Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 280 (2026) 959–964

Procedia Computer Science

www.elsevier.com/locate/procedia

The 15th International Workshop on Agent-based Mobility, Traffic and Transportation Models, Methodologies and Applications (ABMTrans 2026)
April 14-16, 2026, Istanbul, Türkiye

A Kernel-Based Approach for Simulating Synthetic Travel Patterns

Michael Sederlin[a,b,*], Gunnar Flötteröd[a,b]

[a]*Linköping University, Department of Communications and Transport Systems, Bredgatan 33, 602 21, Norrköping, Sweden*
[b]*The Swedish National Road and Transport Research Institute, Malvinas Väg 6, 114 28, Stockholm, Sweden*

Abstract

Modeling dynamic traffic processes requires detailed traveler-level demand unavailable in aggregated national forecasting models. We present a kernel-based approach for generating synthetic disaggregate travel demand using travel survey data. The method links survey respondents to agents in a synthetic population through similarity kernels over socio-demographics, schedule characteristics, and activity–location feasibility, and embeds these in a Metropolis–Hastings sampling framework to generate complete daily plans. The approach provides a modular way to combine heterogeneous information sources and to transfer observed behaviour to new populations without requiring exact matches between individuals. We demonstrate the approach on a stylized grid-world and on real data, where survey records are mapped to a national synthetic population. Results show that the method reproduces key marginal travel patterns, preserves individual variability, and avoids collapse toward modal behaviour, illustrating its potential as a demand synthesis and data-fusion tool in dynamic traffic modelling.

© 2026 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer review under the responsibility of the scientific committee of the Program Chairs

Keywords: Kernel-methods; Synthetic travel demand; Metropolis-Hastings; Travel-Survey; Data-fusion

1. Introduction

Accurate modeling of dynamic urban traffic requires high resolution travel demand data. Congestion arises from overlapping, time dependent trips that static demand representations, such as temporally aggregated OD matrices, cannot capture. Meanwhile, national forecast tools, such as Sampers [7], estimate demand using aggregated data and model simplifications. Using these forecasts in congestion analysis, capacity assessment, or evaluation of time dependent mobility policies therefore requires disaggregation. This disaggregation must preserve observed mobility behavior and physical constraints, remain consistent with the underlying forecast assumptions, and accommodate changes in transport supply and land use. A natural way to do so is to model demand at the level of individual

* Corresponding author. Tel.: +46-8-555-365-11 ; fax: +0-000-000-0000.
E-mail address: michael.sederlin@vti.se

1877-0509 © 2026 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0)
Peer review under the responsibility of the scientific committee of the Program Chairs

travelers and their daily activity patterns, as in agent-based models such as MATSim [1], which model individual decision-making and travel behavior directly. Travel demand is represented as a schedule over the modeling horizon where movement arises from the utility gained from visiting different locations.

Existing work either reconstructs incomplete activity sequences for partially observed populations, or synthesizes synthetic populations and travel plans jointly. Cui et al. [2] impute missing activities in long-term GPS traces using a probabilistic multi-level framework that models visit probabilities from observed sequences, durations, and frequencies. Their approach requires individual-level traces and does not incorporate traveler attributes, limiting its suitability for synthetic populations. Joubert and De Waal [6] use Bayesian Networks to generate trip–activity–time sequences conditioned on socio-demographics, but do not model the spatial dimension and therefore cannot incorporate network locations or aggregated flow information. Tozluoğlu et al. [9] combine mobility surveys with mobile phone location data to synthesize travel patterns for phone users using a generative model. Similarly, Zilske and Nagel [10] construct full-day plans for mobile phone users by augmenting call records with traffic counts within a traffic simulation calibration framework. More generally, Hörl and Balac [5] present an open and reproducible pipeline for joint population and daily travel schedule synthesis (demonstrated for Paris), relying on a sequence of deterministic and randomized steps rather than a unified likelihood formulation.

We propose a kernel-based framework for incorporating individual-level survey information when simulating agent travel plans. Kernel similarity measures combine socio-demographics, trip characteristics, and activity—location information, enabling flexible comparison of heterogeneous objects (travelers, trips, zones) across feature spaces. Since kernel functions remain valid under addition or multiplication, corresponding to OR and AND logic [3], the framework can represent multiple behavioural dimensions jointly while remaining interpretable. This is well-suited to travel-demand synthesis, where cross-domain comparisons are required. The framework builds on the Metropolis-Hastings (MH) travel plan simulator introduced in Flötteröd [4], with inter-zonal OD information incorporated in Rupprecht [8].

2. Method

We consider a given existing synthetic population of size n with socio-demographics $\widehat{X}$ and unobserved travel plans $\widehat{Y}$, and a survey of size $m \leq n$ (typically $m \ll n$) providing attributes X and observed travel plans Y. A travel plan refers to the movements of an individual over an entire day including visited locations and departure/arrival-times. We aim to infer $\widehat{Y}$ by sampling from the target density $\pi(\widehat{Y} \mid \widehat{X}, X, Y)$, transferring observed respondent behavior to unobserved agent behaviours.

We use the MH sampler described in [4] and keep its proposal mechanism unchanged. We require that all trips are roundtrips, starting and ending at the same location. A state corresponds to a set of roundtrips $\widehat{Y}$ for all n agents. At each iteration a proposal $\widehat{Y}'$ is generated by modifying a small subset of roundtrips. Since the MH acceptance probability only depends on the ratio of target densities, normalizing constants cancel. We therefore specify the target density up to proportionality as

$$\pi(\widehat{Y} \mid \widehat{X}, X, Y) \sim W(\widehat{Y}; \widehat{X}, X, Y) P(\widehat{Y} \mid \widehat{X}), \tag{1}$$

where "~" means "equal up to normalization". Here $P(\widehat{Y} \mid \widehat{X})$ constitutes a behavioural prior that weights the density in the absence of survey information. The weights $W(\widehat{Y}; \widehat{X}, X, Y)$ are survey-based compatibility weights used in the MH-sampling scheme. We define the weights using a 'product-of-agents' formulation

$$W(\widehat{Y}; \widehat{X}, X, Y) = \prod_{i=1}^{n} \sum_{j=1}^{m} \kappa((\widehat{X}_i, \widehat{Y}_i), (X_j, Y_j)), \tag{2}$$

where $\kappa((\widehat{X}_i, \widehat{Y}_i), (X_j, Y_j))$ is a soft-matching kernel function measuring how well agent i aligns with respondent j. For each agent i, the inner sum goes over all respondents, meaning that every agent is matched to every respondent. The product over i factorizes the weight W over the full set of roundtrips into per-agent contributions.

Construction of the kernel function

Let $\widehat{X}_i$ and X_j be the socio-demographic attributes of synthetic agent i and survey respondent j, respectively. Let $\widehat{Y}_i$ and Y_j be their simulated and reported travel plans, respectively. We evaluate the similarity between agent i and survey record j using the composite kernel

$$\kappa\Big((\widehat{X}_i, \widehat{Y}_i), (X_j, Y_j)\Big) = \kappa_{\text{trav}}(\widehat{X}_i, X_j)\ \kappa_{\text{sched}}(\widehat{Y}_i, Y_j)\ \kappa_{\text{loc}}(\widehat{Y}_i, Y_j), \tag{3}$$

where $\kappa_{\text{trav}}(\widehat{X}_i, X_j)$ is a *traveler kernel* comparing socio-demographic attributes ($\widehat{X}_i$ vs. X_j), $\kappa_{\text{sched}}(\widehat{Y}_i, Y_j)$ is a *schedule kernel* comparing travel-plan summary characteristics ($\widehat{Y}_i$ vs. Y_j), and $\kappa_{\text{loc}}(\widehat{Y}_i, Y_j)$ is an *activity-location kernel* comparing reported activities in Y_j with the locations visited in $\widehat{Y}_i$.

Traveler Kernel. In these experiments, similarity between travelers is computed as a product of Exponential Kernels measuring deviations in income and age as well as a fixed penalty for individuals of different genders as

$$\kappa_{\text{trav}}(\widehat{X}_i, X_j)(\widehat{X}_i, X_j) = \exp\left(-\frac{(\Delta\text{inc}_{ij})^2}{2\ell_{\text{inc}}^2} - \frac{(\Delta\text{age}_{ij})^2}{2\ell_{\text{age}}^2}\right) \eta^{\mathbb{1}\{\text{gender}(\widehat{X}_i)\neq\text{gender}(X_j)\}} \tag{4}$$

where $\mathbb{1}\{\cdot\}$ is the 0/1 indicator function. The length scales ℓ_{inc}, ℓ_{age} and mismatch factor η each control the strictness of the matching for their respective dimensions.

Schedule Kernel. The schedule kernel represents the similarity between a given respondent's schedule and a simulated trip chain. This is done in an aggregated manner taking into account the total travel time *ttt* and the number of locations $nStops$ visited in a day as follows:

$$\kappa_{\text{sched}}(\widehat{Y}_i, Y_j) = \exp\left(-\frac{(\Delta ttt_{ij})^2}{2\ell_{ttt}^2}\right) \exp\left(-\frac{|\Delta nStops_{ij}|}{\ell_{\text{nStops}}}\right) \tag{5}$$

where Δttt_{ij} and $\Delta nStops_{ij}$ are the differences between total travel times and number of stops between agent i and respondent j.

Location-Activity Kernel. Since we consider soft matching between respondents and agents, we do not rely on direct correspondence and do not require simulated round trips to visit the observed *locations*. Instead, the location-activity kernel presented here compares the activities in observed trip Y_j with the locations l visited in simulated round trip $\widehat{Y}_i$ to ensure feasibility.

Let $L(\widehat{Y}_i)$ denote the locations visited in the simulated round trip $\widehat{Y}_i$. Each location l has a set of possible activities A^l, and we define the set of feasible activities at the locations visited by agent i as $A_i^{\text{sim}} = \bigcup_{l\in L(\widehat{Y}_i)} A^l$. A_j^{obs} are the reported activities of respondent j. A simple similarity measure is the Jaccard kernel (or Jaccard index):

$$J(A_i^{\text{sim}}, A_j^{\text{obs}}) = \frac{|A_i^{\text{sim}} \cap A_j^{\text{obs}}|}{|A_i^{\text{sim}} \cup A_j^{\text{obs}}|}. \tag{6}$$

To control the sharpness of the matching, we apply the exponential transformation

$$\kappa_{\text{loc}}(\widehat{Y}_i, Y_j) = \exp\left\{-\frac{1 - J(A_i^{\text{sim}}, A_j^{\text{obs}})}{\lambda}\right\}, \quad \lambda > 0 \tag{7}$$

which penalizes mismatches more strongly for smaller λ.

Opportunity weighted activity-location kernel. To illustrate how we can modify or add dimensions to the kernel functions, we consider that the Jaccard kernel does not weight certain activity-location-pairs higher than others, a zone with more work opportunities has the same *pull* as one with fewer opportunities. To achieve this we can define the score function

$$S\left(L(\widehat{Y}_i), A_j^{\text{obs}}\right) = \sum_{l \in L(\widehat{Y}_i)} \sum_{a \in A_j^{\text{obs}}} \alpha_{l,a} \mathbb{1}\left\{a \in A^l\right\} \tag{8}$$

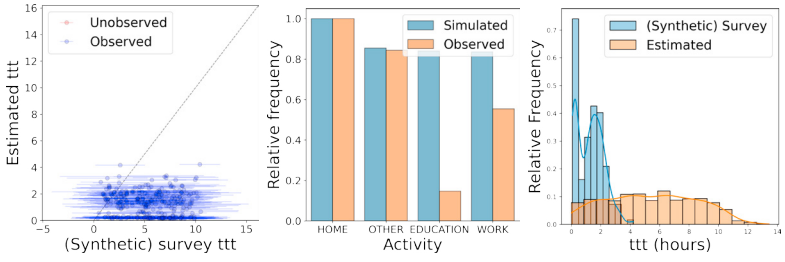
where $\alpha_{l,a}$ represents the attractiveness for performing activity a at location l. As before, the sharpness of this score is controlled using the same transformation as Equation (7). Similar extensions can incorporate factors such as the time spending, ordering of activities, or other structural characteristics of activity–location pairs, as long as they can be expressed through pair-wise scores that are cheap to compute.

3. Computational Experiments

We demonstrate the model on two datasets: a synthetic grid world scenario with 300 travelers and four activity types (EDUCATION, HOME, WORK, and OTHER), and the Swedish household travel survey (national travel survey (RVU)) mapped to a synthetic population used in policy analysis. Each traveler receives an age, income and gender, based on Swedish population with characteristics. Mobility behavior is simulated using a round trip generating model that reacts to socio-demographic, producing synthetic survey data. We assess marginal and individual-level reproduction under different kernel configurations. We report statistics for 5 retained MH states, spaced to mitigate autocorrelation in the Markov chain.

Grid-World

Figure 1 shows the marginal distributions and individual-level statistics generated from the behavioral prior alone, i.e. without applying the survey-based kernel weights.

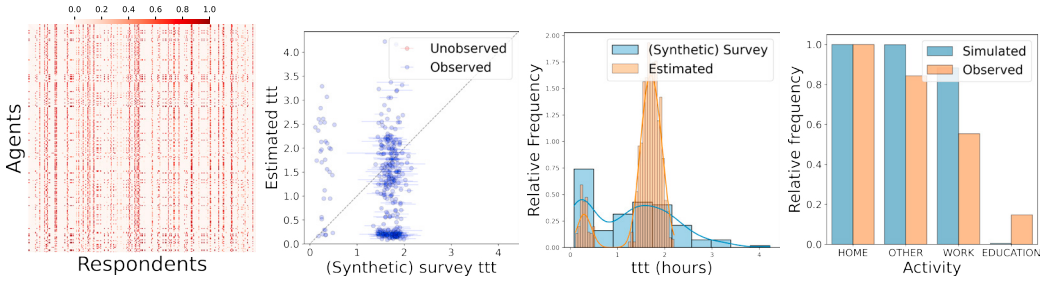


(a) Individual reproduction of ttt (b) Activity frequencies (c) Correspondence of marginal ttt distribution

Fig. 1: Simulations from prior

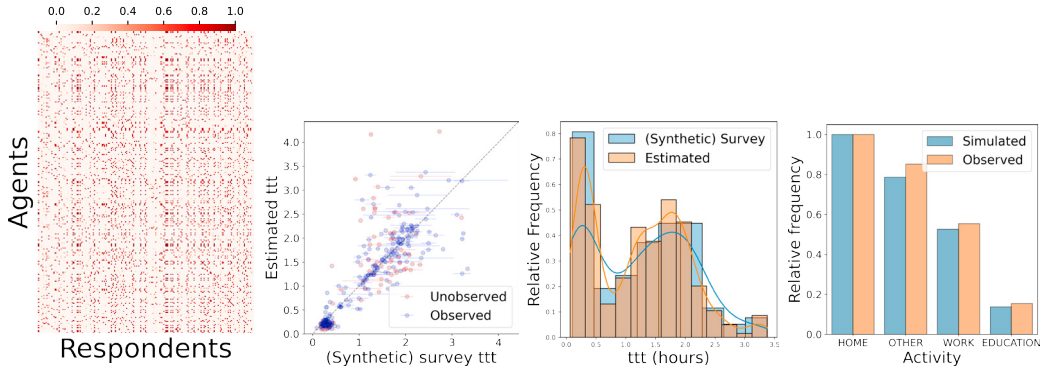
When including kernel weights, we examine the effect of socio-demographics by estimating travel patterns with and without the traveler kernel. Schedule kernel ($\kappa_{\text{sched}}(\widehat{Y}_i, Y_j)$) similarity between agents and respondents is shown in

Figures 2a and 3a, where darker colors indicate higher correspondence. The traveler kernel acts as a population filter by downweighting schedule similarities of socio-demographically dissimilar agent-respondent pairs. As a result, the acceptance ratio becomes more sensitive to schedule changes that increase similarity to socio-demographically similar respondents, as indicated by the sharpened diagonal in Figure 3a. Figure 2b shows the individual-level correspondence between observed and simulated time spent traveling. In Figure 3, we introduce the traveler kernel $\kappa_{\text{trav}}(\widehat{X}_i, X_j)$ and mask observed travel information for 30% of travelers to assess individual- and marginal-level fit for unobserved travelers. Figure 3b shows a clear positive association between observed and estimated travel time, with good agreement at low travel times and higher dispersion at medium and high travel times. Horizontal bars show the 2σ variation across the retained MH states. Figures 2c and 2d show the marginals collapsing towards the modes of the survey distribution and failing to represent the tails. In Figure 3c we see improved reproduction of the tails of the marginal distributions.



(a) $\kappa_{\text{sched}}(\widehat{Y}_i, Y_j)$ (b) Observed v. simulated ttt (c) Marginal ttt distribution (d) Marginal activity distribution

Fig. 2: Survey reproduction without traveler socio-demographics

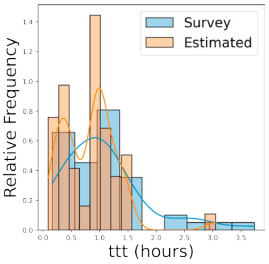


(a) $\kappa_{\text{sched}}(\widehat{Y}_i, Y_j)$ (b) Observed v. simulated ttt (c) Marginal ttt distribution (d) Marginal activity distribution

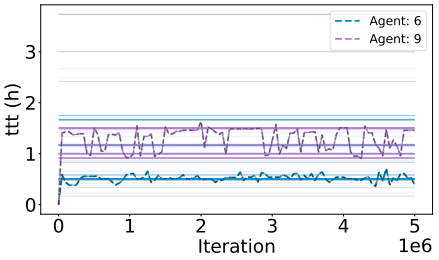
Fig. 3: Traveler kernel incorporated with 70% survey response rate

Real world data

We apply the method to combine survey data with a synthetic population using 50 records from the Swedish household travel survey (RVU) and 200 agents from the synthetic population used for national modelling in Sweden. The RVU provides 45 trip purposes, here grouped into the six categories: WORK, EDUCATION, SERVICES, SOCIAL, LEISURE, and OTHER. Figure 4a shows correspondence of marginal travel times. Figure 4b compares selected agents with survey respondents, with each solid horizontal line representing a single survey respondent, and line opacity and thickness representing traveler-kernel similarity. The x-axis shows the MH iteration index. After an initial burn-in, the chain targets the stationary distribution and successive iterations can be interpreted as samples from the target density. Dashed lines show estimated behaviour of two selected synthetic agents. When a single respondent clearly dominates the traveler-kernel weights, the agent's estimated mobility behaviour remains close to that of the respondent. When an agent has comparable traveler-kernel similarity with multiple respondents each having different mobility behaviour, the simulated mobility behaviour of accepted proposals alternates between these respondents. Sampling across iterations of the Markov chain therefore allows within-group variability to be represented.



(a) Simulated and observed marginal ttt distribution



(b) Agent correspondence of ttt with observations

Fig. 4: Survey characteristics in synthetic population

4. Conclusions and Outlook

We present a kernel-based approach for incorporating mobility information on the level of individual survey respondents into travel plan synthesis. Similarity is defined across three dimensions: traveler attributes, trip characteristics, and activity–location correspondence. Our experiments with synthetic data show that segmentation by traveler attributes improves prediction at both individual and aggregate levels. The traveler kernel steers similarity in other dimensions and prevents collapse toward modal behavior. With partially observed travelers, the method imputes travel plans using this segmentation and produces realistic trip distributions. Applied to real survey data and a synthetic population, the individual-level kernels reproduce marginal patterns and activity structures although there is no perfect match between survey respondents and the synthetic population used in estimation, demonstrating the value of soft matching. Ongoing work focuses on implementation of kernels to capture different aspects of land-use and survey responses, upscale simulations, improve calibration of kernel hyper-parameters, and incorporate additional data sources such as static origin/destination matrices.

Statements and Declarations

Data Availability: The synthetic population and code for the simulated scenarios are available on request. Travel survey data are under access restrictions. Code for the sampling algorithm is available at https://github.com/vtisweden/java-projects/tree/master/roundtrips.

Competing Interests: The authors have no relevant financial or non-financial interests to disclose.

Funding: This research was funded by the Swedish Transport Administration (Trafikverket), grant TRV 2023/33572.

References

[1] Axhausen, K.W., ETH Zürich, 2016. The Multi-Agent Transport Simulation MATSim. Ubiquity Press. doi:10.5334/baw.

[2] Cui, Y., He, Q., Bian, L., 2021. Generating a synthetic probabilistic daily activity-location schedule using large-scale, long-term and low-frequency smartphone GPS data with limited activity information. Transportation Research Part C: Emerging Technologies 132, 103408. doi:10.1016/j.trc.2021.103408.

[3] Duvenaud, D., 2014. Automatic Model Construction with Gaussian Processes. Ph.D. thesis. University of Cambridge.

[4] Flötteröd, G., 2025. An operational alternative to origin/destination matrices, in: hEART 2025, Aalto University, Helsinki, Finland.

[5] Hörl, S., Balac, M., 2021. Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data. Transportation Research Part C: Emerging Technologies 130, 103291. doi:10.1016/j.trc.2021.103291.

[6] Joubert, J.W., De Waal, A., 2020. Activity-based travel demand generation using Bayesian networks. Transportation Research Part C: Emerging Technologies 120, 102804. doi:10.1016/j.trc.2020.102804.

[7] Kristoffersson, I., Berglund, S., Samuelsson, S., Almström, P., Algers, S., 2018. Sampers4 - Skattning av regionala efterfrågemodeller. Technical Report.

[8] Rupprecht, F., 2024. Metropolis-Hastings Based Synthesis of All-Day Round Trips from Origin-Destination Matrices and Limited Population Segmentation. Master's thesis. Technische Universität Wien.

[9] Tozluoğlu, Ç., Liao, Y., Sprei, F., 2024. Mobile Phone Application Data for Activity Plan Generation. arXiv:2410.22386.

[10] Zilske, M., Nagel, K., 2015. A Simulation-based Approach for Constructing All-day Travel Chains from Mobile Phone Data. Procedia Computer Science 52, 468–475. doi:10.1016/j.procs.2015.05.017.