Conference Paper

Route and mode choice models using GPS data

Author[s]:
Montini, Lara; Antoniou, Constantinos; Axhausen, Kay W.

Publication Date:
2017-01

Permanent Link:
https://doi.org/10.3929/ethz-b-000119554

Rights / License:
In Copyright - Non-Commercial Use Permitted
ROUTE AND MODE CHOICE MODELS USING GPS DATA

Lara Montini
IVT, ETH Zürich, CH-8093 Zürich
phone: +41-44-633 30 88
fax: +41-44-633 10 57
lara.montini@ivt.baug.ethz.ch

Constantinos Antoniou
Chair of Transportation Systems Engineering
Technical University of Munich
Arcisstraße 21, D-80333 Munich, Germany
c.antoniou@tum.de

Kay W. Axhausen
IVT, ETH Zürich, CH-8093 Zürich
phone: +41-44-633 39 43
fax: +41-44-633 10 57
axhausen@ivt.baug.ethz.ch

Words: 6170 words + 1 figure + 4 tables = 7420 word equivalents
Submitted for presentation in the 96th Annual Meeting of the Transportation Research Board (January 2017).

TRB 2017 Annual Meeting
ROUTE AND MODE CHOICE MODELS USING GPS DATA

ABSTRACT

GPS data has become almost ubiquitous and data collection and processing technologies have advanced considerably. In transportation research, GPS traces are used, along with other data sources, to construct travel diaries, as they promise higher accuracy of collected data, combined with fewer fatigue effects. In this research, we present such a study, where a sample from Zurich, Switzerland, used GPS devices to monitor their daily trips for a week. A follow-up survey provided complementary information, including sociodemographic characteristics of the participants, as well as a large number of attitudinal parameters. Route choice models were estimated using this data set for car and public transport trips, along with a joint route and mode choice model, which also considered bike and walk stages. To the best of our knowledge, this is the first time that GPS data have been used for the estimation of mode and route choice models. The model fit is reasonable for all models and in line with those obtained with traditional data-collection methods. The coefficient estimates have the appropriate signs and meaningful magnitudes. A number of transformations were performed on some of the variables, improving the model fit significantly, and reducing the number of variables. We discuss some of the difficulties associated with such data collection efforts, and point out the advantages that can come from follow-up studies, based on similar concepts.

Keywords: GPS traces, surveys, route choice, mode choice, Zurich, Switzerland
INTRODUCTION AND RELATED WORK

In recent years, GPS data has become more common; almost ubiquitous. Furthermore, location data is collected as part of travel surveys but also as part of different kinds of smartphone applications, such as journey planners, health or running apps and self-tracking applications. In transportation research, GPS traces are used, along with other data sources, to construct travel diaries (1-3), as they promise higher accuracy of duration and distance and increased reporting of short trips and activities. In addition, it is often argued that GPS-based surveys allow for longer survey periods with fewer fatigue effects as response burden is reduced by automatically generated diaries. The responsibility for data annotation is still given to participants and is mostly handled using self-guided web-based prompted recall approaches (4-9). Location data have primarily been collected using dedicated GPS devices that respondents carry with them during the tracking period. Smartphones are a promising alternative source for GPS data (see e.g., (10)), as they have been equipped with good-quality GPS, accelerometers, gyroscopes and other potentially useful sensor functionality during the last years and, as opposed to dedicated devices, are often carried by participants anyway.

A recent review of GPS-based travel studies and the required processing tools is given in Shen and Stopher (11), who list representative studies using dedicated devices from 14 different countries, as well as four smartphone studies. The first GPS studies were undertaken in the late 1990s (12). GPS devices were at first attached to cars. Later on, hand-held devices were used to capture all modes of travel (13). Initial solutions for mobile phones were implemented in the mid-2000s (14,15). By now, the smartphone-based travel data collection is growing rapidly, as evidenced by several new applications implemented over the last few years by the research community ((16), Quantified Traveler (17), UbiActive (18), Future Mobility Survey (19), CONNECT of the MOVE project (20), SmarTrAC (21), SITTS (22)). Already, commercial tools designed to be used in different mobility studies are implemented, e.g., rMOVE (23) and Studio Mobilita (24) (used in Becker et al. (25)). An application with very similar goals as the PEACOX app (used for data collection in this research) is the GoEco! app launched in March, 2016 (26), a tracking app where usage of sustainable travel modes is encouraged.

When working with revealed preference data and real networks one of the problems is the huge number of possible path alternatives. Furthermore, these paths overlap and models have to deal with resulting similarities. Both problems are analyzed in Schlüssler (27), who introduces the Breadth First Search on Link Elimination algorithm (BFS-LE) for choice set generation, and compares it to the random walk (28), a branch & bound algorithm (29), as well as a stochastic choice set generation, where link costs are randomized, given a probability distribution before each least cost path computation, similar to e.g. (30-33).

To account for route overlaps Schlüssler (27) compared different adjustment terms: two formulations of the C-logit model (34), two path size models (35), which is used here, as well as the path size correction term (36) and two trip part specific path size factors (37). This last factor is used in the public transport models in this research, as it performed best on the PEACOX trip planner data as shown in Fischer (38).

Route choice models for car on Swiss data were estimated e.g. in Bierlaire et al. (39). This data set was also used to estimate bicycle route choice in Menghini et al. (40) using the Path Size Logit approach as was done by Hood et al. (41) and Broach et al. (42). The same approach was also successfully used recently for pedestrian route choice (43-44).

The remainder of this paper is structured as follows. The data collection process is presented next, followed by an overview of the analysis methodology. The model specification and results are presented next, while a concluding section discussed the findings and outlooks.

TRB 2017 Annual Meeting
DATA COLLECTION AND ANALYSIS

Travel behavior is commonly modeled using socio-demographic as well as mobility specific attributes. But there are less easily surveyed latent variables that influence behavior. Examples of such variables are risk propensity, attitude towards the environment as well as search for variety. The goal of this research is to evaluate the influence of these attitudes on route choice behavior with a special focus on public transport. To better observe the route choice behavior, a person-based GPS travel survey was combined with an attitude questionnaire. The responsibility for data annotation was still given to participants. The high resolution of the position data to accurately observe taken routes and the long survey period to investigate behavioral patterns are appealing for such analyses.

GPS Data

Network data

Route choice in this research is restricted to the area around Zurich depicted in the top subfigure of Figure 1. All map data was extracted from OpenStreetMap (45). The network for cyclists and pedestrians includes all links except motorway and trunk links, resulting in approximately 3 million links. For the car network on the other hand, pedestrian- and cycling-only links were excluded resulting in 2.4 million links as depicted with road types in the bottom subfigure of Figure 1. Using the map data, a network in MatSim format (46) was created, mainly based on the highway attribute and taking into account the oneway tag, as the network is directed. The conversion code can be found as part of the POSition DAta Processing project on sourceforge (47).

Elevation model and measures

Elevations for the canton of Zurich were released very recently as Open Source under the GIS-ZH license; the digital terrain model is available with a resolution of 0.5 meters (48) and is therefore used whenever possible. Outside the Zurich canton a digital elevation model with a resolution of 25 meters by swisstopo is used (49). Each node of the network is assigned the elevation of the nearest measurement point. The elevation measures, that is maximum and average rise, as well as maximum and average fall, are then calculated per route. For every link longer than 20 meters the slope is calculated directly; if a link is too short it is joined with the next links until the sum of link lengths is greater than 20 meters, the slope is then calculated for the joined segment. The average rise is then calculated as the average of all positive segment slopes; the average fall is the absolute value of the average of all negative segment slopes. Respectively, the maximum fall is the absolute value of the most negative slope.

Routes

Routes are extracted from a data set collected in and around Zurich in 2012/13 comprising of approximately one week of data for 150 participants (50). Data was collected with dedicated GPS trackers and participants corrected the processed travel diaries using a web-based prompted recall tool, with the ability to change times, travel mode and trip purpose. All tracks were double checked by student assistants after the survey period (discussed in the next subsection). Based on these corrected diaries, 7233 stages -that are part of 5128 trips- are observed. Within the defined area, 2250 car stages remain for modelling. Most kilometres are driven on the motorway, followed by primary, secondary and tertiary roads.

Furthermore, 410 bicycle stages are used for modeling. From the original data 82 stages are filtered as being too short (less than 500 meters), being round-trips with same start and end location or being round-trip suspects (chosen distance > 2.5 * shortest distance). Filtering round-trips is especially important as these caused positive distance parameters. To model pedestrian route choice, 985 stages
are available. From the original data round-trips and trips with an average speed higher than 5 m/s are excluded.

FIGURE 1. Map of study area in Zurich (top), and extracted car network (bottom)
Survey design and execution

Survey design

This survey aimed at collecting one week of GPS data of participants living in and around Zurich. The exact study area contains all municipalities within 22 km of Zurich Bellevue, just including the cities of Winterthur in the north-east and Zug in the south. Addresses including telephone numbers were procured from an address dealer. As age distribution of such address databases are in our experience not representative for the population, in particular older people are over represented, addresses were obtained by age category.

The survey was implemented as online questionnaires with two major parts:

- Psychometric scales for the attitudes towards risk, environment and change, and
- The one-week GPS-based travel diary

The third questionnaire concerning person and household characteristics covered basics such as age, gender, income, education level, was enriched by mobility tool ownership (e.g. cars, bikes, public transport season ticket) and concluded with questions about typical locations (home, work, and two main shopping addresses).

The development and implementation of the psychometric scales has been extensively described in Rieser-Schüssler and Axhausen (51) in the context of a pre-study, and is summarized next. Corresponding to the three attitude domains that are investigated in this study, three separate scales have been developed:

- One measuring the risk propensity of the respondents, which combines a reduced version of the domain specific risk propensity scale by Weber et al. (52) with seven additional items for transport related risks, for a total of 42 items;
- One addressing their attitude towards the environment and environmental protection, combining the scales used by Gatersleben et al. (53) and Kitamura et al. (54) into a 25-item scale; and
- One quantifying the level of variety the persons seek in their life, containing 28 questions, including some of those reported by Mokhtarian and Salomon (55)

Each scale is presented to the respondents with a 5-point agree-disagree scale. To minimize effects resulting from the order of the scale items, their order is determined randomly with three different random orders for each scale.

The main parts of the prompted recall interface are

- Visualization of the collected data,
- Presentation of the travel diary, and
- Editing of activities and stages.

For more details regarding the survey, the interested reader is directed to Montini et al. (50).

Survey execution

Response burden as suggested in Axhausen and Weis (56) was calculated to be approximately 1360 points, split as follows on the three survey parts: 66 points for the socio-demographic questionnaire, 190 points for the three psychometric scales, while the majority of the burden caused by the travel diary with around 1100 points. As no incentives were offered and burden is comparatively high, response rate was expected to be moderate.

In total, 1134 persons were contacted by telephone between 6 and 8 in the evening, of those 176 (16%) agreed to participate, 133 persons (12%) were not reached. Young people were both less likely to be
reached and less likely to participate. From the 176 persons agreeing to participate 156 (14%) collected data for at least three days, and were therefore classified as valid.

It was expected that GPS devices are more accepted by younger people, but this does not seem to be the case as people over 55 are well represented, whereas younger people below 25 are highly underrepresented with 1.3% in this study compared to 20.3% in the Microcensus (57). The most interested group are the 45 - 54 year olds, who are well reached by phone and over 20% accepted to participate. At first sight it seems that females were less willing to participate (42%); however, having a look at the addresses revealed that the share of addresses of females was also 42%, therefore, willingness to participate is very similar. Our respondents were wealthier, better educated and lived in smaller households, than a representative sample of the Swiss population. However, this is a common finding in IVT transport studies. Furthermore, the share of public transport ticket owners is higher than in the Microcensus, this could be due to the study area, but also because of the public transport goals of the study.

ANALYSIS METHODOLOGY

Map-matching and network-based variables

Matching the GPS points to a given network is performed as described in Schüssler and Axhausen (2009); the implemented module is part of the POSDAP (47) framework. To ensure valid map-matching, the map-matched distance is compared to the distance computed from the GPS points. Routes are enriched with information on traffic signals and crossings that were extracted from OSM and turn information is calculated for each route alternative, based on a set of rules.

Route choice model

For route choice modeling the path size logit formulated by Ben-Akiva and Bierlaire (35) is used, a multinomial logit (MNL) model (59) with the path size (PS) as adjustment term to correct for overlapping routes. The MNL model is a discrete choice model where utility $U$ is the sum of a deterministic part $V$ and an identically and independently (i.i.d.) Gumbel distributed error term $\varepsilon$; given this definition the probability of alternative $i$ being chosen out of the choice set $C$ is:

$$P(i|C) = \frac{e^{V_i}}{\sum_j e^{V_j}}$$  \hspace{1cm} (1)

The path size is given in Equation 2, ranging from 1, for paths that do not overlap with any other in the choice set, to very small values if there is a lot of overlap. In essence, this can be thought of as a discounting factor, in order to avoid double-counting of similar paths in the choice set.

$$PS_{route, set} = \sum_{link \in route} \frac{\text{length}_{link}}{\text{length}_{route}} \frac{1}{\# \text{routes in set via link}}$$ \hspace{1cm} (2)

The general form of the deterministic part of the utility function is given in Equation 3. The path size is either transformed logarithmically, such that it is very negative for very overlapping routes and 0 for routes without overlap, or more generally a box cox transformation is applied (which simplifies to the logarithmic transformation for $\lambda=0$).

$$V = \beta_{TT} \times \text{travel time} + \beta_{RT} \times \text{proportion road type} +$$

$$+ \beta_{ELEV} \times \text{elevation measure} + \beta_{PS} \times \ln(\text{path size})$$  \hspace{1cm} (3)
For model estimation the python implementation of BIOGEME version 2.4 was used (60). To correct for panel effects, the log likelihood in all models is defined as sum of the conditional probabilities of a person, and only robust error measures computed with the sandwich estimator are reported (61).

**Choice set generation**

For car, bike and walk stages the Breadth First Search on Link Elimination algorithm (BFS-LE), as described in Rieser et al. (62), is used to generate route alternatives. This algorithm was shown to be computationally very efficient, while as well producing relevant routes, e.g. for bicycle routes in Halldorsdottir et al. (63). Most important for the work with high-resolution OSM networks is the performance optimization of the BFS-LE implementation, where a topologically equivalent network is created before choice set generation. That is, vertices that are not a junction, intersection or a dead-end are removed and links are joined per direction. For the car network this means instead of 2'363'307 links 499'928 segments are processed.

For public transport the network is much more sparse; instead schedules have to be considered when generating alternatives. Two of the approaches described in Rieser-Schüssler et al. (64) are used here: the basic and the via point choice set generation (CSG). For the basic version, connections are generated for all combinations of start and end stops that are within acceptable walking distance around origin and destination. The via point choice set generation further acknowledges that some stops provide well-known transfer opportunities, also to other stops.

For the following models up to 200 alternatives were generated for car routes and up to 100 for bike and walk stages. Additionally, the chosen path was added if not generated by the algorithm. These choice sets were reduced using similarity distribution-based reduction described in Schüssler and Axhausen (65) or random sampling of the paths. For public transport, all alternatives are generated given the run parameters as defined in Rieser-Schüssler et al. (64).

**MODEL SPECIFICATION AND RESULTS**

For model estimation some observations are excluded: Car stages where the map-matched distance is more than 2 kilometers longer or 1.5 times the GPS recorded distance are excluded. Bike and walk trips are removed if they are round-trips or if the chosen route is more than 2.5 times the shortest route, assuming that these are sports trips that can not be explained by the available explanatory variables. Furthermore, very short bike trips below 500 meters are removed as well as walk trips with an average speed above 5 m/s. For the mode choice experiment access and egress walk trips are not considered either, as these are part of the public transport connections.

**Route choice models**

For car routes, first, the influence of the choice set size on the parameter estimates is shown in Table 1, given a rather simple model specification considering all types of input features available, i.e. travel time, the sum of turns and traffic lights, as well as the average speed, which is a rough indicator for the road types, as these are to a great extent distinguishable by their speed limits. To ease comparison, results are presented as values scaled to the time parameter of the choice set with 40 alternatives. Different types of choice set elimination were tested and a random sampling has been selected. Parameters estimates start to stabilize with 40 alternatives in the choice set, and therefore this is chosen as the choice set size for the development of the full models, shown in Table 2. A model for the entire sample is shown first, followed by models by trip purpose. (Transformations of) time proportions on different road types, information on turns and signals, free flow time, as well as path size correction factor are the explanatory variables (similar variables are used in Bierlaire et al. (39) and Schüssler and Axhausen (65), two models also estimated on Swiss data).
Travel time parameters have the correct sign, and proportions of extra-urban roads have the most negative sign, followed by proportions of motorway/trunk links. Higher density of left turns, right turns and signals (in decreasing order of impact) have a negative impact on the utility associated with a path, as expected. Table 2 also shows results of the same model by trip purpose. The best model fit of the combined model is 0.36, and the individual models range between 0.28 and 0.36.

Public transport route choice model results are presented in Table 3 for the basic, for as well as the via point choice set generation (CSG). Tendencies of the two models are the same, and all parameters have appropriate signs. The higher $p^2$ supports the via CSG model. The most important parameters are the number of transfers, in-vehicle travel time as well as the share of tram stages. The (very small) positive transfer time parameter estimate makes sense if tight connections are perceived as negative.

### TABLE 1. Scaled parameter values of car route MNL model for different choice set sizes

<table>
<thead>
<tr>
<th>Size</th>
<th>Time [min]</th>
<th>Turns and lights [1/km]</th>
<th>Speed [km/h]</th>
<th>ln(PS)</th>
<th>$\rho^2$</th>
<th>Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>-1.050</td>
<td>-1.245</td>
<td>-0.035</td>
<td>0.480</td>
<td>0.214</td>
<td>1.197</td>
</tr>
<tr>
<td>20</td>
<td>-1.050</td>
<td>-1.127</td>
<td>-0.032</td>
<td>0.412</td>
<td>0.172</td>
<td>1.063</td>
</tr>
<tr>
<td>40</td>
<td>-1.050</td>
<td>-1.100</td>
<td>-0.027</td>
<td>0.359</td>
<td>0.144</td>
<td>1.000</td>
</tr>
<tr>
<td>60</td>
<td>-1.050</td>
<td>-1.069</td>
<td>-0.026</td>
<td>0.350</td>
<td>0.132</td>
<td>0.963</td>
</tr>
<tr>
<td>80</td>
<td>-1.050</td>
<td>-1.069</td>
<td>-0.026</td>
<td>0.338</td>
<td>0.125</td>
<td>0.938</td>
</tr>
<tr>
<td>100</td>
<td>-1.050</td>
<td>-1.050</td>
<td>-0.025</td>
<td>0.325</td>
<td>0.119</td>
<td>0.921</td>
</tr>
<tr>
<td>120</td>
<td>-1.050</td>
<td>-1.041</td>
<td>-0.025</td>
<td>0.320</td>
<td>0.116</td>
<td>0.913</td>
</tr>
<tr>
<td>140</td>
<td>-1.050</td>
<td>-1.050</td>
<td>-0.025</td>
<td>0.319</td>
<td>0.114</td>
<td>0.913</td>
</tr>
<tr>
<td>all</td>
<td>-1.050</td>
<td>-1.033</td>
<td>-0.024</td>
<td>0.305</td>
<td>0.109</td>
<td>0.875</td>
</tr>
</tbody>
</table>

### TABLE 2. Passenger car route choice models for different trip purposes

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
<th>All Rob. t-test</th>
<th>Value</th>
<th>Work Rob. t-test</th>
<th>Value</th>
<th>Leisure Rob. t-test</th>
<th>Value</th>
<th>Other Rob. t-test</th>
<th>Value</th>
<th>t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>ln(Free flow time [min])</td>
<td>-8.41</td>
<td>-13.25</td>
<td>-9.70</td>
<td>-8.56</td>
<td>-8.00</td>
<td>-5.20</td>
<td>-9.44</td>
<td>-14.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sqrt{#\text{left turns} / \text{km}}$</td>
<td>-3.30</td>
<td>-14.07</td>
<td>-1.97</td>
<td>-6.28</td>
<td>-1.36</td>
<td>-4.41</td>
<td>-1.69</td>
<td>-11.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sqrt{#\text{right turns} / \text{km}}$</td>
<td>-3.05</td>
<td>-13.29</td>
<td>-1.05</td>
<td>-2.75</td>
<td>-1.43</td>
<td>-4.58</td>
<td>-1.58</td>
<td>-9.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sqrt{#\text{traffic signals} / \text{km}}$</td>
<td>-1.43</td>
<td>-5.10</td>
<td>-1.16</td>
<td>-2.79</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>-6.45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T(prop. motorway)</td>
<td>-2.72</td>
<td>-6.69</td>
<td>-2.86</td>
<td>-3.36</td>
<td>-3.26</td>
<td>-4.28</td>
<td>-2.52</td>
<td>-5.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(path size)</td>
<td>0.65</td>
<td>10.94</td>
<td>0.91</td>
<td>7.57</td>
<td>0.47</td>
<td>3.47</td>
<td>0.66</td>
<td>9.41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Num. of observations</td>
<td>2035</td>
<td>532</td>
<td>416</td>
<td>1087</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Init log likel. $\mathcal{L}(\beta_0)$</td>
<td>-7506.87</td>
<td>-1962.48</td>
<td>-1534.57</td>
<td>-4009.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final log likel. $\mathcal{L}(\hat{\beta})$</td>
<td>-4824.32</td>
<td>-1328.49</td>
<td>-1097.91</td>
<td>-2545.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho^2$</td>
<td>0.36</td>
<td>0.32</td>
<td>0.28</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: $T()$: arcus sinus transformation $\bar{y} = \arcsin(\sqrt{y})$

### TABLE 3. Public transport route choice models for different choice set methods

<table>
<thead>
<tr>
<th>Name</th>
<th>Basic CSG</th>
<th>VIA CSG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Egress travel time</td>
<td>-0.0047</td>
<td>-2.31</td>
</tr>
<tr>
<td>Prop. tram stages []</td>
<td>2.02</td>
<td>6.2</td>
</tr>
<tr>
<td>In vehicle travel time [min]</td>
<td>-0.232</td>
<td>-4.66</td>
</tr>
<tr>
<td>Number of transfers</td>
<td>-2.11</td>
<td>-8.18</td>
</tr>
<tr>
<td>Path size (stage times)</td>
<td>1.70</td>
<td>3.25</td>
</tr>
<tr>
<td>Transfer time [min]</td>
<td>0.0032</td>
<td>4.58</td>
</tr>
<tr>
<td>Num. of observations</td>
<td>268</td>
<td>273</td>
</tr>
<tr>
<td>Init log likel. $\mathcal{L}(\beta_0)$</td>
<td>-667.78</td>
<td>-800.181</td>
</tr>
<tr>
<td>Final log likel. $\mathcal{L}(\hat{\beta})$</td>
<td>-300.519</td>
<td>-318.058</td>
</tr>
<tr>
<td>$\rho^2$</td>
<td>0.541</td>
<td>0.595</td>
</tr>
</tbody>
</table>
Combined mode and route choice model

Parameter estimates for the combined mode and route choice model are given in Table 4. The influence of person-based variables is fixed for car, in this case the alternative specific constant (ASC) and whether a person has no car or only seldom access to one. The alternative specific constants for the other modes are highly significant and very negative. Having no car or only seldom access to one has a positive effect on the other modes, as expected. Of the attitude scores only being in agreement with CO₂ reduction measures (Env. score 1) has a significant positive effect on choosing to cycle. The other attitude scores, gender, age, income, household size as well as trip purpose were tested, but not significant (with the exception of walk leisure trips). The route choice parameters for the modes are similar to the separate model estimations.

<table>
<thead>
<tr>
<th></th>
<th>Car</th>
<th>Rob. t-test</th>
<th>PT</th>
<th>Rob. t-test</th>
<th>Bike</th>
<th>Rob. t-test</th>
<th>Walk</th>
<th>Rob. t-test</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASC</td>
<td>0 (ref)</td>
<td>N/A</td>
<td>-7.33</td>
<td>-10.71</td>
<td>-5.96</td>
<td>-10.19</td>
<td>-3.33</td>
<td>-5.01</td>
</tr>
<tr>
<td>No car or rarely</td>
<td>0 (ref)</td>
<td>N/A</td>
<td>3.62</td>
<td>5.81</td>
<td>2.68</td>
<td>4.03</td>
<td>3.31</td>
<td>5.11</td>
</tr>
<tr>
<td>Prop motorway</td>
<td>-1.83</td>
<td>-4.53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop extra urban</td>
<td>-1.82</td>
<td>-2.94</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prop track/other</td>
<td>-4.07</td>
<td>-4.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nr left turns/km</td>
<td>-0.952</td>
<td>-9.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nr right turns/km</td>
<td>-0.714</td>
<td>-5.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nr traffic lights</td>
<td>-0.606</td>
<td>-7.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ln(Car time [min])</td>
<td>-2.77</td>
<td>-10.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Egress travel time [min]</td>
<td>-0.105</td>
<td>-2.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>In vehicle travel time [min]</td>
<td>-0.159</td>
<td>-6.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share tram [-]</td>
<td>2.19</td>
<td>7.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Envl score</td>
<td>0.336</td>
<td>2.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distance [km]</td>
<td>-0.289</td>
<td>-2.69</td>
<td>-2.79</td>
<td>-6.69</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum rise</td>
<td>-11.4</td>
<td>-5.34</td>
<td>-3.65</td>
<td>-4.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leisure trip</td>
<td>1.23</td>
<td>4.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample size: 2861
Init log likel. $L(\beta_0) = -8245.42$
Final log likel. $L(\hat{\beta}) = -5536.975$
$\rho^2 = 0.326$

DISCUSSION AND CONCLUSION

Route and mode choice is an important input to transportation planning and engineering. Traditionally, these models have been estimated using stated-preference questionnaires or (costly and difficult to execute) revealed preferences studies. The advancement of data collection technologies allows for a paradigm shift that could both make data collection for transportation model estimations easier, and allow for the acquisition of richer and more reliable data.

In this research, we present such a study, where a sample from Zurich, Switzerland, used GPS devices to monitor their daily trips for a week. A follow-up survey provided complementary information, including sociodemographic characteristics of the participants, as well as a large number of attitudinal parameters. Conducting such an extensive survey needs a lot of effort and attention. Realization with part-time assistants and therefore spreading it over a long time period is not ideal. Our main finding, that might be obvious to experts, is that full-time workers are needed, and we consequently recommend concentrating survey periods to predefined weeks.

One of the goals when using GPS data is reduction of response burden; achieving this strongly depends on the display and handling of the prompted-recall diary. The user-friendliness and clarity of
the prompter-recall interface is important in improving the response rate and quality of the responses. Finally, the quality of the GPS tracks and the quality of the stage and stop point detection are key to the reduction of response burden. As in the future probably more and more GPS data sets are likely to be collected by Smartphones, ensuring good quality and homogeneity will be a major challenge, especially if different Smartphones and therefore different GPS sensors will be used.

As overall the quality of the corrections was very diverse, all diaries were double checked by the researchers. The majority of the diaries needed some additions, which is definitely not cost-effective for large survey samples, while some people simply handed back hand-written notes, which were entered into the system by the researchers. Support and reminders during the survey period are therefore extremely important.

Route choice models were estimated using this data set for car and public transport trips, along with a joint route and mode choice model, which also considered bike and walk stages. To the best of our knowledge, this is the first time that GPS data have been used for the estimation of mode and combined mode and route choice models. The model fit of all models is reasonable and in line with those obtained with traditional data-collection methods. The coefficient estimates have the appropriate signs and meaningful magnitudes. A number of transformations were performed on some of the variables, improving the model fit significantly, and reducing the number of variables.

While the data collection included a large number of sociodemographic and attitudinal variables, very few of them were retained in the final model specifications. Incorporation of such variables in the models could have improved their fit (if they were found significant in the model formulations); on the other hand, the need to collect such data might have decreased the ease of applicability, as well as the transferability of these models.

In conclusion, in this research we have used a novel data collection technique to collect rich data about the travel patterns of a sample of the population of Zurich Switzerland. We discuss some of the difficulties associated with such data collection efforts, and point out the advantages that can come from them. Discrete choice models are estimated and the results are discussed. The main findings of these models are intuitive, and illustrate the vast opportunities that could come from follow-up studies, based on similar concepts.

ACKNOWLEDGEMENTS
"Data collection was funded by the Swiss State Secretariat for Education and Research as part of the project "Route choice in urban public transport systems" within the COST Action "TU0603 - Buses with high levels of service". We would like to thank our student assistants: Ilona Imoberdorf, Nathalie Schenk, Sebastian Sele, Lara Carisch, Fabian Binder and especially Patrice Frei and Tobias Karrer who assisted us during the entire survey." The third author is thankful for the financial support from ETHZ during his visiting professorship there during the period of March to June 2016.

REFERENCES


