Extracting Activity Schedules from the Continuous Mobility Panel Data

Anne-Valérie Preto

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne

3rd July 2026

TRANSP-OR

EPFL

Outline

Project Objectives

SBB CFF FFS

RQ1: Typology of mobility behavior

How can we identify interpretable mobility patterns from long-term tracking data?

Key challenge

- Travelers are highly heterogeneous
- Daily mobility is variable and noisy
- Stable patterns must be separated from observation-level fluctuations

Continuous Mobility Panel Data

Data collection:

- Managed by Intervista's Footprint Panel
- Purchased by SBB for mobility analysis
- Multimodal tracking across all transportation modes

Sample and tracking:

- 3–12+ months of continuous tracking per person
- Smartphone app with Wi-Fi, Bluetooth, geofencing
- Modes and purposes automatically assigned

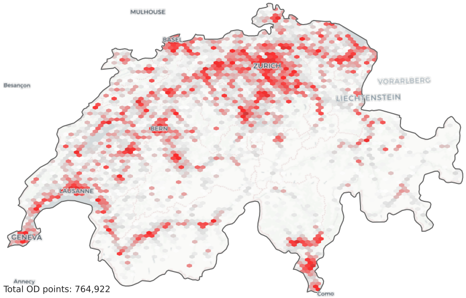


Figure: Switzerland OD density in the CMP.

Continuous Mobility Panel Data

intervista

Sample	Tracking	Respondents	Mobility
1,992 respondents	**96%** tracking completeness	**50 / 50** women / men	**28.4 km** median daily distance
Oct. 2024 focus month	**330k** trips in October	**47** median age	**11** median trips / day
Switzerland-wide panel	**658k** legs in October	**69%** employed	**66%** car available

Socio-demographic information comes from the survey.
Long-term mobility indicators are derived from passive GPS tracking.

Outline

Research Question 1: Methodological Exploration

Approaches tested and lessons learned

ML clustering with DTW	$\rightarrow$	Distance-sensitive; low robustness
Daily latent classes	$\rightarrow$	High variance; weak user identification
Hierarchical model with individual random effects	$\rightarrow$	Costly; prior/init-sensitive (STRC)
Tour-level latent classes	$\rightarrow$	Reliable; interpretable tour patterns
Hierarchical latent tour model	$\rightarrow$	Traveler types from tour distributions

Current position

- RQ1 outcome: a tour-based representation that supports interpretable traveler typologies

Rethinking the Modeling Unit

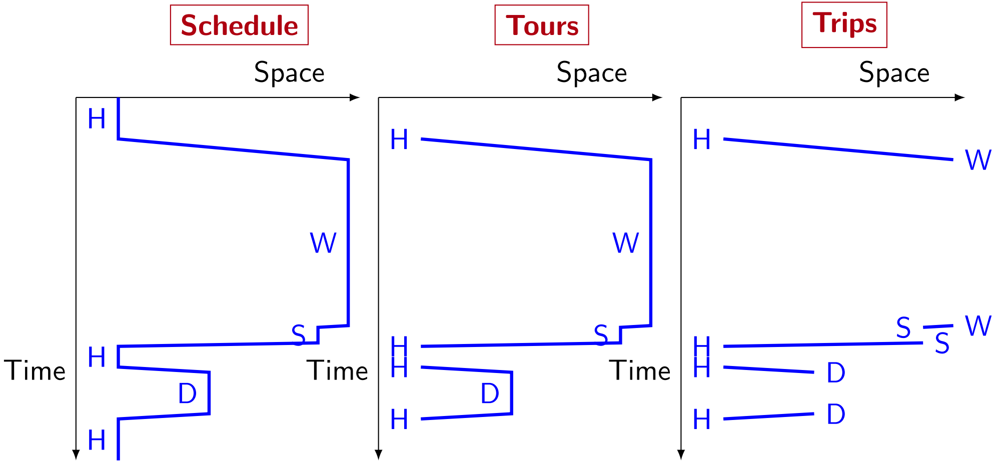
Traveler-centered approach

- Directly classify individuals
- Strong behavioral heterogeneity
- Large day-to-day variability
- Difficult to obtain stable personas

Tour-centered approach

- First characterize individual tours
- Tours are more homogeneous and interpretable
- Reduce within-person variability
- Traveler profiles emerge from tour distributions

Why Tour-Level Indicators?



H: Home, W: Work, S: Shop, D: Dining out [Source: M. Ben-Akiva]

From Days to Tours

Daily-level limitations

- Too granular for stable behavior identification
- High day-to-day variability (noise)
- Difficult to separate signal from random fluctuations

Tour structure

- Tours naturally aggregate activities and trips over a coherent sequence
- Reduces noise compared to daily-level indicators
- Better captures behavioral patterns (e.g., work-based subtours, leisure chains)

Latent tour model

Explanatory variables (class membership) x_{ri} → Latent tour class s_r ⇢ Tour indicators (measurement) y_r

ω_r ε_r

1. Explanatory variables

2. Latent class membership

3. Measurement model

$$V_{rs} = \alpha_s + \gamma_s x_r + \delta_s z_r$$

1. Explanatory variables

a. Socio-economic variables
- age (piecewise)
- gender (0: male)
- employment status (0: employed)

b. Time variables
- day of week (0: weekend)
- time of day (0: night)

c. Weather variables
- precipitation (0: no rain)
- temperature (0: mild)
- wind speed (0: low wind)

$$P(s_r = s \mid x_r, z_r) = \frac{\exp(V_{rs})}{\sum_{k=1}^{S} \exp(V_{rk})}$$

2. Latent class membership

We estimate the distribution parameters for each class

Parameters estimated from 3.

1. μ_{js}: class-specific means
2. σ^2_{js}: class-specific variances
3. p_{js}: class-specific probabilities

$$f(y_r \mid s) = \prod_{j=1}^{J_c} \mathcal{N}(y_{rj} \mid \mu_{js}, \sigma^2_{js}) \prod_{j=1}^{J_b} \mathrm{B}(y_{rj} \mid p_{js})$$

3. Measurement indicators

a. Spatio-temporal dispersion
- tour duration (log)
- max dist home (log)
- max act duration (log)

c. Modal composition
- multimodal tour (binary)
- car share
- public transport share

b. Tour complexity
- total distance (log)
- has subtour (binary)
- number of legs (log)
- numb distinct POIs (log)

d. Activity composition
- work share
- shopping share
- leisure share

Overall likelihood (all tours)

Observation-level likelihood: $L_r = \sum_{s=1}^{S} P(s_r = s \mid x_r, z_r) f(y_r \mid s)$

Model likelihood: $L = \prod_{r=1}^{R} L_r$

Log-likelihood: $\ell = \sum_{r=1}^{R} \log(L_r)$

Latent tour model: Results

Table: Measurement model estimates.

Parameter	C0	C1	C2	C3
Std.: Log Total Distance	0.944*** (0.0538)	0.705*** (0.0225)	0.873*** (0.038)	0.968*** (0.0376)
Mean: Log Total Distance	-0.0147 (0.0819)	-0.611*** (0.0396)	0.622*** (0.0759)	-0.228*** (0.0526)
Std.: Log Tour Duration	1.04*** (0.0325)	0.555*** (0.01)	0.813*** (0.0138)	0.86*** (0.013)
Mean: Log Tour Duration	0.314*** (0.0459)	-0.847*** (0.0329)	0.645*** (0.0264)	-0.29*** (0.0298)
Std.: Log Max. Distance From Home	0.936*** (0.0537)	0.643*** (0.0264)	0.905*** (0.0388)	0.93*** (0.0342)
Mean: Log Max. Distance From Home	0.0168 (0.0851)	-0.701*** (0.0639)	0.638*** (0.0703)	-0.209*** (0.0557)
Std.: Log Max. Activity Duration	0.748*** (0.0144)	0.78*** (0.0222)	0.869*** (0.0176)	0.985*** (0.0188)
Mean: Log Max. Activity Duration	0.599*** (0.02)	-0.925*** (0.0638)	0.209*** (0.0223)	0.007 (0.0219)
Std.: Log Number Of Legs	0.662*** (0.0155)	0.585*** (0.0252)	0.876*** (0.0196)	0.644*** (0.0106)
Mean: Log Number Of Legs	-0.274*** (0.0418)	-0.737*** (0.0414)	1.02*** (0.0357)	-0.391*** (0.0276)
Std.: Log Distinct POIs	0.766*** (0.0186)	0.718*** (0.0243)	0.906*** (0.0138)	0.789*** (0.0145)
Mean: Log Distinct POIs	-0.109*** (0.0344)	-0.663*** (0.0408)	0.84*** (0.026)	-0.352*** (0.0222)
Std.: Car Distance Share	0.991*** (0.102)	1.04*** (0.131)	0.899*** (0.0416)	1.02*** (0.0782)
Mean: Car Distance Share	0.282*** (0.0759)	-0.119*** (0.0434)	-0.207*** (0.0388)	0.11*** (0.0362)
Mean: PT Distance Share	-0.451*** (0.00866)	-0.453*** (0.0142)	1.02*** (0.0567)	-0.451*** (0.00711)

Notes: Standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Class C0 is the reference class for the structural equations.

Latent tour model: Results

Table: Measurement model estimates.

Parameter	C0	C1	C2	C3
Std.: Work Share	0.101*** (0.00265)		0.981*** (0.0524)	
Mean: Work Share	1.65*** (0.0143)	-0.65*** (0.0172)	0.162** (0.0687)	-0.649*** (0.0213)
Mean: Shopping Share	-0.448*** (0.0105)	0.836*** (0.217)	0.248*** (0.0779)	-0.458*** (0.0156)
Mean: Leisure Share	-0.942*** (0.0135)	-0.97*** (0.0101)	-0.135** (0.0537)	1.15*** (0.0209)
Binary: Multimodal Tour	-0.586*** (0.0815)	-1.27*** (0.101)	2.19*** (0.0818)	-0.755*** (0.0618)
Binary: Has Subtour	-1.01*** (0.102)	-2.26*** (0.163)	1.64*** (0.0938)	-1.39*** (0.0792)
Std.: Shopping Share		1.47* (0.814)	1.07*** (0.0409)	
Std.: PT Distance Share			1.33*** (0.054)	
Std.: Leisure Share			0.851*** (0.0402)	

Notes: Standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Class C0 is the reference class for the structural equations.

Latent tour model: Results

Table: Explanatory variable estimates.

Parameter	C0	C1	C2	C3
Const	—	-1.05*** (0.148)	0.659*** (0.123)	-0.0406 (0.125)
Age35 65	—	0.689*** (0.091)	0.0504 (0.0728)	0.592*** (0.0763)
Age65 Plus	—	2.05*** (0.147)	1.02*** (0.138)	1.61*** (0.136)
Female	—	-0.136* (0.0721)	0.067 (0.0628)	-0.069 (0.064)
Work	—	-0.268*** (0.0846)	-0.312*** (0.0745)	-0.29*** (0.0749)
Monday	—	-0.49*** (0.122)	-0.393*** (0.113)	-0.932*** (0.11)
Tuesday	—	-0.441*** (0.121)	-0.149 (0.111)	-0.8*** (0.108)
Wednesday	—	-0.506*** (0.121)	-0.335*** (0.111)	-0.977*** (0.108)
Thursday	—	-0.381*** (0.125)	-0.0805 (0.113)	-0.771*** (0.112)
Friday	—	-0.367*** (0.12)	-0.193* (0.111)	-0.826*** (0.108)
Morning	—	1.11*** (0.102)	0.232*** (0.0849)	1.07*** (0.0893)
Midday	—	1.33*** (0.129)	0.106 (0.121)	1.21*** (0.118)
Afternoon	—	1.5*** (0.116)	-0.00645 (0.108)	1.76*** (0.103)
Evening	—	1.51*** (0.224)	-1.04*** (0.272)	1.57*** (0.199)
Weather Rain	—	0.191** (0.0908)	0.11 (0.0807)	0.0956 (0.0823)
Weather Heavy Rain	—	-0.0491 (0.15)	-0.0866 (0.131)	-0.103 (0.137)
Weather Cold	—	0.018 (0.0722)	0.127** (0.0631)	0.0209 (0.0643)
Weather Windy	—	-0.0955 (0.248)	0.164 (0.219)	0.0764 (0.223)

Notes: Standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Class C0 is the reference class.

Tour Class Details

Class 0 – **Work commuting tours**

- Moderate travel distance
- Moderately long duration
- Car-oriented, few stops

Typical tour
Commute to work with a long primary activity (57% work tours vs. 28% overall).

Class 1 – **Local shopping / errand tours**

- Very short distance
- Very short duration
- Few stops, mostly local travel

Typical tour
Short shopping or errand trips within 5 km lasting 15–30 minutes.

Class 2 – **Complex multimodal activity tours**

- Longest distance
- Longest duration
- Most complex tours (many stops)

Typical tour
Multimodal activity chains combining work, shopping and leisure.

Class 3 – **Leisure-oriented tours**

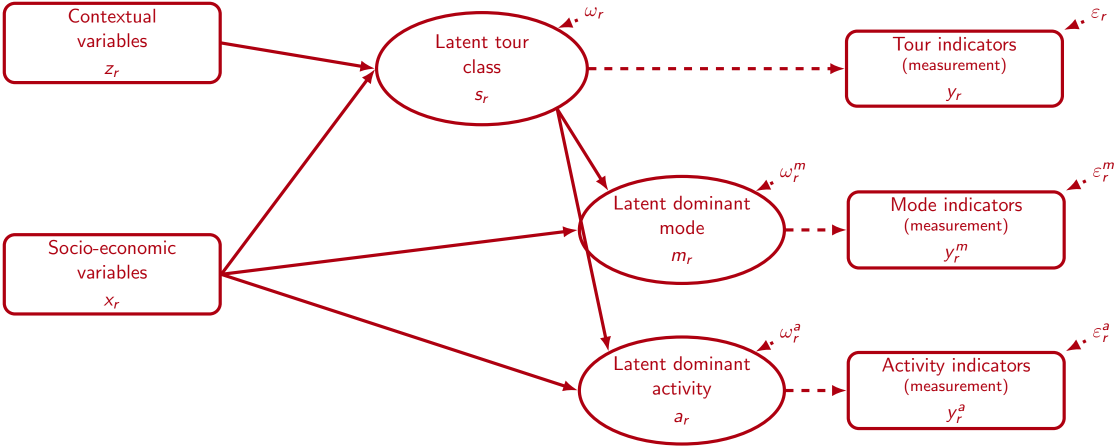
- Medium distance and duration
- Leisure-focused
- Moderate number of stops

Typical tour
Leisure-oriented outings such as parks, restaurants or entertainment.

Limitation of the Flat Latent Class Model

- Tour classes are primarily distinguished by their **dominant activity purpose**:
  - C0: Work
  - C1: Shopping
  - C2: Mixed activity chains
  - C3: Leisure
- Spatial, temporal and modal indicators contribute much less to the class separation.
- **Idea:** explicitly model activity and mobility as separate latent variables.

Hierarchical Latent tour model



Why Model Dominant Mode and Activity as Nested Latents?

Data Interpretability

- Observable from mode and activity shares
- Easy to validate against raw data

Predictive Power

- Estimate from socio-economic factors and context
- Move from ”what happened” to ”what explains it”

Structural Insight

- Tours explain mode/activity tendencies
- Integrated in a single likelihood

$$L_r = \sum_{s=1}^{S} P(s_r = s \mid x_r, z_r) f(y_r \mid s) \left[\sum_{m=1}^{M} P(m_r = m \mid s_r = s, x_r) f(y_r^m \mid m) \right] \left[\sum_{a=1}^{A} P(a_r = a \mid s_r = s, x_r) f(y_r^a \mid a) \right]$$

Hierarchical Latent tour model

1. Structural equations

$$\begin{aligned} V_{rs} &= \alpha_s + \gamma_s x_r + \delta_s z_r \\ V_{rms} &= \alpha_m + \beta_m x_r + \eta_{ms} s_r \\ V_{ras} &= \alpha_a + \beta_a x_r + \eta_{as} s_r \end{aligned}$$

2. Latent probabilities

$$\begin{aligned} P(s_r = s \mid x_r, z_r) &= \frac{\exp(V_{rs})}{\sum_{k=1}^{S} \exp(V_{rk})} \\ P(m_r = m \mid x_r, s_r = s) &= \frac{\exp(V_{rms})}{\sum_{q=1}^{M} \exp(V_{rqs})} \\ P(a_r = a \mid x_r, s_r = s) &= \frac{\exp(V_{ras})}{\sum_{h=1}^{A} \exp(V_{rhs})} \end{aligned}$$

3. Measurement equations

$$f(y_r \mid s_r) = \prod_{j=1}^{J_c} \mathcal{N}(y_{rj} \mid \mu_{js}, \sigma_{js}^2) \prod_{j=1}^{J_b} \mathrm{B}(y_{rj} \mid p_{js})$$

$$\begin{aligned} f(y_r^m \mid m_r) &= \prod_{j=1}^{J_c} \mathcal{N}(y_{rj}^m \mid \mu_{jm}, \sigma_{jm}^2) \\ f(y_r^a \mid a_r) &= \prod_{j=1}^{J_c} \mathcal{N}(y_{rj}^a \mid \mu_{ja}, \sigma_{ja}^2) \end{aligned}$$

1. Explanatory variables

a. Socio-economic variables x_r
- age, piecewise continuous
- gender
- employment status
- car availability
- bike availability
- public transport subscription
- household size / region, if used

b. Contextual variables z_r
- day of week
- time of day
- weather conditions
- tour context variables

2. Latent variables

a. Latent tour class s_r
Main behavioral class of the tour.

b. Latent dominant mode m_r

c. Latent dominant activity a_r

Class-to-layer effects
- η_{ms}^m: effect of tour class s on dominant mode m
- η_{as}^a: effect of tour class s on dominant activity a

Estimated class-specific parameters
- tour indicator means and variances
- binary indicator probabilities
- dominant mode probabilities
- dominant activity probabilities

3. Measurement indicators

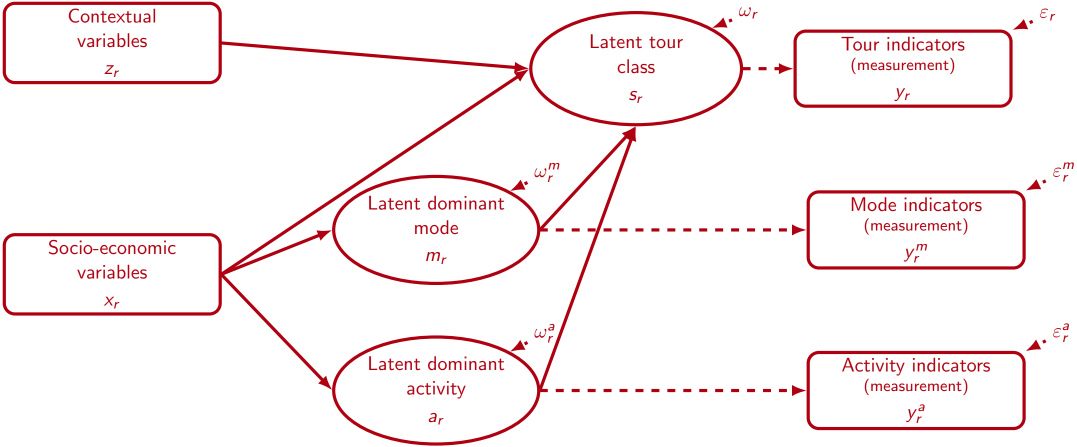
a. Tour indicators y_r
- total distance
- tour duration
- max distance from home
- max activity duration
- number of legs
- number of distinct POIs
- multimodal tour

b. Mode indicators y_r^m
- observed dominant car/PT mode
- car distance share
- PT distance share

c. Activity indicators y_r^a
- observed dominant activity
- work share
- shopping share
- leisure share
- other share

d. Posterior summaries
- $P(m_r = \mathrm{PT} \mid \cdot)$
- $P(a_r = \mathrm{work} \mid \cdot)$
- $P(a_r = \mathrm{shopping} \mid \cdot)$
- $P(a_r = \mathrm{leisure} \mid \cdot)$

Change the causality structure ?



Add a latent class of users ?

$$\boxed{\text{Person} \rightarrow \text{Tour}}$$

Rather than clustering heterogeneous individuals directly, we model the behavioral building blocks of mobility.

Results: Activity-choice model

Table: Dominant activity structural and measurement model estimates.

Parameter	Work	Shopping	Leisure	Other
ASC	0.736* (0.436)	-0.467 (0.519)	0.716 (0.438)	–
Age 18–35	0.278*** (0.0331)	0.247*** (0.0378)	0.293*** (0.0333)	–
Age 35–65	-0.0682*** (0.0187)	-0.0371* (0.0191)	-0.0552*** (0.0186)	–
Age 65+	-0.273*** (0.0508)	-0.00378 (0.0482)	-0.0159 (0.0468)	–
Female	0.166 (0.310)	0.322 (0.319)	0.264 (0.309)	–
Measurement scale		1.950*** (0.00138)		

Notes: Standard errors in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$. Other is the reference alternative in the structural model.

Results: Mode-choice model

Table: Dominant mode structural and measurement model estimates.

Parameter	Car	PT	Bicycle	Foot
ASC	0.207 (0.401)	-0.0218 (0.461)	-0.132 (0.553)	–
Age 18–35	0.128*** (0.0233)	0.0621** (0.0260)	0.00536 (0.0355)	–
Age 35–65	-0.0193*** (0.00746)	-0.0238*** (0.00865)	-0.00852 (0.0101)	–
Age 65+	-0.0317 (0.0247)	-0.00981 (0.0300)	-0.113*** (0.0428)	–
Car always available	0.235 (0.156)	-0.208 (0.166)	–	–
Car by arrangement	-0.0756 (0.165)	0.0768 (0.172)	–	–
Bike always available	0.0159 (0.102)	-0.0395 (0.130)	–	–
Bike by arrangement	0.175 (0.217)	0.0469 (0.260)	–	–
Half-fare pass	0.0818 (0.107)	-0.178 (0.179)	–	–
Half-fare combo	-0.0727 (0.179)	0.204 (0.220)	–	–
GA pass	-0.0828 (0.160)	0.212 (0.201)	–	–
Regional pass	-0.0588 (0.259)	0.187 (0.306)	–	–
Measurement scale		0.755*** (0.000456)		

Conclusion

- The tour is a more appropriate behavioral unit than the individual for characterizing long-term mobility.
- Latent tour classes reveal meaningful mobility patterns but remain strongly influenced by activity purpose.
- Introducing hierarchical latent variables might improve interpretability, but there's still the question of the causality.

Outline

1. Project Overview
2. Methodology
3. Next Steps: RQ2

Next Step: Multi-Day Activity-Based Modeling

From typology to simulation

- **RQ1**: typology of mobility behavior.
- **RQ2**: multi-day activity-based model.

Modeling objective

- Identify multi-day activity patterns
- Capture weekly routines such as groceries, home office and leisure
- Model timing, duration and location chains

Data fusion

- **Microcensus**: representative one-day travel behavior
- **CMP**: repeated observations and multi-day patterns
- Joint basis for simulation and scenario analysis

Thank You

Questions?

anne-valerie.preto@epfl.ch