Bevorzugter Zitierstil für diesen Vortrag

Axhausen, K.W. (2016) How to capture long-distance travel?, presentation at the DLR, Berlin, January 2016.

How to capture long-distance travel?

KW Axhausen

IVT
ETH
Zürich

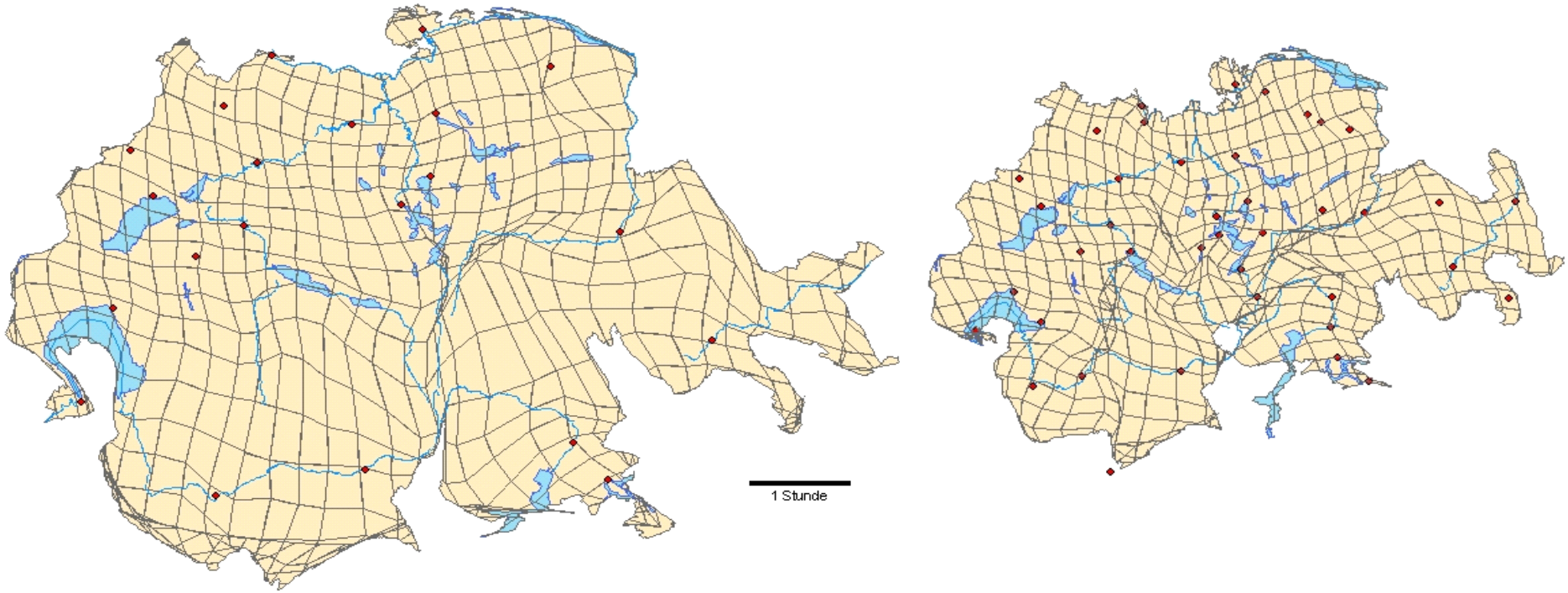
January 2016

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

What is the issue?

Road based – Switzerland 1950 and 2000



Scherer, 2004

Data challenges

Surveys, observations are „talk“

Two speakers

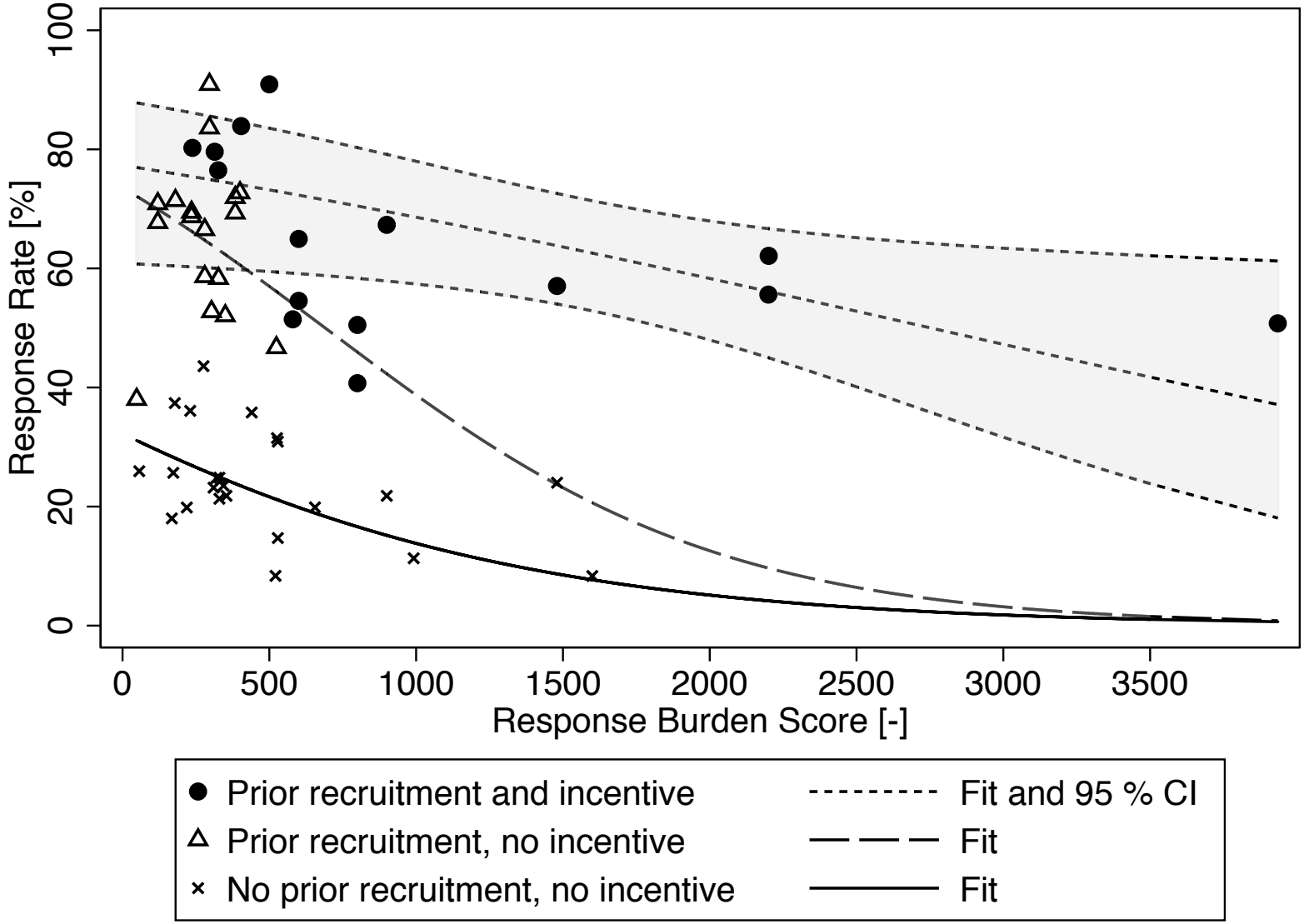
managing their „image“
staying within the rules of talking
staying within their socially allocated/identified role
fulfilling social expectations

talk and report with/to each other

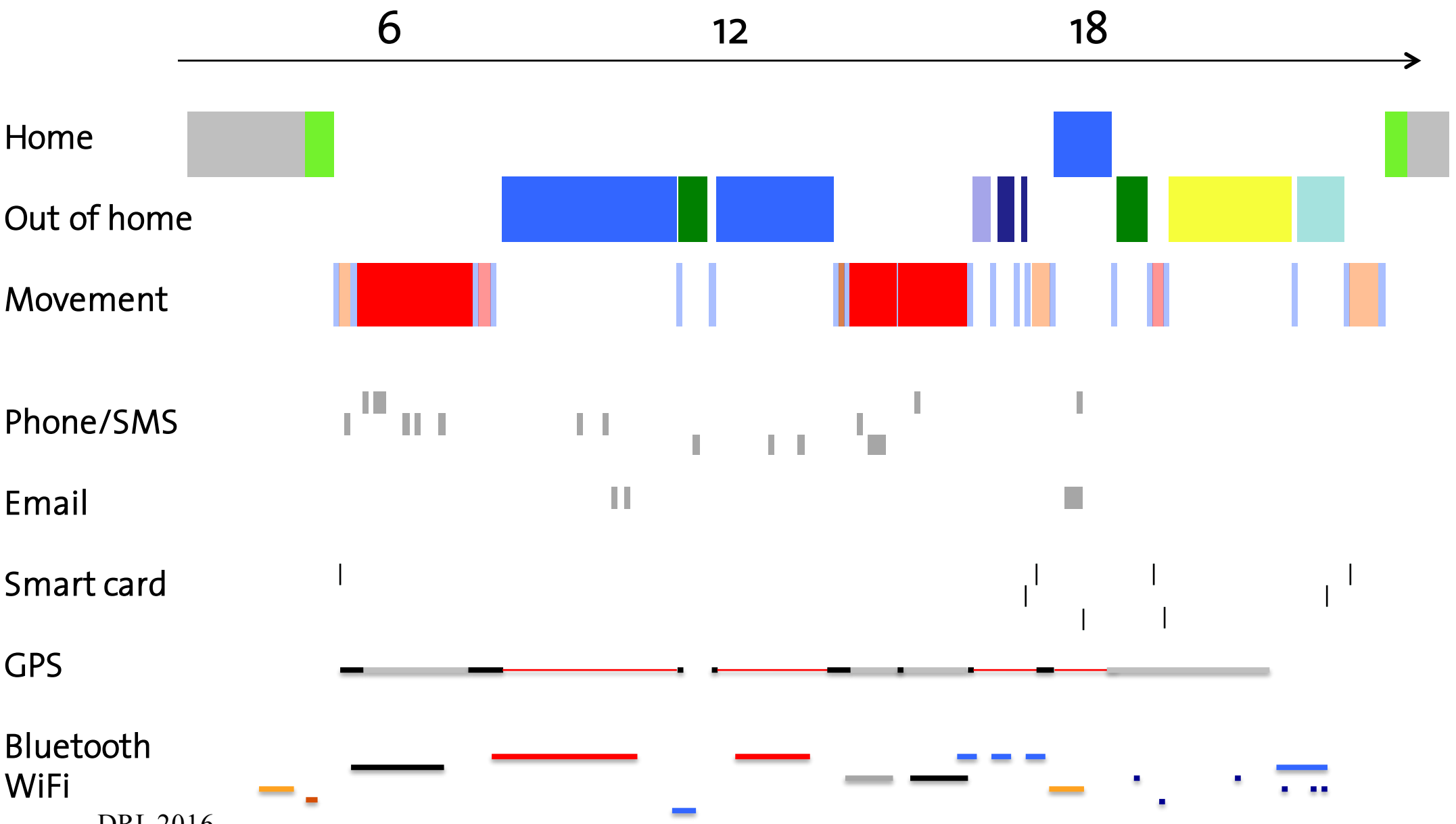
=>

„Maintaing the willingness of the respondent to report“

Response as a function of response burden @IVT, 2015

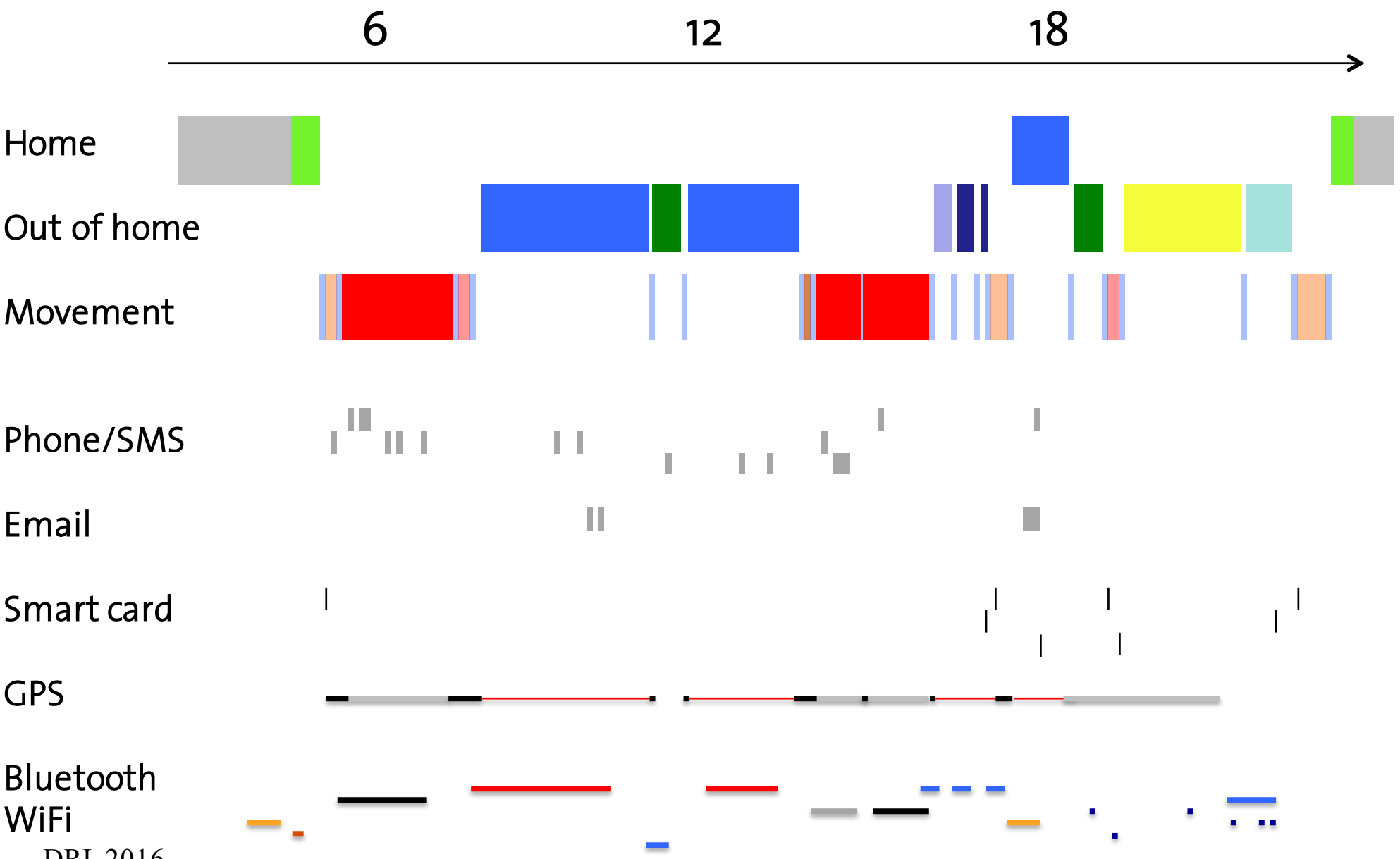


Activities, movement and traces: A full example record

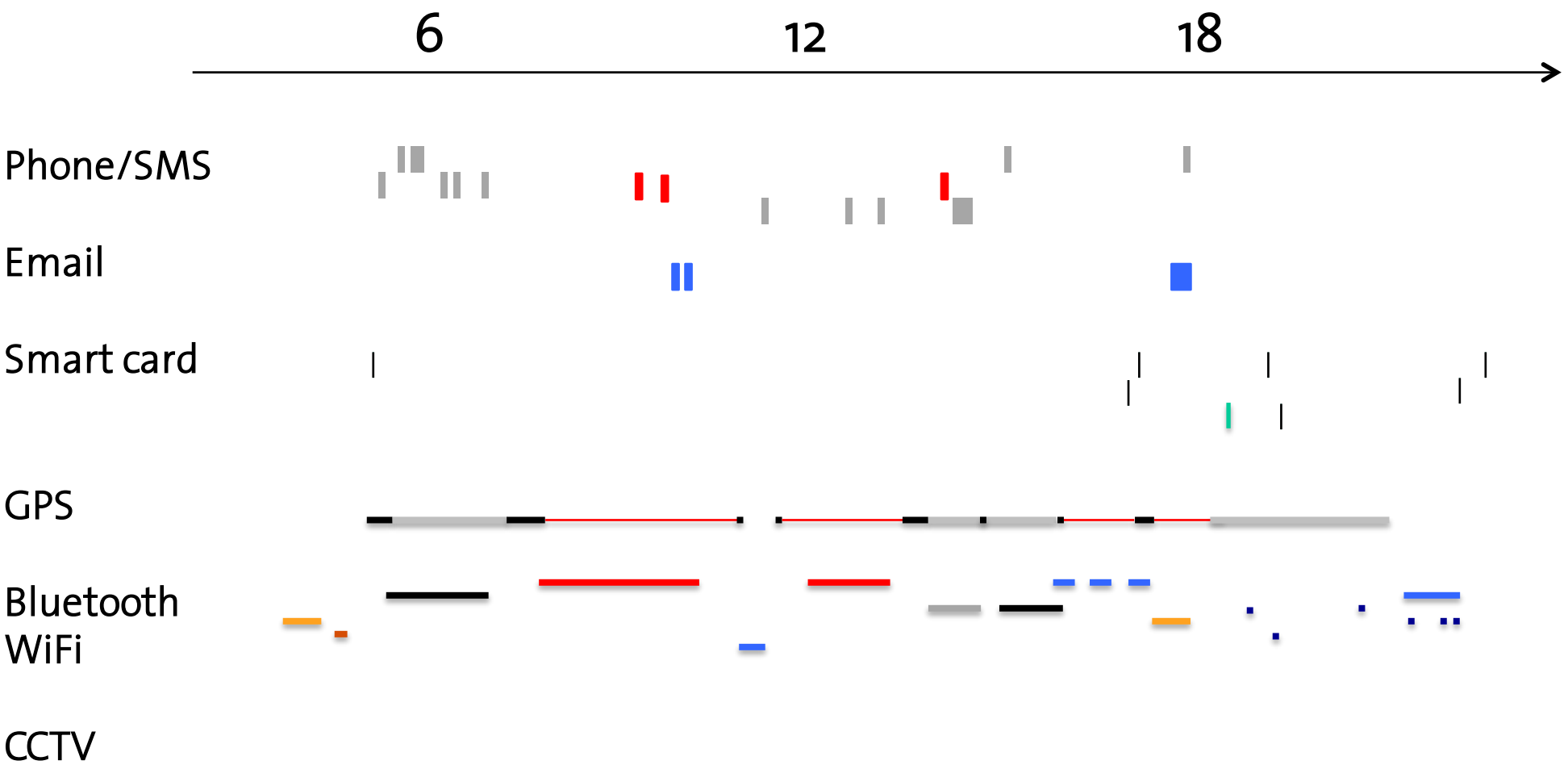


Known „error“ generating processes

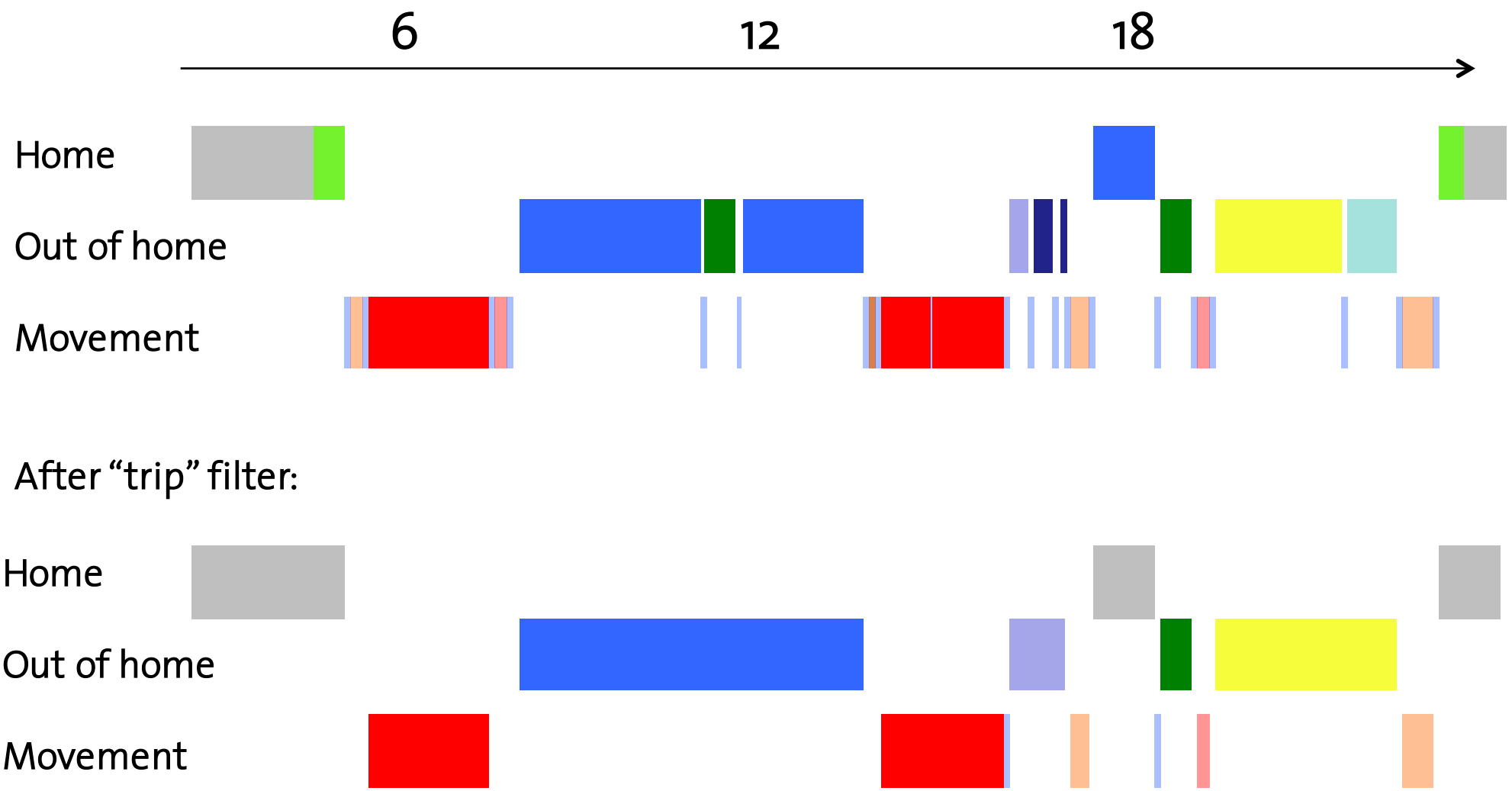
Activities, movement and traces: A full example record



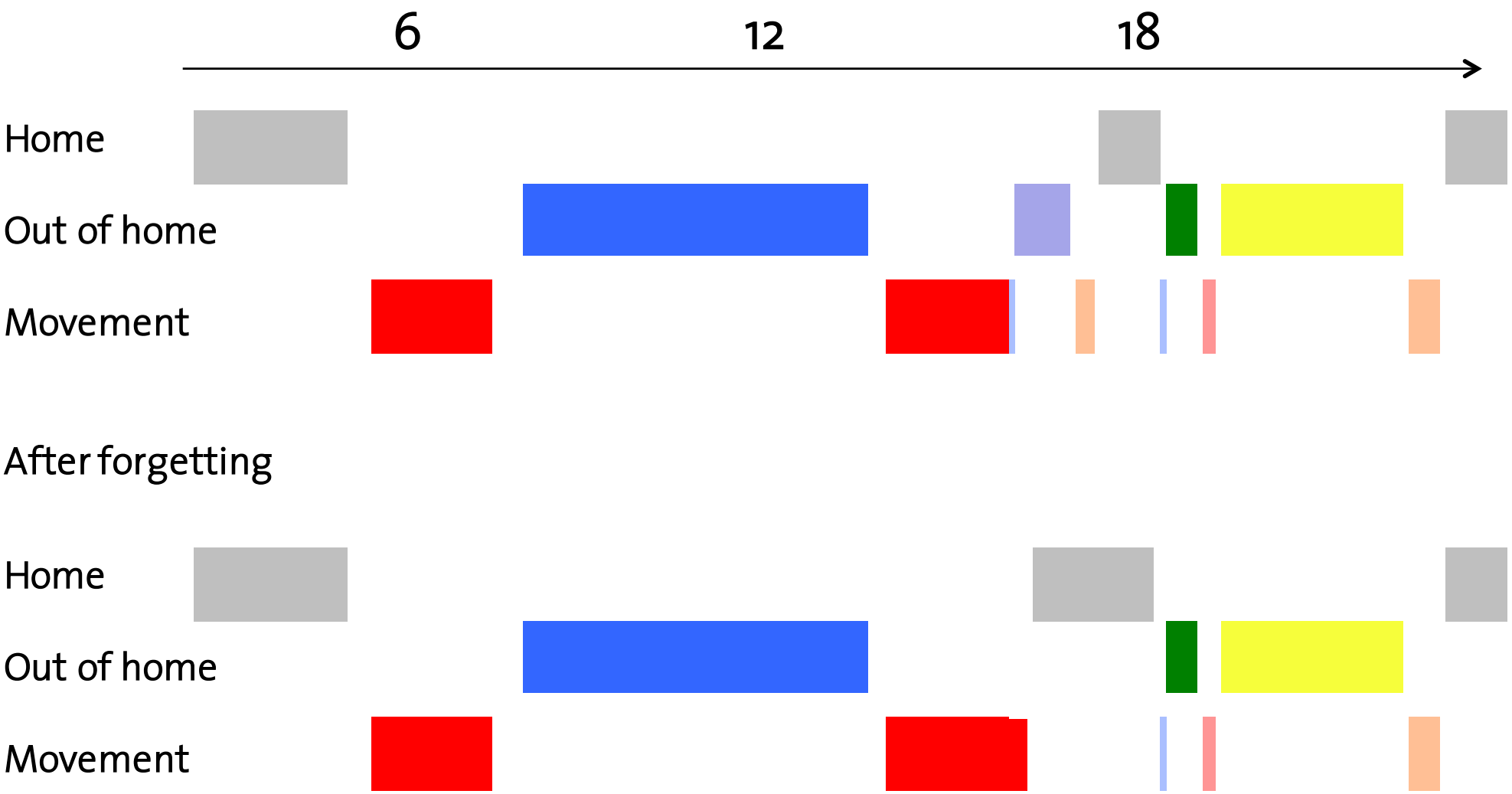
Active/passive tracing: Many owners, locations, quality levels



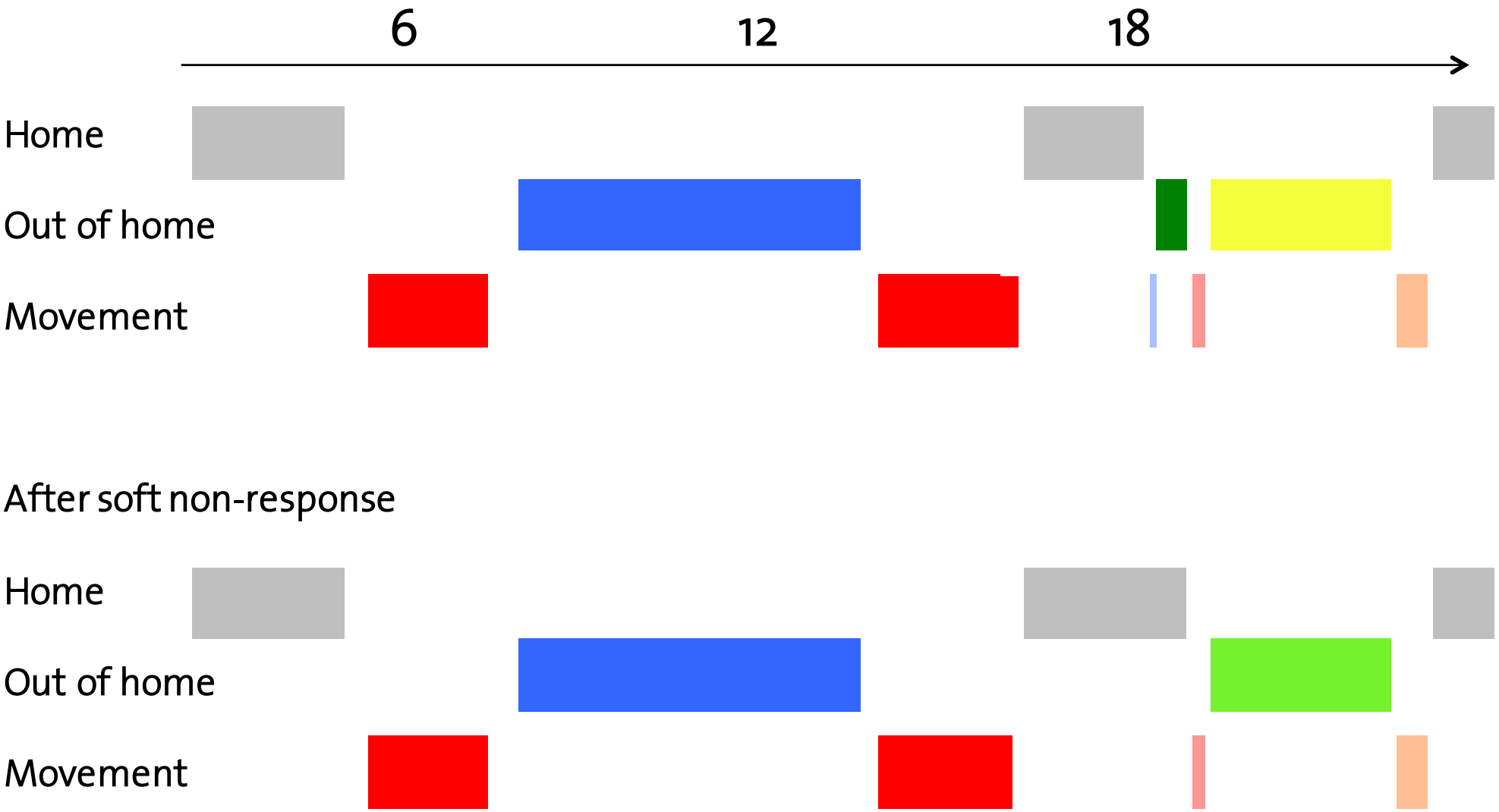
Filters imposed/suggested by the study: „Trips“



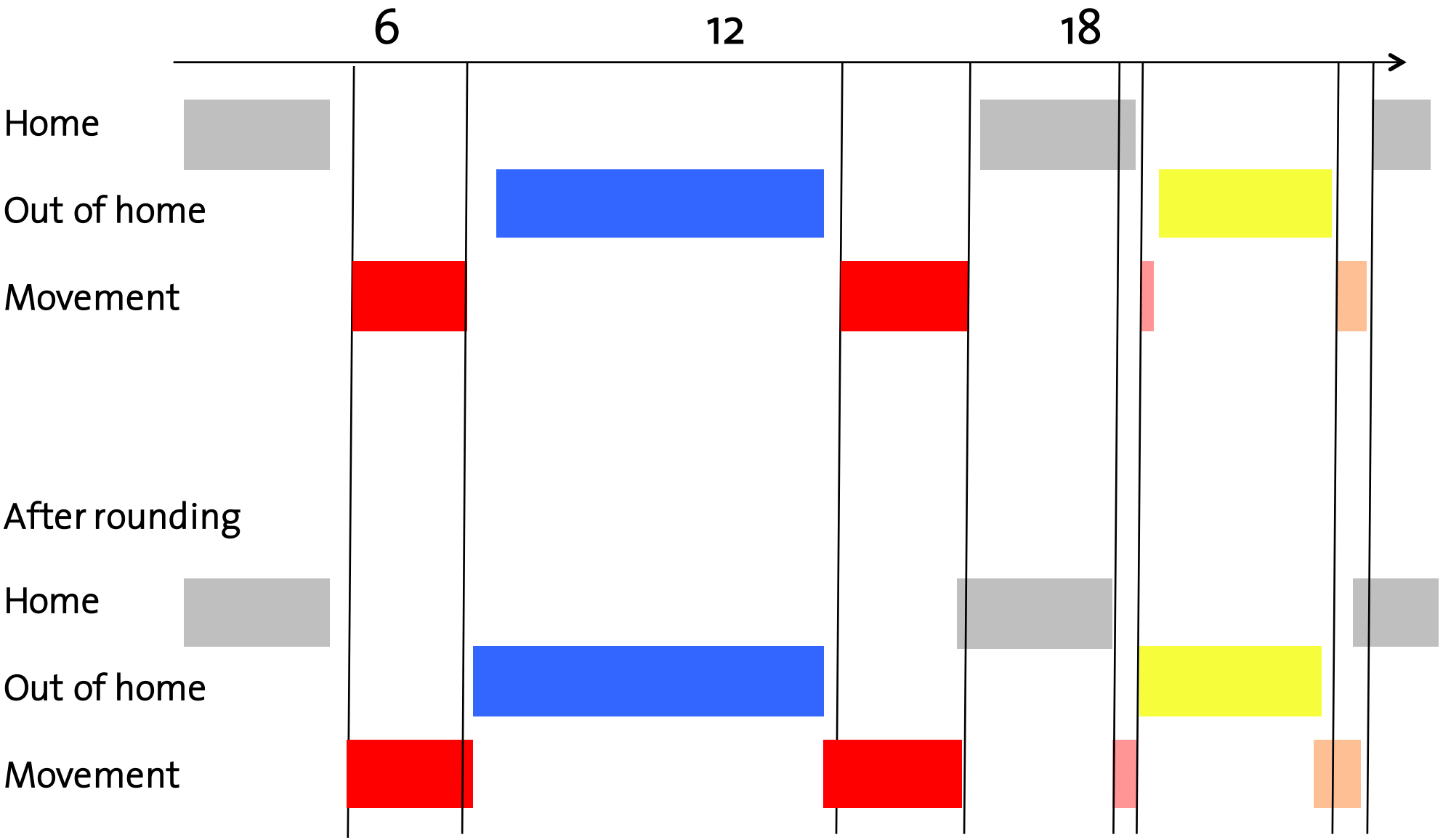
Filters due to the respondent: Forgetting



Filters imposed by the respondent: Soft non-response



Filters due to the respondent: Rounding



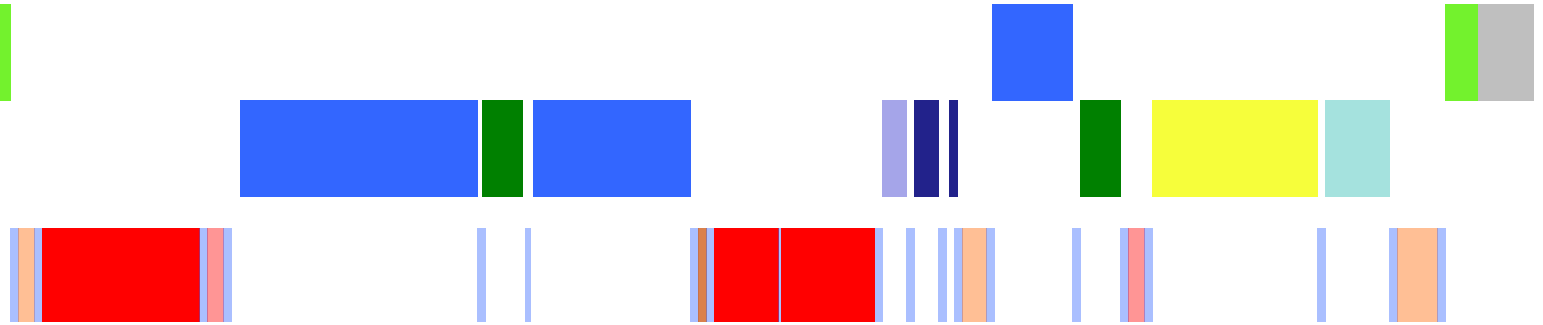
What is left ?

True

5 at home

9 Out of home

26 Stages,
11 trips,
1 subtour,
2 tours

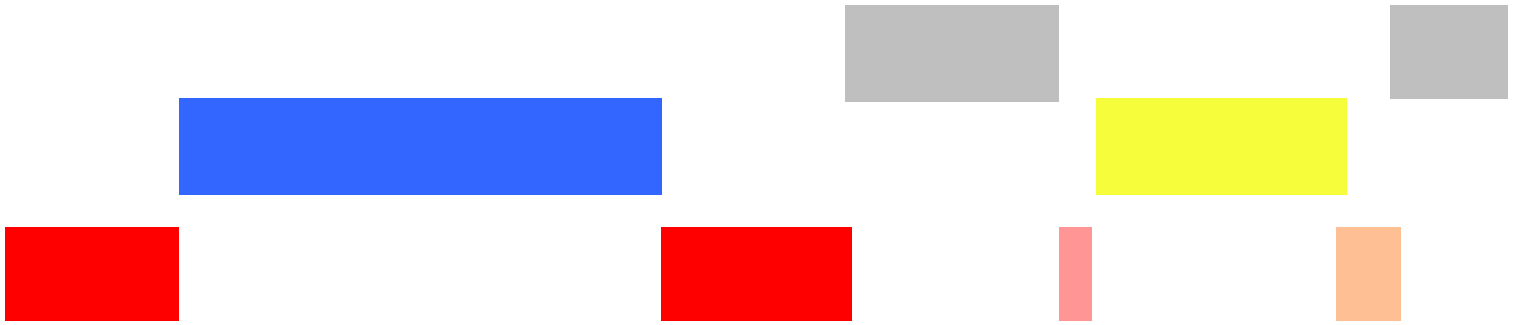


After all processes

3 at home

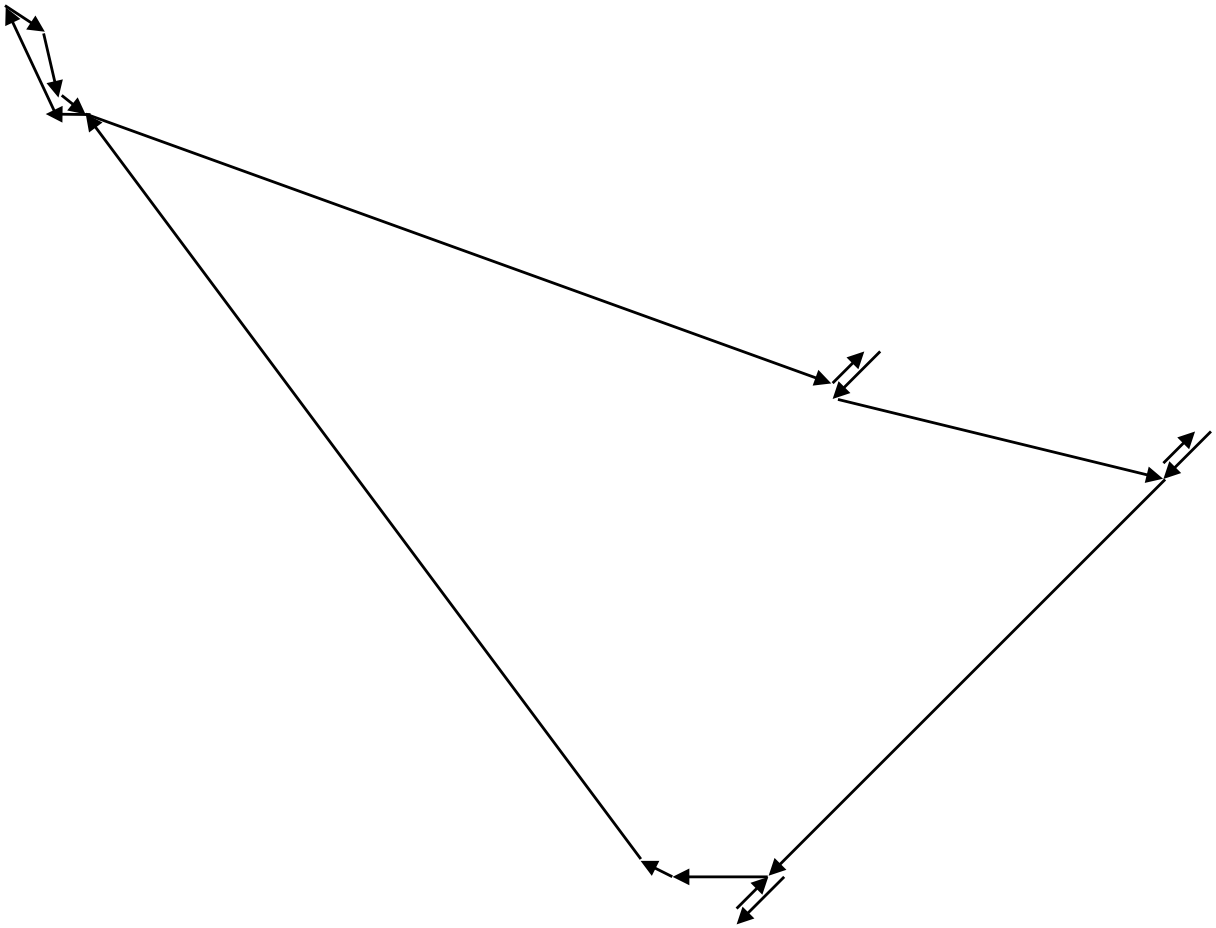
2 Out of home

4 trips,
2 tours

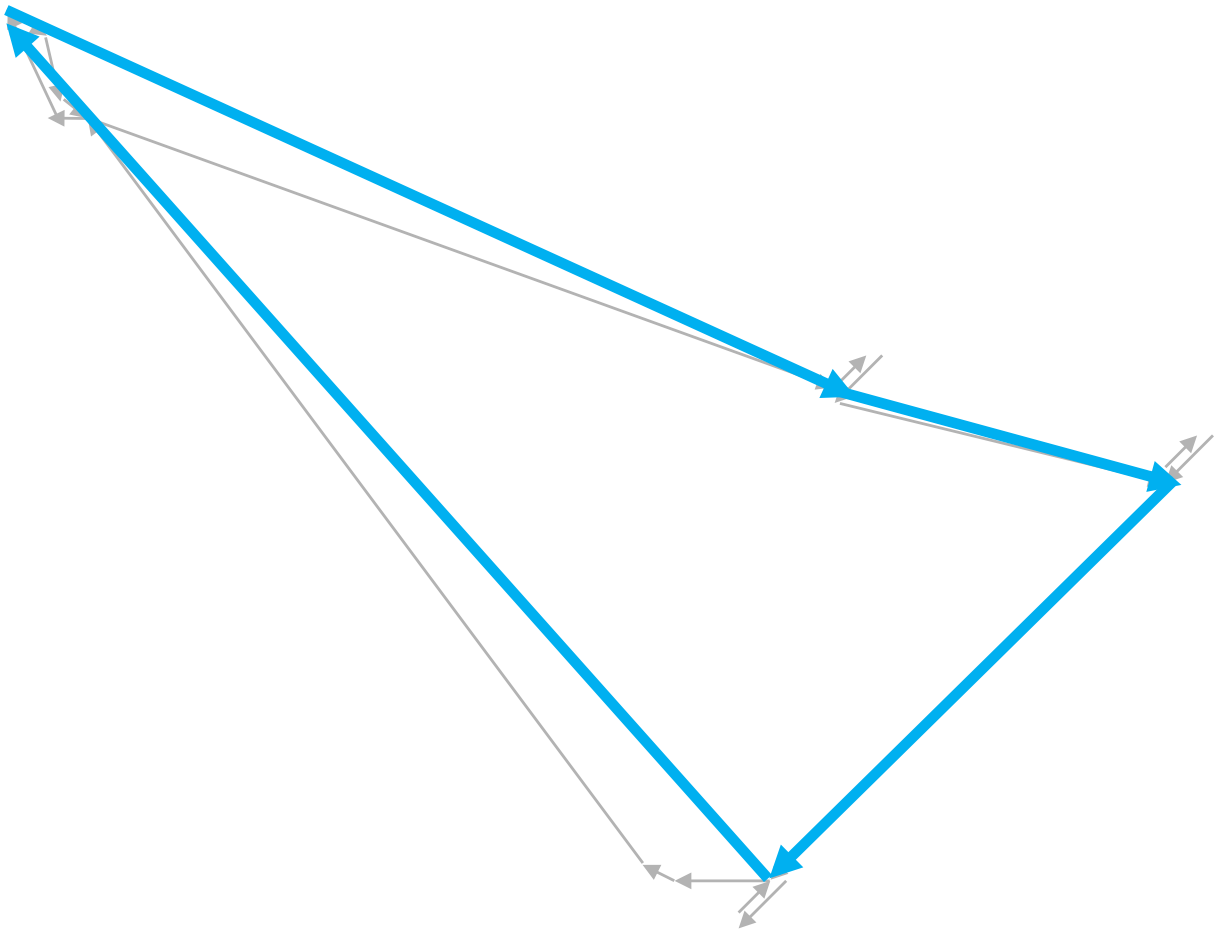


Long-distance data collection

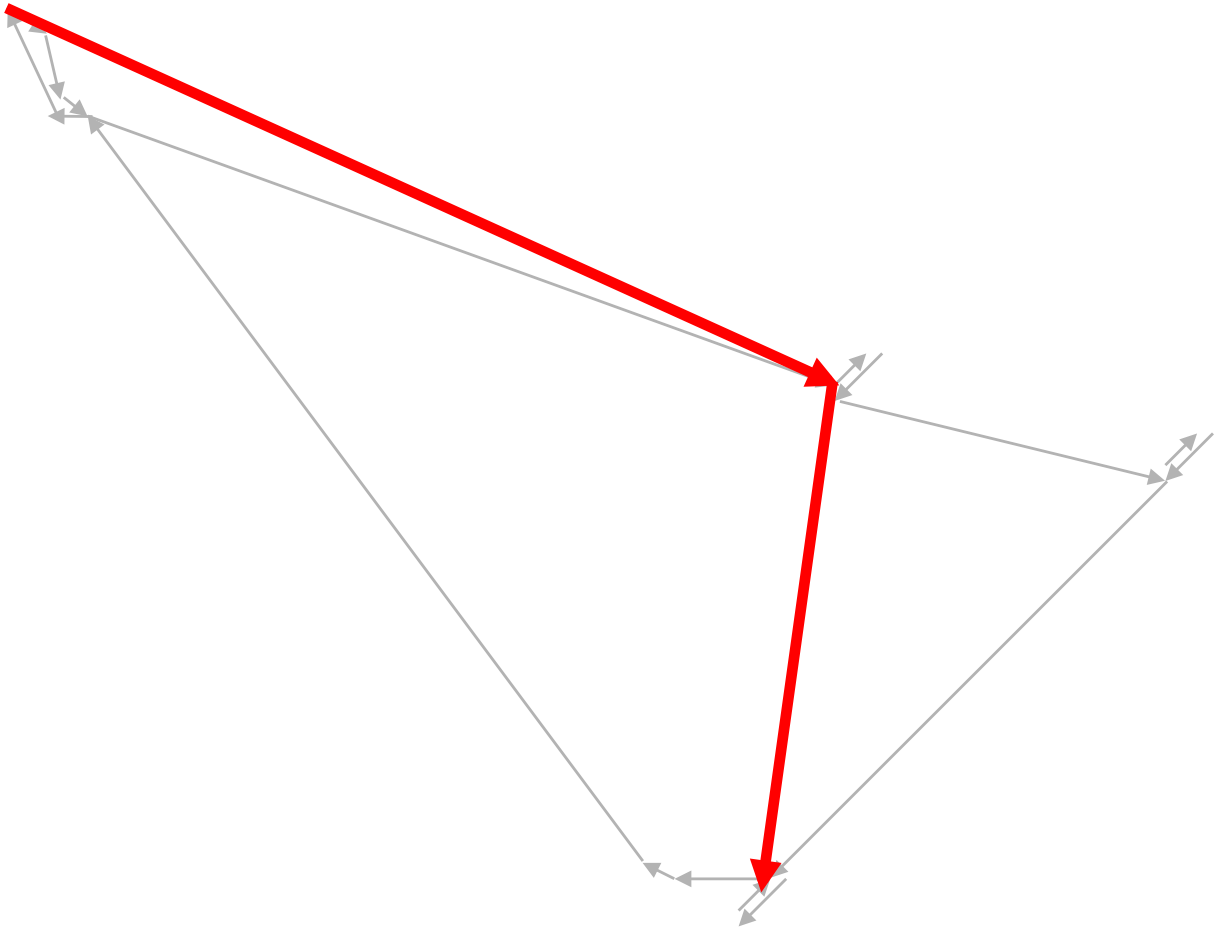
Aggregation: All stages



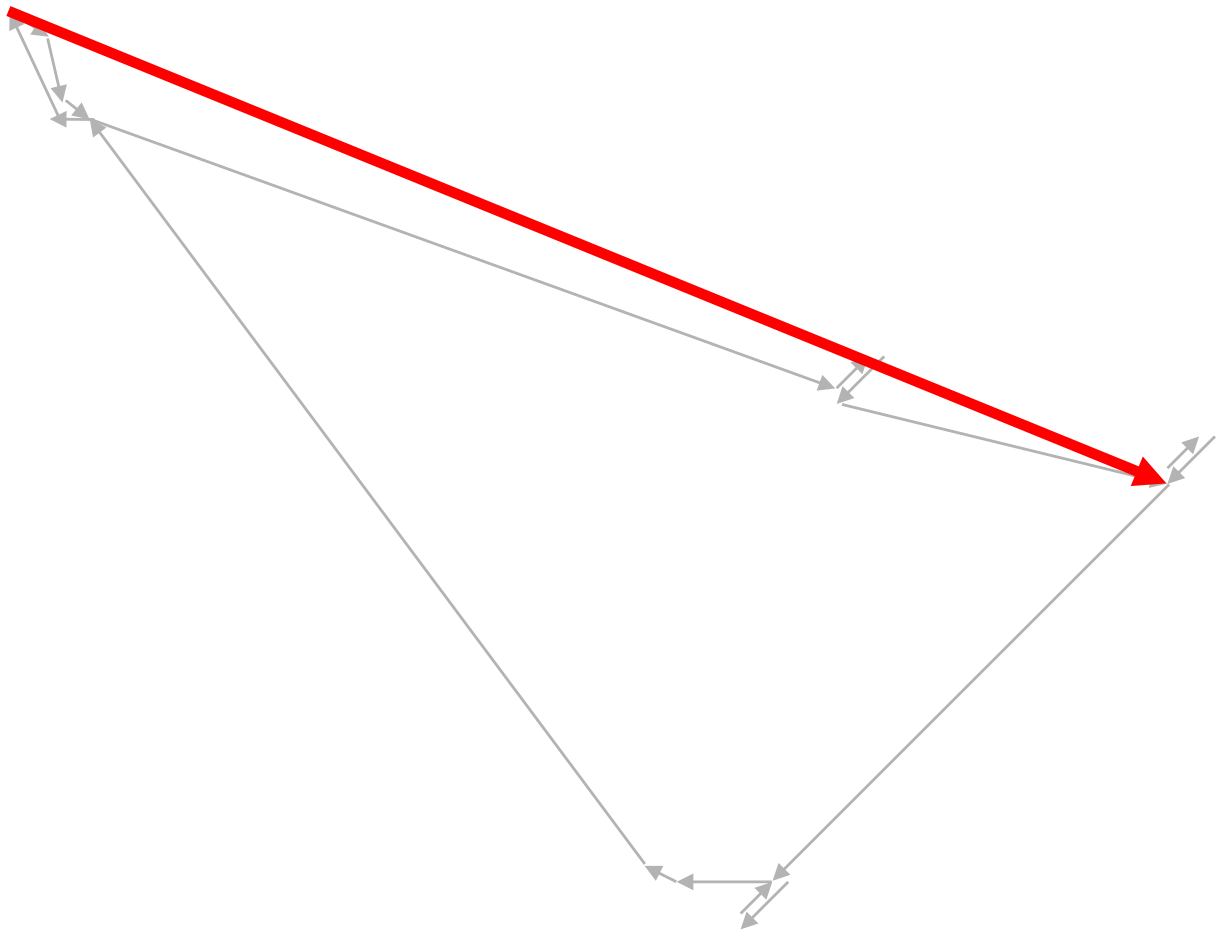
Aggregation: moves to locations 50+ km away



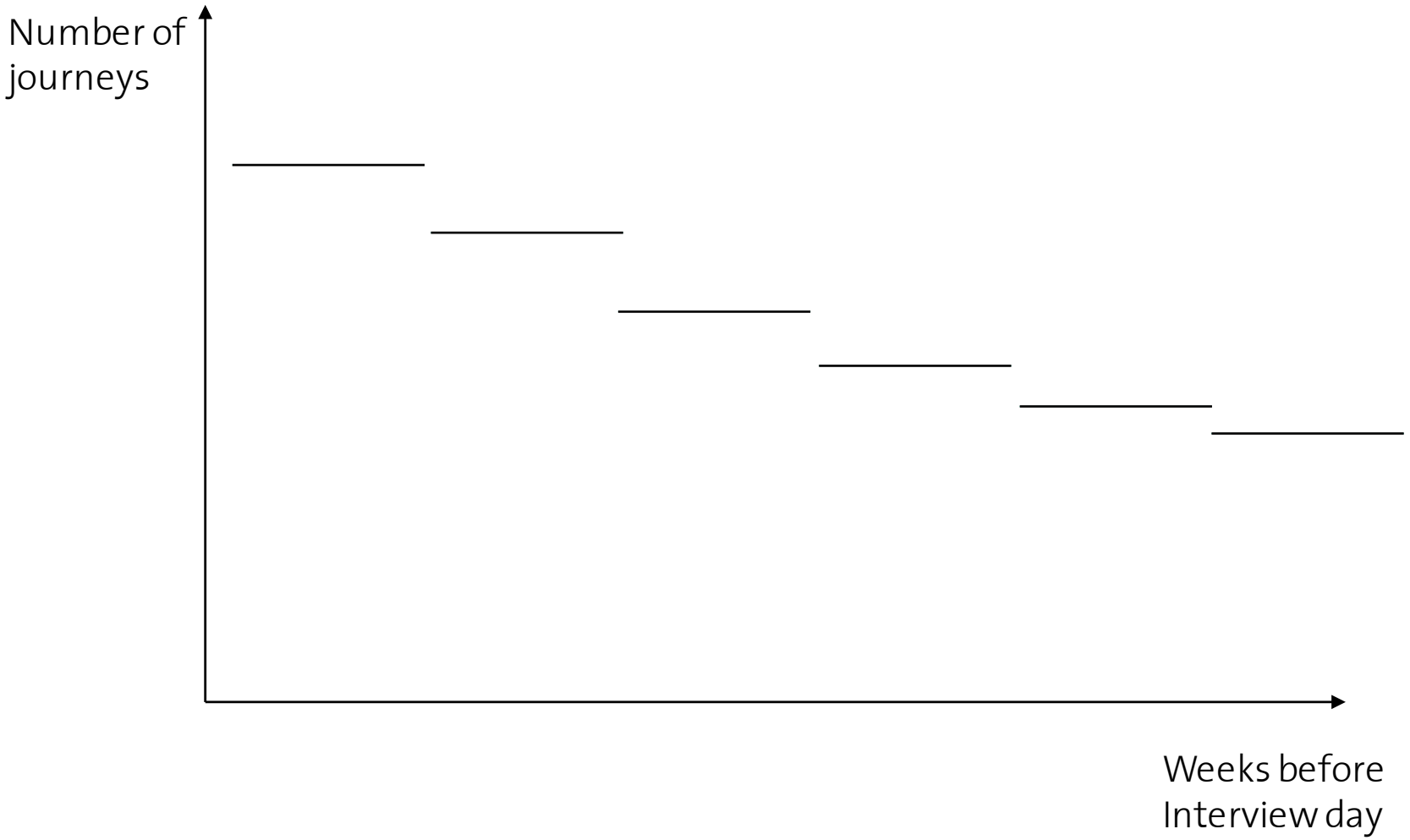
Aggregation: displacement to the furthest location of that day



Level of aggregation: displacement to the furthest location



Correction for forgetting



Capturing long distance traffic with GSM data

Capturing long distance traffic with GSM data

with:

Maxim Janzen (IVT, Zürich)

Maarten van Hoof (Orange Labs, Paris)

Data base

Some facts:

- reports all GSM actions (originating/terminating calls/SMS)
- in Orange network
- for each action a Call Data Record (CDR) appears in the data
- users are anonymised
- covers the time period: 16 May 2007 till 15 October 2007
- in total 22.3 million customers
- in total 15.5 billion CDRs

Approach

Approach:

1. select municipalities

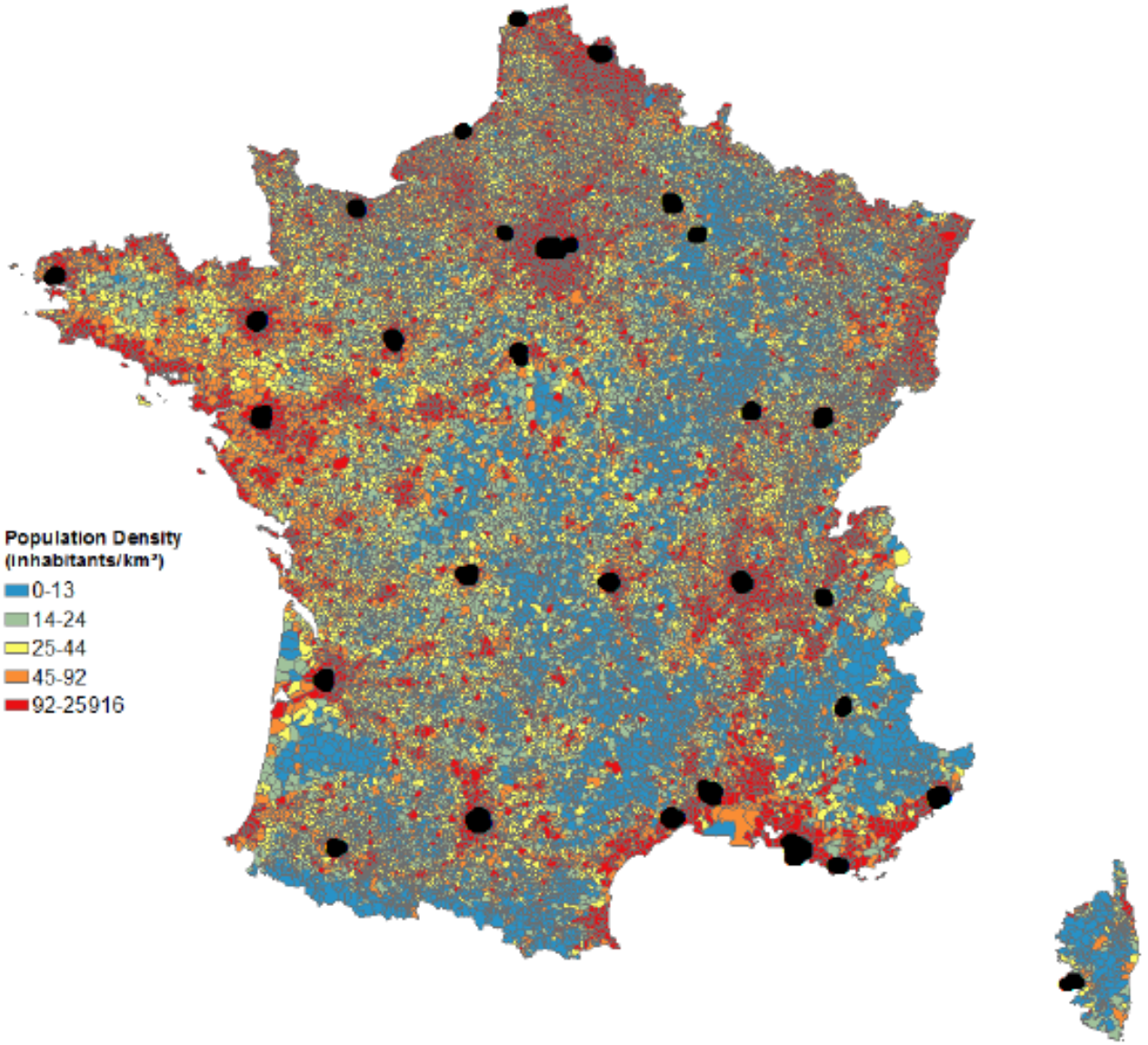
2. select mobile residents
   - identify home locations (Ahas, 2008)
   - identify mobile customers

3. extract data for selected customers

4. reconstruct long distance tours

5. store the tours

Selected locations

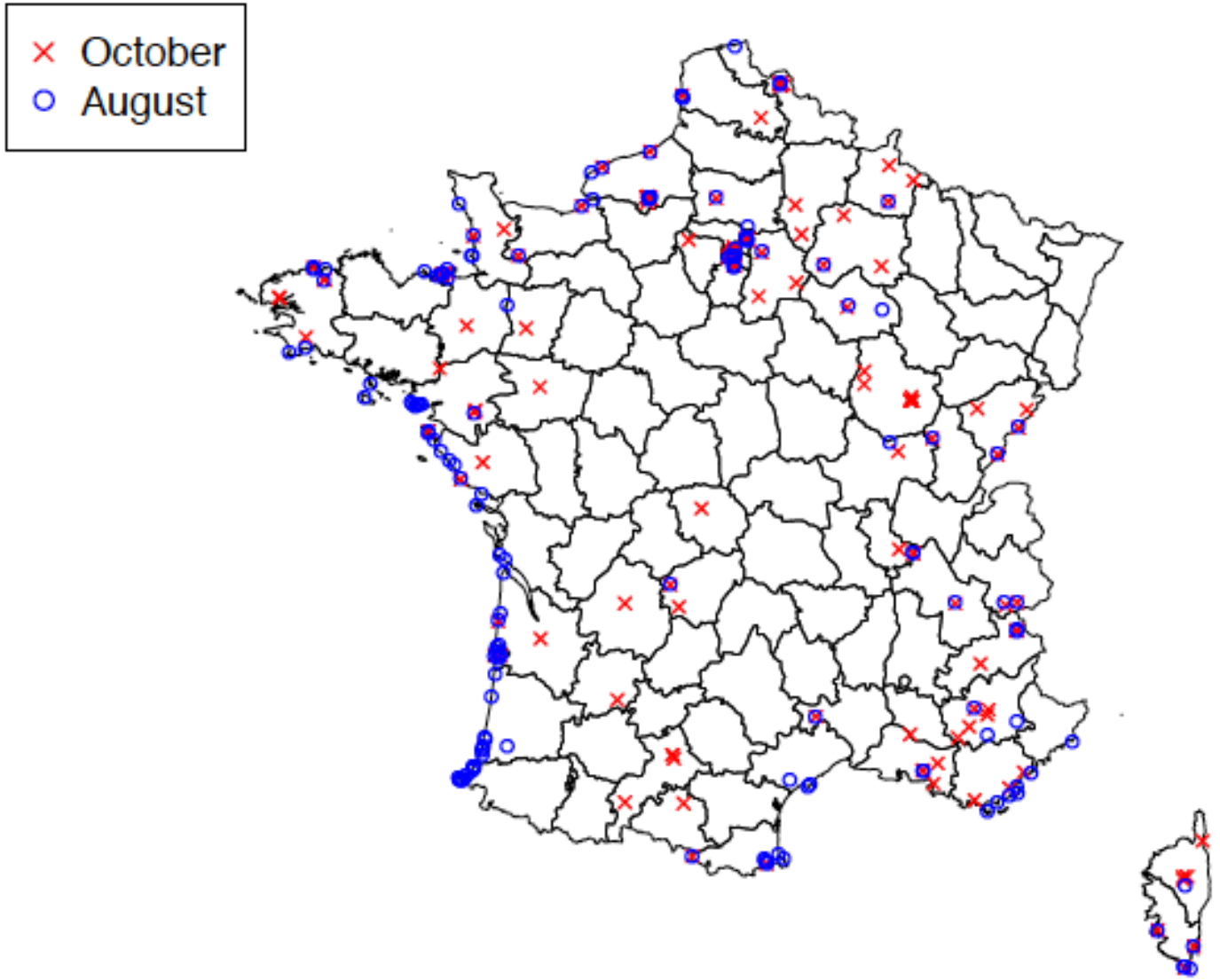


Sample frame

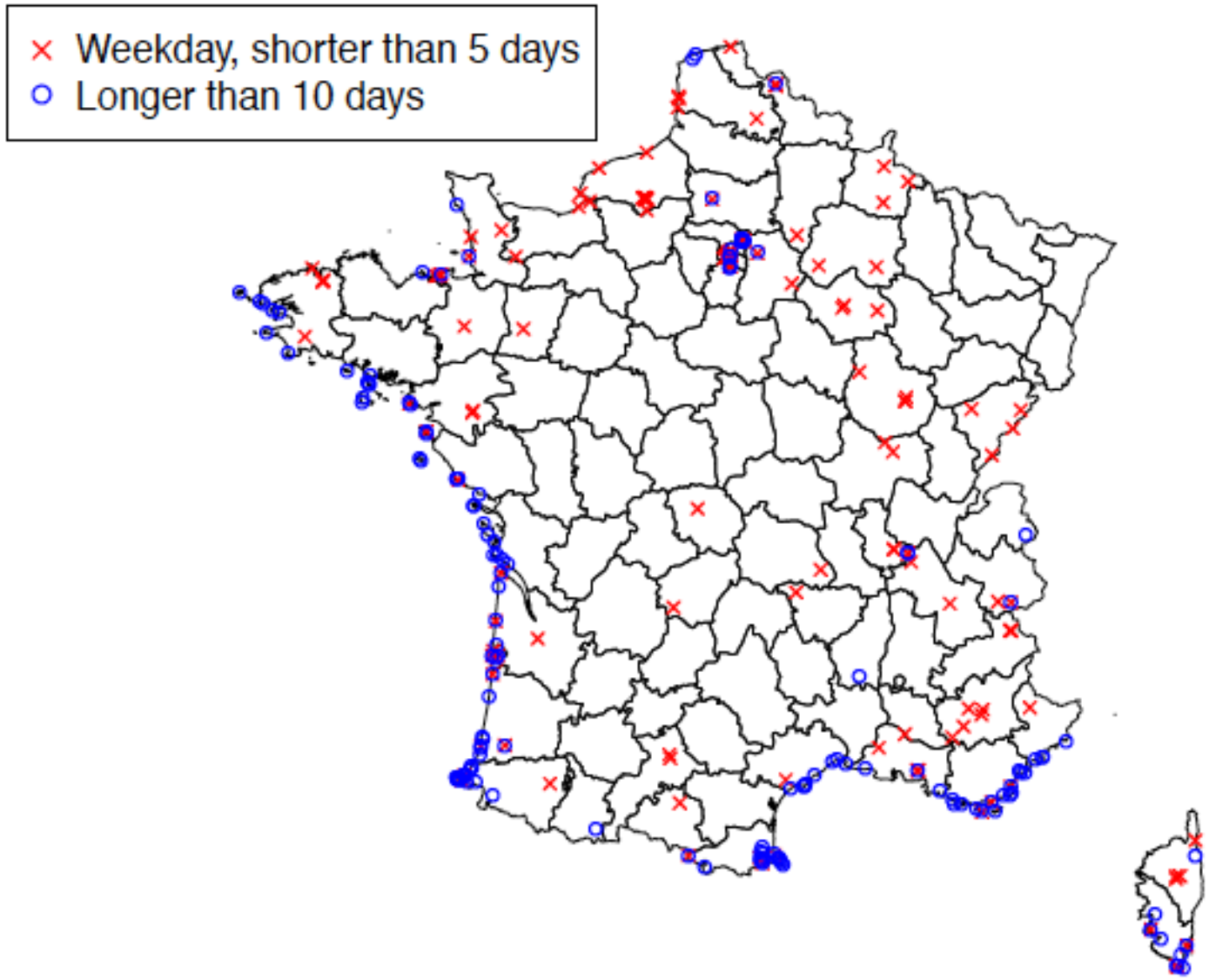
All customers	22'300'000
With at least one home anchor	18'000'000
Three home anchors within one of the 31 municipalities	1'360'000
Without m2m (non-human customer)	1'290'000
Mobile customer (went > 50km from home anchor in June)	793'900

For each municapility 2000, Paris 5000

Most visited towers by month



Most visited towers by type of tour



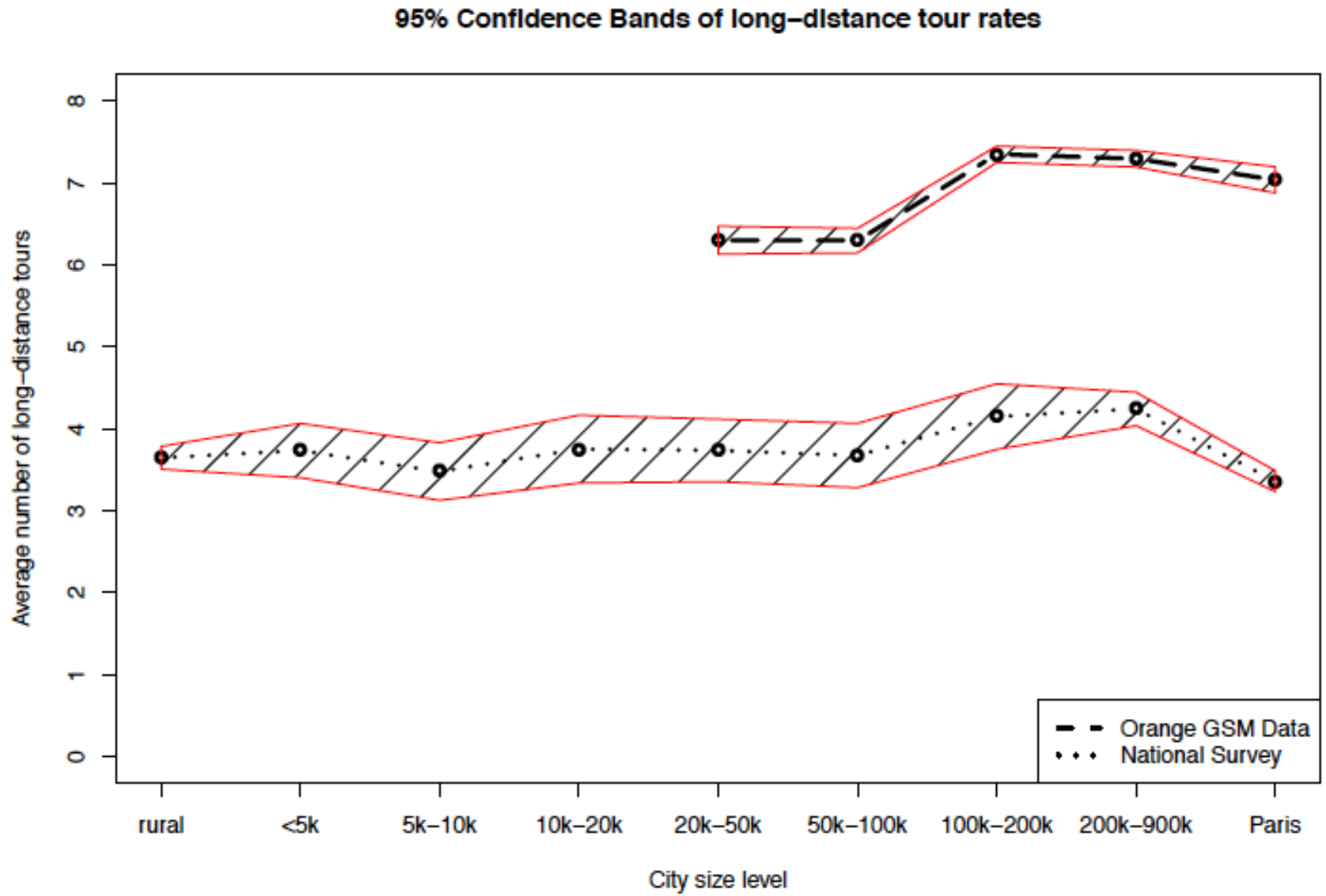
Comparison with ENTD

	ENTD		Orange	
Population	Resp.	Tours	Resp.	Tours
rural	1'503	2'156	0	0
Up to 5k	202	318	0	0
5k-10k	187	233	0	0
10k-20k	150	202	0	0
20k-50k	160	248	7'455	100'771
50k-100k	202	297	9'579	94'551
100k-200k	204	329	23'343	272'412
200k-900k	965	1'554	17'486	202'842
Paris	1'223	1'641	4'951	42'660

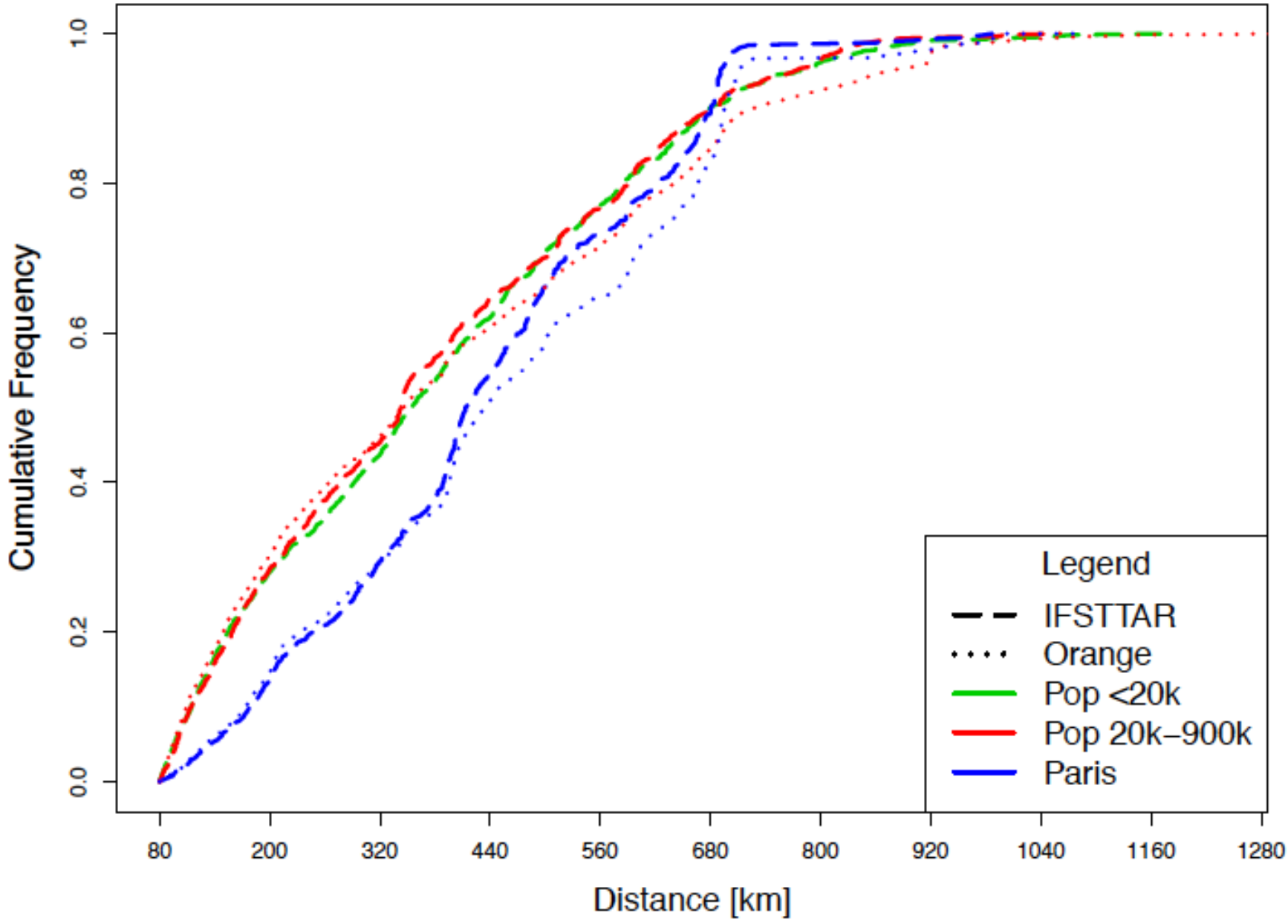
Comparison with ENTD: Number of journeys in the period

Population	ENTD		Orange	
	Winter	Summer	Summer	Factor
rural	4.25	3.65		
Up to 5k	4.36	3.73		
5k-10k	4.06	3.48		
10k-20k	4.38	3.75		
20k-50k	4.36	3.74	6.30	1.68
50k-100k	4.29	3.67	6.30	1.72
100k-200k	4.84	4.15	7.35	1.77
200k-900k	4.95	4.24	7.30	1.72
Paris	3.92	3.36	7.04	2.10

Comparison with ENTD: Number of journeys in the period

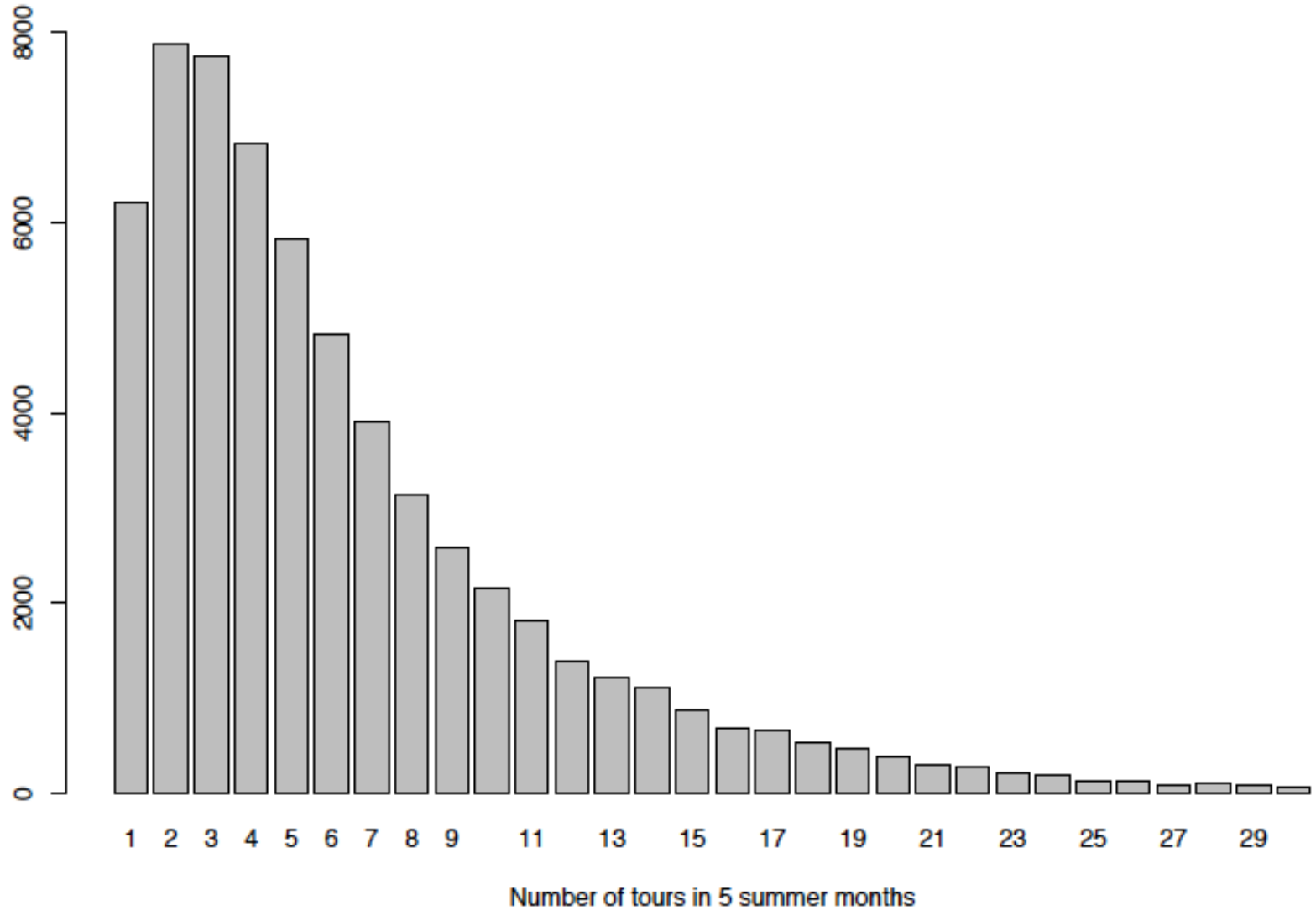


Comparison with ENTD: Tour distance distribution



DI

Tour frequency distribution during the observed months



Potential errors and biases in the Orange data

- As mentioned before: Frequency of GSM data points.
- Selection of customers might be biased (frequent callers are more likely to be chosen)
- Computation of home (and work) anchors.
- International tours (we assumed that there are none in the data).
- In the comparison: Error in weighting the ENTD data

What should we do ?

What do we know ?

Variable	Diary	GPS (logger or mobile) (no prompted recall)	Mobile
Participation	Self-selected	Self-selected	(Random)
Duration	1 day (- 6 weeks)	1 day (- 6 weeks)	1 day (Unlimited)
Stage	Yes, underreported	(Yes)	No
Trip	Yes, underreported	Yes	(Yes)
Journey	Yes	Yes	(Yes)
Time	Rounded	Exact	Imputed
Location	Rounded	Exact	Imputed
Mode	Yes	Imputed	Imputed
Purpose	Yes	Imputed	Imputed
Group	Yes	No	No
Expenditure	Yes	**No**	**No**

What do we know ?

Variable	Diary	GPS (logger or mobile) (no prompted recall)	Mobile
$/reported day	High	High-medium	Low
Data availability	Months	Week	Daily
Corrections	Needed	No	No
Imputations	Needed	Needed	Needed
Choice models	Yes	Yes	Difficult
Socio-demographics	Yes	Yes	Imputed

Next steps

- Query what we really need for
  - Cost-benefit analysis
  - Planning of prices and services
  - Planning for the slow modes
  - Social accounting

- High-quality multi-modal surveys to establish the measurement errors (add bluetooth and wifi senders, noise profile)
- Error correction models
- Cross check against third party sources

- Treat survey data as indicators in a measurement model
- Treat traces as indicators in a measurement model

, but especially for surveys

- Remember, that observation/surveys are ‘talk’ and
- Treat respondents as partners in a talk, discussion:

  - Frame your request in a way which addresses them in a clearly defined social role (citizen, driver, customer, etc.)
  - Match your role and the questions
  - Account for their constraints (readability of text, full guidance through the forms, require no calculations – unless necessary, speak their ‘language’)
  - Be as complex, as the topic warrants, requires, but not more so
  - Don’t surprise them with unannounced requests
  - Don’t ask them to do work you can do for them
  - If appropriate, provide an incentive, acknowledgement

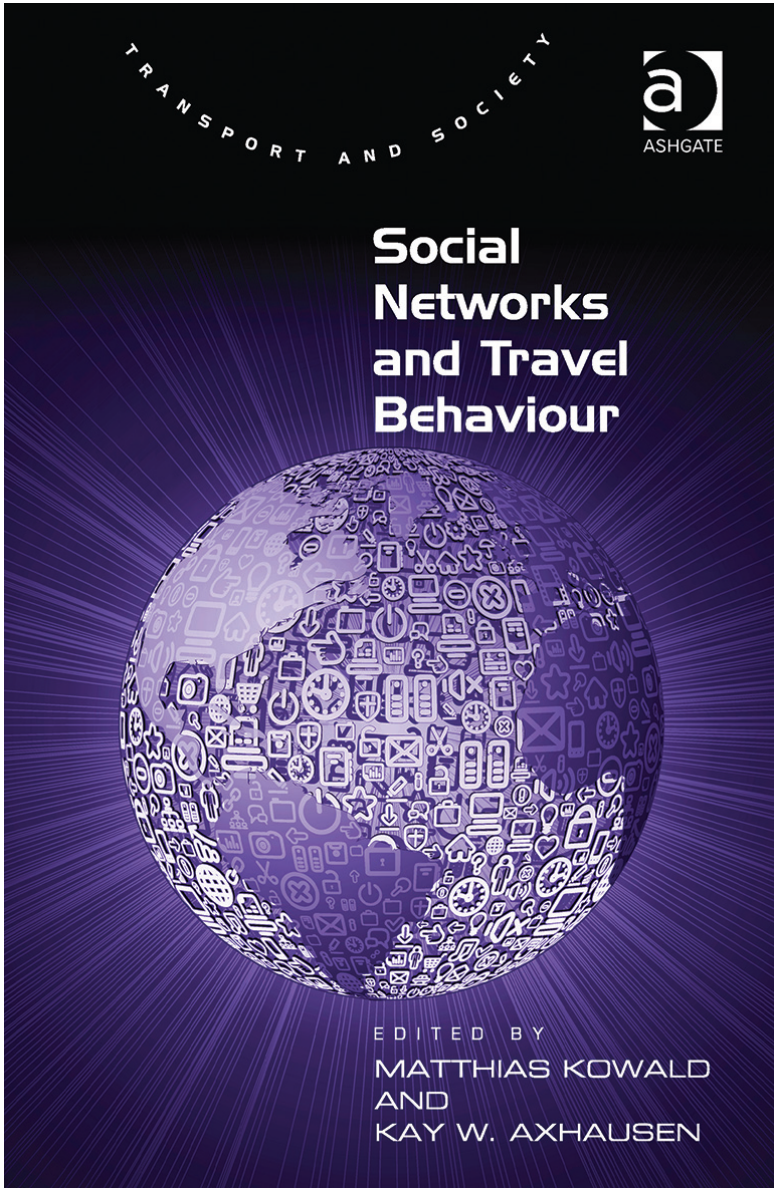
Questions ?

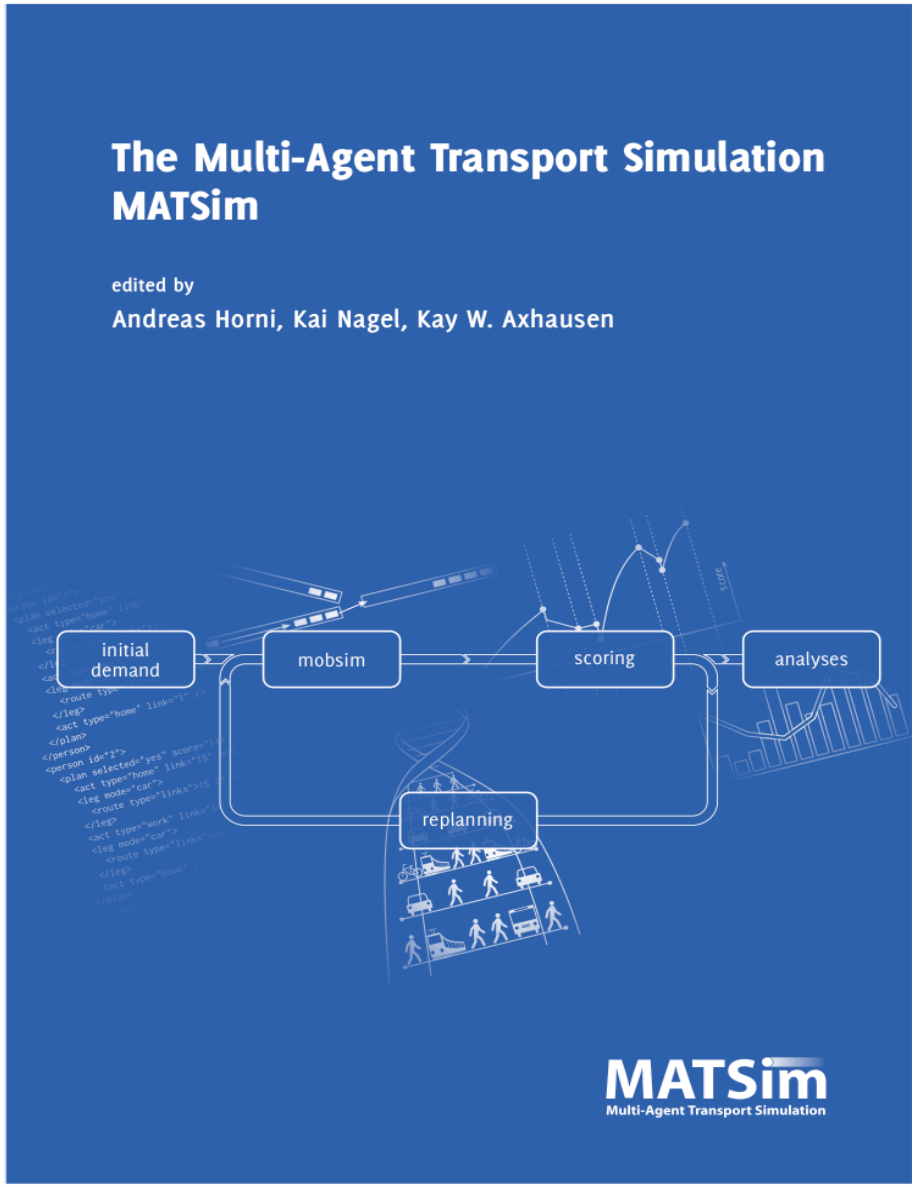
www.matsim.org

www.ivt.ethz.ch
www.futurecities.ethz.ch

www.senozon.com

Questions ?


TRANSPORT AND SOCIETY
ASHGATE
Social Networks and Travel Behaviour
EDITED BY
MATTHIAS KOWALD
AND
KAY W. AXHAUSEN


The Multi-Agent Transport Simulation MATSim
edited by
Andreas Horni, Kai Nagel, Kay W. Axhausen
initial demand
mobsim
scoring
analyses
replanning
MATSim
Multi-Agent Transport Simulation

References

Chikaraishi, M., A. Fujiwara, F. Nishikawa, H. Seya and J. Zhang (2015) Modeling shopping behavior in a neighborhood with endogenous representation of retail attractiveness, presentation at the The 4th International BinN research seminar "Dynamic modelling based on bounded rationality to understand diversifying travel behavior", University of Tokyo, June 2015.

De Palma, A., N. Picard and P. Waddell (2007) Discrete choice models with capacity constraints: An empirical analysis of the housing market of the greater Paris region, Journal of Urban Economics, 62 (2) 204-230.

Manski, C.F. (1993) Identification of Endogenous Social Effects: The Reflection Problem. The Review of Economic Studies 60, 531-542.

Rust, J. (1987) Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher, Econometrica 55, 999-1033.

Schmutz, Simon (2015) Auswirkung von analytische Einheiten und Aggregationsregeln auf die Verkehrsmittelwahlmodellierung, MSc thesis, Zürich, January 2015.

Schuessler, N. (2010) Accounting for similarities between alternatives in discrete choice models based on high-resolution observations of transport behaviour, ETH Zürich, Zürich.

Vrtic, M. (2003) Simultanes Routen- und Verkehrsmittelwahlmodell, PhD Dissertation, Fakultät für Verkehrswissenschaften, TU Dresden, Dresden.

Walker, J.L., Ehlers, E., Banerjee, I., Dugundji, E.R. (2011) Correcting for endogeneity in behavioral choice models with social influence variables. *Transportation Research Part A: Policy and Practice* 45, 362-374.

Zilliak, S. and D. McCloskey (2008) *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*, University of Michigan Press, Ann Arbor.

Jäggi, B. (Forthcoming) Decision modeling on the household level for energy, fleet choice and expenditure, , Dissertation, ETH Zürich, Zürich.