<http://www.diva-portal.org>

Postprint

This is the accepted version of a paper presented at *International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2015, 3 June 2015 through 5 June 2015*.

Citation for the original published paper:

Prelipcean, A., Gidófalvi, G., Susilo, Y. (2015)
Comparative framework for activity-travel diary collection systems.
In: *2015 International Conference on Models and Technologies for Intelligent Transportation Systems, MT-ITS 2015* (pp. 251-258). IEEE conference proceedings
<http://dx.doi.org/10.1109/MTITS.2015.7223264>

N.B. When citing this work, cite the original published paper.

Permanent link to this version:

<http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-181605>

THIS IS A COPY OF THE POST-PRINT MANUSCRIPT.

THE PUBLISHED VERSION:

Prelipcean, A.C., Gidófalvi, G. and Susilo, Y.O. (2015) Comparative framework for activity-travel diary collection systems, 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), 251-258 (DOI: 10.1109/MTITS.2015.7223264), found at <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?reload=true&arnumber=7223264>

Preferred citation:

Prelipcean, A.C., Gidófalvi, G. and Susilo, Y.O. (2015) Comparative framework for activity-travel diary collection systems, 2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), 251-258 (DOI: 10.1109/MTITS.2015.7223264)

Comparative framework for activity-travel diary collection systems

Adrian C. Prelipcean

Department of Geoinformatics
Royal Institute of Technology (KTH)
Drottning Kristinas väg 30
Stockholm, Sweden
Email: acpr@kth.se

Győző Gidófalvi

Department of Geoinformatics
Royal Institute of Technology (KTH)
Drottning Kristinas väg 30
Stockholm, Sweden
Email: gyozo.gidofalvi@abe.kth.se

Yusak O. Susilo

Department of Transport Science
Royal Institute of Technology (KTH)
Teknikringen 10
Stockholm, Sweden
Email: yusak.susilo@abe.kth.se

Abstract—The needs for cheaper and less intrusive ways to collect activity-travel diaries led scientist to pursue new technologies, e.g., positioning technologies like GPS. While a fully, reliable and widely accepted automatic activity-travel diary collection system is yet to be developed, scientists have presented systems that automate parts of an activity-travel diary collection. In the advent of automated systems, it is important to discuss how to analyse the potential of such systems and how to compare different activity-travel diary collection systems. To achieve this objective, this paper introduces a parallel survey design and a comparison framework for collection systems. The framework can be used as a development tool to optimise system design, to report and monitor progress of different system designs, to objectively weigh benefits in decision making, and to automate systematic analysis. In particular, the framework can be used as a comparison tool to reveal the qualitative difference in the data gathered using different collection systems. To achieve this, the framework defines: 1) a number of activity-travel diary measurement entities (trips and triplets), entity attributes (e.g., trip purpose, origin / destination, etc.), 2) similarity functions between instances of the same entities, and 3) spatial and temporal quality indices to establish a notion of ground truth. The utility of the proposed framework is demonstrated by analysing the results of a trial survey where data is collected via two activity-travel collection systems: a web-based system (PP) and a smartphone-app-based system (MEILI). PP was collected for one day period and MEILI was used for one week period (with one day overlapping). The results show that half of the trips are captured by both systems, while each system roughly captures the same number of trips as the other. The strengths and weaknesses of MEILI are analysed using the framework on the entire week dataset.

Keywords—activity-travel diary collection; comparison of collection methods; ground truth; spatial and temporal quality indices;

I. INTRODUCTION

A comprehensive activity-travel diary gives valuable insight into the day-to-day variability of individual's activity-travel engagements and needs, which are essential in urban and transportation planning analyses. An activity-travel diary contains information about individual's trip characteristics in the given day, which include: the travel time and combination of travel mode that the traveller chose, the selected route

to a destination, the destination itself, and the purpose of performing a trip. This information is used to investigate the reasons and mechanisms that underlie an individual's activity-travel decision making processes and how these vary for different time and space contexts. This understanding then can be further aggregated in order to either predict the effect of implementing new transportation policies or changing the transportation infrastructure, or just to understand the dynamic of transportation movement within the given study areas.

Although activity-travel diaries are undeniably useful, the classical collection methods are burdensome and prone to error. The participants are asked to fill in a standard travel diary (originally, as a paper form, and nowadays as a web-site that mimics the form) or are asked to report their trips via computer-assisted telephone interviews, commonly known as CATIs. These methods are vulnerable to under-reporting [4,19] – i.e., participants might forget to report a trip or disregard a trip that does not fall into their personal definition of a trip – and to a low response rate due to a high cognitive demand on participants [11,24]. In light of the aforementioned limitations and with the emerging mobile and GPS technologies, there have been a surge in the number of recent studies which investigated new methods that could be used to reduce the burden of participants and the number of under-reported trips.

The development of new Internet technologies, together with the pervasiveness of position-aware mobile and wearable devices allows scientists to deepen their understanding of human mobility. Recent progress towards identifying the transportation mode of a traveler [14,20–23] and inferring the purpose of a trip [1,3,9,18], from location traces, pave the way for solving an important task: the automatic generation of travel activity-diaries. To overcome misclassification errors of automatic inference methods, a number of researchers have proposed semi-automatic systems that collect location traces and through web GIS technologies allow the users to annotate the collected traces (e.g., [8,13]). No matter what the proposed collection system is, a framework is needed to objectively assess a collection system's benefits and shortcomings. On one hand, the framework can be used to optimise system design, to report and monitor progress of different system designs, to

objectively weigh benefits in decision making, and to automate systematic analysis. On the other hand, the framework can be used to compare different activity-diary collection systems. One particular challenge that the comparison framework must address is the question of ground truth, i.e., what were the actual travels and activities that the individual has performed?

To this extent, a framework that can tackle the challenge of identifying the ground truth has been developed and tested. Two travel-diary collection systems have been designed and deployed simultaneously, according to the proposed framework. One collection system prompts users to fill in an online survey, which is a simple electronic adaptation of the classical way of obtaining activity-travel diaries. The other collection system is MEILI, which is a travel-diary collection and annotation system built on the open source Mobility Collector [13]. In MEILI, a user installs an application that collects his/her GPS trace which then the user subsequently annotates with trip and activity information via a web GIS based MEILI web application. To compare the data generated by both systems, the framework defines a number of activity-travel diary measurement entities (trips and triplegs) and entity attributes (e.g., trip purpose, start / end times of a trip / tripleg, etc.). The framework can then measure the similarity between instances of the same entities and find the corresponding trips and triplegs in the two systems. The framework then facilitates the analysis and comparison of subsets of the data (trips) from both systems in terms of a number of relevant analytical measures. Furthermore, the framework allows for a clear definition of error, i.e., differentiating between entities that are candidates for ground truth and those that are affected by error. The distinction is made according to new spatial and temporal indices that are based on the relative spatio-temporal distribution of consecutive samples of a trip route and a pairwise dissimilarity based quality index.

This paper explores concepts necessary for the “fair” comparison between activity-travel diary collection systems. Section II presents related work. Section III discusses the general methodology associated with matching trips from different activity-travel diary collection systems and introduces spatial and temporary quality indices that give valuable insight regarding what segments of a trip can be considered ground truth. Section IV presents the results of the case study. Section V discusses and analyses the effect of the proposed methodology with regards to the case study. Finally, Section VI concludes the paper and suggests future research directions.

II. RELATED WORK

A. Travel diary collection methods and systems

1) *Mode Inference*: Inferring the transportation mode of a user is a task that is of interest for two research areas: transportation science [2,7,10,13] and Location Based Services (e.g., GeoLife [21–23], and others [14,20]). Scientists collect different data sets – such as GPS-only data [21–23], accelerometer-only data [7,20], GPS traces fused with accelerometer data [10,13,14] or GPS traces complemented by GIS information [2,15] – that are annotated by users

and afterward use machine learning or rule based systems to train classifiers that automatically determine the transportation modes of future data sets. High classification accuracies have been achieved, which are suitable for travel diary generation: e.g., Prelipcean et al. [13] 90.8% (seven classes), Reddy et al. [14] 93.6% (five classes), Stenneth et al. [15] 93.5% (five classes), and Yu et al. [20] 90.6% (five classes).

2) *Destination and Purpose Inference*: On one hand, scientists investigated trip destinations. Dill and Broach [5] explore how travel destinations can be defined using stated and revealed data, and show that 46% of the collected trips’ destinations belong to the “stated” dataset. Axhausen et al. [1] use Trip Identification and Analysis System (a black-box GPS travel data processing software system developed by GeoStats) with a set of heuristics for determining a trip’s destination.

On the other hand, scientists tried to correctly infer a trip’s purpose. Bohte and Maat [3] use GIS data to allocate trip purposes. They identify the closest point of interest to a trip’s end and, based on its type, derive the purpose of the trip. All the potential trip purposes are classified into 13 categories and the reported overall accuracy is 43%. Oliveira et al. [12] evaluate two methods for purpose inference that are based on choice modeling and decision tree analysis. These methods rely on GIS land use and points of interest datasets, and the reported overall accuracy is above 70% for 12 categories. Wolf et al. [18] show in their pilot study that it is feasible to derive trip purpose by combining GPS point data with a spatially accurate GIS land use database. While the reported prediction accuracy is over 90%, the study contains a small data set of 151 trips and the approach is highly heuristical, where land uses are “mapped” to one or more plausible purposes. Montini et al. [9] use random forests to infer trip purposes for a 1 week travel survey in Switzerland, 2012, that involved 156 participants and they found that the accuracy varies between 80% and 85%.

3) *Travel-Diary Collection Systems*: In the first studies [1,18,19] scientists used to collect GPS data and then use different processing techniques and software to test the feasibility of using such systems. However, as this approach is prone to errors that cannot be accounted for, the next logical step was to use systems that allow travelers to annotate their data. In this sense, travel-diary collection systems transitioned from using fully automated surveys to using semi-automated surveys. These semi-automated systems allow users to annotate trips and they have similar characteristics: they present full or partial travel diaries to users via a map interface and whenever users find discrepancies or missing information in the data, they are allowed to correct and / or complement them. Examples of such systems are [3] and [8], and MEILI (used in this study), which uses the open source Mobility Collector [13] to collect GPS traces fused with accelerometer values and allows users to interact with their trips via a travel-diary web app.

B. Comparative case studies

Brika and Bhat [4] attempted to understand trip underreporting by analysing trips of drivers that reported their travel patterns via CATIs and drove a GPS-equipped vehicle. The

data was extracted from the Kansas City Household Travel Survey conducted in the spring of 2004, which contained details (both GPS and CATI) on 228 households - corresponding to 377 drivers and 2539 vehicle trips. However, the details of the matching of trips from CATIs and those derived by GPS are not publicly available.

Forrest and Pearson [6] compared trips determined from GPS readings (as obtained by using the GeoStats GeoLogger, thus restricting the collection to drivers) with trips declared in CATI from a data set of 103 households. A trip, as determined by CATI consists of: start time, end time, origin, destination, purpose, and vehicle miles traveled (VMT), which is computed using the shortest path between origin and destination along the street network. A trip, as determined by GPS data, has the same attributes, and is heuristically determined:

- the start of the trip is determined based on consecutive reads where the velocity is greater than 0, using a dwell time threshold of 120 seconds,
- the end of the trip is identified visually using ArcGIS plots of a suggested trip,
- VMT is calculated with the Haversine formula, and
- the trip purpose is inherited from CATI.

To clearly identify trips, authors grouped the purposes declared in CATI into home-based work (HBW), home-based non-work (HBNW) and non-home based (NHB). When comparing trips declared in CATI with those determined by GPS data, the authors used a time threshold of 15 minutes and a direct match of purposes. Because of trip linking, a GPS trip is first classified as direct (it fits the start, stop time and trip purpose of a CATI trip) or linked (a series of multiple GPS trips that, when combined, are comparable in start, stop time and trip purpose). The authors found, after linking trips, that 79 more HBW trips were reported in CATI than GPS, 132 less HBNW trips in CATI than in GPS and 304 less NHB trips in CATI than in GPS. The authors conclude that the data collected by CATI are very detailed and there were, at the time when the paper was written, no alternative methods capable of acquiring that level of detail in a similar time frame and cost. The visual identification of trip ends is subjective and not scalable for automated comparison systems, such as the proposed framework of the current paper. Furthermore, the proposed framework is more flexible in terms of the entities and the entity attributes that it can compare.

The most in-depth comparison study to date was done by Stopher and Li [16] in Melbourne, Victoria, Australia – the survey collected 1104 trips –, where the authors analyse travel diary records with their equivalents (as derived from GPS records) to check whether the two groups of data match. The matching is performed in four main steps:

- 1) the geographic comparison step, where a the origin and destination of consecutive trips (for GPS and diary) are visually checked for spatial consistency using TransCAD and Google maps,
- 2) the time comparison step, where the start (stop) time of the GPS trip is compared with the start (stop) time of the

Fig. 1. Entities, Attributes and Attribute Measurement Units. One trip can contain several triplegs. The measurement units of attributes are presented in parentheses, in italics. The time period and metric measurement units can be represented at any suitable scale, e.g., hours for time period and kilometers for metric. The geometry is a polyline for routes and a point for destinations.

- 3) duration comparison step, where the duration of a GPS trip is compared with the duration of a diary trip and, by using a 15 minute acceptance threshold, and
- 4) the distance comparison step, where the distance between a GPS origin (destination) and the diary origin (destination) is compared with a 920 meters (vehicle) / 500 meters (walk) acceptance threshold.

Out of the 1,104 trips, 587 have GPS data and diary data that contain enough information to be compared. For the remaining trips, the authors identified reasons why trips were not present in either one of the system and categorized the reasons as: 71 trips are only recorded by GPS (memory mistakes of the respondent, unwillingness to report trips, failure to report trips in the diary), 77 trips are only in the diary (respondents forgot carrying the device, GPS device problems) and 110 trips are mismatched (GPS cold starts, and other reasons). The remaining 329 trips were recorded by both the GPS devices and the diary, and they could be matched. Similar to the present work, the authors also define approximate criteria and investigate the characteristics of different subsets based on the match. In addition, the proposed framework also: 1) incorporates quality indices for the GPS data that can be used to establish a notion of ground truth, 2) defines error measures compared to this ground truth, and 3) automatically evaluates the subsets qualitatively using the error measures.

III. FRAMEWORK

This section presents the design of parallel surveys, in this case *Sys1* and *Sys2*, introduces the entities stored in an activity-travel diary, provides definitions for ground truth candidates, and describes how to find corresponding entries between *Sys1* and *Sys2* and their comparison.

A. Parallel survey design – Choices and motivations

When comparing any two activity-travel diary collection system, it is important to collect the same data at the same time with the same group of respondents and the same set

Fig. 2. Temporal (horizontal axis) and Spatial Indicators (vertical axis). The intersection of the vertical and horizontal axes denotes the threshold between low and high values for the indices. A grey box is the representation of a trip / trajectory in space (upper side of the box) and in time (lower side of the box) and the associated grey text box provides a brief summary of the case that the trajectory falls in. Two consecutive points are connected by a blue line when the distance / time between them is smaller than or equal to $\Delta d_{min} / \Delta t_{min}$ and by a red line when the distance / time between them are greater than $\Delta d_{min} / \Delta t_{min}$.

of activity-travel parameters so that the data can be compared (see Figure 1). Depending on how data is collected, if the two systems use different collection techniques, i.e., manual / declarative vs automatic, it is recommended that the participants are prompted to answer the manual collection system first, as to avoid declaration bias.

B. Definition of activity-travel diary measurements

The framework distinguishes between two main entities, namely trips and triplegs, which are shown together with their attributes in Figure 1. The temporal aspects of a trip / tripleg is described by the following attributes: 1) *start time*, 2) *end time*, 3) *duration* (the difference between end time and start time), and 4) *waiting time* (the difference between the end time of the previous trip / tripleg and its start time).

The spatial aspect of a trip / tripleg is described by the: 1) *destination*¹, which is the point that the user identified as the place / POI where the trip / tripleg² ends, 2) *route*, which is the sequence of locations available in between origin and destination (only origin and destination of a trip / tripleg in case of the classical surveys), and 3) *length* (calculated as the cumulative distance of the sequence of locations between the origin and the destination of a trip / tripleg). Additionally, nominal variables contains information about a trip's *purpose* and about a tripleg's *mode of transportation*.

C. Similarity functions

Trips from two systems are matched using two constraints:

- 1) temporal co-occurrence, which implies that the start time and stop time of a trip from *Sys1* has to be within T_T of the start and stop time of a trip from *Sys2*, and
- 2) identical purposes, which restricts two trips to be wrongfully match solely on temporal co-occurrence – the case of short trips.

While it is possible to incorporate a proximity constraint, i.e., the origin (and destination) of a trip as recorded in *Sys1*

¹The destination of a trip is the origin of the subsequent trip.

²Possible POI types of triplegs are transit stops / statoins, parking places.

has to be within a certain distance of the origin (and destination) of a trip as recorded in *Sys2*, this is not recommended because of errors in the accuracy of recording one location, the fuzzy string matching between addresses (that is employed when mapping names of destinations to a known database of geocoded names) and / or the inherent errors in geocoding (the cumulative extent of these errors is measured in Section IV).

D. Ground truth indicators

One important aspect of a comparison between entities is establishing the ground truth. However, due to the fact that there is no activity-travel diary collection system that captures the ground truth, it is imperative to asses the quality of the entities and their attributes as collected by one system (intrinsically) to find trip representations that can be considered candidates for ground truth. Furthermore, it is important to compare the entities pairwise (extrinsically) to understand the amount of agreement / disagreement between trips as captured by two systems. Combining intrinsic and extrinsic indicators can be used to understand which collection system offers a more accurate description of an entity, and whether the two system can complement each other on any dimension.

1) Intrinsic indicators - Spatial and Temporal Quality Indices: The intrinsic indicators pertain to how well a trip / tripleg and its entities are captured by a system. The attribute that can offer such insight is the route of the trip / tripleg. Usually, the route is a sequence of locations where each location has a timestamp associated to it. While the route can be inferred (e.g., by using the shortest path between the origin and destination of a trip [6]), it is common practice for new technologies to sample a user's location by using a location receiver (e.g., by using positioning technologies based on GPS, WiFi, Bluetooth, etc.). When this type of sampling is used, there is an expectancy regarding the resolution of the samples: 1) the consecutive samples are approximately Δd_{min} meters apart (equidistance sampling), 2) the consecutive samples are recorded every Δt_{min} seconds (equitime sampling), or 3) the consecutive samples are recorded based on logical rules that combine Δd_{min} and Δt_{min} (hybrid sampling). For more information about sampling types, see Prelipcean et al. [13].

Using any of the aforementioned sampling parameters, the framework proposes spatial (S_{idx}) and temporal (T_{idx}) indices that determine the fraction of an entity that was sampled according to the spatial and / or temporal resolutions.

The spatial index S_{idx} is computed as shown in Equation 1, where $dist(i+1, i)$ represents the distance in meters between two consecutive locations and Δd_{min} represents the expected distance between any two consecutive locations. In the equidistance sampling case, Δd_{min} is the specified spacing between consecutive samples and in the equitime sampling case, Δd_{min} can be established by simple heuristics, e.g., the expected distance between two consecutive locations when the user is traveling by the fastest transportation mode.

$$S_{idx} = \frac{\sum dist(i+1, i) * I(dist(i+1, i) \leq \Delta d_{min})}{\sum dist(i+1, i)} \quad (1)$$

The temporal index T_{idx} is computed as shown in Equation 2, where $\delta t_{i,i+1}$ is the time difference between two consecutive locations and Δt_{min} represents the expected temporal distribution of the sampled locations. In the equitime sampling case, Δt_{min} is the sampling rate and in the equidistance sampling case, Δt_{min} can be established by simple heuristics, e.g., the expected time to pass between the recording of two consecutive locations when the user is traveling by the slowest transportation mode.

$$T_{idx} = \frac{\sum \delta t_{i,i+1} * I(\delta t_{i,i+1} \leq \Delta t_{min})}{\sum \delta t_{i,i+1}} \quad (2)$$

In Equations 1 and 2, $I(condition)$ is an indicator function that returns 1 when the condition is fulfilled and 0 otherwise.

While a high value for S_{idx} indicates a good capture of a trip's geometry, when it is accompanied by a low value for T_{idx} it implies that the time between two (or more) consecutive locations is more than expected, which suggests that a trip / tripleg was merged with the next trip / tripleg. Similarly, a high value for T_{idx} accompanied by a low value of S_{idx} indicates a "noisy" trip. How the proposed spatial and temporal indices can be used to categorize samples is presented in Figure 2.

2) *Extrinsic indicators - Pairwise Entity Comparison:* The extrinsic indicators are suitable for comparing the same entities as captured by both systems, $Sys1$ and $Sys2$.

While the comparison of nominal, period and metric attributes is straightforward (i.e., in terms of equality), the comparison of time period and geometry attributes relies on subjective interpretation.

Depending on the robustness of data, the analysis may ignore the date and route attributes comparison and only rely on comparing the measures derived from them (i.e., length and duration). However, one interesting observation can be made in terms of comparing destinations, namely that a destination³ can

³This observation is also valid for origins because the destination of one trip is the origin of the following trip.

Fig. 3. Trip subsets based on matching functions and quality indicators.

be determined in *different and independent* ways: 1) *manually*, by the user, as specifying a point geometry on a map interface, or an address, which is afterward geocoded to a point geometry, and 2) *automatically*, as the last point of the polyline that represents the route. Assuming that the measurements are independent, one can compare them pairwise, according to a relevant similarity measure, and favour the measurement that agrees more to the others by using a statistical aggregate, such as mean and / or standard deviation. This case is presented in detail and discussed in Section IV.

E. Relating entities and defining subsets based on matching function and quality indices

One important part of the analysis is determining the trips that are only present in $Sys1$, the trips that are only present in $Sys2$ and the trips that have been captured by both $Sys1$ and $Sys2$ (Figure 3a). Considering that the matching of trips was performed as described in Section III-C, the spatial and temporal indices presented in Section III-D1, and the pairwise entity comparison in Section III-D2, the different cases that can be encountered are presented with regards to Figure 3. For simplicity, in the presented cases, $Sys1$ does not include spatial or temporal indices.

The initial step is to analyse the spatial and temporal indices of all trips and establish threshold values $-T_S$ for the spatial index and T_T for the temporal index— that discriminate between trips that should be considered reliable in space (Figure 3b) or in time (Figure 3c). In the second step, trips from $Sys1$ are matched with trips from $Sys2$ based on the

similarity functions proposed previously (Section III-B). In the third step, the following trip subsets are generated, namely:

- subset of matched trips (Figure 3a),
- subset of trips that are candidates for ground truth (high S_{idx} and high T_{idx} , as in Figure 3e),
- subset of noisy trips (low S_{idx} and high T_{idx}),
- subset of merged trips (high S_{idx} and low T_{idx}),
- subset of unreliable trips (low S_{idx} and low T_{idx}), and
- subset of partially matched trips (the matching is approximate and it is based on a non-exhaustive entity set, namely start / stop date and purpose of trips, which can result in partially matched trips as seen in Figure 3f).

F. Analysis of subsets based on measures and relative set cardinality

After performing the steps mentioned in the previous section, it is imperative to explore the computed subsets. The spatial adequate trip subset⁴ (Figure 3b) and the temporal adequate trip subset (Figure 3c) allow for the assessment of the collection quality of the system, i.e., the intersection of the spatial adequate trip subset and the temporal adequate trip subset, which can be considered the ground truth candidate subset. Another dimension of the collection quality of the system is the size of the subset of collected trips (Figure 3a), which, for a perfect case, should be equal to all trips, i.e., ground truth. However, since obtaining ground truth is not possible, the greater the subset of collected trips, the better the system.

As for understanding why trips are missed or wrongfully captured by one system, it is important to analyse the subsets of trips present in one system only (i.e., in Figure 3a the left and right circles excluding the intersection) and the subset of partially matched trips (Figure 3f) to identify potential flaws in the system design process.

IV. CASE STUDY

This section provides a description of the case study that was performed to test the data quality of the MEILI system. Two different activity-travel diary collection systems are used in the case study: a web version of a standard paper and pen activity-travel diary, further referred to as PP, and the MEILI system, which is a modern annotation-based activity-travel diary collection system.

A. Trial survey

The survey took place from Monday, 29th of September 2014 to Sunday, 5th of October 2014 and it consisted of two parts:

- 1) the PP part, where users received via email a link to a web-based form that mimics classical travel diaries (according to the Swedish baseline form for travel activity-diaries) and they declared their trips for the first day of the survey period, (29th of September 2014), and

⁴An adequate trip subset in space / time is the subset of trips that have the spatial / temporal index greater than the threshold value.

Fig. 4. The Entities, Attributes and Attribute Measures for MEILI and PP. In the setup used in this case study, MEILI offers more information about the route of a trip and more detailed information about triplegs. The temporal and spatial indices are not relevant for PP because the route is not sampled.

- 2) the MEILI part, which took place over the whole time frame, and where users installed the MEILI Mobility Collector on their Android smartphone and then annotated their data via the MEILI web-app.

First, to avoid declaration bias (as presented in Section III-A), the users were sent an email containing a link to the first part of the survey and a link to the MEILI Mobility Collector download page. Second, after filling in the paper and pen travel diary, the users gained access to the MEILI web-app, which also contained instructions on how to annotate their trips. Finally, the travel diaries obtained from both systems were centralized and compared. The participants were asked to use MEILI for the whole week to better understand the consequences of using a semi-automatic collection system.

B. Data sets

Initially, 51 persons showed interest, out of which 42 answered the travel diary and 30 collected and annotated their data with MEILI. During the time frame of the case study, 1055 trips were collected with MEILI, out of which 718 were annotated. For the one day where the users generated travel diaries with both systems, the paper and pen method generated 116 trips and the MEILI system generated 97 trips. However, out of the 116 trips, only 104 were generated by users who also used MEILI to collect trips.

C. Analytical results of application of the framework

This section analyses MEILI and PP according to the framework proposed in Section III.

In terms of the collected entities and their attributes, MEILI collected trip and tripeg attributes that were not present in PP, most noticeable trip and tripeg routes. Furthermore, MEILI captures more information at a tripeg level. This is due to the consideration taken into account when designing PP, i.e., mimicking the Swedish national travel survey.

In PP, users specified their destination via an open question of address, while in MEILI, users specified their destination

Fig. 5. Agreement between origins (blue) and destinations (red) as they are recorded by the two systems: declared in PP, declared in MEILI and the end of trajectory in MEILI. The MEILI declared origins and destination agree with the trajectories and disagree with the origins and destinations declared in PP.

via map interaction and the destination was automatically determined as the end of the route of a given trip. The chosen similarity function (as indicated in Section III-D2) was the Euclidean distance. After measuring the distance between the three cases, the MEILI specified destination has the minimum disagreement with the MEILI automatically determined destination and the PP specified destination, as shown in Figure 5. This was expected because of known errors in geocoding and user's lack of attention when specifying the address in PP, which indicates that the MEILI manually revealed destinations can be considered ground truth candidates and the PP manually revealed destinations are affected by errors.

When comparing the subsets derived using the proposed framework, we noticed that statistics of the trips collected by both systems are similar in both representations (MEILI and PP). One noticeable difference lays in the standard deviation of the duration (higher in MEILI) and the length (higher in PP) of a trip, which is expected due to the automatic collection of MEILI data. Furthermore, MEILI offers more information about waiting time and allows for the calculation of spatial and temporal indices (see columns 1 and 2 of Table I).

The trips present in PP only are shorter than the average trip captured by both systems, with the same duration and containing approximately two triplegs. The trips present in MEILI only appear to be longer in duration and shorter in distance, while mostly using only one mean of transportation, thus also a low value for waiting time. This might be consistent with short and / or long walking trips.

Using the one week data collected by MEILI, we derived values for each transportation mode used in this case study, as presented in Table II. While the data is not sufficient to derive conclusions based on the provided values for all transportation modes, the MEILI collection has reasonable ($\geq 50\%$) spatial and temporal indices for the modes that are well represented in the data set (≥ 50 registered triplegs for almost all travel modes considered). One expected problem is the collection of triplegs performed via subway, where the low spatial quality index is explained by the fact that the GPS receiver loses connection while underground. The low values for the spatial quality index

TABLE I. THE PAIRWISE COMPARISON OF ENTITIES ATTRIBUTES AS RECORDED BY BOTH SYSTEMS: PP (USING PP TRIP ATTRIBUTES - PP I.) AND BY MEILI (USING MEILI TRIP ATTRIBUTES - MEILI I.), AND THE TRIPS RECORDED ONLY BY ONE SYSTEM: MEILI (MEILI ONLY) AND PP (PP ONLY). THE VALUES ARE PRESENTED AS AVERAGE $\pm$ STD DEV.

	I. PP	I.MEILI	PP Only	MEILI Only
Duration (min)	23 $\pm$ 15	24 $\pm$ 19	23 $\pm$ 20	64 $\pm$ 85
Length (m)	6300 $\pm$ 7600	6300 $\pm$ 6000	4500 $\pm$ 5000	3800 $\pm$ 5100
Trip segments	1.8 $\pm$ 1.1	1.8 $\pm$ 1	1.7 $\pm$ 1.1	1.2 $\pm$ 0.3
Waiting time (min)	-	6 $\pm$ 10	-	1 $\pm$ 3
T_idx	-	85% $\pm$ 30%	-	83% $\pm$ 29%
S_idx	-	61% $\pm$ 36%	-	60% $\pm$ 32%
#trips	43	43	51	44

^a The median value is 20 minutes

TABLE II. THE TRIPLEG ATTRIBUTES DERIVED FROM MEILI ON THE DATASET THAT WAS COLLECTED DURING THE WHOLE WEEK. THE PRESENTED VALUES ARE THE TRIMMED MEAN, I.E., THE MEAN OF THE VALUES THAT FALL WITHIN THE AVERAGE $\pm$ 2 STD. DEV. NOTE THAT THE STATISTICS HAVE TO BE PERFORMED ON A LARGER DATA SET.

	Walk	Bike	Moped	Car	Taxi	Bus	Subway	Tram	Commute	Train
Durat	6	17	14	15	7	6	8	45	13	73
%dur	41	86	71	88	54	18	29	85	25	61
Len	0.5	2.5	6.5	6	3.5	1.5	3	8	11	175
%len	31	91	99	97	59	30	53	96	51	89
Wait	1	2	-	5	0	5	8	-	17	20
%dur	3	6	-	5	0	13	19	-	27	23
T_idx	93	99	100	99	71	82	86	100	72	100
S_idx	71	91	92	81	71	71	44	23	39	64
#tlegs	414	113	6	140	2	63	104	3	26	6

for tram and commuter train triplegs is due to the low number of triplegs, which could mean that they were taken by the same user whose smartphone might not be fully compatible with MEILI.

V. DISCUSSION

After applying the proposed framework for the scenario presented in the case study, we reached the following conclusions: 1) half of the trips are captured by both PP and MEILI, 2) the number of trips captured only by MEILI is roughly the same with the number of trips captured only by PP, 3) MEILI captures more information about triplegs, 4) the spatial and temporal indices allows us to assess the quality of the collected MEILI data, and 5) the location attributes derived from the route (destinations) agrees with the user annotated data more than with the address specified in PP.

While the size of the participants in the case study is relatively small, it demonstrates the applicability of using the framework. First, it is clear that the MEILI system captures more attributes of trips and triplegs than PP. Second, the utility of the spatial and temporal indices is clearly illustrated by identifying which trips are candidate for ground truth. Furthermore, the spatial and temporal indices offer valuable insight when it comes to identifying whose transportation mode's capture is problematic for MEILI, thus indicating a direction for improvements. Finally, the framework is generalizable to any number and type of activity-travel diary collection systems.

VI. SUMMARY AND FUTURE WORK

In this paper, we have proposed a framework that compares activity-travel diary collection systems. The framework uses new spatial and temporal indices to clearly distinguish between the capture of entities that can be regarded as either error or ground truth candidates. Furthermore, the framework has successfully identified the strengths and weaknesses of each activity-travel diary collection system, thus indicating future developmental / implementation prospects for a system. The utility of the framework is tested in the case study, where trips and triplegs collected by two systems, MEILI and PP, are matched to test which system is the better alternative.

We envision the adoption and adaptation of the framework for the following uses: 1) in an iterative design process to optimise designs, 2) to report performance to the research community to gain a better understanding about progress in the field, 3) to present results to decision makers that need to objectively weigh costs against benefits, and finally 4) to automate the systematic analysis of results. In particular we intend to use the framework in the iterative redesign of the MEILI system within the SPOT project entitled “Trialling and comparing smartphone based travel data collection with paper-and-pencil method”. Based on our previous experiences, due to the complex interdependence of system parameters and dynamics in deployments, we argue that a solid comparative framework is imperative for an effective design process. Thus we plan to implement fully automatic evaluations based on the proposed framework and report results of future activity-travel surveys in the SPOT project. Finally, we plan to actively refine the definition of the derivable subsets to better identify strengths and weaknesses of diary collection systems.

ACKNOWLEDGMENT

This work was partly supported by the Trafikverket (Swedish Transport Administration) under grant “TRV 2014/10422”.

REFERENCES

- [1] K. W. Axhausen, S. Schönfelder, J. Wolf, M. Oliveira and U. Samaga. “80 weeks of GPS-traces: Approaches to enriching the trip information”, *Arbeitsbericht Verkehrs- und Raumplanung*, number 178, 2003.
- [2] F. Biljecki, H. Ledoux and P. van Oosterom. “Transportation mode-based segmentation and classification of movement trajectories”, *International Journal of Geographical Information Science*, number 2, volume 27, pp. 385–407, 2013.
- [3] W. Bohte and K. Maat. “Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands”, *Transportation Research Part C: Emerging Technologies*, number 3, volume 17, pp. 285–297, 2009.
- [4] S. Bricka and C. R. Bhat. “Comparative Analysis of Global Positioning SystemBased and Travel SurveyBased Data”, *Journal of the Transportation Research Board*, number 1972, pp. 9–20, 2006.
- [5] J. Dill and J. Broach. “Travel to Common Destinations: An Exploration Using Multiday GPS Data”, *Journal of the Transportation Research Board*, number 2413, pp. 84–91, 2014.
- [6] T. L. Forrest and D. F. Pearson. “Comparison of Trip Determination Methods in Household Travel Surveys Enhanced by a Global Positioning System”, *Journal of the Transportation Research Board*, number 1917, pp. 63–71, 2005.
- [7] S. Hemminki, P. Nurmi, and S. Tarkoma. “Accelerometer-based transportation mode detection on smartphones”, *Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems*, number 13, pp. 13:1–13:14, 2013.
- [8] L. Montini, N. Rieser-Schüssler and K. W. Axhausen. “Field Report: One-Week GPS-based Travel Survey in the Greater Zurich Area”, *13th Swiss Transport Research Conference (STRC)*, 2013.
- [9] L. Montini, N. Rieser-Schüssler, A. Horni and K. W. Axhausen. “Trip Purpose Identification from GPS Tracks”, *Journal of the Transportation Research Board*, pp. 16–23, 2014.
- [10] L. Montini, N. Rieser-Schüssler, A. Horni and K. W. Axhausen. “Personalisation in multi-day GPS and accelerometer data processing”, *14th Swiss Transport Research Conference (STRC)*, 2014.
- [11] J. Ogle, R. Guensler, and V. Elango. “Georgia’s commute Atlanta value pricing program: Recruitment methods and travel diary response rates.” *Journal of the Transportation Research Board*, number 1931, pp.28–37, 2005.
- [12] M. S. Oliveira, P. Vovsha, J. Wolf, M. Mitchell. “Evaluation of Two Methods for Identifying Trip Purpose in GPS-Based Household Travel Surveys”, *Journal of the Transportation Research Board*, number 2405, pp. 33–41, 2014.
- [13] A. C. Prelipcean, G. Gidófalvi and Y.O. Susilo. “Mobility Collector”, *Journal of Location Based Services*, volume 8, issue 4, pp. 229–255, 2014.
- [14] S. Reddy, M. Mun, J. Burke, D. Estrin, M. Hansen, M. Srivastava. “Using Mobile Phones to Determine Transportation Modes”, *ACM Transactions on Sensor Networks*, number 2, volume 6, pp. 13:1–13:27, 2010.
- [15] L. Stenneth, O. Wolfson, P. S. Yu and B. Xu. “Transportation mode detection using mobile phones and GIS information”, *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pp. 54–63, 2011.
- [16] P. Stopher and S. Li. “In-depth Comparison of Global Positioning System and Diary Records”, *Journal of the Transportation Research Board*, number 2046, pp. 32–37, 2011.
- [17] D. P. Wagner. “Lexington area travel data collection test: GPS for personal travel surveys”. *Final Report, Office of Highway Policy Information and Office of Technology Applications, Federal Highway Administration, Battelle Transport Division, Columbus*, 1997.
- [18] J. Wolf, R. Guensler and W. Bachman. “Elimination of the Travel Diary: An Experiment to Derive Trip Purpose From GPS Travel Data”, *Journal of the Transportation Research Board*, number 1768, pp. 125–134, 2001.
- [19] J. Wolf, M. Oliveira and M. Thompson. “The impact of trip underreporting on VMT and travel time estimates: preliminary findings from the California statewide household travel survey GPS study”, *Journal of the Transportation Research Board*, number 1854, pp. 189–198, 2003.
- [20] M. Yu, C. Lin, T. Yu, E. Y. Chang and S. Wang. “Big data small footprint: The design of a low-power classifier for detecting transportation modes”, *The Proceedings of the VLDB Endowment*, volume 7, pp. 1429–1440, 2014.
- [21] Y. Zheng, Q. Li, Y. Chen, X. Xie and W. Ma. “Understanding Mobility Based on GPS Data”, *Proceedings of the 10th International Conference on Ubiquitous Computing*, pp. 312–321, 2008.
- [22] Y. Zheng, L. Liu, L. Wang and X. Xie. “Learning Transportation Mode from Raw Gps Data for Geographic Applications on the Web”, *Proceedings of the 17th International Conference on World Wide Web*, pp. 247–256, 2008.
- [23] Y. Zheng, Y. Chen, Q. Li, X. Xie and W. Ma. “Understanding Transportation Modes Based on GPS Data for Web Applications”, *ACM Transactions on the Web*, number 1, volume 4, pp. 1:1–1:36, 2010.
- [24] M. Zimowski, R. Tourangeau, R. Ghadialy and S. Pedlow. “Nonresponse in household travel surveys”, *Report to the Federal Highway Administration*, 1997.