ETH zürich

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Mind the gap!
How tragic are interruptions in GPS traces for time use and mobility studies?

Winkler, Mesaric, & Axhausen

STRC 2021, Monte Verità, Ascona

Introduction

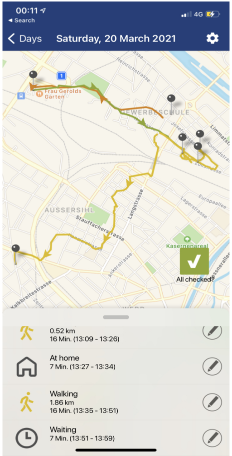
- GPS tracking studies have become an invaluable tool for mobility studies, and are making their way into time use research.
- The MOBIS and MOBIS:COVID-19 studies employed the Catch-My-Day (CMD) app by MOTIONTAG that passively tracks participants and returns a diary of the locations and the modes he or she used to travel there.
- Though these diaries should be a 24-hour account of a person's activity patterns, the data are often incomplete when aggregated at a daily level.
- **These gaps in participants' timelines interfere with the estimation of choice models of activities and their duration that rely on complete days.**

Research Questions

- What proportion of a person's 24-hour day is accounted for in the data? Do the remaining gaps reveal any patterns in terms of:
  - Duration
  - Distance
  - Start/end time of day
  - Day of the week
  - Mode or trip purpose before/after a gap
- Do smartphone operating systems (OS; Android vs. iOS) affect the gaps?
- What are potential reasons gaps occur?
- Should these gaps be filled? If yes, which methods are reasonable to apply?

Method

Data collection



- 4,173 individuals participated in MOBIS and MOBIS:COVID-19 between September 2019 and March 2021.
- Participants used CMD to track between one and 546 days total, with a mean of 129 days (SD = 115, median = 81).
- The total number of gaps amounts to 650,838. Taking all other recorded events into consideration, this is 1 in 10 events (and 14.31% of the total duration).

Method
Data preparation

1. Aggregate respondents' daily events (tracks and stays) on a daily level

2. Insert gaps as placeholders to represent unaccounted portions of a person's day

3. Disaggregate the events and merge gaps that span over two days

Descriptive statistics for the full set of gaps

N = 650,838 gaps

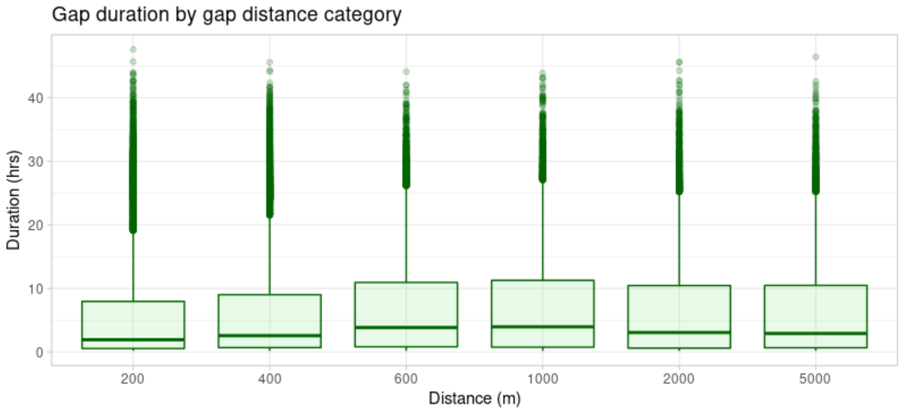
	distance (km)	**duration (hrs)**
minimum	0.00	0.00
median (IQR)	0.02 (0.00, 0.23)	0.07 (0.02, 1.48)
mean (sd)	17.85 ± 319.44	2.71 ± 5.96
maximum	16,098.21	47.62

$\rightarrow$ **Median near zero for both dimensions suggests that at least half of the gaps are short**

$\rightarrow$ **Indeed, 65% of gaps are within 100m and 89% are within one kilometer**

A closer look at gaps between 100m and 5km (>10 min duration)

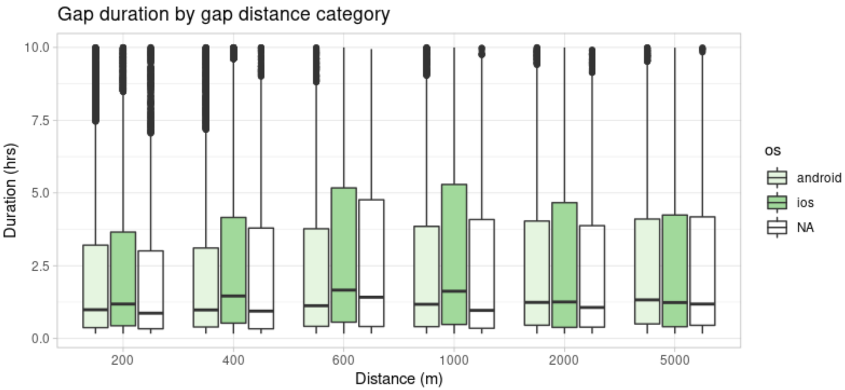
N = 142,181 gaps (22% of total gaps)



→ **No increase in duration with increasing distance!**

How does this look for gaps shorter than 10 hours?

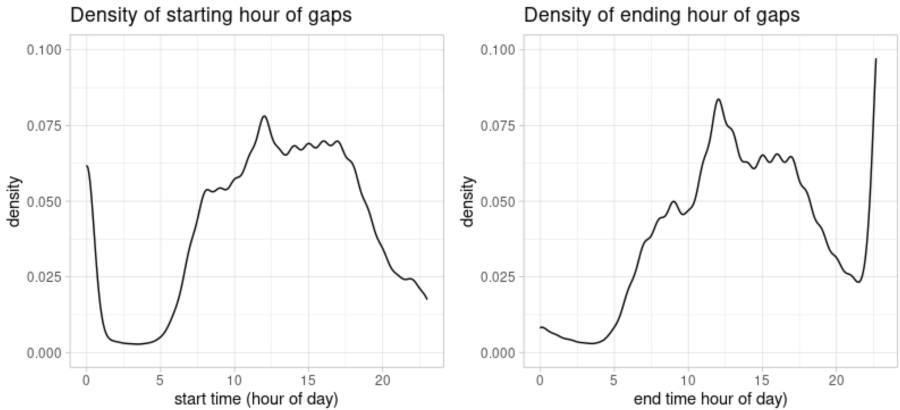
N = 108,194 gaps (17% of total gaps)



→ **Median duration consistently below 1.3 hours**

→ **Gaps often longer for iOS devices**

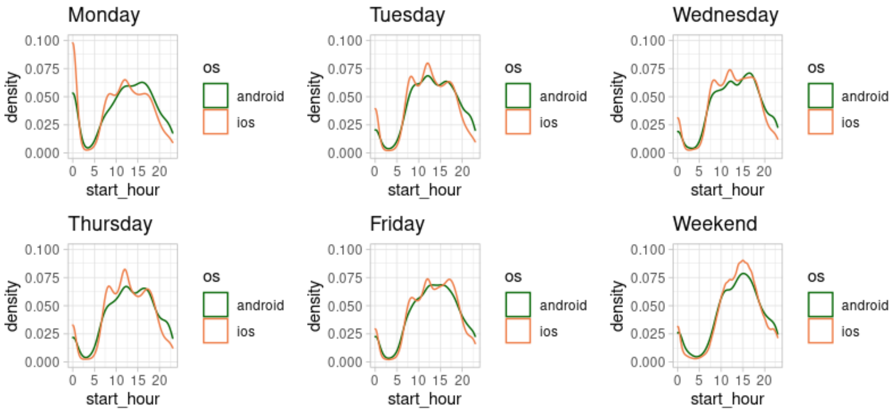
When do gaps start and end throughout the day?



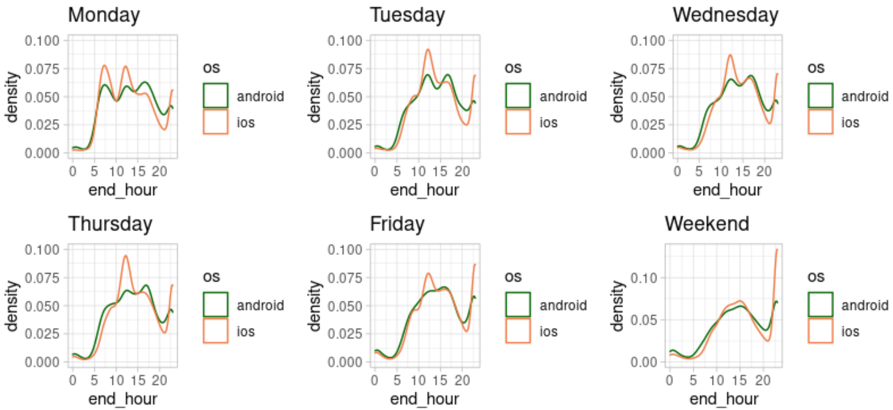
→ **??? Hard to explain!**

Do gaps begin at the same time on every day of the week?

N = 102,310 gaps (5,884 gaps excluded b/c unknown OS)



Do gaps *end* at the same time every day?



Takeaways from day of the week comparisons

- Android and iOS start and end times *do* overlap for the most part, but iOS devices seem to have a much more distinctive pattern.
- The iOS patterns suggest that there are certain times of day when iOS devices are not recording tracks (legs) or stays.
- IOS devices seem to force a gap (reset or similar) at midnight from Sunday to Monday, as well as exactly at midday daily.
- Tuesday through Friday show similar patterns, and these strongly differ from gap start and end times on weekends.
  → **They also look oddly similar to everyday travel patterns**

Summary of Findings

Presented findings suggest that gaps are not tragic, and potentially random:

- The majority of gaps are truly a missing chunk of time at or around the same location as the events that precede and follow the gap in time (within 100m). These can easily be assigned to the prior event, for instance, but this decision becomes trickier when it comes to larger distances.
- Duration of gaps does not increase with increases in distance. That gaps at one or even five kilometers show similar duration distributions to all gaps that are shorter in distance supports the idea of the gaps occurring randomly.
- When looking at the start and end time of day of gaps per day of the week, it is also interesting that the incidence of gaps loosely follows patterns of daily out-of-home activity participation.

Why do gaps likely arise?

- Gaps may occur because CMD goes into "sleep" mode to save battery when the phone remains idle for a long period of time.
- Alternatively, GPS signal may be lost when:
  - a participant passes through a tunnel or is in a dense area with many buildings
  - a phone loses battery
  - a user disables tracking/turns his or her phone off
  - a user connects to a VPN (cause for cases with extremely long duration and no change in distance)

Should we fill the gaps?

- It is likely reasonable to assign very short gaps to the event that precedes them. Also, these may vanish as the algorithms behind CMD improve.
- For the 20% of gaps whose median duration is around 1.5 hours: those with stay events at two different locations before and after the gap can have some time assigned to travel (time can be routed).
- Very long gaps are more difficult to handle. They can be dealt with by clustering similar respondents' activity chains and use them to predict a person's gap $\rightarrow$ **Only works if enough data are available; still adds error/complexity to the data set.**

Conclusion

- At the end of the day, data quality is high because of the complex machine learning algorithms applied by CMD. These are powerful, but never 100% accurate.
  $\rightarrow$ **Arguably more accurate than traditional travel diary collection modes and provides minimal burden for participants.**

- Mobility studies will benefit from filling the small gaps, but are generally not critically affected, as most travel events are captured in the (then) over 95% of a person's day that will be accounted for.

- For time use studies, gaps can be shortened/reduced and trip purpose assignment improved if an active diary component is added to the CMD app. As it stands, only complete days should be used for discrete choice behavior modeling, and more flexible models need to be used/developed to handle the minimal amount of time use data still missing from respondents' days.

ETH zürich

Thank you for your attention!
Questions?

STRC | 21th Swiss Transport Research Conference
Monte Verità / Ascona, September 12 – 14, 2021

Literature

- Molloy, J., A. Castro Fernández, T. Götschi, B. Schoeman, C. Tchervenkov, U. Tomic, B. Hintermann and K. W. Axhausen (2020) A national-scale mobility pricing experiment using GPS tracking and online surveys in Switzerland: Response rates and survey method results, *Arbeitsberichte Verkehrs-und Raumplanung*,**1555**.

- Molloy, J., C. Tchervenkov, T. Schatzmann, B. Schoeman, B. Hintermann and K. W. Axhausen (2021) Mobis-covid19/53: Results as of 02/08/2021, *Arbeitsberichte Verkehrs-und Raumplanung*,**1654**.