

Generating Realistic Individual Activity Schedules via Activity Location Allocation Based on Simulated Travel Times

Tatsuya Mitomi ✉ 

AI for Collective Intelligence Hub, University of Glasgow, Glasgow, UK
Fujitsu Limited, Kawasaki, Japan

Yahya Gamal ✉ 

AI for Collective Intelligence Hub, University of Glasgow, Glasgow, UK

Esra Suel ✉ 

Urban Analytics, Department of Geography, University of Zurich, Zurich, Switzerland

Gary Polhill ✉ 

Information and Computational Sciences Department, The James Hutton Institute, Aberdeen, UK

Daniel Konioukhov ✉ 

AI for Collective Intelligence Hub, University of Glasgow, Glasgow, UK

Alison Heppenstall ✉ 

AI for Collective Intelligence Hub, University of Glasgow, Glasgow, UK

Abstract

Individual level daily activity schedules are essential for a wide range of applications, including infectious disease control, urban transportation planning, and policy design. In practice, such schedules are typically generated by combining population data with travel survey data. These data sources are used because they are often publicly available, whereas observed individual activity schedules are difficult to obtain due to privacy concerns. However, because of the complexity of mobility modelling, it is difficult to generate realistic activity schedules that also preserve travel times consistent with those reported in travel surveys. To address this issue, we propose a framework for generating activity schedules that iteratively applies a dynamic programming method to allocate activity locations based on simulated travel times. Numerical experiments with dummy data show that the proposed method reduces the discrepancy between simulated travel times and those reported in travel surveys by 52.2% relative to the first iteration through iterative refinement.

2012 ACM Subject Classification Computing methodologies → Modeling and simulation

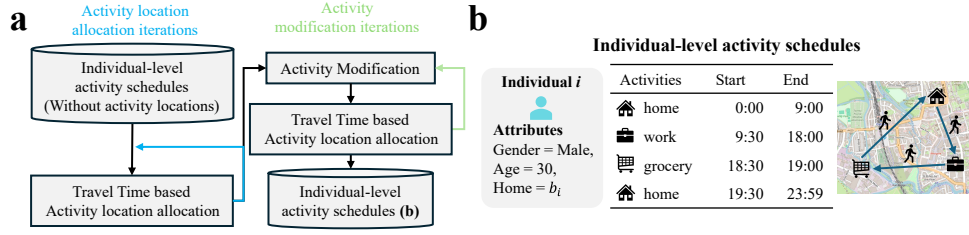
Keywords and phrases Activity based modelling, Synthetic population

1 Introduction

Datasets of individual level daily activity schedules, together with sociodemographic attributes such as gender and age, are critical for a wide range of applications, including infectious disease control, urban transport planning, and policy formulation [11]. However, such datasets are difficult to obtain in practice due to privacy concerns and limited observability. These data are typically generated using other observable data sources.

To this end, several studies have proposed methods for generating individual level activity schedules (Figure 1b) using population data and travel survey data [6, 7]. These generation methods often incorporate observed static travel times between location pairs rather than mobility simulation-based travel times. However, reliance on static travel times may overlook

This is the author version of a short paper accepted for presentation in the poster session at the 17th Conference on Spatial Information Theory (COSIT 2026).



■ **Figure 1** Overview of the proposed framework. (a) Overall workflow of the method. The location allocation iteration is the process proposed in this study, whereas the activity modification iteration extends an existing activity modification framework [5]. (b) Example individual-level activity schedules.

uncertainty in activity locations arising from the complex and dynamic nature of traffic conditions. Because activity locations determine an individual’s travel path, they also influence congestion patterns and travel times – these are often unaccounted for by existing methods. Therefore, relying solely on static travel time data may lead to unrealistic activity locations and, consequently, implausible travel times that do not match those observed in the travel survey. Since activity locations are endogenous to the generation of daily activities, they can be modified by using mobility simulation-based travel time in order to generate more realistic individual level activity schedules.

To address this issue, we propose a framework for generating activity schedules that combines iterative refinement with a dynamic programming method for allocating activity locations based on simulated travel times. Additionally, we investigate how this framework can be integrated into an existing activity modification framework [5] to further improve the matching of travel times between the simulated and the travel survey data. Numerical experiments with dummy data demonstrate that the proposed framework can generate activity schedules and, through iterative refinement, reduce the discrepancy between simulated and travel survey data by 52.2%. These results demonstrate the potential of this approach to produce realistic individual level activity schedules, enabling more accurate simulations for transport planning.

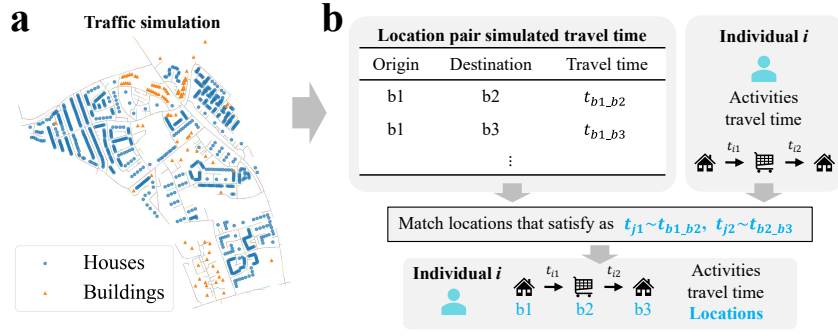
2 Method

We use population data and travel survey data as the starting point, as these data sources are often publicly available as open data [10, 3, 6]. As a first step, a statistical matching procedure is applied to combine the population data with the travel survey data [6]. This matched dataset is used as the input to the proposed method. In general, population data include individual attributes such as gender, age and home locations, whereas travel survey data include daily activity types, start times and durations, as well as travel times and travel modes between activities. As a result, the combined dataset contains all the necessary information to generate activity schedules except activity locations.

The specific objective of this framework is to estimate activity locations in the combined dataset to ensure that simulated travel times between activities match those observed in the travel survey data. Figure 1 illustrates the overall workflow of the proposed method.

2.1 Simulated Travel Time Based Activity Location Allocation

We introduce a dynamic programming based procedure for activity location allocation, similar to the Viterbi algorithm on a weighted graph [9], where dynamic programming refers to a



■ **Figure 2** Overview of the simulated travel time based activity location allocation. (a) Example traffic simulation on a map of Aberdeen, UK, used to pilot the application of the proposed method. (b) Overview of activity location allocation using simulated travel times between location pairs.

method for solving the allocation problem through sequential multistage decomposition [2]. This procedure computes a closed path of locations $(l_0, \dots, l_T)$, such that the simulated travel times along the route are as close as possible to the survey travel times $(dur_1, \dots, dur_T)$ (Figure 2). The algorithm is deterministic and it always identifies a single best route among all feasible route combinations. To compute this route, the algorithm requires simulated travel times and individual specific travel survey information. For the simulated travel times, let V denote the set of all locations on a map, and let $W = \{(u, v) \mid u, v \in V, u \neq v\}$ denote the set of paths between pairs of locations, with an associated travel time function $w : W \rightarrow \mathbb{R}$ such that $w(u, v) = w_{uv}$. Regarding individual specific information, the inputs are: the individual's activity schedule $a_1, \dots, a_T$, the survey travel times between these activities $dur_1, \dots, dur_T$, and the individual's home location l_{home} . Algorithm 1 presents this procedure. In the initial step of the algorithm, simulated travel times cannot be used because the traffic simulation requires activity locations (i.e., the origins and destinations of individual trips), which are themselves generated by this algorithm. Therefore, in the initial step, we use travel times between locations calculated based on the speed limits assigned to each road segment instead of simulated travel times.

We then apply an iterative procedure consisting of the activity location allocation and traffic simulation (Figure 1a) to address the uncertainties that arise from activity locations. These locations determine the individuals' travel paths, which in turn affect congestion patterns and travel times. By iterating between the activity location allocation and traffic simulation, the simulated travel times associated with the matched locations are repeatedly fed back into the activity location allocation process. This ensures consistency between the travel times reported in the travel survey and those produced in the simulation.

2.2 Sociodemographic based Activity Modification

In addition to introducing a dynamic programming procedure for activity location allocation, we investigate how the proposed method can be integrated into an existing activity modification framework [5]. Specifically, this framework swaps activities between pairs of individuals with similar sociodemographic attributes to enable more flexible activity location allocation. This allows for flexible activity location allocation while preserving the empirical activity sequences and their distributions within each sociodemographic cluster. This framework is extended and integrated into the methodology in this work.

The extended activity modification framework first follows the existing method to identify candidate pairs of individuals with similar sociodemographic attributes based on the *attribute space difference* [5]. In contrast to the existing activity modification framework, our approach

■ **Algorithm 1** Viterbi algorithm based activity location allocation

Require: Map travel-time function W with edge travel times w_{uv} ; activity schedule $a_1, \dots, a_T$; home location l_{home} ; observed travel times $dur_1, \dots, dur_T$

```

1: Initialize  $DP[t][v] \leftarrow \infty$  for all  $t, v$  ▷  $DP[t][v]$ : minimum cost to reach node  $v$  at step  $t$ 
2:  $DP[0][l_{\text{home}}] \leftarrow 0$ 
3: Initialize  $Back[t][v]$  for all  $t, v$  ▷  $Back[t][v]$ : previous node before arriving at  $v$  at step  $t$ 
4: Initialize  $Time[t][v]$  for all  $t, v$  ▷  $Time[t][v]$ : edge travel time used to arrive at  $v$  at step  $t$ 
5: for  $t = 0$  to  $T - 1$  do
6:    $d \leftarrow dur_{t+1}$ 
7:   for all  $v \in V$  do
8:     if  $DP[t][v] = \infty$  then
9:       continue
10:    end if
11:    for all  $(v, u) \in W$  do
12:       $cost \leftarrow DP[t][v] + (d - w_{vu})^2$ 
13:      if  $cost < DP[t+1][u]$  then
14:         $DP[t+1][u] \leftarrow cost$ 
15:         $Back[t+1][u] \leftarrow v$ 
16:         $Time[t+1][u] \leftarrow w_{vu}$ 
17:      end if
18:    end for
19:  end for
20: end for
21:  $l_T \leftarrow l_{\text{home}}$ 
22:  $curLoc \leftarrow l_{\text{home}}$ 
23: for  $t = T$  down to  $1$  do
24:    $prevLoc \leftarrow Back[t][curLoc]$ 
25:    $l_{t-1} \leftarrow prevLoc$ 
26:    $curLoc \leftarrow prevLoc$ 
27: end for
28: for  $t = 1$  to  $T$  do
29:    $dur_t^{sim} \leftarrow Time[t][l_t]$ 
30: end for
31: return  $(l_0, \dots, l_T)$  and  $(dur_1^{sim}, \dots, dur_T^{sim})$ 

```

does not account for household structure or the geospatial locations of candidates undergoing activity switching. The procedure then calculates, for each candidate pair (i, j) , the error function defined below for both the original pair and the pair obtained by swapping their activity sequences, activity start times and survey travel times. The resulting errors are denoted by $e_{i,j}^{orig}$ and $e_{i,j}^{swapped}$, respectively. Unlike the function used in the existing activity modification framework, this study defines a new error function. The error function is defined as follows:

$$e_{i,j} = \sum_{t \in T} (dur_{i,t}^{sim} - dur_{i,t})^2 + \sum_{t \in T} (dur_{j,t}^{sim} - dur_{j,t})^2. \quad (1)$$

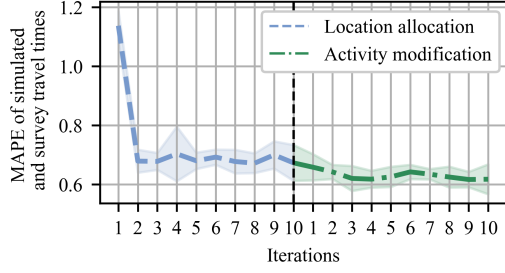
Subsequently, for each pair of individuals satisfying the following condition, their activity sequences, activity start times and survey travel times are swapped:

$$e_{i,j}^{swapped} < e_{i,j}^{orig}. \quad (2)$$

The framework also employs an iterative procedure consisting of traffic simulation, activity modification and activity location allocation, as shown in Figure 1a.

2.3 Traffic Simulation

We deploy the agent-based traffic simulation framework MATraM [4]. This simulation framework models a population of individuals moving within a geospatial environment that



Method	Iter.	MAPE
Location allocation	1st	1.14
Location allocation	10th	0.673
Activity modification	1st	0.658
Activity modification	10th	0.618

■ **Figure 3** Change in the mean absolute percentage error (MAPE) between simulated and survey travel times, shown as the mean over five runs, with a band indicating the minimum and maximum values. The table reports the MAPE values between simulated and survey travel times at specific iterations.

includes a road network and building locations. Individuals execute daily activity schedules, travel between activity locations, and interact indirectly through congestion effects. Each activity is assigned to a building and trips are defined as origin-destination pairs between consecutive activities. In this study, we use a version of MATraM in which only car agents are simulated.

3 Pilot Application

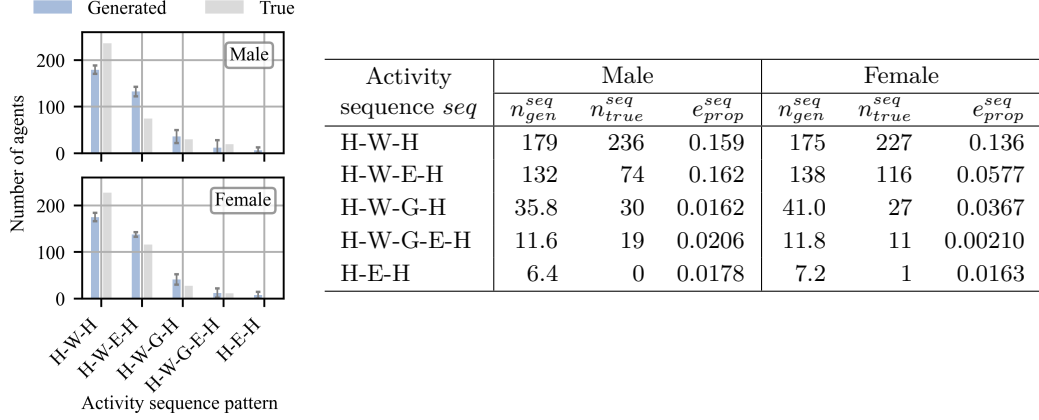
3.1 Data and Traffic Simulation Settings

The proposed method is verified using dummy activity schedules and a dummy population, together with a traffic map of Aberdeen, Scotland (Figure 2a), downloaded from OpenStreet-Map [8]. In these dummy activity schedules and population, sociodemographic attributes are sampled from empirical distributions for employment status (unemployment, full-time employment, and part-time employment), gender and age, which are based on the National Travel Survey [3] in England and Wales. For the dummy activity schedules, we adopt the following assumptions for data generation: (1) Activity sequences, activity types, start times, and durations are sampled from the corresponding empirical distributions. The activity types considered are **Work**, **Home**, **Grocery**, and **Entertainment**. (2) Travel times between activity locations are drawn from a lognormal distribution $\text{Lognormal}(\mu, \sigma)$, where $\sigma^2 = \ln(1 + \frac{0.005}{0.08^2})$ and $\mu = \ln(0.2) - \frac{\sigma^2}{2}$. For simplicity, vehicles are assumed to travel along road segments at empirically specified speed limits. The number of iterations for both activity location allocation and activity modification is set to 10 each. The experiment is repeated five times. All experiments are performed on a machine equipped with an Intel Core Ultra 7 165H CPU (3.10 GHz) and 32 GB of RAM.

3.2 Results

3.2.1 Travel Time Matching

Figure 3 illustrates the change in mean absolute percentage error (MAPE) between simulated and survey travel times over every iteration, aggregated across all trips and individuals. The results for activity location allocation indicate a 46.7% reduction in MAPE from the first to the tenth iteration. This shows that the proposed activity location allocation process achieves better alignment of activity locations in terms of travel time. In addition, the results



■ **Figure 4** Number of individuals by activity sequence pattern for full-time-employed males and females aged 18–75, after 10 iterations of activity location allocation and 10 iterations of activity modification. The generated data are shown as the mean over five runs, with bands indicating the minimum and maximum values. H means **Home**, W means **Work**, E means **Entertainment**, and G means **Grocery**.

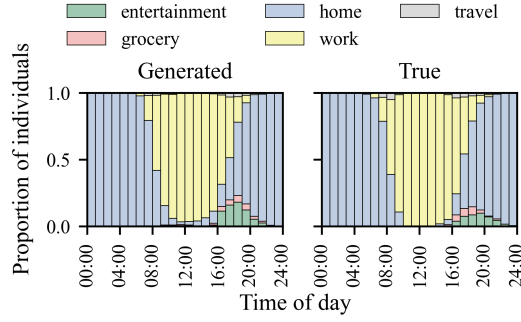
for activity modification indicate a further 5.5% reduction in MAPE. Overall, the proposed method achieves a total 52.2% reduction in MAPE, confirming that integrating activity modification with the proposed activity location allocation method further improves the consistency between simulated and survey travel times.

3.2.2 Sociodemographic Attributes

Although the extended activity modification framework enables flexible activity location allocation, it can affect the relationship between sociodemographic attributes and activity sequence patterns. Therefore, we analyze the effect of the modification framework on this relationship.

Figure 4 shows an example of the number of individuals by activity sequence pattern for full-time-employed males and females aged 18–75, after 10 iterations of activity location allocation and 10 iterations of activity modification. We calculate the proportional error of activity sequence seq as $e_{prop}^{seq} = \frac{1}{n_{cat}} |n_{gen}^{seq} - n_{true}^{seq}|$ to assess distortion in activity sequence patterns, where n_{cat} shows the total number of individuals in each category for male and female, n_{gen}^{seq} and n_{true}^{seq} show the number of individuals in each type of activity sequence seq for the generated activity schedule and the true travel survey data, respectively. The generated data show a distortion in activity sequence patterns with a proportional error of up to 16.2% for males and 13.6% for females. This is expected because activity modification alters the activity sequences. Nevertheless, the distortion remains limited as the ranking of the most common activity sequences within each category is preserved.

Figure 5 presents an example of the activity distribution across time slots for full-time-employed males aged 18–75. We calculate the proportional error as $e_{prop}^{act} = \sum_t |prop_{gen}^{act,t} - prop_{true}^{act,t}|$ to assess distortion in the activity distribution, where $prop_{gen}^{act,t}$ and $prop_{true}^{act,t}$ show the proportion of individuals who engage in activity act at time of day t in the generated activity schedule and the true travel survey data, respectively. A comparison between the generated data and the travel survey data indicates that the temporal activity distribution is not substantially altered, with a proportional error of 2.0% on average.



Activity act	e_{prop}^{act}
Home	0.032
Work	0.027
Grocery	0.0081
Entertainment	0.015
Overall average	0.020

Figure 5 Activity distribution across time slots for full-time-employed males aged 18–75, after 10 iterations of activity location allocation and 10 iterations of activity modification. The generated data are shown as the mean over five runs. The table reports the proportional error $e_{prop}^{act} = \sum_t |prop_{gen}^{act,t} - prop_{true}^{act,t}|$.

4 Discussion

The results show that, through iterative refinement, the first part of the proposed simulated travel time based activity location allocation can reproduce realistic activity schedules while improving the matching between survey travel times and those generated by the traffic simulation by 46.7% relative to the first iteration and incurs no distortion of activity sequence pattern. A further, 5.5% improvement is gained by applying a simplified version of an existing activity modification framework [5] – limited to activity switching between individuals, rather than households, without considering the geospatial locations of switching candidates. However, the application of such framework incurs a proportional error in activity sequence patterns of up to 16.2% in certain categories. Nonetheless, the whole process preserves the characteristic activity sequences of individuals within specific sociodemographic groups, with an average change in proportion limited to 2.0%.

As future work, the proposed method should be applied while considering households during activity switching. The representation of errors to flag activity modifications should be extended to indicate the geospatial areas where activity switching will lead to minimizing errors [5]. Further, the proposed method should be validated using real world data. Previous studies have validated similar approaches using commuting origin–destination matrices, real world population data, and travel survey data, confirming the accuracy of activity schedule methods in real world case studies [6]. Another important issue is computational cost. The location matching algorithm (Algorithm 1) has a time complexity of $O(TV_n^2)$, where T denotes the length of an individual’s activity sequence and V_n denotes the number of locations on the map. This computational burden may become problematic when the method is applied to larger maps or large scale simulations with many individuals. One possible solution is beam search, a heuristic method that prunes unlikely candidates during the search process. Because beam search is widely used to reduce the search space of the Viterbi algorithm and lower computational costs [1], incorporating it into Algorithm 1 may improve scalability.

References

- 1 Sherif Abdou and Michael S. Scordilis. Beam search pruning in speech recognition using a posterior probability-based confidence measure. *Speech Communication*, 42(3):409–428, 2004. doi:10.1016/j.specom.2003.11.002.
- 2 Richard Bellman. *The Theory of Dynamic Programming*, volume 60. 1954. doi:10.1090/S0002-9904-1954-09848-8.
- 3 Department for Transport. National travel survey, 2002–2024. UK Data Service, 2026. 19th Edition. SN: 5340. doi:10.5255/UKDA-SN-5340-15.
- 4 Yahya Gamal, Ricardo Colasanti, Gary Polhill, Tatsuya Mitomi, Esra Suel, and Alison Heppenstall. Matram: A multi-activity transport and mobility agent-based model for activity modifications, 2026. URL: <https://arxiv.org/abs/2605.30547>, arXiv:2605.30547.
- 5 Yahya Gamal, Tatsuya Mitomi, Alison Heppenstall, Esra Suel, and Gary Polhill. A geospatial activity modification framework for calibrating mobility models. In *The Geospatial Artificial Intelligence (GeoAI)*. Zenodo, June 2026. doi:10.5281/zenodo.20542665.
- 6 Sebastian Hörl and Milos Balac. 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, 2021. doi:10.1016/j.trc.2021.103291.
- 7 Hussein Mahfouz, Sam F. Greenbury, Bowen Zhang, Stuart Lynn, and Tao Cheng. A reproducible pipeline for activity-based travel demand generation in england. *Environment and Planning B: Urban Analytics and City Science*, 52(9):2326–2339, 2025. doi:10.1177/23998083251379620.
- 8 OpenStreetMap contributors. Planet dump retrieved from <https://planet.osm.org> . <https://www.openstreetmap.org>, 2017.
- 9 L.R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989. doi:10.1109/5.18626.
- 10 Hugh P. Rice, Andreas Höhn, Petra Meier, Alison Heppenstall, and Nik Lomax. An inclusive economy dataset for wards in great britain using administrative and synthetic data sources. *Scientific Data*, 12(1):1230, 2025. doi:10.1038/s41597-025-05502-x.
- 11 Ruoxi Wang, Xinyuan Zhang, and Nan Li. Zooming into mobility to understand cities: A review of mobility-driven urban studies. *Cities*, 130:103939, 2022. URL: <https://www.sciencedirect.com/science/article/pii/S026427512200378X>, doi:10.1016/j.cities.2022.103939.