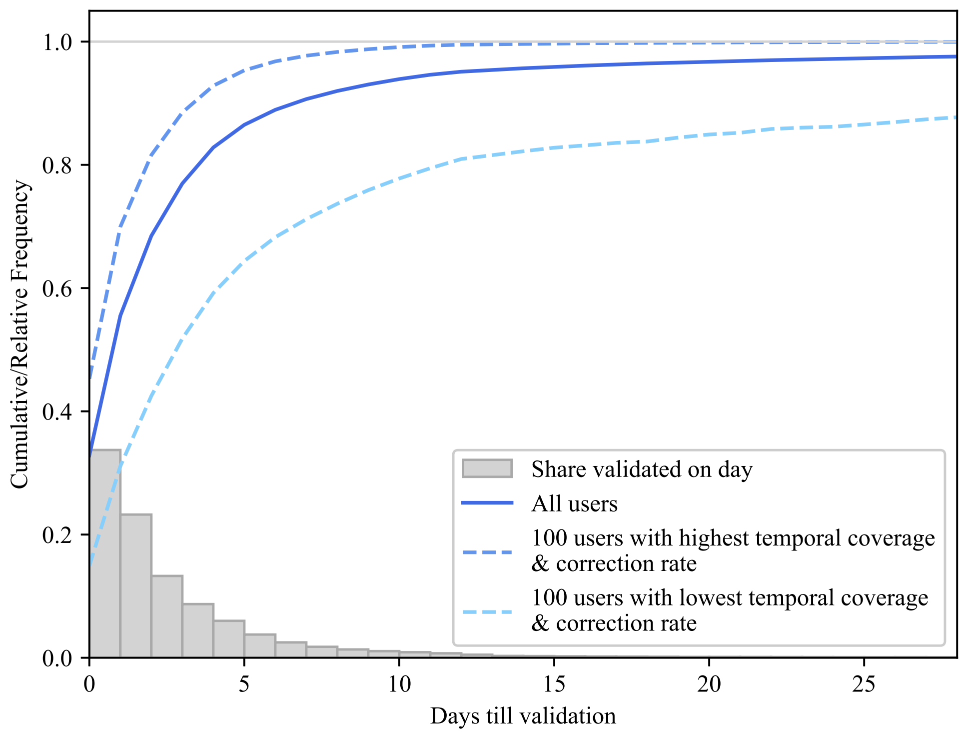


FIGURE 7: Share of activities *validated* since *activity* end for different user groups

and 96.4% of all users were *correcting* (in a two week time period) and *validating*, respectively. We argue that *corrections* are a better indicator of user involvement, as inattentive users will accept the passively generated drafts but not make the effort to look for errors. Moreover, while the processing pipeline aims to correct faulty or incomplete tracking, it is nonetheless of interest that users rectify faulty *activities*, given that they know the ground truth. Similarly, for the purpose assignment, we have observed that a large number of participants never annotate frequently visited locations, which impoverishes the overall quality of the data. Since the app follows an iterative learning approach and learns from the user's previously tagged locations, it would have been enough if the users had tagged them just once. We propose that, for example, when the app detects that a location with *unknown* purpose is visited regularly, it displays a pop-up notification demanding the user to annotate its purpose. In this way, key user-locations would be identified without overburdening the user.

We summarize our recommendations as follows:

1. Emphasize the importance of annotating the purpose of key locations, particularly *home*, within the first days of a study.
2. When using the data for further analysis, do not presume that *validated* diaries are necessarily correct. Many users pay little attention to improve the *automatically-generated travel diaries*. Instead, the frequent correction of activities is a better indicator to identify good observations.
3. Owing to the previous recommendation, encourage users to *correct* their *activities* (e.g.,

make the participation-reward dependent on the diaries' quality).

4. Aim for smart, interactive, and engaging app designs (e.g., use pop-up notifications) to benefit from the synergies between the app (which handles the most demanding work, i.e., tracking the user and generating a draft diary) and the participant (who performs minor corrections that significantly improve the quality of the travel diaries).

CONCLUSION AND FUTURE RESEARCH

In this paper we detail our experience working with data from a long-duration *semi-passive* mobility tracking app and present a data quality enhancement method, hence contributing to fill the existing gap in the literature. Furthermore, we discuss the implications of the results and make recommendations for future studies. Our approach involves three stages: (i) *Cleansing and processing*, (ii) *Data enrichment*, and (iii) *Integration of external data sources and assessment of tracking-quality*. The data quality enhancement results in a high-quality dataset that is rich in information and greatly increases the suitability for further mobility analyses, both in terms of reliability and versatility (due to the wide range of attributes/information). We make recommendations for future studies that focus on the importance of user-involvement and optimal app design.

To further improve the quality of the *enhanced semi-passive travel diaries*, future research should explore the incorporation of stay purpose imputation. This could span from simple rule-based home-imputation, to advanced imputation models (*16*). A limitation of this work is the lack of ground truth, hence the quality of the generated diaries cannot be measured quantitatively. Thus, it would be interesting for future studies to have a subset of participants who additionally self-report their trips (i.e., as in a traditional travel survey), such that these data can be used as ground truth to improve the data enhancement method. Another promising area of research is the study of mobility behavior on the basis of our enhanced data. In particular, light could be shed on the effectiveness of transport policy instruments such as the *9-Euro ticket* and the *Deutschlandticket*.

In closing, it can be seen that data collected in studies involving *semi-passive GPS travel diaries* can be informative and easily scaled over several months with low marginal costs for additional days. Considering the dynamics and heterogeneity of travel behavior in the 21st century, household travel surveys and their travel diaries would highly benefit from data collected using such an app, nevertheless, our paper showed that not all data can be used and that meaningful *activities* have to be identified and their data enriched. To facilitate the data quality enhancement in future studies involving semi-passive travel diaries, we are planning to make our method open-access.

AUTHOR CONTRIBUTIONS

The authors confirm their contribution as follows: study conception: VD, SAO, AL, KB; background: VD, SAO; data collection: VD, SAO, AL, KB; processing and analysis: VD, SAO; manuscript: VD, SAO, AL, KB. All authors reviewed the results and approved the final manuscript.

ACKNOWLEDGEMENTS

This research is supported by the TUM Georg Nemetschek Institute. AL acknowledges funding by the Bavarian State Ministry of Science and the Arts in the framework of the *bidt* Graduate Center for Postdocs. The authors thank the TUM Think Tank at the Munich School of Politics and Public Policy for their financial and organizational support and TUM Board of Management for supporting personally the genesis of the project. Additionally, the authors express their gratitude to the

MDSI and *Mobilität.Leben* participants.

REFERENCES

1. Axhausen, K. W., *Travel diaries: An annotated catalogue*. Institut für Straßenbau und Verkehrsplanung, Universität Innsbruck, 1995.
2. Prelipcean, A. C., Y. O. Susilo, and G. Gidófalvi, Collecting travel diaries: Current state of the art, best practices, and future research directions. *Transportation Research Procedia*, Vol. 32, 2018, pp. 155–166.
3. Wolf, J., S. Schönfelder, U. Samaga, M. Oliveira, and K. W. Axhausen, Eighty Weeks of Global Positioning System Traces: Approaches to Enriching Trip Information. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1870, No. 1, 2004, pp. 46–54.
4. González, M. C., C. A. Hidalgo, and A.-L. Barabási, Understanding individual human mobility patterns. *Nature*, Vol. 453, No. 7196, 2008, pp. 779–782.
5. Clarke, M., M. Dix, and P. Jones, Error and uncertainty in travel surveys. *Transportation*, Vol. 10, No. 2, 1981, pp. 105–126.
6. Storesund Hesjevoll, I., A. Fyhri, and A. Ciccone, App-based automatic collection of travel behaviour: A field study comparison with self-reported behaviour. *Transportation Research Interdisciplinary Perspectives*, Vol. 12, 2021, p. 100501.
7. Faghih Imani, A., C. Harding, S. Srikukenthiran, E. J. Miller, and K. Nurul Habib, Lessons from a Large-Scale Experiment on the Use of Smartphone Apps to Collect Travel Diary Data: The "City Logger" for the Greater Golden Horseshoe Area. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2674, No. 7, 2020, pp. 299–311.
8. Lynch, J., J. Dumont, E. Greene, and J. Ehrlich, Use of a Smartphone GPS Application for Recurrent Travel Behavior Data Collection. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2673, No. 7, 2019, pp. 89–98.
9. Thomas, T., K. T. Geurs, J. Koolwaaij, and M. Bijlsma, Automatic Trip Detection with the Dutch Mobile Mobility Panel: Towards Reliable Multiple-Week Trip Registration for Large Samples. *Journal of Urban Technology*, Vol. 25, No. 2, 2018, pp. 143–161.
10. Zhao, F., F. C. Pereira, R. Ball, Y. Kim, Y. Han, C. Zegras, and M. Ben-Akiva, Exploratory Analysis of a Smartphone-Based Travel Survey in Singapore. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2494, No. 1, 2015, pp. 45–56.
11. Molloy, J., A. Castro, T. Götschi, B. Schoeman, C. Tchervenkov, U. Tomic, B. Hintermann, and K. W. Axhausen, The MOBIS dataset: a large GPS dataset of mobility behaviour in Switzerland. *Transportation*, 2022, pp. 1–25.
12. Loder, A., F. Cantner, L. Adenaw, M. Siewert, S. Goerg, M. Lienkamp, and K. Bogenberger, *A nation-wide experiment: fuel tax cuts and almost free public transport for three months in Germany – Report 1 Study design, recruiting and participation*, 2022.
13. Prelipcean, A. C., G. Gidófalvi, and Y. O. Susilo, Mobility Collector. *Journal of Location Based Services*, Vol. 8, No. 4, 2014, pp. 229–255.

14. Prelipcean, A. C., G. Gidofalvi, and Y. O. Susilo, Measures of transport mode segmentation of trajectories. *International Journal of Geographical Information Science*, Vol. 30, No. 9, 2016, pp. 1763–1784.
15. Marija Nikolic, M. B., Review of transportation mode detection approaches based on smartphone data. In *17th Swiss Transport Research Conference*, 2017.
16. Montini, L., N. Rieser-Schüssler, A. Horni, and K. W. Axhausen, Trip Purpose Identification from GPS Tracks. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2405, No. 1, 2014, pp. 16–23.
17. Huang, Y., L. Gao, A. Ni, and X. Liu, Analysis of travel mode choice and trip chain pattern relationships based on multi-day GPS data: A case study in Shanghai, China. *Journal of Transport Geography*, Vol. 93, 2021, p. 103070.
18. Perroud, S., *Lake Geneva consumers surveyed as part of a study on Climate Change*. https://actu.epfl.ch/news/lake-geneva-consumers-surveyed-as-part-of-a-study-/, 2022.
19. Montini, L., S. Prost, J. Schrammel, N. Rieser-Schüssler, and K. Axhausen, Comparison of Travel Diaries Generated from Smartphone Data and Dedicated GPS Devices. *Transportation Research Procedia*, Vol. 11, 2015.
20. Martin, H., Y. Hong, N. Wiedemann, D. Bucher, and M. Raubal, Trackintel: An open-source Python library for human mobility analysis. *Computers, Environment and Urban Systems*, Vol. 101, 2023, p. 101938.
21. Pappalardo, L., F. Simini, G. Barlacchi, and R. Pellungrini, scikit-mobility: A Python Library for the Analysis, Generation, and Risk Assessment of Mobility Data. *Journal of Statistical Software*, Vol. 103, No. 4, 2022, p. 1–38.
22. infas, DLR, IVT, and infas 360, *Mobilität in Deutschland – MiD 2017 Regionalbericht Stadt München*, 2020.
23. Safi, H., B. Assemi, M. Mesbah, and L. Ferreira, Trip Detection with Smartphone-Assisted Collection of Travel Data. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 2594, No. 1, 2016, pp. 18–26.
24. Hongmian Gong, Cynthia Chen, Evan Bialostozky, and Catherine T. Lawson, A GPS/GIS method for travel mode detection in New York City. *Computers, Environment and Urban Systems*, 2012.
25. Meert, W. and M. Verbeke (eds.), *HMM with Non-Emitting States for Map Matching*, 2018.
26. OpenStreetMap contributors, *OSM Planet dump*. `https://www.openstreetmap.org`, 2017.
27. Tsoleridis, P., C. F. Choudhury, and S. Hess, Deriving transport appraisal values from emerging revealed preference data. *Transportation Research Part A: Policy and Practice*, Vol. 165, 2022, pp. 225–245.
28. *TomTom Routing API*. `https://developer.tomtom.com`, 2023, accessed: 2023-07-15.