GPS, GSM, Diary: How to capture travel behaviour?

KW Axhausen

IVT
ETH
Zürich

July 2012
Acknowledgements

Stated choice and diaries:
  • Claude Weis
  • Alexander Erath

Social networks:
  • Timo Ohnmacht
  • Andreas Frei
  • Matthias Kowald

GPS surveys and analysis:
  • Nadine Rieser – Schüssler
  • Lara Montini
Starting point
Hypotheses for travel behaviour

Wages

Activities

Fleet comfort

Housing consumption

Energy costs

vttss et al.

Elasticity > 0

Elasticity < 0

Specialisation

Tours

pkm

vkm

Migration

Professional and personal activity space

Number of networks

Network overlap

Network geography

Local anomie

- Elasticity < 0

+ Elasticity > 0
Evidence
Productivity growth in Western Europe

Adapted from Galor and Weil (2000)
Road based – Switzerland 1950 and 2000

Scherer, 2004
VTTS all purposes Switzerland 2001 - 2006
Real prices of telekommunication

US International and interstate average revenue per minute

Quelle: nach FCC (2001)
What do we need?
Stages – trips – tours - activities

At home
Stages – trips – tours - activities

Breakfast
Stages – trips – tours - activities
Stages – trips – tours - activities

Trip 1

Work
Stages – trips – tours - activities

Tour 1
What should we capture?

Elements of the generalised costs of the movement:

- Duration of the stages
- Routes of the stages
- Circumstances of the stages (congested; parking search)
- Monetary (decision relevant) costs of the stages
- Joint activities during the stages
- Joint travel with whom
- Time pressure of the stages
What should we capture?

Elements of the generalised costs of the non-chosen alternatives (modes * routes/connections):

- Duration of the stages
- Routes of the stages
- Circumstances of the stages (congested; parking search)
- Monetary (decision relevant) costs of the stages
Elements of the generalised costs of the activity:

- Congestion
- Price levels
- Price worthiness (value for money)
- Social milieu

- Purpose of the activity
- Joint activities during the activity
- Joint activity with whom and expenditure sharing, if any
- Planning horizon of the activity

What should we capture?
What should we capture?

Elements of the generalised costs of the **non-chosen** alternatives (locations * purposes):

- Congestion
- Price levels
- Price worthiness (value for money)
- Social mileu
But how?
<table>
<thead>
<tr>
<th>Dimension</th>
<th>Diary</th>
<th>GSM</th>
<th>GPS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trips</td>
<td>Easy</td>
<td>Impossible</td>
<td>Post-processing</td>
</tr>
<tr>
<td>Completeness</td>
<td>Respondent dependent</td>
<td>No</td>
<td>Yes, but data loss possible</td>
</tr>
<tr>
<td>Duration</td>
<td>Rounded to the next 5, 15 min</td>
<td>No</td>
<td>Exactly</td>
</tr>
<tr>
<td>Destinations</td>
<td>(Exactly)</td>
<td>Cell tower</td>
<td>Exactly</td>
</tr>
<tr>
<td>Purpose</td>
<td>Yes</td>
<td>No</td>
<td>Imputation</td>
</tr>
<tr>
<td>Company</td>
<td>Partially</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Routes</td>
<td>Expensively</td>
<td>(Impossible)</td>
<td>Exactly</td>
</tr>
<tr>
<td>Recruitment</td>
<td>F(response burden)</td>
<td>(Easy)</td>
<td>F(response burden)</td>
</tr>
<tr>
<td>Response period</td>
<td>1 (-42) days</td>
<td>(Unlimited)</td>
<td>(1-) 7-14 (-) days</td>
</tr>
</tbody>
</table>
Response rate = f(response burden) @ IVT
Diaries
Format: Swiss Mikrozensus (MZ) 2010

- Geocoded CATI interview (LINK)
  - Person- and household socio-demographics
  - Stage-based travel diary
  - Routes of car/motorcycle-stages were identified with two way-points (interviewer had map interface)
- Add-on modules for sub-samples (LINK)
  - One-day excursions
  - Long-distance travel
  - Attitudes to transport policy
- Integrated, but independent SC questionnaire (IVT)
  - SC mode choice
  - SC route and departure time choice
Sample: MZ 2010

- Households (59’971)
- Persons (62’868)

- Stages (310’193)
  - Car/mc > 3km with 2 way-points
  - Railway connections with HAFAS
  - Recruitment of the SC-experiments (IVT)
Protocol: MZ 2010

- CATI with multiple calls (no incentives) (72% response rate)
- Recruitment for the SC: 50% willingness
- Customized SCs for 85% of the recruited within 12 days
- 70% response rate for the SC experiments
Research and development needs: Diaries

- Integration of the “non-chosen” alternatives (Routes, transit service levels, flight costs, etc.)
- Trade-off between incentive and participation (“young men with smatz phones only”)
- Integration of relevant and standardised scales (Environment, risk, variety seeking)
- Integration of social networks
- Integration of multi locality
Why social networks in transport/spatial planning?
Example: Number of accompanying travellers

- Short vacation
- Excursion: nature
- Other
- Excursion: culture
- Meeting friends
- Further education (leisure)
- Garden/ cottage
- Voluntary work
- Disco, pub, restaurant, cinema
- Meeting relatives/family
- Window shopping
- Pick up/drop off/attendance
- Group/club meeting
- Family duty
- Cemetery
- Active sports
- Education
- Long-term shopping
- Walk or stroll
- Daily shopping
- Private business
- Private business (doctor,...)
- Work

Axhausen et al., 2007
Example: Required travel for leisure meetings of ego-alter

Distance between home locations [km]

Percent [%]

Important contact

- No
- Yes

Schlich et al., 2002
Heterogenity in choice

Location choice

- WTP
- Taste
- Joint choice with family, friends, persons to meet
- Schedule constraints
- Social constraints

For mode choice in addition

- Luggage
- Company
## Example: Residential location choice in Kt. Zürich

<table>
<thead>
<tr>
<th>Variable</th>
<th>Beta</th>
<th>t-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rent/Income</td>
<td>-5.51</td>
<td>***</td>
</tr>
<tr>
<td>log(m²/head)</td>
<td>0.98</td>
<td>***</td>
</tr>
<tr>
<td>Frequency weighted mean distance to friends</td>
<td>-8.16</td>
<td>*</td>
</tr>
<tr>
<td>Exponent (friends)</td>
<td>0.22</td>
<td>**</td>
</tr>
<tr>
<td>Mean distance to work/school</td>
<td>-1.59</td>
<td>**</td>
</tr>
<tr>
<td>Exponent (distance to work)</td>
<td>0.37</td>
<td>**</td>
</tr>
<tr>
<td>Travel time to Bürkliplatz</td>
<td>0.02</td>
<td>**</td>
</tr>
<tr>
<td>log(transit accessibility) * &quot;No car&quot;</td>
<td>0.41</td>
<td>**</td>
</tr>
<tr>
<td>log(car accessibility) * “Car&quot;</td>
<td>-0.30</td>
<td>**</td>
</tr>
<tr>
<td>Share of equally sized HH within 1 km</td>
<td>0.02</td>
<td>*</td>
</tr>
<tr>
<td>Population density within 1 km</td>
<td>0.01</td>
<td>**</td>
</tr>
<tr>
<td>Share of empty flats in municipality</td>
<td>-0.11</td>
<td></td>
</tr>
</tbody>
</table>

N = 683, ρ₀² = 0.2128; * > 0.1; ** > 0.05; *** > 0.01

Belart, 2011
Travel and social networks
Benchmarking the current state

- Numbers of contacts
- Distance distributions
- Geographies
- Frequency and mode of contact

- “Productivity”
- Levels of local anomie
- Levels of local trust
- Level of place attachment
Empirical strategy

- Surveys of social geographies & mobility biographies
  - Egocentric
  - Snowball

- Travel diaries
  - One-Day
  - Multiple days

- With/without information about the presence of others
- With/without named co-travellers, co-present persons
Social network surveys @ IVT

- Ohnmacht: 50 egos qualitative/quantitative in Zürich
- Larsen/Urry: 24 egos qualitative/quantitative in NE England
- Frei: 300 egos quantitative in Zürich
- Kowald: snowball; 750 egos quantitative worldwide (starting with 40 egos in Kanton Zürich)(12000 alters in total) (8 day diary included)
Number of contacts reported

Frei and Axhausen, 2007
Distances between home locations

![Histogram of great circle distances](image)
Great circle distances between “leisure” contacts

Daten: Schneeballbefragung IVT, Siehe Kowald et al. 2012
Example of a social network geography
Size of network geometries

95%-confidence ellipse of the social network geography

Frei and Axhausen, 2007
Interactions by mode and distance between homes

![Graph showing interactions by mode and distance between homes](image)

- Face-to-face visits/year
- Phone calls/year
- Email messages/year
- SMS messages/year

Frei and Axhausen, 2007
2010/11 Snowball survey
Challenges of snowball sampling

Challenges:

- Start with representative seeds
- Avoid selection bias
- React to homogeneous clusters
- Correct the overrepresentation of 'socializers' and underrepresentation of 'isolates'

Kowald and Axhausen, 2011
Response rate and response burden (IVT surveys)

Kowald and Axhausen, 2011
Behind egos’ horizons: The connected ‘snowball’-graph

Kowald and Axhausen, 2011
Comparisons
Transport motivated social network surveys

East York, Ontario (Wellman, Carrasco et al.)

Eindhoven, Netherlands (Arentze, Van der Berg)

Concepcion, Chile (Carrasco)

City of Zürich (Frei)

Kanton Zürich snowball (Kowald)
Contact “density” – shares by distance class

Zurich
Eindhoven
Switzerland
Concepcion
Toronto
Shares of contact by mode

**Face-to-face**

**Telephone**

**Internet**

- Zurich
- Eindhoven
- Switzerland
- Concepcion

Great circle distance [km]
Multi-locality
Multi-locality: Multiple reference points

• Parental “home”

• “Home”

• Student digs
  • Pied-a-terre
  • Company dormitory
  • Hotel (chain)

• Weekend home(s) (of others)
  • RV (camp ground) (of others)

• Living apart together
GPS self-tracing
### Current examples

<table>
<thead>
<tr>
<th>Captured with</th>
<th>Where</th>
<th>What</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPS logger</td>
<td>Cincinatti, OH</td>
<td>Regional transport study; purpose imputation (Ohio DOT)</td>
</tr>
<tr>
<td></td>
<td>Jerusalem</td>
<td>Regional transport study; (Israel DOT)</td>
</tr>
<tr>
<td></td>
<td>Kanton Zürich</td>
<td>Transit route choice (IVT, ETH)</td>
</tr>
<tr>
<td>Smartphones</td>
<td>Singapore</td>
<td>Capturing activities within and without building (SMART)</td>
</tr>
<tr>
<td></td>
<td>Bay Area</td>
<td>Capturing trips (Joan Walker and UC Berkeley Kollegen)</td>
</tr>
</tbody>
</table>
COST and peacox 7th framework project: GPS based diary at IVT

GPS unit:
- Interval: 1Hz
- 3D position
- Date and time
- HPOD and other measures of accuracy

Accelerometer
- Interval 10 Hz
- 3D acceleration

Battery
- Multiple days

GSM:
- Savings every 4 hours on SQL database server
GPS-based prompted recall survey

300 participant for 7 days

Web-survey
- Socio-demographics
- Attitudes to risk, environment, variety seeking
- Checking and correction of the automated processing
Data processing

Filtering and smoothing

Identify stages and stops

Stages

Mode detection

map-matching

Purpose imputation

Stops

Analysis and application
Filtering and smoothing

Filtering

• VDOP > 5
• Unrealistic elevations
• Jumps with $v > 50\text{m/s}$

Smoothing

• Gauss Kernel smoother over the time axis
• Speed as first derivate of the positions
• Raw data are kept
Detection of stages and stops

Clusters of high density

Longer breaks without points
  • Accelerometer data
  • GPS points

No movement
  • $V \approx 0$ km/h
  • No accelerations

Mode change
  • Walk stage as signal
Number of trips/day in comparison with MZ 2005

Source
- MZ 2005
- GPS study

Quelle: Schüssler, 2010 (ohne Beschleunigungsdaten)
Trip durations and length in comparison with MZ 2005

Quelle: Schüssler, 2010 (ohne Beschleunigungsdaten)
## Comparison with MZ 2005

<table>
<thead>
<tr>
<th></th>
<th>ZH</th>
<th>WI</th>
<th>GE</th>
<th>MZ 2005</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of persons</td>
<td>2,435</td>
<td>1,086</td>
<td>1,361</td>
<td>2,940</td>
</tr>
<tr>
<td>Days per persons</td>
<td>6.99</td>
<td>5.96</td>
<td>6.51</td>
<td>1</td>
</tr>
<tr>
<td>Trips per day</td>
<td>4.50</td>
<td>3.40</td>
<td>4.26</td>
<td>3.65</td>
</tr>
<tr>
<td>Trip lengths [km]</td>
<td>7.72</td>
<td>7.37</td>
<td>7.19</td>
<td>8.79</td>
</tr>
<tr>
<td>Daily trip lengths [km]</td>
<td>34.74</td>
<td>23.20</td>
<td>29.25</td>
<td>32.13</td>
</tr>
<tr>
<td>Trip duration [min]</td>
<td>15.17</td>
<td>13.71</td>
<td>15.05</td>
<td>26.21</td>
</tr>
<tr>
<td>Stages per day</td>
<td>1.40</td>
<td>1.31</td>
<td>1.47</td>
<td>1.68</td>
</tr>
</tbody>
</table>

Quelle: Schüssler, 2010 (ohne Beschleunigungsdaten)
Mode detection

Speed profile

Accelerometer profile

Home | Walk | Waiting | Bus
---|---|---|---

Bus | Car | Tram | Stop point | Walk
Trip length by mode in comparison with MZ 2005

Source: Schüssler, 2010 (ohne Beschleunigungsdaten)

Zu Fuss

Fahrrad

Pkw

Städtischer ÖV

Bahn
Map matching

Car and bike stages
Selection from a set of possible routes
• At nodes all possible on-going links become new candidate routes (branches)
• If the tree has enough branches, it is pruned based on their total errors
Each candidate branch is assessed by
• Squared error between GPS and path
• Deviation between GPS speed and posted speeds

Transit-stages
• Route identified as for cars
• Line identified by time table
Map-Matching: First branches

initial route candidate
Map-Matching: Number of branches against computing times

Quelle: Schüssler, 2010 (ohne Beschleunigungsdaten)
Research needs: GPS post processing

• Test dataset with true values

• Automatic calibration of the fuzzy logic parameters

• Further integration between map-matching and mode detection

• Integration of the logic of the stage – mode – sequences

• Better imputation of the movement during signal loss

• More and better purpose imputation (POI – data base; land use data; frequency over multiple days)
What now? Cost as a function

- Rate of usable addresses
- Recruitment rate
- Response rate
- Rate of usable returns
- Correlation between household members
- Correlation between days
- Correlation between tours of a day
- Correlation between trips of tour
- Number of waves with the GPS loggers
- Rate of loss of the GPS loggers
## What now?

<table>
<thead>
<tr>
<th><strong>Advantages</strong></th>
<th><strong>Diaries</strong></th>
<th><strong>GPS-self tracing</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All variables</td>
<td>All movements (but data loss)</td>
</tr>
<tr>
<td></td>
<td>Social contacts</td>
<td>Exact times, routes, locations</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Longer observation periods</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Disadvantages</strong></th>
<th><strong>Diaries</strong></th>
<th><strong>GPS-self tracing</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>≈15% under reporting of trips</td>
<td>Post-processing and imputation (Effort of the post-processing)</td>
<td></td>
</tr>
<tr>
<td>Rounded times</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximate routes only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decreasing response rates / expensive response</td>
<td>Unknown response rates (Costs of the units and their distribution/collection)</td>
<td></td>
</tr>
</tbody>
</table>
Questions?

www.ivt.ethz.ch

www.matsim.org

www.futurecities.ethz.ch
Literature and references


Literature and references


