



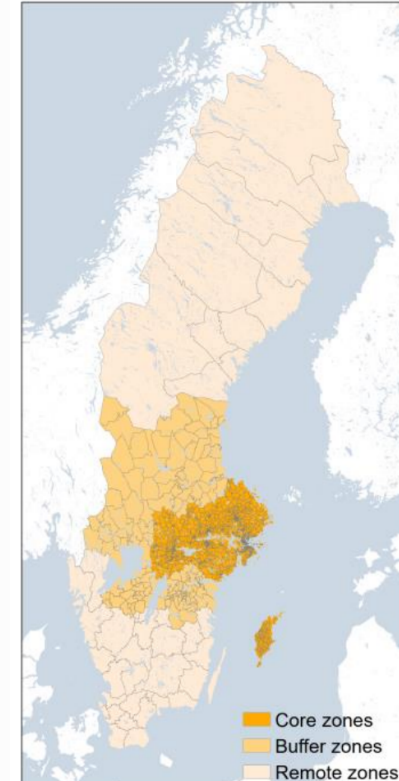
A Kernel-Based Approach for Survey-Informed Generation of Synthetic Travel Plans

Michael Sederlin^{1,2} & Gunnar Flötteröd^{1,2}

Thursday 15th October, 2026 — LiU¹, VTI²

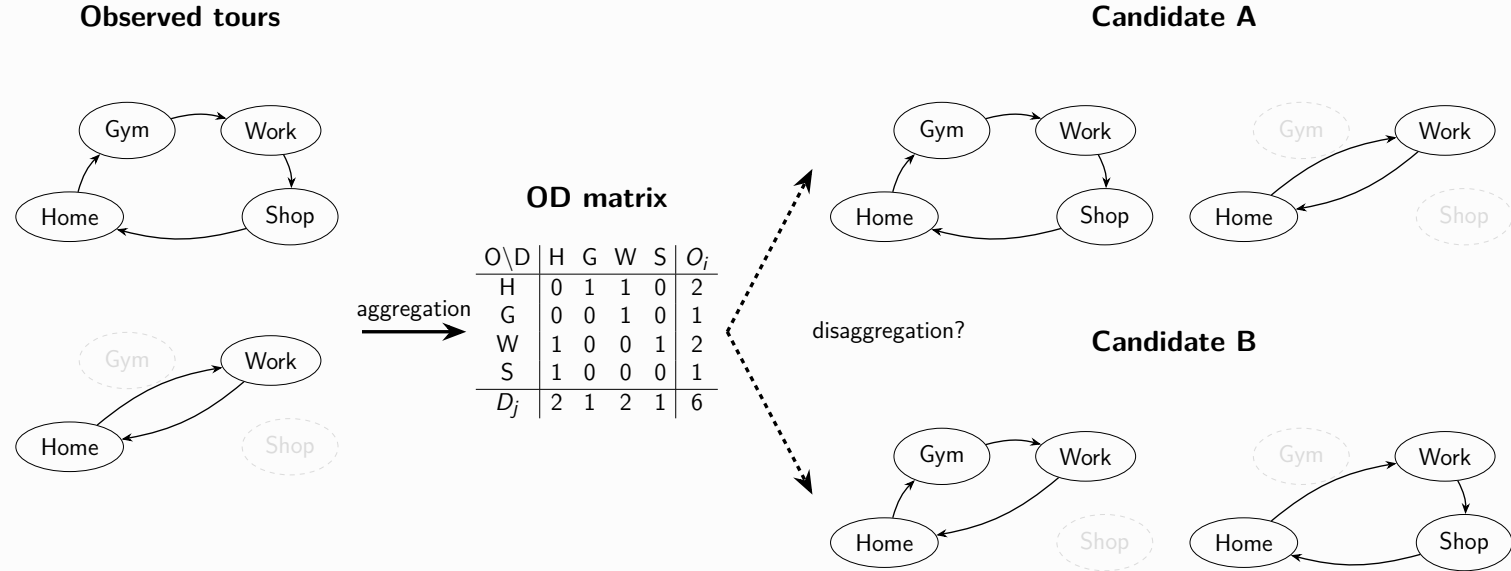
From national forecasts to urban dynamics

- National models (e.g. **Sampers** in Sweden) forecast travel demand at large scales with strong simplifications → static, aggregated results.
- Urban analysis requires less temporal and spatial aggregation for:
 - Analysis of dynamic congestion patterns
 - Evaluation of time-dependent policies
 - Certain equity-related analyses and accessibility
- Disaggregation should remain:
 - behaviourally consistent (survey evidence),
 - physically feasible (time/activity constraints),
 - transferable across populations/scenarios.
- **Consequence:** high-resolution dynamic demand representations are needed for detailed analysis.



Sampers study area

Many daily plans fit the same data



Same OD matrix, different tours: **which one is right?**

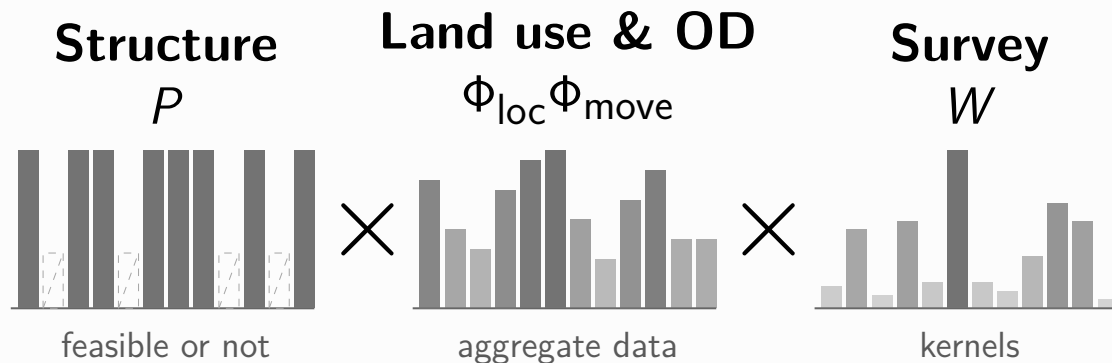
State and sampling

A daily plan of agent i is $Y_i^{\text{sim}} = (l_i, d_i)$: visited locations $l_i = (l_i^{(1)}, \dots, l_i^{(J_i)})$ and departure times $d_i = (d_i^{(1)}, \dots, d_i^{(J_i)})$; X_i^{sim} are its socio-demographics. One MH state is a complete set of daily plans for all synthetic agents, $Y^{\text{sim}} = (Y_1^{\text{sim}}, \dots, Y_N^{\text{sim}})$.

i	age [yr]	X_i^{sim} income [EUR/yr]	l_i	d_i
1	22	0	[home ⁽¹⁾ , university ⁽¹⁾ , supermarket ⁽¹⁾]	[9, 18, 19]
2	78	38,000	[home ⁽²⁾ , store ⁽²⁾ , home ⁽²⁾ , cafe ⁽²⁾]	[7, 9, 15, 17]
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
N	52	60,000	[home ^(N) , office ^(N)]	[7, 18]

Narrowing down the possible schedules

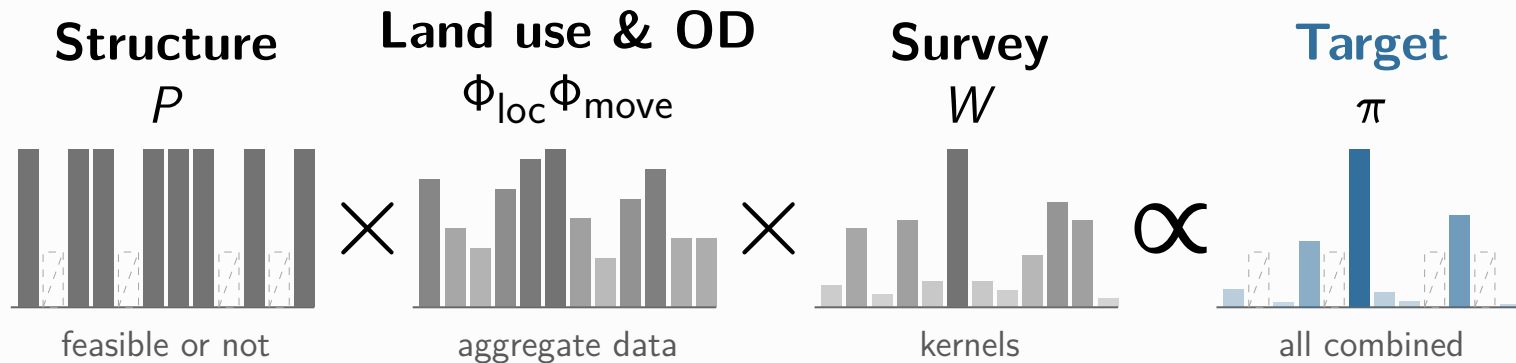
Instead of picking one, we **score all candidate plans**.



Schematic. P : round trips from a fixed home, uniform over plan sizes.

Narrowing down the possible schedules

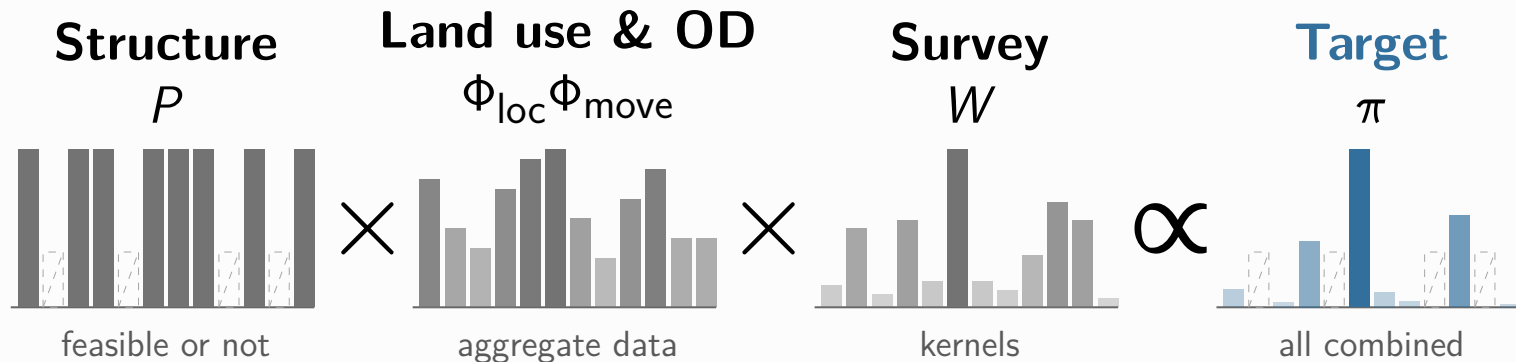
Instead of picking one, we **score all candidate plans**.



Schematic. P : round trips from a fixed home, uniform over plan sizes.

Narrowing down the possible schedules

Instead of picking one, we **score all candidate plans**.



$$\pi \propto \underbrace{W}_{\text{survey}} \cdot \underbrace{\Phi_{loc} \Phi_{move}}_{\text{land use \& OD}} \cdot \underbrace{P}_{\text{structure}}$$

Schematic. P : round trips from a fixed home, uniform over plan sizes.

Behavioural similarity via kernels

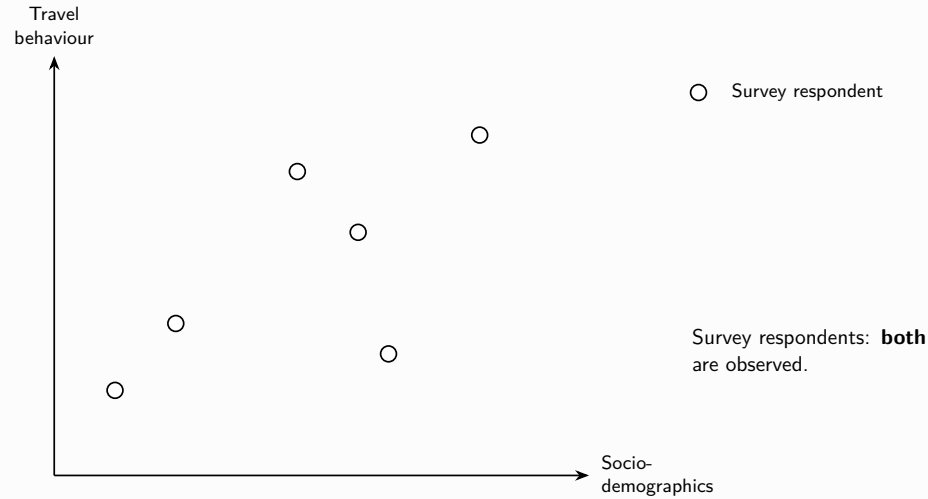
Travel
behaviour



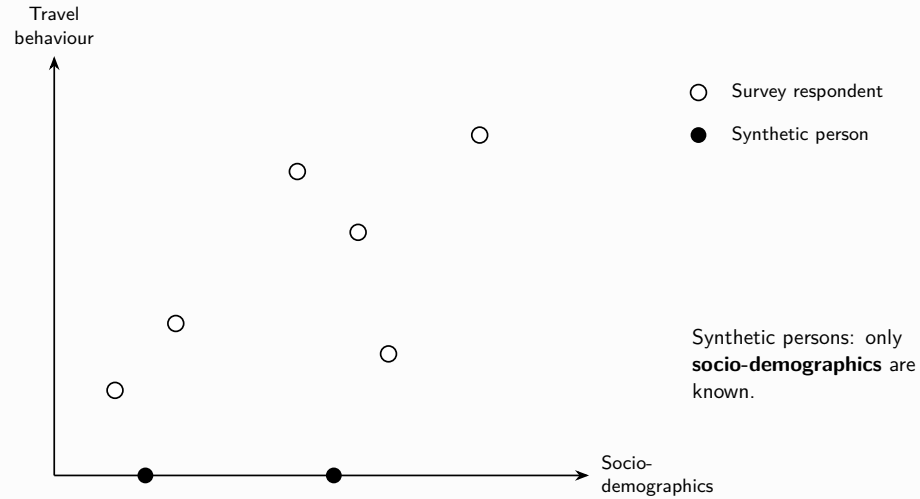
Socio-
demographics

Each person has socio-
demographics and travel be-
haviour.

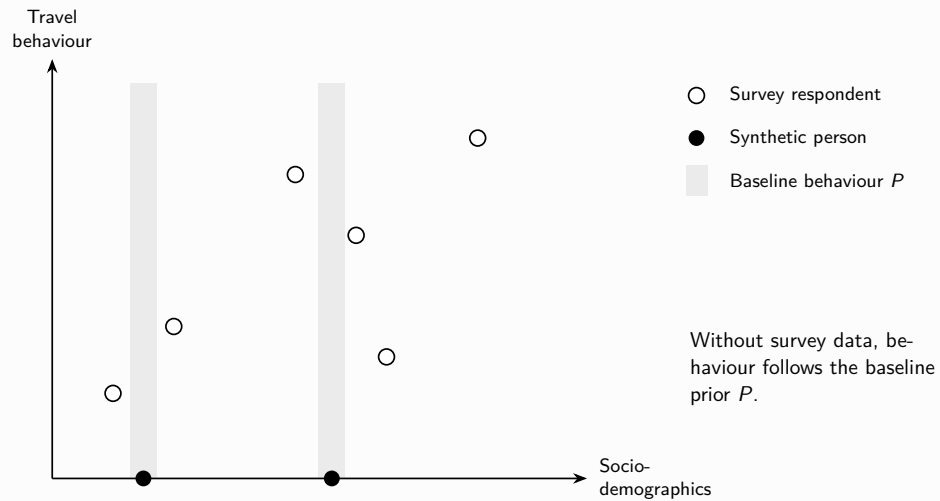
Behavioural similarity via kernels



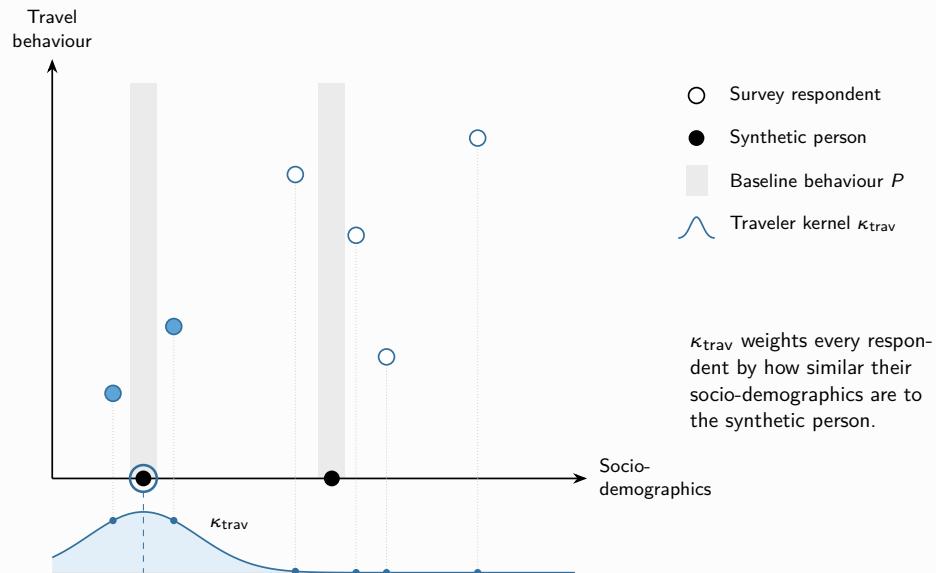
Behavioural similarity via kernels



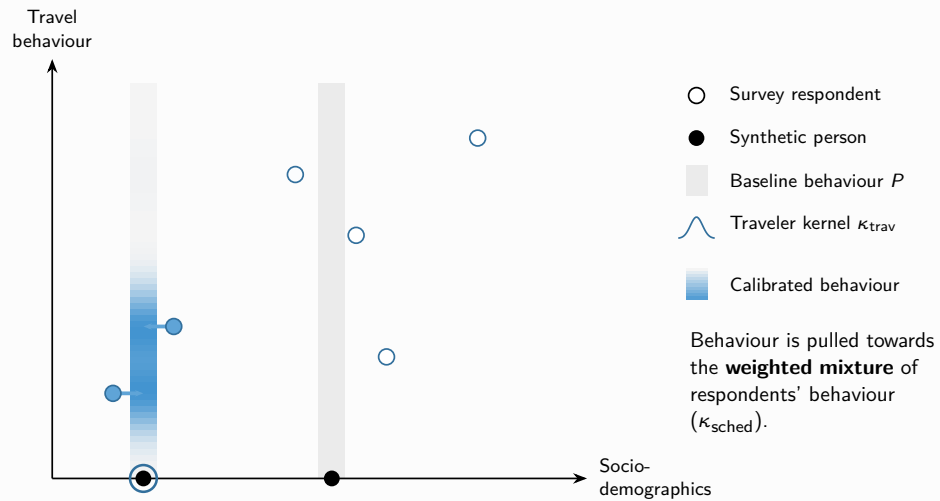
Behavioural similarity via kernels



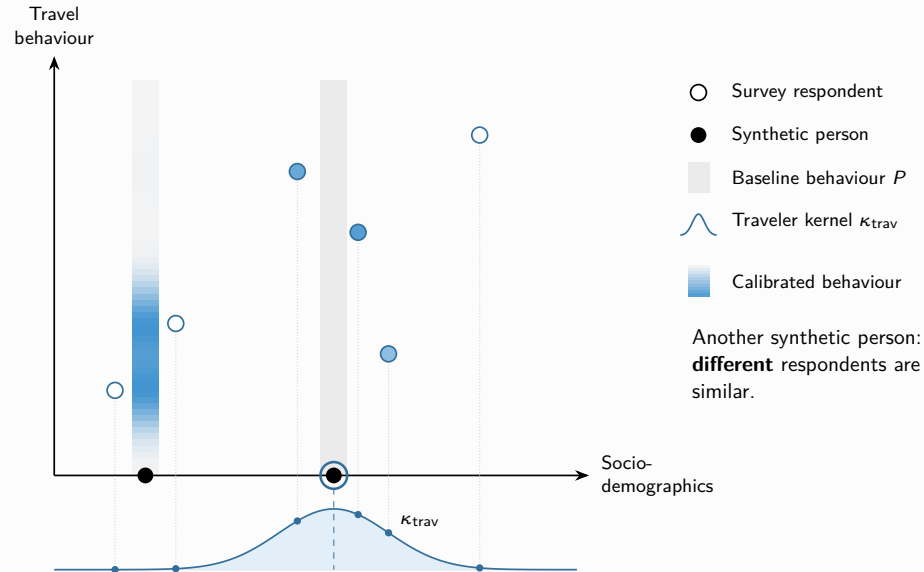
Behavioural similarity via kernels



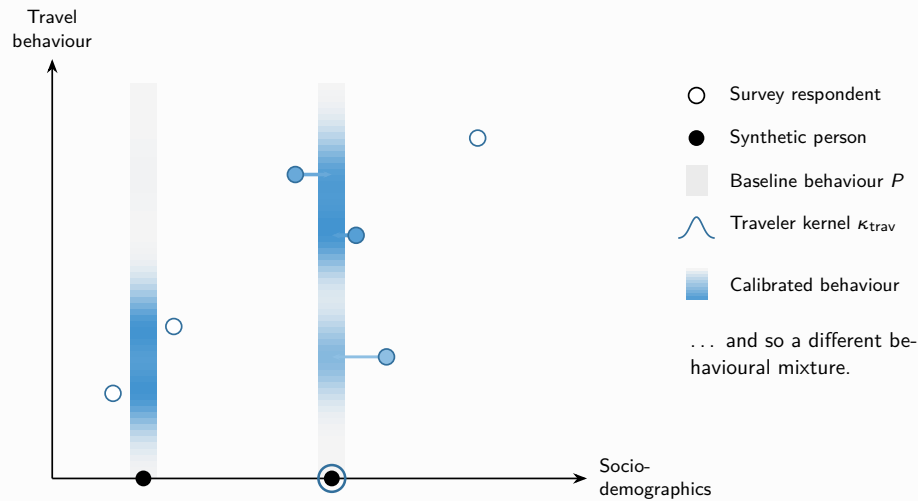
Behavioural similarity via kernels



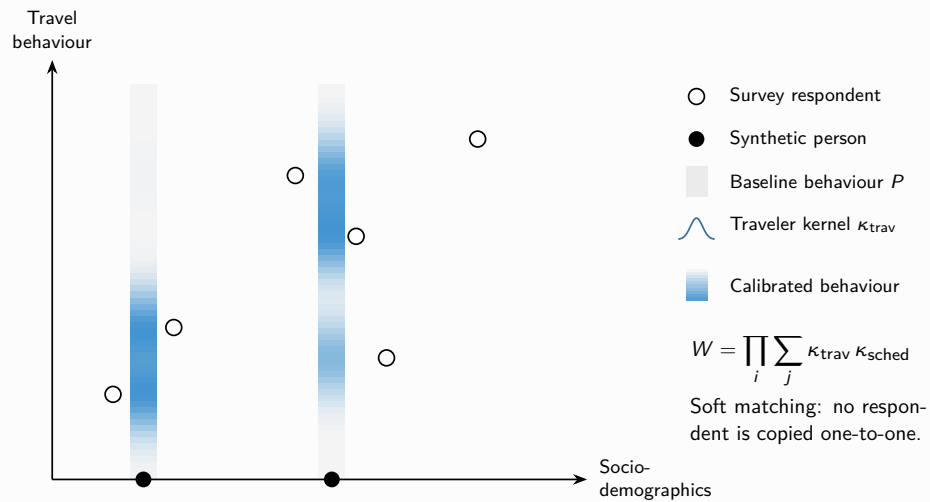
Behavioural similarity via kernels



Behavioural similarity via kernels



Behavioural similarity via kernels



Target distribution

The sampler draws the plans Y^{sim} of all agents from

$$\pi \left(Y^{\text{sim}} \mid \cdot \right) \propto \underbrace{W \left(Y^{\text{sim}}; X^{\text{sim}}, X^{\text{obs}}, Y^{\text{obs}} \right)}_{\text{survey}} \cdot \underbrace{\Phi_{\text{loc}}(Y^{\text{sim}}) \Phi_{\text{move}}(Y^{\text{sim}})}_{\text{land use \& OD}} \cdot \underbrace{P \left(Y^{\text{sim}} \mid X^{\text{sim}} \right)}_{\text{structure}}$$

- $W = \prod_{i=1}^N \sum_{j=1}^M \kappa_{\text{trav}}(X_i^{\text{sim}}, X_j^{\text{obs}}) \kappa_{\text{sched}}(Y_i^{\text{sim}}, Y_j^{\text{obs}})$: every agent i is softly matched to every respondent j .
- $\Phi_{\text{loc}}, \Phi_{\text{move}}$: aggregate data (zone scores, OD matrix), scored per plan.
- P : plans in the absence of data: round trips, fixed home location, uniform over plan sizes.
- Metropolis–Hastings only needs ratios of π , so normalizing constants cancel.

Traveler & schedule kernels

i : agent, j : respondent
 J : # visited locations

Traveler kernel

$$\kappa_{\text{trav}}(X_i^{\text{sim}}, X_j^{\text{obs}}) = \kappa_{\text{inc}}(\Delta \text{inc}) \cdot \kappa_{\text{age}}(\Delta \text{age}) \cdot \kappa_{\text{gender}}(\Delta \text{gender})$$

Income deviation Age deviation Gender difference

Schedule kernel

$$\kappa_{\text{sched}}(Y_i^{\text{sim}}, Y_j^{\text{obs}}) = \kappa_J(\Delta J) \cdot \kappa_{\text{overlap}}(Y_i^{\text{sim}}, Y_j^{\text{obs}})$$

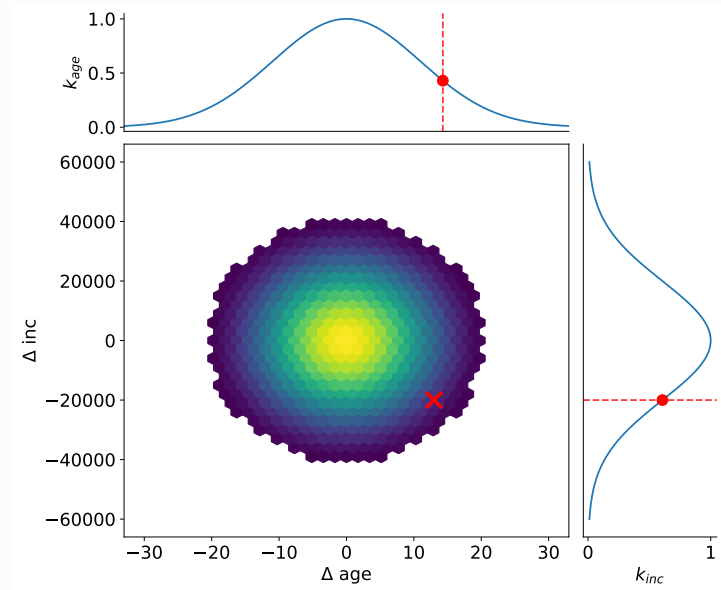
Number of stops Schedule overlap

Example of Traveler Difference

Comparing two persons p_1 and p_2

- $p_1 = (32, 130,000)$
- $p_2 = (45, 110,000)$
- $\Delta \text{age} = 13$
- $\Delta \text{inc} = 20,000$

Similarity decreases as age and income differences increase.

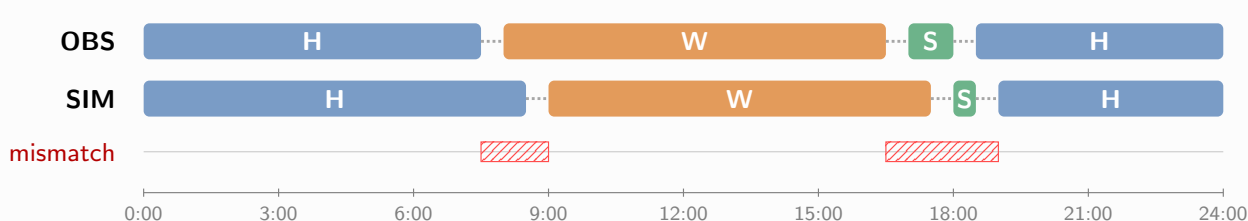


Schedule kernel: when are activities performed?

Compare **when** the agent and the respondent do each activity.

$$\kappa_{\text{overlap}}(\text{plan}, \text{diary}) = \prod_{\text{activities}} \gamma(\text{mismatch in activity time windows})$$

γ shrinks as the mismatch (red) grows; tolerance set per activity



$$\delta_a = |\vartheta_{ia}^{\text{sim}} \cup \vartheta_{ja}^{\text{obs}}| - |\vartheta_{ia}^{\text{sim}} \cap \vartheta_{ja}^{\text{obs}}|$$

home: $\delta_H = 1.5$ h

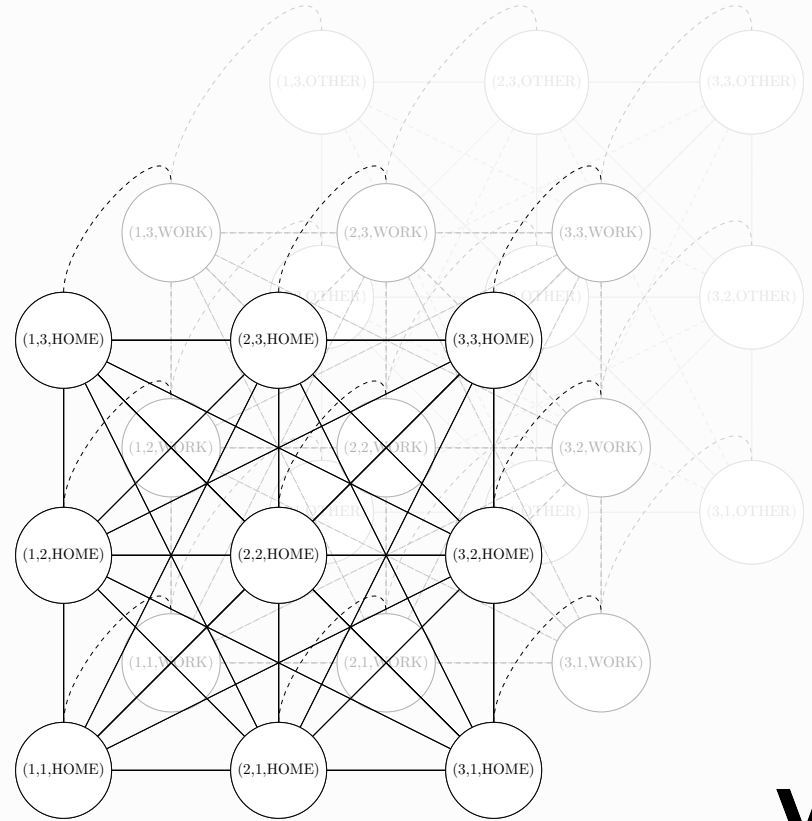
work: $\delta_W = 2.0$ h

shopping: $\delta_S = 1.5$ h

$$\kappa_{\text{overlap}} = \prod_a \gamma(\delta_a, \omega_a)$$

Activities via network expansion (no exact locations needed)

- **activity-expanded network** with duplicated nodes for each possible activity (work, shopping, leisure, ...).
- simulated trip visiting certain locations implies a set of **performed activities**.
- **not** trying to reproduce exact survey destinations.
- seek to have respondent's activities **possible** in simulated plans.

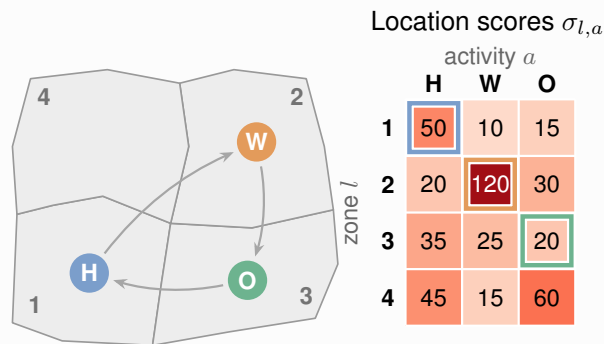


Land use and OD data enter as weights

A plan scores higher if it visits zones with **many opportunities** and uses **frequent OD relations**.

Location weight

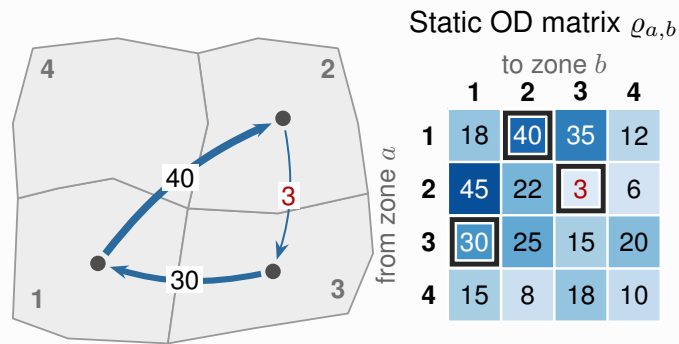
$$\ln \Phi_{\text{loc}}(\text{plan}) = \sum_{\text{visited zones}} \ln (\text{zone score})$$



Zone score: e.g. building floor area

Movement weight

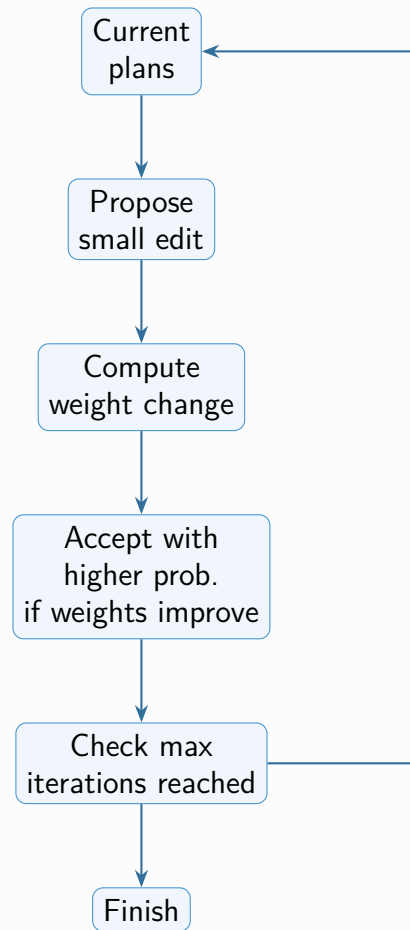
$$\ln \Phi_{\text{move}}(\text{plan}) = \sum_{\text{transitions}} \ln (\text{OD flow})$$



OD flow: any static OD matrix, e.g. Samplers

Algorithm

- We keep a **complete set of daily plans** for the synthetic population.
- **Propose:** make a small edit to one/few agents (trip chain / times / stops).
- **Survey compatibility weights** (higher = more consistent with survey),
- **Accept:** accept the edit with a probability that increases when the **kernel weights improve**.
- Only ratios of π are needed, so normalizing constants cancel.



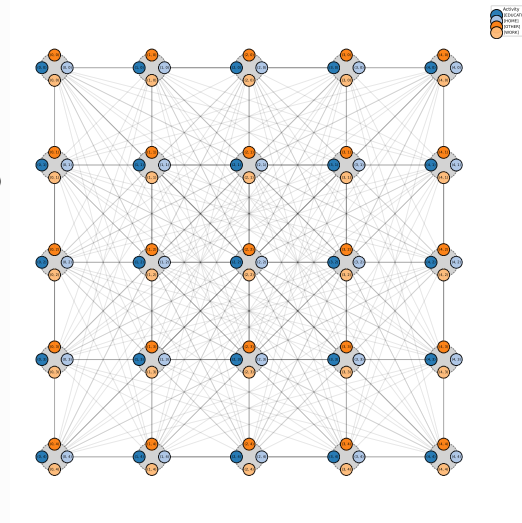
EXPERIMENT 1

Controlled grid world

Round trips of a 1000-person synthetic population calibrated against 100 individual-level survey records, in a grid world with known ground truth.

Grid-world scenario

- **25 zones** on a 5×5 grid.
- Each zone is **activity-expanded** into home, work, education and other locations.
- Hourly departure times over one day.
- Homes spread evenly: **40 persons per zone**.



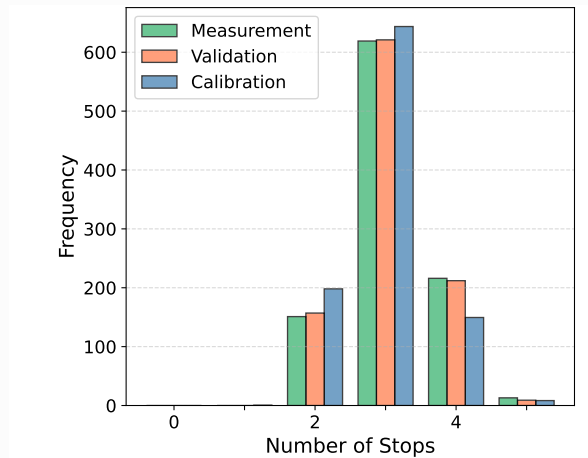
Activity-expanded 5×5 grid

40	40	40	40	40
40	40	40	40	40
40	40	40	40	40
40	40	40	40	40
40	40	40	40	40

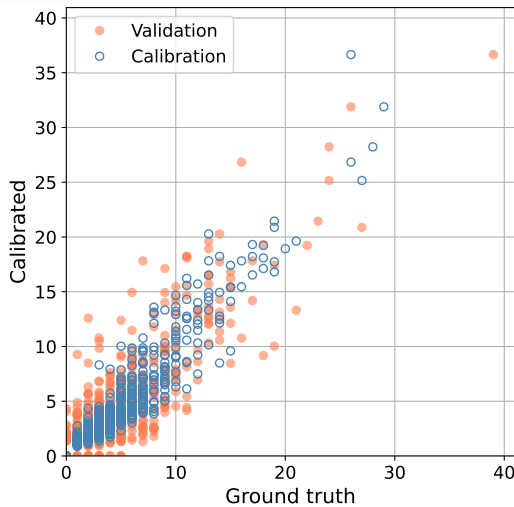
31	44	52	38	58
82	30	23	41	25
28	24	35	67	27
24	116	26	37	33
47	32	25	26	29

Activity locations per zone: home (left), work (right)

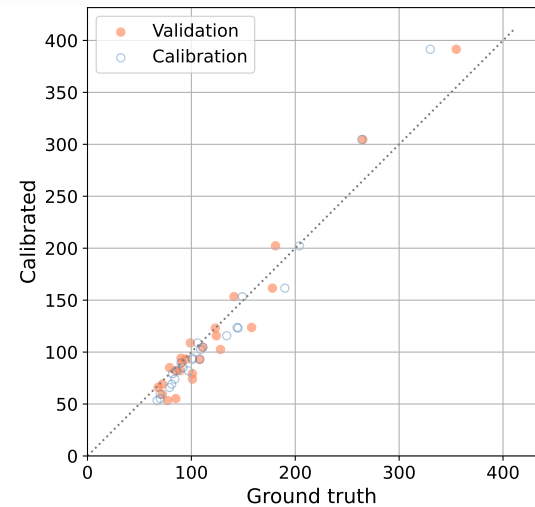
Results: round trips, OD flows, zonal visits



Round trip sizes



OD flows



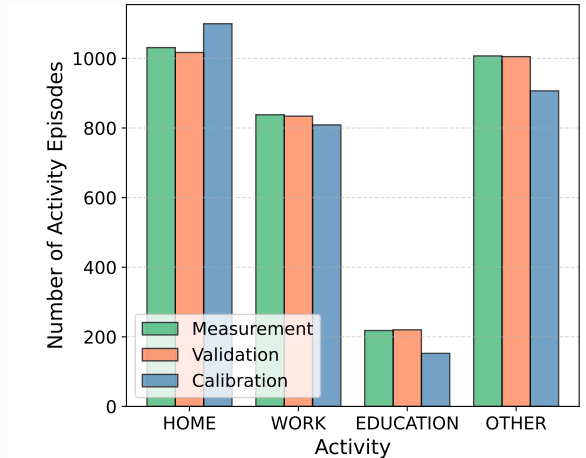
Zonal visits

Measurement: data used for calibration

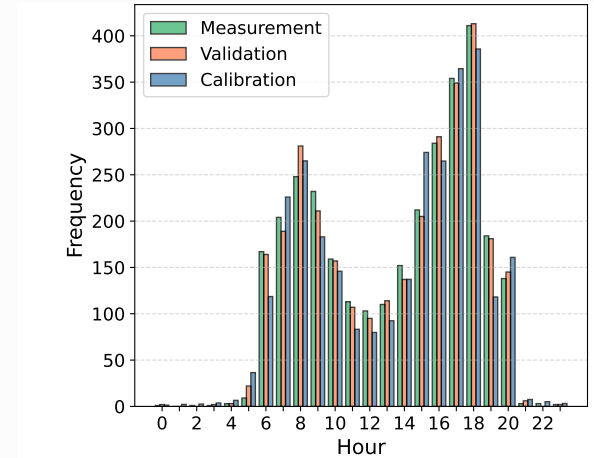
Validation: independent draw from the synthetic reality

Calibration: sampled round trips

Results: activity participation and departures

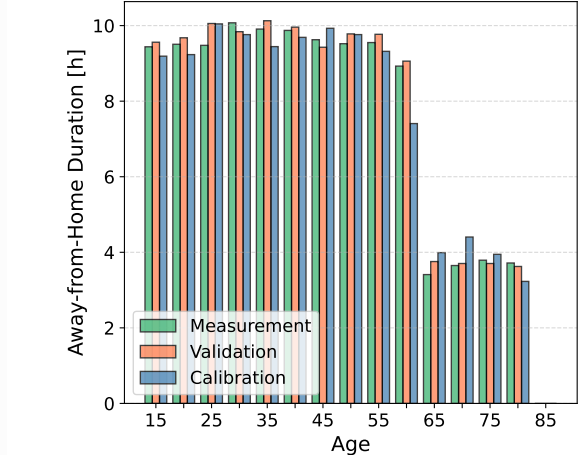


Activity participation

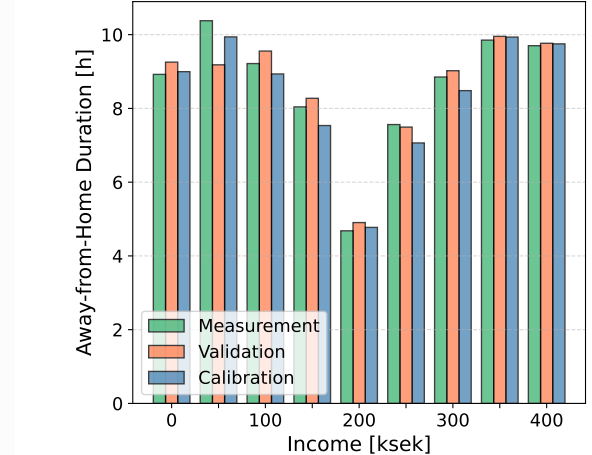


Within-day trip starts

Results: durations by socio-demographic group

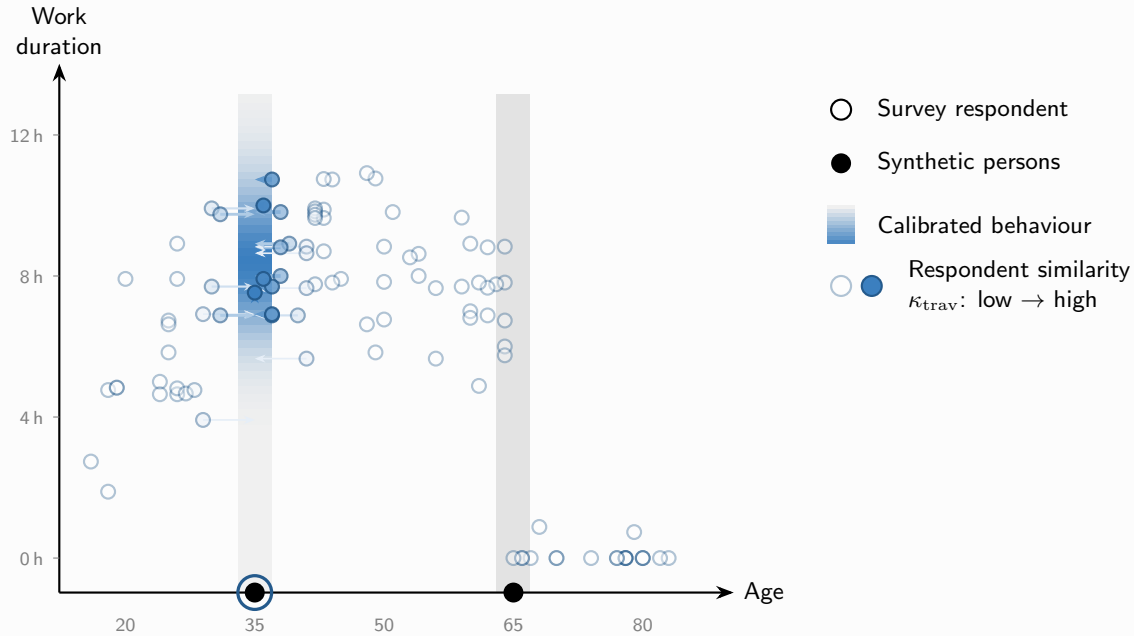


Activity duration (all but home) by age

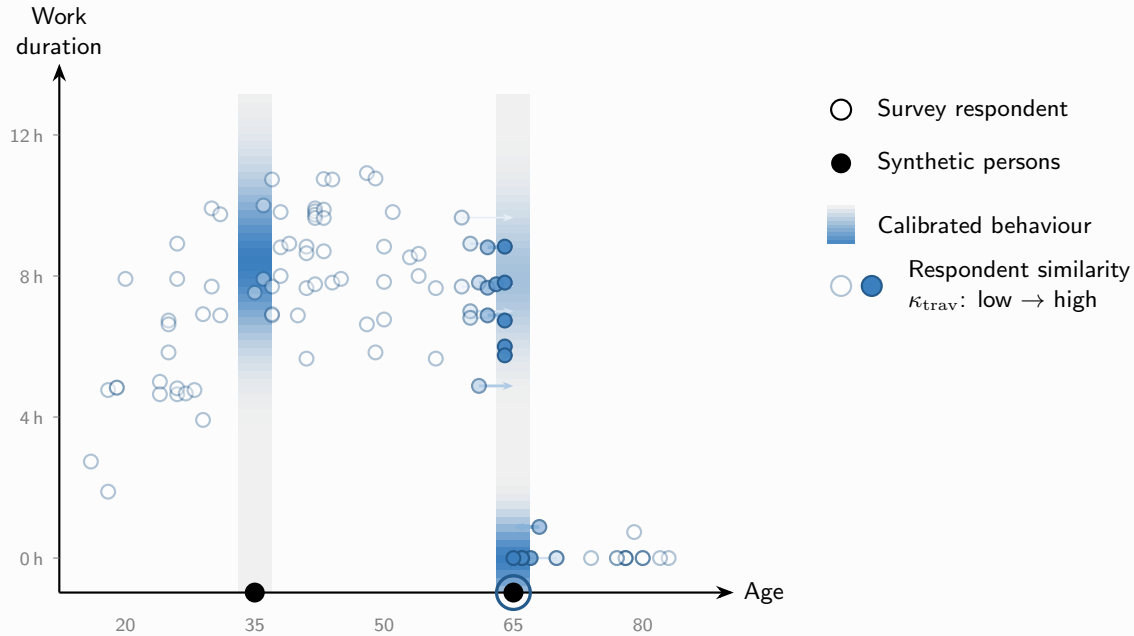


Activity duration (all but home) by income

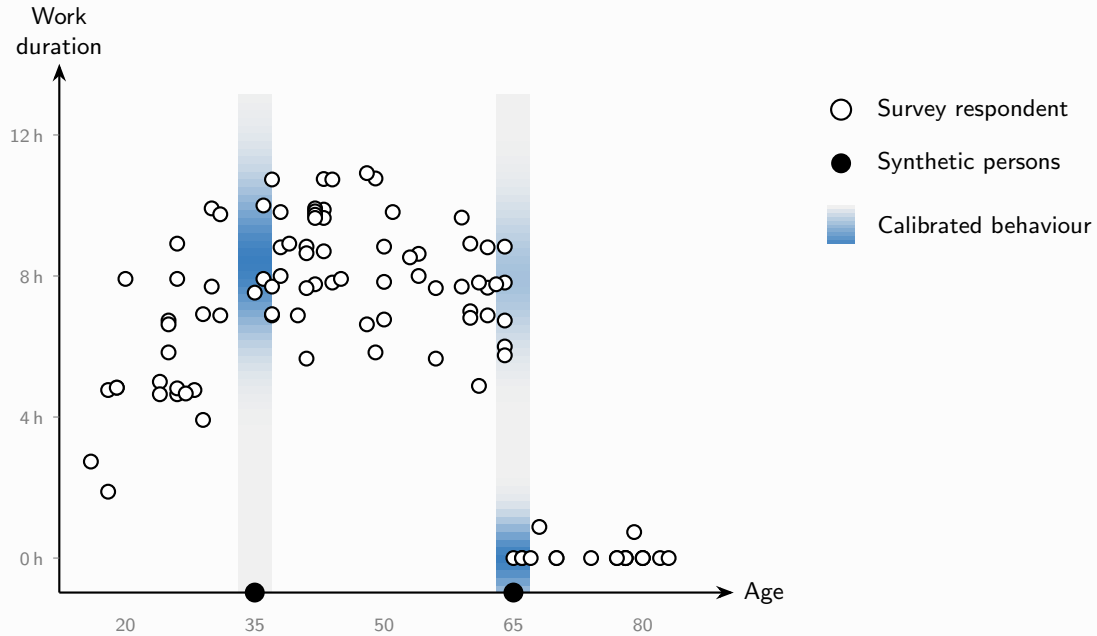
Results: Illustration work duration distribution



Results: Illustration work duration distribution



Results: Illustration work duration distribution



EXPERIMENT 2

Real-world case: Stockholm

Travel survey diaries (RVU) transferred to Trafikverket's synthetic population for policy analysis (Sampers)

Two separate populations, no shared individuals.

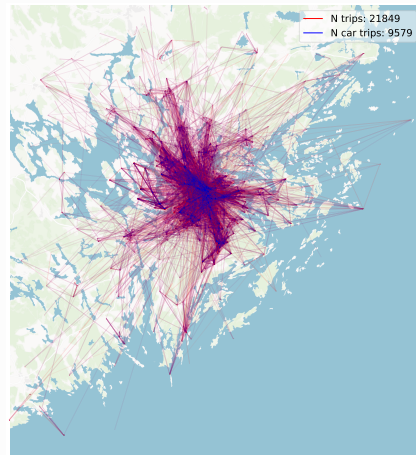
Stockholm: data and setup

Data

- RVU travel survey: car round trips within the study area
- Trip purposes grouped into five activity types
- Trafikverket's synthetic population (Sampers); agents keep their home location

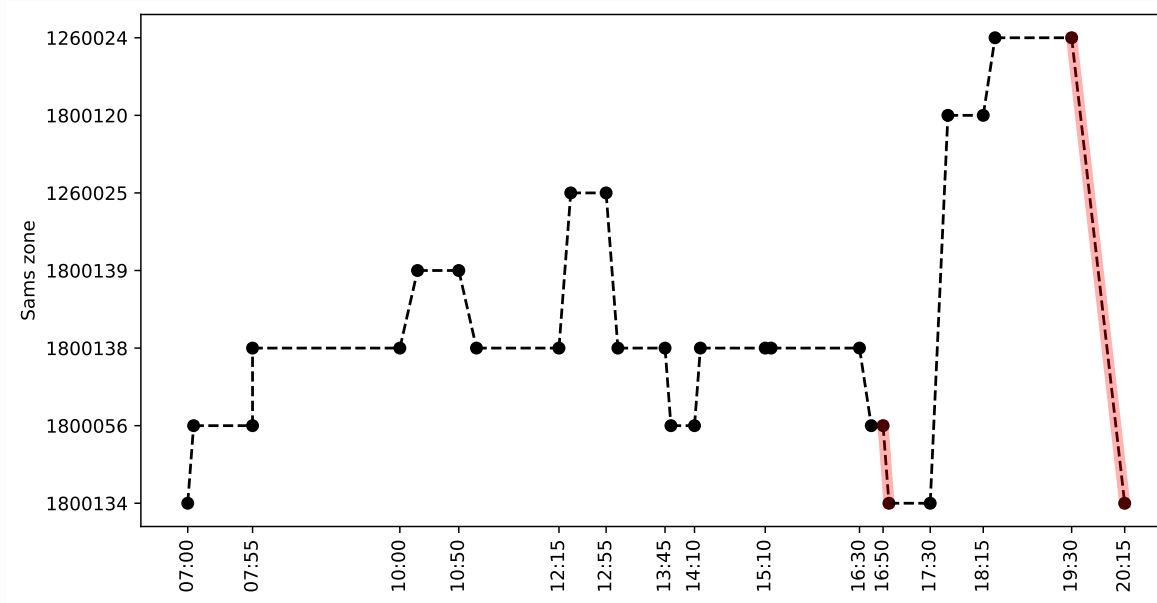
Model

- Activity-expanded zone network
- Opportunity scores from building floor area



RVU trips within the Stockholm region

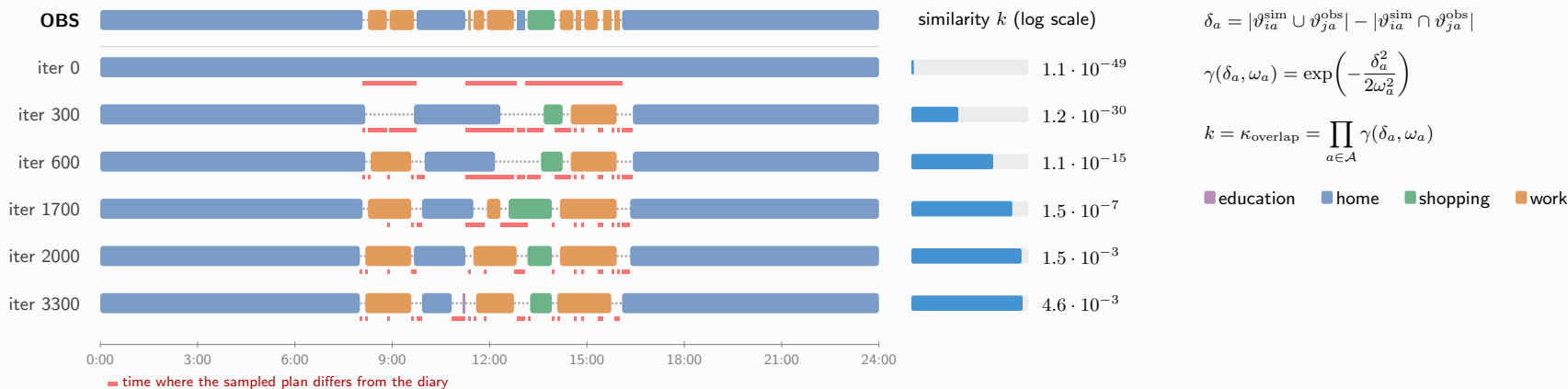
Example observed daily schedule (diary view)



Diary view provides (i) ordering, (ii) timing, (iii) purposes — used to build schedule and feasibility kernels.

Sampling moves the plan towards the diary

One observed diary (OBS) and six plans sampled for the same agent: the mismatch shrinks and the similarity k grows.



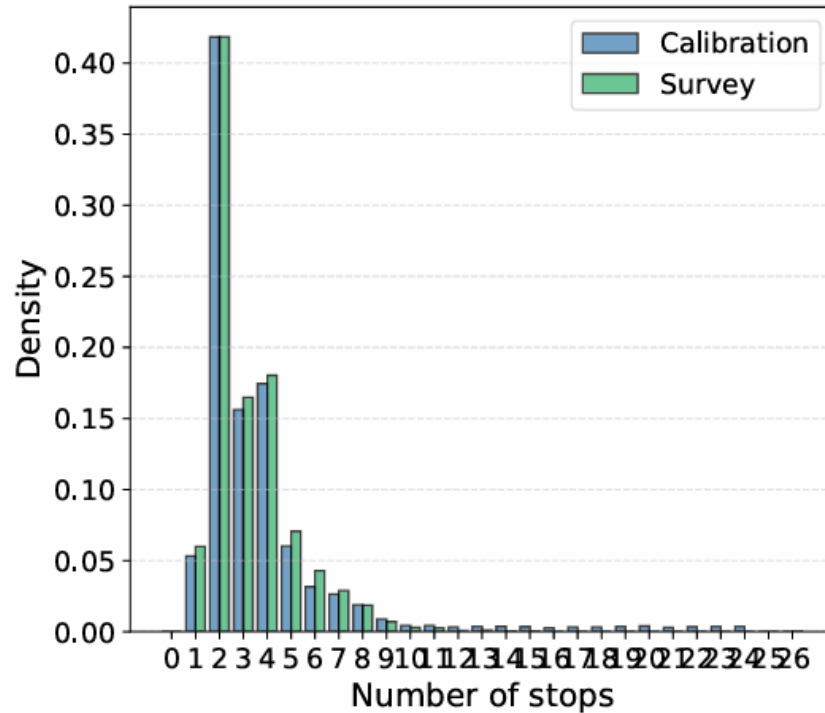
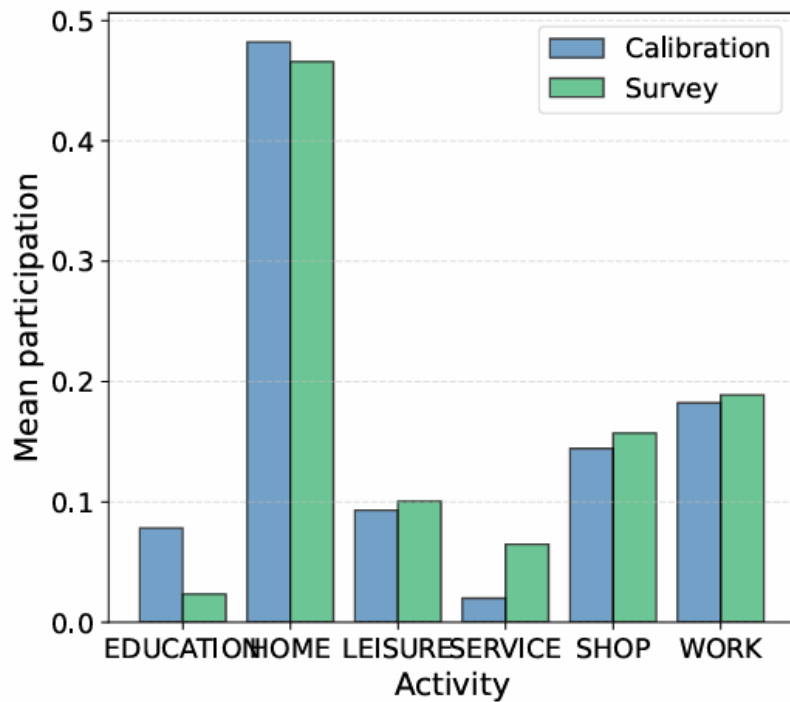
$$\delta_a = |\vartheta_{ia}^{\text{sim}} \cup \vartheta_{ja}^{\text{obs}}| - |\vartheta_{ia}^{\text{sim}} \cap \vartheta_{ja}^{\text{obs}}|$$

$$\gamma(\delta_a, \omega_a) = \exp\left(-\frac{\delta_a^2}{2\omega_a^2}\right)$$

$$k = \kappa_{\text{overlap}} = \prod_{a \in \mathcal{A}} \gamma(\delta_a, \omega_a)$$

One synthetic agent and one RVU diary; selected samples from one Metropolis–Hastings run.

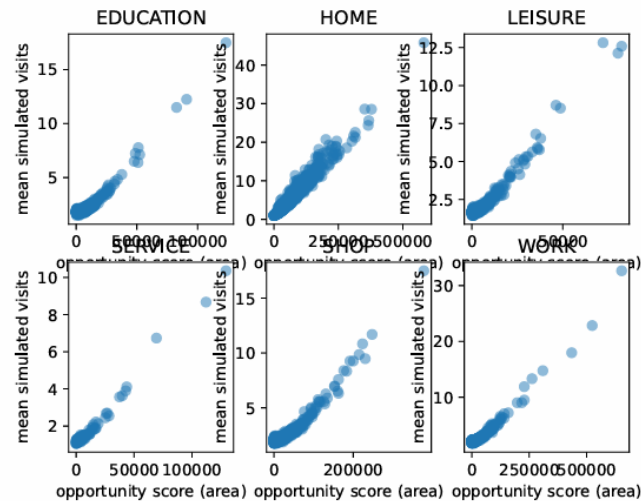
Stockholm: activity participation and number of stops



Stockholm: activities follow land-use opportunities



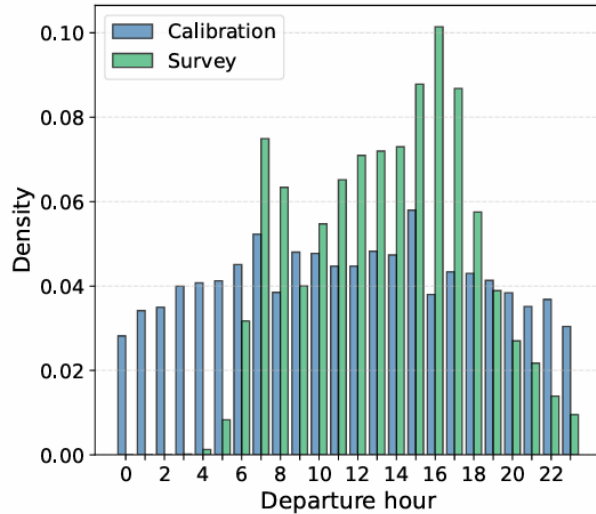
WORK attractiveness per zone from building floor area



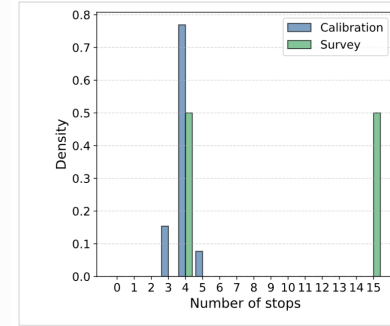
Correlation of visited zones and the zone score for that given activity in that

zone

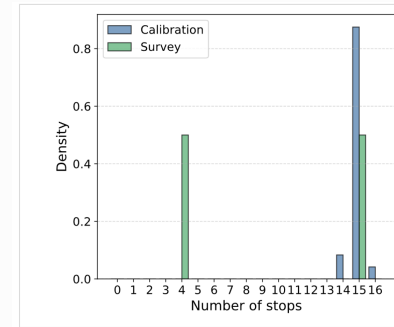
Departure times and mixing



Departure hour: survey vs. simulated



Number of stops: cold start



Number of stops: warm start

Summary and outlook

Summary

- Complete daily plans for a synthetic population, scored by **structure**, **aggregate data** and **survey** at once; sampled with Metropolis–Hastings.
- **Grid world:** trip sizes, OD flows, zonal visits, departures and durations by group reproduced against held-out data.
- **Stockholm:** activity participation and stops broadly reproduced; visits follow land use. Departure times not yet.

What this enables

- **Dynamic demand for existing models:** aggregate the plans into time-resolved OD matrices.
- **Completing today's demand:** add a few hours of observations as one more weight; linked trips let the morning inform the afternoon → short-term forecast.

Next

- Full regional run, better mixing, calibration of kernel parameters

Thank you

Any questions and comments are appreciated!

Paper



papers.ssrn.com/abstract_id=7468019

Code



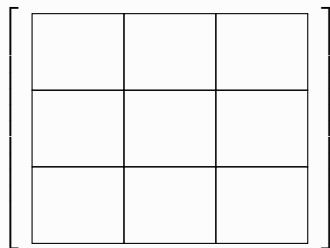
github.com/vtisweden/java-projects

Same framework, other uses

- Delivery vehicle charging (Skellefteå)
- Drone delivery fleets
- Freight vehicle choice sets
G. Flötteröd et al.
- Disaggregation of VISUM OD matrices
F.-X. Rupprecht, BOKU
- Weekly schedules in a 15-minute city
G. Charalampidou, BOKU

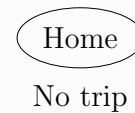
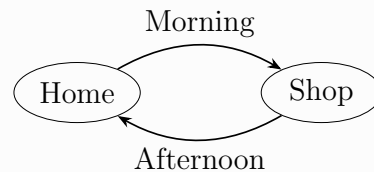
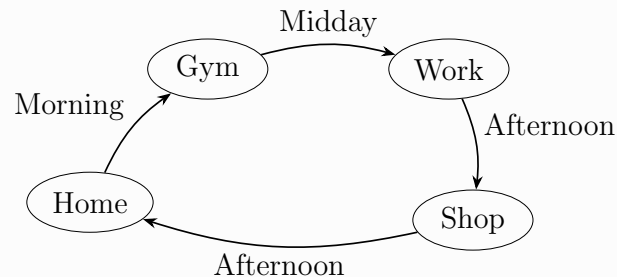
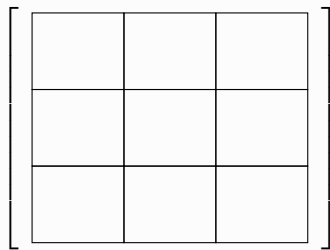
Aggregate v. Disaggregate demand representations

Morning



No connection

Evening



Proposed approach

Approach

- Generate complete daily travel plans for a synthetic population
- Use survey responses as soft behavioral evidence
- Avoid one-to-one copying of real individuals

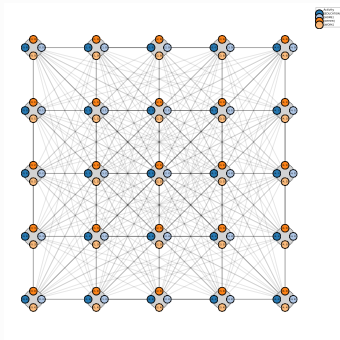
Mechanism

- Compare synthetic agents to survey respondents with kernels
- Combine traveler and schedule similarity; land use and OD flows enter as separate weights
- Use Metropolis–Hastings to sample better plan sets

Results overview

Controlled grid world

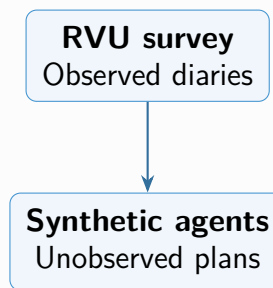
Known ground truth, independent validation data



- 1000 synthetic persons, 100 survey records
- Calibrated plans compared with held-out data

Stockholm: RVU → Trafikverket synthetic population

Transfer between two *separate* populations

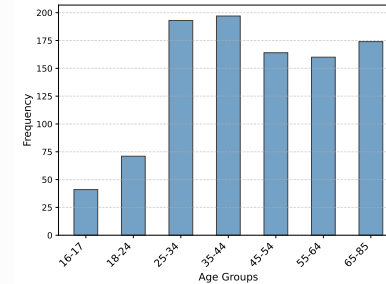


- No true behaviour for the synthetic agents.

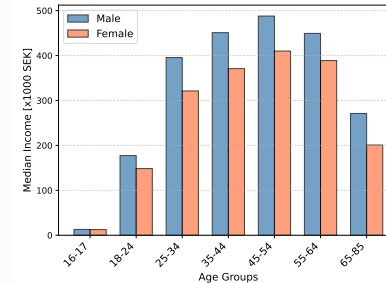
Synthetic population (1000 persons)

gender	age	income [kSEK]	n	mean daily hours			
				home	work	edu.	other
male	0–17	0–50	112	11.0	3.7	4.3	5.0
female	0–17	0–50	102	11.7	1.5	4.3	6.5
male	18–29	100–200	32	9.7	7.7	2.0	4.7
female	18–29	100–200	34	10.3	5.5	2.0	6.2
male	18–29	≥300	39	8.7	8.7	1.5	5.2
female	18–29	≥300	48	9.3	6.5	1.5	6.7
male	30–44	≥300	98	10.0	9.3	0.0	4.7
female	30–44	≥300	80	10.7	7.2	0.0	6.2
male	45–64	≥300	125	10.3	8.8	0.0	4.8
female	45–64	≥300	119	11.0	6.7	0.0	6.3
male	65–100	200–300	99	12.2	6.2	0.0	5.7
female	65–100	200–300	112	12.8	4.0	0.0	7.2

Population groups and their time allocation in the synthetic reality



Age distribution



Median income by age group and gender

Other applications of the underlying simulation framework

Vehicle-to-grid in Skellefteå, Sweden



- Explore travel/(un)charging episodes of a single representative traveler.
- Charging modeled by network expansion: each location i is split into two: i_{charging} and $i_{\text{no-charging}}$.
- $N = 1$, $\hat{J} = 4$, $K = 24$ (hours), $L = 7$.
- Sampling weights from minimalistic modeling assumptions:
 - Gravity model destination choice.
 - Uniform time choice within opening times.
 - Penalty for battery running empty.

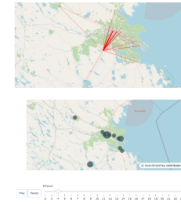
→ Gunnar's talk at hEART 2024.

Freight vehicle choice set sampling



- Construct truck round-trips between Swedish freight terminals. → Joint sampling of alternatives for many decision makers.
- Example: $N = 1e3$, $\hat{J} = K = 6$, 7 , $L = 329$, prior-info takes less than one hour to converge.
- Sampling weights:
 - Ensure connectivity of terminals.
 - Align with estimated shipment demand.
- Ongoing work for Swedish Transport Administration.

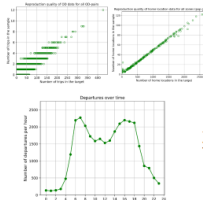
Drone fleet over Northern Sweden



- One central hub, Poisson distributed number of pickup/delivery requests per $2\text{km} \times 2\text{km}$ cell, to be served within-day.
- Drones can combine pickups and deliveries.
- $N = 1e3 \dots 1e4$, $\hat{J} = 100$, $K = 4 \cdot 24$, $L = 329$.
- Sampling weights:
 - satisfy demand
 - do not run out of battery
 - comply with local night flight bans
 - (avoid unnecessary trips)

Delivery vehicle charging; Drone deliveries; Freight modelling (G Flötteröd et. al)

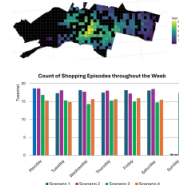
MATSim population from static OD matrix



- Create 24-h round trips for a representative Vienna population.
- $N = 1e4 \dots 1e5$, $\hat{J} = K = 24$, $L = 250$.
- Sampling weights:
 - Likelihood for static (Visum) OD reproduction.
 - Location choice informed by land use & population class data.
 - Departure time / activity duration informed by travel survey.

→ Franz' talk at MATSim user meeting, Session 3, 12/6, 17:00-17:20.

Weekly travel schedules in 15-minutes-city

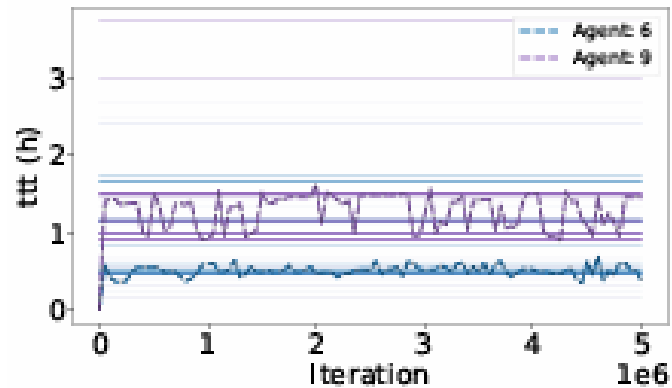


- Explore weekly activity/travel pattern in a 15-minutes city.
- Activity participation by network extension, one location duplicate per activity type (home, work, shop, other).
- $N = 7$ (days), $\hat{J} = 4$, $K = 24$, $L = 877$ (before expansion).
- Time use utility function a la MATSim for sampling weights.

→ Georgia's talk at Session 7A, 12/6, 10:00-10:30.

Disaggregation of VISUM OD matrices (Fx Rupprecht BOKU); 15-minute city (G Charalampidou, BOKU)

Frame Title



(b) Agent correspondence of ttt with observations

Frame Title

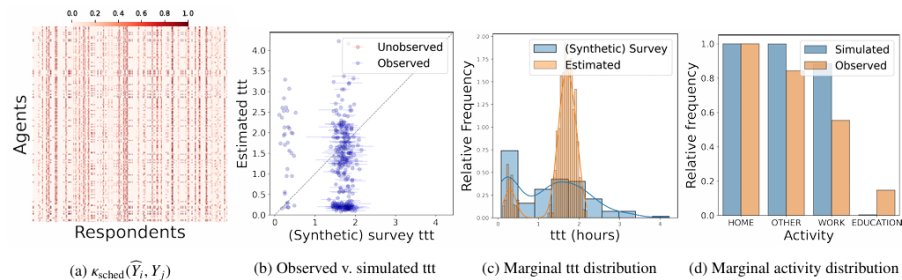


Fig. 2: Survey reproduction without traveler socio-demographics

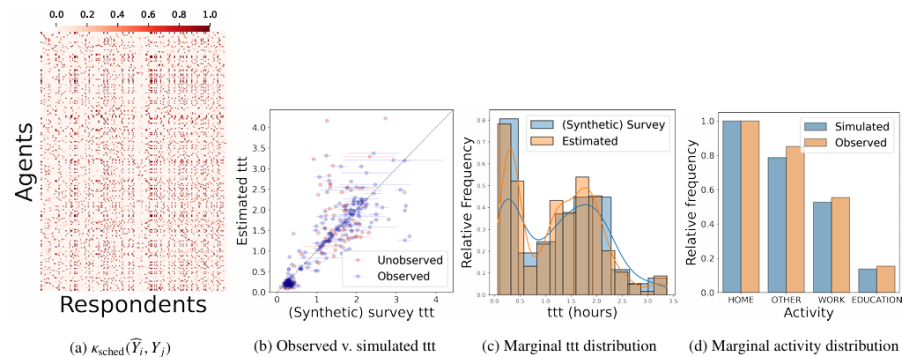


Fig. 3: Traveler kernel incorporated with 70% survey response rate