# A Markov Chain Monte Carlo Approach for Estimating Daily Activity Patterns

Ariane Scheffer[a]*, Claudia Bandiera[a], Ernesto Cipriani[b], Guido Cantelmo[c], Francesco Viti[a]

[a]*Department of Engineering, University of Luxembourg, Esch-sur-Alzette, Luxembourg;*
[b]*Department of Engineering, Roma Tre University, Rome, Italy*
[c]*Chair of Transportation Systems Engineering, Technical University of Munich, Munich, Germany*

* Corresponding author: ariane.scheffer@uni.lu

Determining trip purposes is fundamental to evaluate the travel demand and predict the impact of transport policies and traffic management strategies on travel behaviour. While trip-based representation looks at trip origin and destination for individual trips, the concept of tours is traditionally adopted to model individual daily scheduling and travel choice interdependencies. However, the efforts required to collect high quality data and interpret results of advanced models are major challenges. The objective of this study is to enhance the representation of a trip-based model and infer the trip purpose of a population, at a zonal level, by employing utility theory and by applying a Markov Chain Monte Carlo approach. A utility-based probability function for departure time is proposed and calibrated using aggregated data (e.g. traffic counts) and, in order to ensure consistency, an activity duration constraint is added in the calibration phase. This methodology is characterised by low requirements in terms of data and the model is shown to be flexible and easy to implement.

Keywords: Markov chain Monte-Carlo; travel demand estimation; utility theory; trip purpose; tours; activity-based models

## INTRODUCTION

The inherent complexity of people's mobility has direct consequences on understanding and modelling their travel behaviour. Driven by this reason, sophisticated demand models emerged since the 1990's (*1*). While traditional Trip-Based Models (TBM) currently remain widely adopted to forecast travel demand (*2*), they usually provide a coarse representation of the demand, which makes them inadequate for planning purposes (*3*), especially when within-day dynamics need to be properly modelled. The main problem is that, while researchers agree that travels are derived from the demand for activities and services (*4*), conventional macroscopic TBM do not explicitly account for activity-travel scheduling and duration constraints (*5*). However, TBM models provide essential application and interpretability opportunities, as well as they offer the possibility to be calibrated using observed traffic flows, which still remain the most popular and widely available data in practice. For this reason, trip-based origin-destination demand flows are the typical input for advanced Dynamic Traffic Assignment models (DTA), which are the most established tools for planning, optimizing and managing transportation networks (*6*).

On the other hand, the last decades have witnessed intensive research efforts in developing Activity-Based Models (ABM), which are capable of representing individual mobility on large scale systems (*7*). Theoretically attractive, they propose an in-depth representation of the demand but tend to be hard to apply due to few consistent calibration tools (*8*) and interpretability issues (*9*). Furthermore, ABM typically use synthetic populations for aggregate forecast or microsimulation. The more representative and consistent the generated synthetic population is, the more reliable results will be provided by the model at an aggregated level (*10*). Even though new models are being developed to achieve higher quality and suitable populations (*11*, *12*), this step of creating a realistic population remains crucial and demanding in term of data (*13*).

This drawback is the main restraint for using advanced models such as ABM, while the easier-to-use TBM may have difficulties to properly capture the complexity of the travel demand and the actual correlation between trips. In order to take advantage from these two methods, we present a flow-based approach which correlates successive trips through the activity executed in between. It builds on the work of (*14*) where Monte Carlo sampling methods were proposed to estimate purpose-specific travel-demand along the day. In order to enhance its representativeness and consistency in time and space, the daily demand is separated into functions derived from the utility-based departure time choice model proposed by (*15*). Markov Chain Monte Carlo (MCMC) is used to calibrate parameters, in accordance to observed flows and a-priori belief on their expected values.

After examining the background and components of this model, it is applied on a controlled experiment. We show that demand flows can be represented through tour-specific flows and that purpose-dependent macroscopic demand can be identified, which once aggregated is consistent with the observed traffic flows.

## MODEL FORMULATION

For modelling traffic demand, the area of study is usually subdivided into zones of origin and destination (Traffic Analysis Zones, TAZ), where the demand is assumed aggregated. The continuous-in-time mobility patterns are discretised into matrices, which characterise the OD flows within a certain time period, where the distribution of the mobility patterns is assumed in some way stationary for the time-period. In the proposed methodology, purpose-dependent OD flows are described by "primitives" of the complete demand, i.e. functions representing the number of users travelling for a specific purpose, by time of the day. These primitives are shaped by a departure time choice model estimated for trips to arrive and leave a specific activity-zone, based on the assessment of utility trade-off with the previous and following occupation. Their convolution results in (aggregated) OD flows. While the model proposed by (*14*) focused on testing this approach to identify flows, we provide here two additional extensions:

- The calibration of a **utility-based departure time choice model**
- A **correlation model**, based on typical activity durations to better identify and characterise the different purpose-specific components of the demand.

The necessary input is limited to aggregated information, at the level of the TAZ, for which a time-dependent demand is observed. In addition to this, sample data can be used as ground truth for building the prior information, used for the calibration of the function.

### Utility-based departure time choice model

Following the general framework proposed in (*16*), we define the overall utility U of an individual as the sum of two components:

$$U = (U^T + U^A) \tag{1}$$

where $U$ is the overall net utility accumulated during the reference time period, $U^T$ represents the disutility of travelling and $U^A$ the utility of performing one or more activities. For sake of simplicity, trip costs are still at the time considered as a constant factor that reduces the potential positive utility. While a share of works actually focus on the trip cost, neglecting the positive part, we focus instead on the positive component. Preliminary work has shown that it can already generate many different departure time profiles consistent with utility-maximization principles (*17*). The utility can further be expressed as:

$$U^A = \sum_n U^{A,n} \tag{2}$$

where $U^{A,n}$ is the utility of performing a certain activity *n*, usually formulated as a time-dependent function $U^{A,n}(t)$. The utility associated to a certain time interval *t* can be mathematically calculated using different functional forms. In our study, we assume that the travellers choose a departure time that maximizes the utility derived from the activities defined in their schedule as in the following equation (*15*):

$$U^{A,n}(t) = \frac{\gamma_n \beta_n (U_n^{max})}{exp[\beta_n(t-\alpha_n)].(1+exp[-\beta_n(t-\alpha_n)])^{\gamma_n+1}} \tag{3}$$

- $U_n^{max}$ : maximal utility accumulated for a certain activity;
- $\alpha_n$ : position on the temporal axis;
- $\beta_n$ : variance around the saturation point;
- $\gamma_n$: affects the position of saturation.

In the context of trip chain estimation, the utility for a sub-tour of three activities is calculated according to equation (4) assuming a travel time $t_t$ and departure time intervals $t_1$, $t_2$, $t_3$ and $t_4$ :

$$U(t_1, t_2, t_3, t_4) = \int_{t_1}^{t_2} U_1^A(t)dt + \int_{t_2+t_t}^{t_3} U_2^A(t)dt + \int_{t_3+t_t}^{t_4} U_3^A(t)dt \tag{4}$$

Each activity is fully described by two parameters: arrival and departure time. For tours consisting of two trips (e.g. home-work-home) only two variables remain ($t_1$ =0, $t_4$=24). For any more complex tour, we can separate sub-components of the tour and use Equation (4) to study sub-tours, therefore simpyfing the problem. This allows to work on single traffic zones, using generated and attracted demand flows, instead of modelling tour-based demand flows at a network level. This leads to a more computationally efficient formulation. As $U_1^A$ and $U_3^A$ depend on other activities performed within the tour we need certain assumption. However, under proper assumptions, the utility function becomes s eparable meaning that treating each function as separate does not introduce any further error (*17*).

Time intervals are considered to make the departure time decision a discrete choice process represented by a simple multinomial logit model. In equation (5), $U_k$ is a generic marginal utility calculated for a set of time intervals (arrival and departure time at activity location) and $P_k$ is the probability of choosing it in a set of all considered time intervals $j$ in the modelled time period.

$$P_k = \frac{exp(U_k)}{\sum_j exp(U_j)} \tag{5}$$

As mentioned before, when Equation (5) is applied to a single traffic zone, the model will focus only on modelling two departure time intervals.

**MCMC - Markov Chain Monte Carlo**

Given the departure time choice model, we aim at calibrating the associated utility function (equation 3) through MCMC. The Bayesian approach is a common tool in the field of synthetic population generation (*18–20*), however, it has not been used yet to calibrate the departure time choice process, to the best of the authors' knowledge.

Figure 1 shows the interaction between the model components and their purpose.

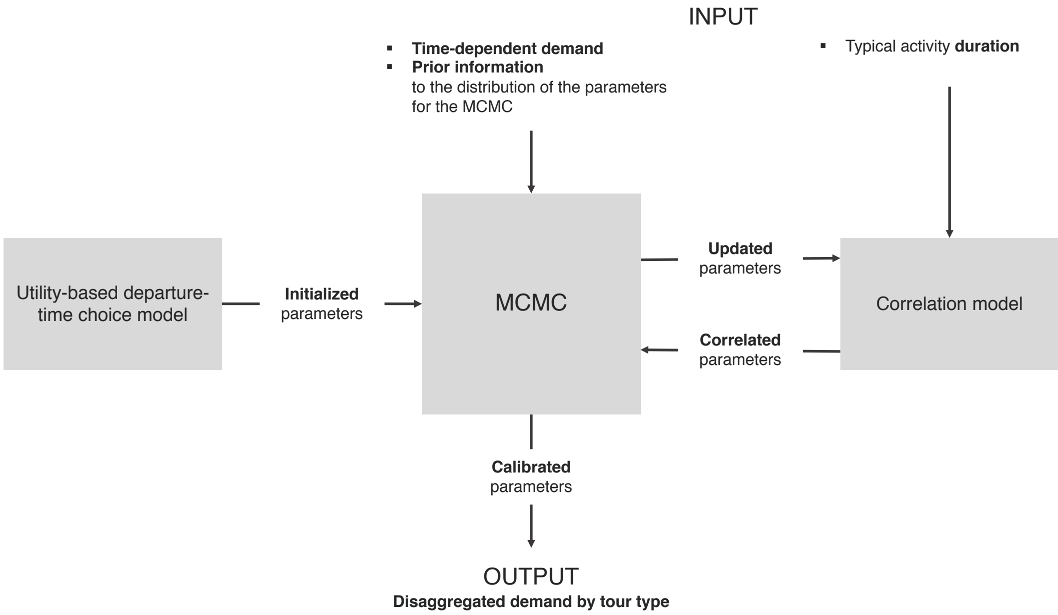


**Figure 1: Methodological Framework**

Instead of sampling parameters of the law describing observed data in a standard Monte Carlo manner, we use a random walk Metropolis-Hastings algorithm. Starting from an arbitrary set of parameters, we generate new parameters and decide upon their conservation in the posterior such that the constructed Markov Chain's stationary distribution is equal to the target distribution.

The **prior** is a law $\pi(\Theta)$ defined for each of these parameters. The **conditional distribution** of the sample given the parameters is $f(x|\Theta)$ and the **posterior** distribution of the parameter $\Theta$ is $\pi(\Theta|\mathrm{x})$. It is calculated using Bayes' formula (equation 6).

$$\mathbb{P}(\Theta|x) = \frac{\mathbb{P}(x|\Theta).\,\mathbb{P}(\Theta)}{\mathbb{P}(x)} \tag{6}$$

The normalizing constant $\mathbb{P}(x)$ being often ignored, the Bayesian modelling approach can be generally summarized by $posterior \sim likelihood \times prior$.

The MCMC is advantageous for two reasons. Firstly, the estimation allows to impose an idea on the behaviour of the parameters, that can include information from the literature or from surveys. This aspect is extremely useful when the amount of data available is relatively low. Secondly, we can describe the behaviour of estimated parameters by means of a probability distribution where either the expectation is used in order to get a point estimate of the theoretical average or the given distribution could be used for calculating reliable ranges.

## Correlation constraint

Because the utility functions contain many parameters, the MCMC is likely to over-fit the data and provide a poor estimation of the actual mobility demand. The "minimum activity duration constraint" contributes to overcome this issue and to avoid unrealistic solutions, under the postulate that duration of activities at destination influences the departure time of successive trips.

For each activity type, a minimal duration is introduced, meaning that the departure time of the second trip of the sequence has to follow the rule:

$$t_2 \ \geq \ t_1 + t_t + d_{min_{activity}}$$

Only after this lower bound, all departure time intervals are assessed for the second trip, without any penalty.

## CASE STUDY

A controlled experiment was designed to examine the performance of the model and assess the proposed methodology.

Based on a previous work (*14*) relying on data collected in Belgium (*21*) the trip types were grouped in the following way:

*A1.* "Work" are mandatory, repeated activities, for career or education purpose

*A2.* "Duty" are necessary activities, mostly done by constraint

*A3.* "Leisure" are non-mandatory activities, for recreational purpose

A4."Home" is always the anchor point of a tour in this case study, every chain starts and ends at home (i.e. $t_1$ =0, $t_4$=24 in equation (4))

Demand profiles are generated by using the same equation type as the target estimation (equation 3). This way, we show how the procedure can distinguish six components of the demand and showcase how precise the calibration can be. Even though the proposed model can be generalized to all kinds of trip chains, the given case study only handles part of the demand that corresponds to home-based tours.

The synthetic reference demand used for comparison (figure 2) consists of three pairs of departure time functions to and from a given activity that are parametrized by a total of 39 parameters.

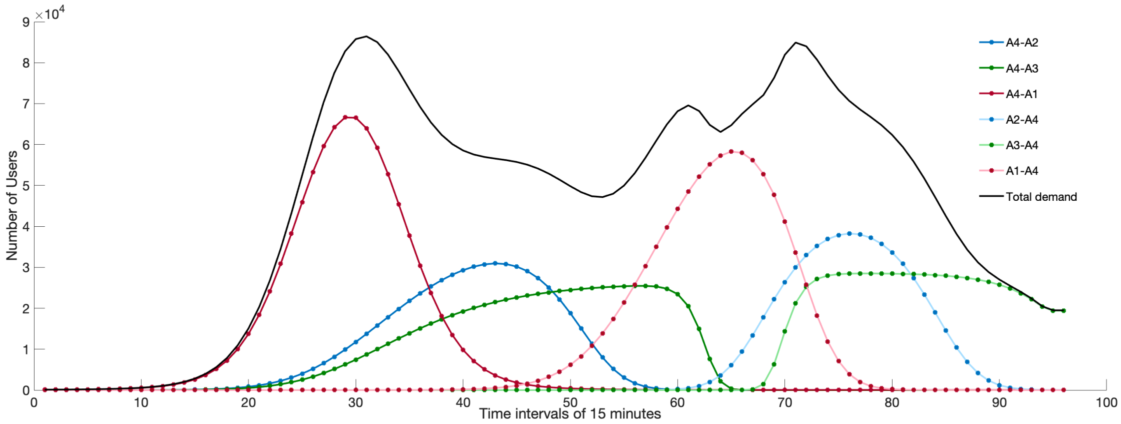


**Figure 2: Reference demand**

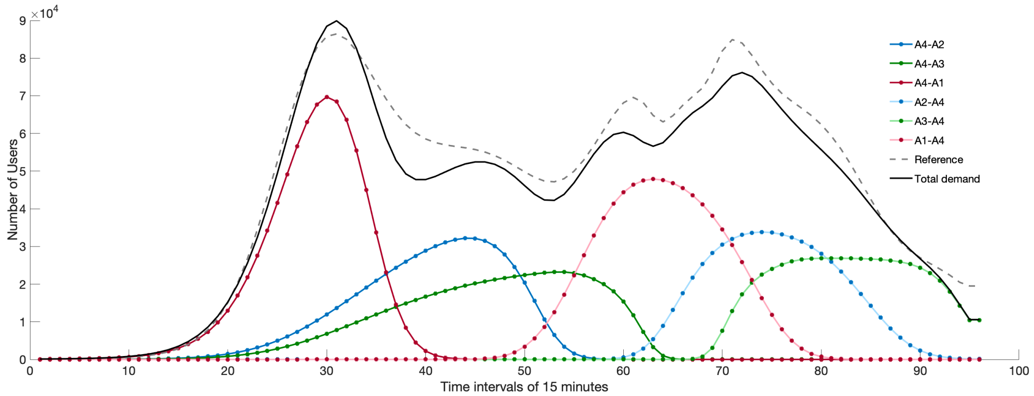


**Figure 3: Decomposition with the duration constraint**

Figure 3 shows a decomposition of the demand obtained after 10.000 iterations. It is close to the reference one (dashed grey line) and primitives were properly configured and attributed to each tour component. On figure 4 we can see more in details that demand has been underestimated and, overall, 7.58% of the daily demand has not been caught by any of the primitives. Indeed, overestimation is prevented by a reduction of the initial demand and can be interpreted as the portion that does not correspond to any of the proposed patterns. Still it is a good starting point for then using an OD estimation method that would further improve the fitting. However, the median relative error between the reference trip type and the estimated share among the six options during a given time interval is 13.1%.

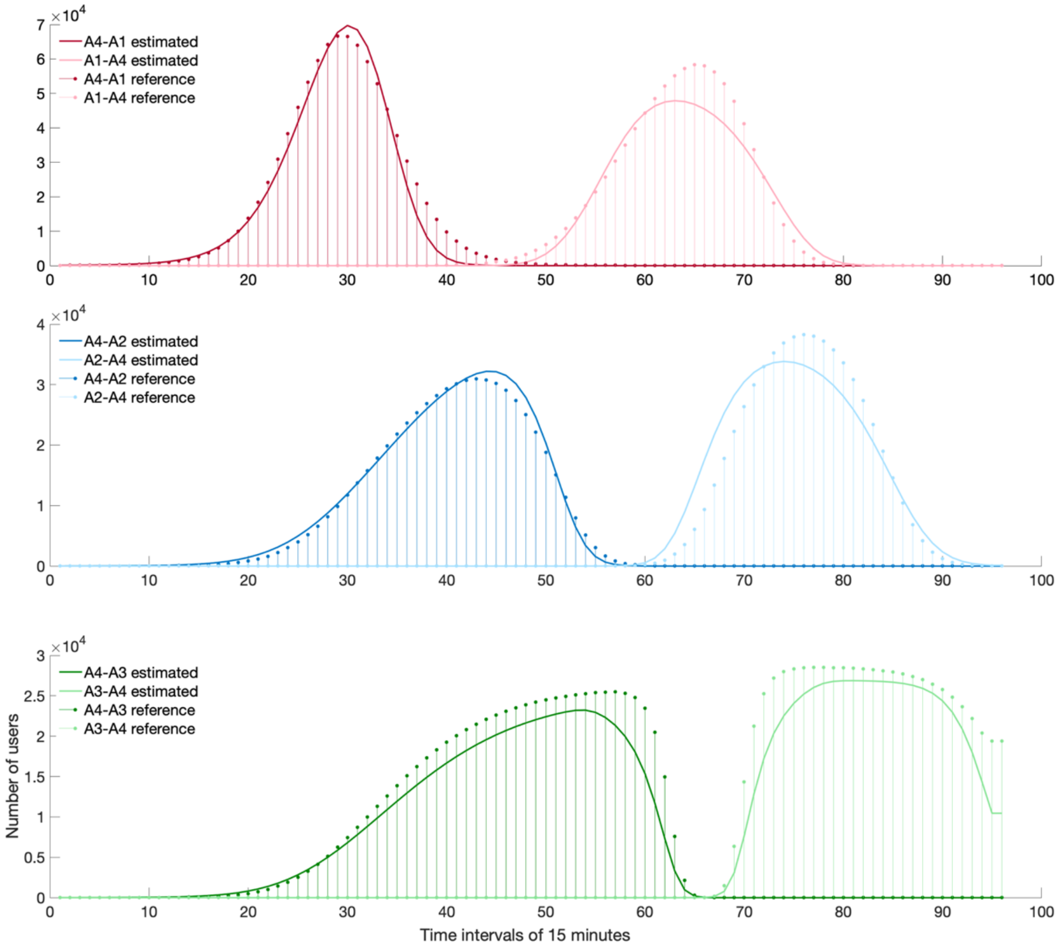


**Figure 4 Estimation and reference by tour type**

The fully controlled experiment provides also a comparison parameter by parameter (figure 5). Coherently, the parameter having a prior following a uniform distribution has more variations but still results in a good approximation. A noticeable outlier is the $\alpha$ parameter for the evaluation of $U_1^{A1}$. In spite of that the overall estimation of the distribution for the A4-A1 trip is very accurate.

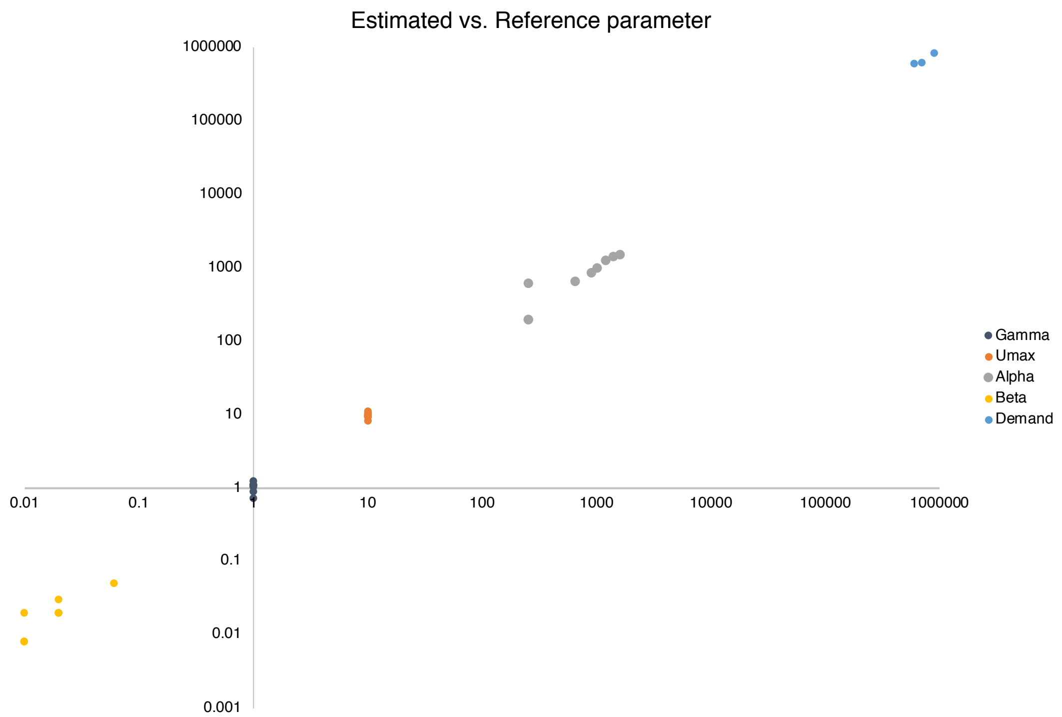


**Figure 5 Estimated vs. reference parameters**

## CONCLUSIONS

In this paper we proposed a model based on advanced sampling methods, specifically MCMC, in order to determine activity-specific demand based on traffic data through the calibration of a departure time choice model. The concept of trip chaining is handled by first dividing the tour into sub tours and then estimating the two-parameters departure time choice. The model further simplifies the estimation process by integrating an activity duration constraint within the departure time choice model. The proposed MCMC is shown to be able to estimate activity-specific flows from the aggregated demand and the utility-based probabilities prove to be adequate for reproducing a whole day traffic pattern. Inserting constraints on the probability form allows to have a better interpretation of the results, however, these constraints make the model unable to reproduce distributions being away from their inherent form. Nonetheless, as the probability curves are calculated with the current model, the results when combined with the actual dynamic OD matrices, can give a useful interpretation to the flows. Future research direction include the application of the method on a network including multiple zones and their interactions. Further development of the model shall consider more in details the travel time and related cost, which would allow to include mode choice as another level of detail for the estimation of OD flows.

## REFERENCES


1. Timmermans, H., T. Arentze, and C.-H. Joh. Analysing Space-Time Behaviour: New Approaches to Old Problems. *Progress in Human Geography*, Vol. 26, No. 2, 2002, pp. 175–190. https://doi.org/10.1191/0309132502ph363ra.
2. McNally, M. G. The Four-Step Model. In *Handbook of Transport Modelling*, Emerald Group Publishing Limited, pp. 35–53.
3. Lindveld, C. D. R. Dynamic O-D Matrix Estimation: A Behavioural Approach. 2003.
4. Axhausen, K. W., and T. Gärling. Activity-based Approaches to Travel Analysis: Conceptual Frameworks, Models, and Research Problems. *Transport Reviews*, Vol. 12, No. 4, 1992, pp. 323–341. https://doi.org/10.1080/01441649208716826.
5. Cantelmo, G., F. Viti, E. Cipriani, and M. Nigro. A Utility-Based Dynamic Demand Estimation Model That Explicitly Accounts for Activity Scheduling and Duration. *Transportation Research Part A: Policy and Practice*, Vol. 114, 2018, pp. 303–320. https://doi.org/10.1016/j.tra.2018.01.039.
6. Peeta, S., and A. K. Ziliaskopoulos. Foundations of Dynamic Traffic Assignment: The Past, the Present and the Future. *Networks and Spatial Economics*, 2001, p. 33.
7. Balmer, M., K. Meister, M. Rieser, K. Nagel, and K. W. Axhausen. *Agent-Based Simulation of Travel Demand: Structure And . . .* 2008.
8. Flötteröd, G. Cadyts a Free Calibration Tool for Dynamic Traffic Simulations. *Infoscience*. https://infoscience.epfl.ch/record/152358. Accessed Feb. 13, 2020.
9. Moeckel, R., L. Huntsinger, and R. Donnelly. From Macro to Microscopic Trip Generation: Representing Heterogeneous Travel Behavior. *The Open Transportation Journal*, Vol. 11, No. 1, 2017, pp. 31–43. https://doi.org/10.2174/1874447801711010031.
10. Barthelemy, J., and E. Cornelis. Synthetic Populations: Review of the Different Approaches. CEPS/INSTEAD, , 2012.
11. Sun, L., A. Erath, and M. Cai. A Hierarchical Mixture Modeling Framework for Population Synthesis. *Transportation Research Part B: Methodological*, Vol. 114, 2018, pp. 199–212. https://doi.org/10.1016/j.trb.2018.06.002.
12. Borysov, S. S., J. Rich, and F. C. Pereira. Scalable Population Synthesis with Deep Generative Modeling. *Transportation Research Part C: Emerging Technologies*, Vol. 106, 2019, pp. 73–97. https://doi.org/10.1016/j.trc.2019.07.006.
13. Guo, J. Y., and C. R. Bhat. Population Synthesis for Microsimulating Travel Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, Vol.

2014, No. 1, 2007, pp. 92–101. https://doi.org/10.3141/2014-12.

14. Scheffer, A., G. Cantelmo, and F. Viti. Generating Macroscopic, Purpose-Dependent Trips through Monte Carlo Sampling Techniques. *Transportation Research Procedia*, Vol. 27, 2017, pp. 585–592. https://doi.org/10.1016/j.trpro.2017.12.111.
15. Ettema, D., and H. Timmermans. Modeling Departure Time Choice in the Context of Activity Scheduling Behavior. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1831, 2003, pp. 39–46. https://doi.org/10.3141/1831-05.
16. Yamamoto, T., S. Fujii, R. Kitamura, and H. Yoshida. Analysis of Time Allocation, Departure Time, and Route Choice Behavior Under Congestion Pricing. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1725, 2000, pp. 95–101. https://doi.org/10.3141/1725-13.
17. Cantelmo, G., and F. Viti. Incorporating Activity Duration and Scheduling Utility into Equilibrium-Based Dynamic Traffic Assignment. *Transportation Research Part B: Methodological*, 2018. https://doi.org/10.1016/j.trb.2018.08.006.
18. Sun, L., and A. Erath. A Bayesian Network Approach for Population Synthesis. *Transportation Research Part C: Emerging Technologies*, Vol. 61, 2015, pp. 49–62. https://doi.org/10.1016/j.trc.2015.10.010.
19. Casati, D., K. Müller, P. J. Fourie, A. Erath, and K. W. Axhausen. Synthetic Population Generation by Combining a Hierarchical, Simulation-Based Approach with Reweighting by Generalized Raking. *Transportation Research Record*, Vol. 2493, No. 1, 2015, pp. 107–116. https://doi.org/10.3141/2493-12.
20. Saadi, I., A. Mustafa, J. Teller, B. Farooq, and M. Cools. Hidden Markov Model-Based Population Synthesis. *Transportation Research Part B: Methodological*, Vol. 90, 2016, pp. 1–21. https://doi.org/10.1016/j.trb.2016.04.007.
21. Castaigne, M. Behaviour and Mobility during the Week “BMW.”