OASIS: Optimisation-based Activity Scheduling with Integrated Simultaneous choice dimensions

Janody Pougala * † Tim Hillel ‡ Michel Bierlaire *

November 24, 2022

Report TRANSP-OR 221124
Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne
transp-or.epfl.ch

*École Polytechnique Fédérale de Lausanne (EPFL), School of Architecture, Civil and Environmental Engineering (ENAC), Transport and Mobility Laboratory, Switzerland

†Corresponding author, {janody.pougala@epfl.ch}

‡University College London (UCL), Department of Civil, Environmental and Geomatic Engineering, United Kingdom

Abstract

Activity-based models offer the potential of a far deeper understanding of daily mobility behaviour than trip-based models. However, activity-based models used both in research and practice have often relied on applying sequential choice models between subsequent choices, oversimplifying the scheduling process. Recent work (Pougala et al., 2022) has established a new modelling approach for activity scheduling which integrates the different choice dimensions (activity participation, ordering, scheduling, etc) into a single optimisation framework, based on utility maximisation principles. In this paper we introduce a complementary methodology for the estimation of model parameters from historic data, allowing for the generation of realistic and consistent daily mobility schedules. In combination, the optimisation-based scheduling simulation and the parameter estimation framework, is hereby referred to as OASIS (Optimisation-based Activity Scheduling with Integrated Simultaneous choice dimensions). The estimation framework introduced in this paper consists of two main elements: (i) choice set generation, where we generate a sample of competitive alternative schedules by applying the Metropolis-Hasting algorithm to historic schedules, and (ii) discrete choice parameter estimation, where the scheduling process is formulated as a discrete choice problem, in which each individual chooses a full daily schedule from a finite set of possible schedules. We validate our approach by estimating parameters for a sample of individuals from the 2015 Swiss Mobility and Transport Microcensus (Office fédéral de la statistique and Office fédéral du développement Territorial, 2017), and evaluating the output of the OASIS model against realised schedules from the data. The results demonstrate the ability of the new framework to estimate stable and significant parameters from historic data that are consistent with behavioural theory. Furthermore, the schedules outputted by the OASIS model are evaluated as being both feasible and realistic at a disaggregate level, and correspond well to the aggregate trends in the historic data. This work opens the way for future developments of activity-based models, where a great deal of constraints can be explicitly included in the modelling framework, and all choice dimensions are handled simultaneously.

Keywords: Activity-based modelling, discrete choice modelling, parameter estimation, choice set generation, maximum likelihood estimation, simulation

1 Introduction

Activity-based models have been the focus of increasing research efforts in a variety of domains, including transport research, energy demand, and epidemiology. In transportation, they provide a behaviourally realistic alternative to traditional trip-based models and aggregate analyses.

In previous work (Pougala et al., 2022) we have introduced an activity-based model to simultaneously estimate choices of activity participation, scheduling, travel mode and location. The model is utility-based and uses mixed-integer optimisation to simulate realisations of feasible activity schedules. The major benefit of the simultaneous approach over traditional sequential approaches (that describe the activity-travel process as a sequence of individual choices, with varying degrees of interaction), is that the simultaneous approach inherently captures trade-offs between activity scheduling decisions. This opens the way for a flexible integration of behavioural extensions, including complex context-specific constraints and interactions.

A significant limitation and challenge of the simultaneous approach is the estimation of stable and significant parameters. In Pougala et al. (2022), the parameters are not estimated, instead a set of accepted values from the literature are used to illustrate the principles of the framework. Parameter estimation is generally a challenging task in activity-based models, due to the size of the problem, the complexity of the structure due to the spatio-temporal constraints, and, often, the lack of appropriate data. In sequential models, the set of parameters can be estimated in stages (e.g., Bowman and Ben-Akiva, 2001, Chen et al., 2020) which considerably simplifies the problem, but at the expense of model flexibility and behavioural realism. Choice sets are also usually considered given, or constructed with mostly arbitrary decision rules. Considering each choice dimension simultaneously makes the estimation problem significantly more complex, as the resulting combinations cannot be fully observed or enumerated, and the correlations between choice dimensions and between alternatives are difficult to properly account for within a tractable mathematical process.

In this paper, we introduce a methodology to estimate the behavioural parameters of the simultaneous model, consisting of two elements: (i) choice set generation, where we generate a sample of competitive alternative schedules by applying the Metropolis-Hasting algorithm to historic schedules, and (ii) discrete choice parameter estimation, where the scheduling process is formulated as a discrete choice problem, in which each individual chooses a full daily schedule from a finite set of possible schedules.. We test different model specifications and evaluate the quality of the parameter estimations and their impact on the simulations for a sample of individuals of the Swiss Mobility and Transport Microcensus (Office fédéral de la statistique and Office fédéral du développement Territorial, 2017).

The integration of simultaneous activity scheduling simulation and the parameter estimation form OASIS (Optimisation-based Activity Scheduling with Integrated Simultaneous choice dimensions): a flexible activity-based framework able to accommodate the requirements and context-specific constraints of different application domains, and thus provide tailored behavioural insights.

2 Relevant literature

Activity-based models originally emerged in the 1970s as a response to the shortcomings of traditional 4-step models (Vovsha et al., 2005, Castiglione et al., 2014), namely:

1. trips are the unit of analysis and are assumed independent, meaning that correlations between different trips made by the same individual are not accounted for properly within the model;
2. models tend to suffer from biases due to unrealistic aggregations in time, space, and within the population; and
3. space and time constraints are usually not included.

The early works of Hägerstraand (1970) and Chapin (1974) established the fundamental assumption of activity-based models, that the need to do activities drives the travel demand in space and time. Consequently, mobility is modelled as a multidimensional system rather than a set of discrete observations. Rasouli and Timmermans (2014) and Axhausen (2000) provide in-depth reviews of the state of research and practice in activity-based modelling.

A significant challenge in activity-based modelling is the estimation of the model parameters. This is especially crucial for utility-based models: while the activity-based problem can be solved taking advantage of random utility maximisation theory and econometric concepts and properties, calibrating the model to data is not straightforward - often due to the lack of available data. In addition, the methodology and assumptions of classical discrete choice modelling cannot easily be transferred to an activity-based context. When the scheduling of activities and travel across time and space is formulated as a choice between discrete alternatives, the problem is multidimensional (involving continuous and discrete choice dimensions such as activity participation, scheduling, mode, destination, route...) and combinatorial. The full set of solutions cannot be enumerated or fully observed by the modeller or the decision maker. In addition, while the schedules in the choice set are overall distinct, they might present significant overlaps in their components. Finally, the constraints further increase the complexity of the problem, limiting the derivation of closed form probabilities (Recker et al., 2008). These issues are even more challenging when the choice dimensions are considered simultaneously.

There are therefore two main issues to address: generating a choice set for the purpose of parameter estimation, and formulating a tractable model specification which is able to capture multidimensional correlations.

The combinatorial nature of the problem prevents a full enumeration of the possible alternatives. There exist strategies to estimate parameters on subsets of alternatives (e.g., Guevara and Ben-Akiva, 2013), but the challenge is to form said set of alternatives to be informative enough to estimate the parameters and varied enough to minimise bias.

Both deterministic and stochastic models exist for the generation of spatio-temporal choice sets (Pagliara and Timmermans, 2009) for the purpose of parameter estimation. Models that use a deterministic approach typically include a choice set predefined by the modeller, or samples of alternatives obtained from decision rules reflecting the domain knowledge. On the other hand, stochastic approaches do not assume that the choice set is universal and known, but rather model the uncertainty associated with it. Deterministic choice sets are used in early activity-based models (e.g., Bowman and Ben-Akiva (2001) enumerate the feasible combinations of primary activity, primary tour type, and number and purpose of secondary tours). In some rule-based models, the choice set generation process involves generating a limited set of activities based on rules, and then enumerating the combinations (e.g., Arentze and Timmermans, 2000).

Stochastic models for choice set generation have been thoroughly investigated in route choice modelling (e.g., Flötteröd and Bierlaire, 2013, Frejinger et al., 2009). However, these methods are not straightforward to apply to activity-based models because of their multidimensionality. Danalet and Bierlaire (2015) adapt and apply the methodology proposed by Flötteröd and Bierlaire (2013) to sample alternatives in an activity-based context. The alternatives are activity schedules, which are represented as paths in a defined network. The nodes of the network are activities potentially performed for a unit of time, and the edges connecting them represent successful performance and succession between activities. In order to include attractive alternatives in their choice set, the authors define an attractivity measure for each node based on their frequency of observation and the frequency of the length of activity-episodes in the network. They validate the method on a synthetic network and on a real dataset describing pedestrian behaviour, and by calibrating the parameters of a discrete choice model with a utility associated with each activity path. It is established that importance sampling with the Metropolis-Hastings algorithm provides a better model fit than randomly sampling the choice model.

Nijland et al. (2009) estimate the parameters of a need-based model to predict multiday activity patterns. The need-based model was first formulated by Arentze and Timmermans (2009), under the assumption that utilities of activities are a function of needs of individuals and households, and that these needs grow over time following a logistic function.The modelled choice is the choice of performing an activity on a specific day d, given that the activity was last performed on day s. The utility function is composed of a term for the satisfaction of needs which builds up between s and d and a term capturing the preference for performing the activity on day d. Nijland et al. define a logit model. The parameters include socio-economic characteristics (e.g. gender, household composition and income, education level, needs).They fit a model for each of six groups of activities (daily and non-daily shopping, social visits, going out, sports and walking/cycling), using the findings of a purpose-designed survey. Arentze et al. (2011) also estimate the parameters of the need-based model. They estimate a mixed logit model with error components to capture randomness and unobserved factors in the need-building process and day-to-day conditions. The duration of the activity is included implicitly through a utility threshold constraint: an activity can only be selected when its associated utility reaches a given threshold. The set-up of both models greatly simplifies the choice set considerations: as only one choice dimension is considered (day of week of participation), the choice set can easily be enumerated. In addition, as they do not model explicitly activity duration and timing decisions, they do not consider the effect of activity-travel interactions (e.g. timing trade-offs between activities).

Xu et al. (2017) propose a discrete choice estimation of the utility parameters of Recker's Household Activity Pattern Problem (HAPP). Similarly to our problem, the utility function of the HAPP defines the objective function of a maximisation problem subject to individual spatio-temporal constraints. Xu et al. try to improve the behavioural interpretation of the model simulations with estimated parameters while preserving the constraints of the optimisation problem. They define a choice problem of representative activity patterns, which they solve using a Path Size logit model. The representative patterns are obtained with pattern clustering of the observed schedules (Allahviranloo et al., 2014). The choice set is constructed by sampling one pattern from the unchosen clusters with a genetic algorithm and adjusting the sampled patterns according to individual constraints. The final set is the combination of alternatives that leads to the minimal D-error. They formulate and estimate the parameters of 12 different model specifications. Their methodology is one of the first applications of discrete choice estimation for an optimisation-based activity-travel model, and shows the added behavioural value of their approach to the framework.

However, it does not ensure unbiased estimators: indeed, they do not correct their maximum likelihood estimation to account for the calibration on a sample of alternatives and not the full choice set. In addition, the methodology to generate choice sets creates endogeneity and is biased towards alternatives with high probability of being chosen: the unchosen alternatives are representative patterns from the observed sample, and the final choice set maximises the information gain. This leads to overfitting, which would reduce the ability of the model to be applied to different contexts and datasets.

Chen et al. (2020) estimate the parameters of their simulation-based activity-based model by using gradient descent methods. The model is a nested logit model where each level contains one or more choice models π. The choice models at lower levels interact with upper level models through logsum terms. They separate the population in mini-batches in order to estimate the gradient of the objective function, which is the sum of the distances between expected values of simulated aggregate statistics and the observed statistics, and between the *a priori* values of the parameters and the calibrated ones, as obtained for each choice model in the problem. The calibration procedure is a minimisation of the objective function, which converges when a maximum number of iterations is reached or the objective function is below a given threshold. They illustrate their method on a sequential activity-based model with 3 levels (day pattern, tour, and intermediate stops). Each level contains several choice models, and the models of the lower levels are dependent on the decisions made at the upper levels. On the presented case study, their approach outperforms traditional gradient descent methods such as the simultaneous perturbation stochastic approximation (SPSA). However, the behavioural insights gained from their method are limited to count aggregate statistics. In addition, the approach can be suited to sequential ABMs to analyse single days and individuals but cannot easily be extended to more complex interactions (simultaneous choices, multiday analyses, household interactions, etc.). Finally, the choice set is assumed to be known and enumerable.

There are examples in the literature where authors use a method other than discrete choice modelling. For instance, Recker et al. (2008) use a genetic algorithm to estimate the parameters of the utility function of their household activity-based model. They introduce distance metrics to compute the errors between observed and predicted multidimensional sequences (Euclidian norm for continuous values such as time variables and Levenshtein distance for discrete components such as travel decisions). The fitness function of the genetic algorithm is derived from these errors, and the vector of parameters is modified at each iteration (through mating, crossover and mutations). The algorithm is run until an optimal set of parameters is found. This set results in a population that is close enough to the observed population. This heuristic approach enables the navigation of the complex activity-travel solution space with relative efficiency. However, the general limitations of the genetic algorithm not only do apply here, but they are actually exacerbated by the nature of the problem: genetic algorithms are slow to converge due to the repeated evaluations of the fitness function. At each iteration, the authors perform a multi-dimensional sequence alignment to compute the respective distance between predicted and observed schedules in both the continuous and discrete dimensions. This involves an enumeration of element-wise combinations, which can be especially costly with increasing complexity. Genetic algorithms are prone to convergence towards local optima, which emphasises the importance of the fitness function for the quality of the solution. As the algorithm is searching for optimal solutions with respect to the fitness function, it is important to purposefully ensure the diversity of the population (e.g. by modifying certain hyperparameters such as the rate of mutations, or introduce random sequences in the populations) in order to reduce the bias in the utility parameters estimated with this method. The estimation of the parameters of the household activity pattern problem is also tackled by Chow and Recker (2012). They formulate an inverse optimisation problem to find the combination of parameters for which the schedule is optimal. A limitation of this approach is the one-to-many nature of the inverse problem which means that, as the problem is underidentified, the found solutions and associated parameters might not be behaviourally interpretable.

In this paper, we propose a parameter estimation procedure for the simultaneous activity-based model presented in Pougala et al. (2021). The model simulates daily schedules of activities for a given individual by maximising the utility they gain from participating to activities. The output is a distribution of schedules conditional on the distribution of the random error terms. The first iteration of the model demonstrated the ability of the approach to generate realistic activity schedules while explicitly accounting for scheduling trade-offs. However, the parameters of the utility function were not estimated, and we used instead values from the literature. The methodology we present here is based on Maximum Likelihood Estimation (MLE). Similarly to other state-of-the-art approaches, we take advantage of the theoretical robustness and flexibility of discrete choice models for this task, but applied to a framework where all of the activity-travel choices are considered simultaneously. This allows to capture trade-offs and interrelations between choices, but with the added cost of complex solution spaces and combinatorial choice sets. We apply here a methodology for choice set generation based on the Metropolis-Hastings based on the works of Flötteröd and Bierlaire (2013) and Danalet and Bierlaire (2015).

Table 1 summarises the papers described in this section, and the methodologies developed or applied by the authors for the generation of individual choice sets and for the estimation of parameters.

Paper	Type of ABM	Choice set generation	Parameter estimation
Recker et al. (2008)	HAPP	-	Genetic Algorithm
Nijland et al. (2009)	Needs-based model	Full enumeration	Logit model
Arentze et al. (2011)	Needs-based model	Full enumeration	Mixed logit
Chow and Recker (2012)	HAPP	-	Inverse optimisation problem
Danalet and Bierlaire (2015)	Network-based	Metropolis-Hastings sampling	-
Xu et al. (2017)	HAPP	Pattern clustering and importance sampling	Path Size logit
Chen et al. (2020)	Sequential ABM	-	Nested logit model
Current paper	Simultaneous ABM	Metropolis-Hastings sampling	Logit model

Table 1: Relevant literature

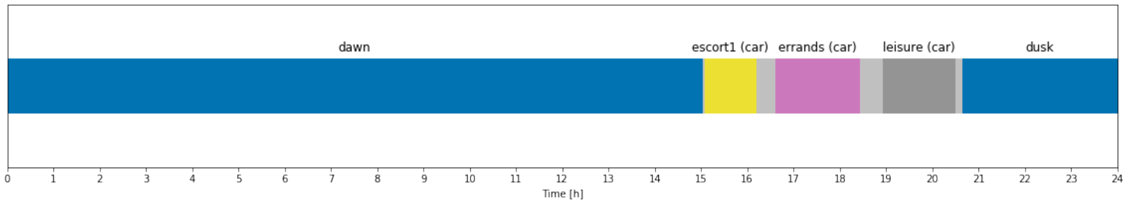


Figure 1: Example of a daily schedule. The light gray patches between activities indicate travel. *Dawn* and *dusk* are the first and last *home* activities of the day.

3 Methodology

We present a methodology to estimate the parameters of an activity-based model where all scheduling choice dimensions (activity participation, timing decisions, mode, location, etc.) are considered simultaneously within a mixed integer optimisation framework. The estimation process consists of two elements: (i) choice set generation, and (ii) discrete choice parameter estimation. The model, presented in Section 3.1, outputs a distribution of feasible schedules for given individuals, each with socio-demographic characteristics and timing preferences (desired start time and duration for each activity or group of activities). These features impact the utility each individual gains from their daily schedule, according to the estimates of the parameters. These estimates are obtained by defining the scheduling process as a discrete choice problem, and deriving the parameters that maximise the likelihood function. This procedure is explained in Section 3.2. The likelihood function, as defined by Train (2009), requires an enumeration of the alternatives of the choice set. We present a methodology to generate an appropriate choice set in Section 3.3.

3.1 Scheduling framework

We use the same definition of a schedule as Pougala et al. (2022): it is a sequence of *activities*, starting and ending at home, over a time horizon T. An activity a is uniquely characterised by a location ℓ_a, a start time x_a, a duration τ_a, a cost of participation c_a and an outbound trip to the location of the next activity with a mode of transportation m_a. The boundary conditions (start and end of the schedule at home), are modelled as two dummy activities "dawn" and "dusk".

Figure 1 shows an example of schedule for one person, which includes 3 out-of-home activities (escort, errands, and leisure). The trips between each location are made by car.

Each schedule S is associated with a utility function U_S, which captures the preferences of the individual for the schedule. For example, time sensitivity is included through the scheduling preferences: a desired start time x_a^* and duration τ_a^*(single values or time intervals) for the activity.

As defined in Pougala et al. (2022), the schedule utility U_S is the sum of a generic utility U associated with the whole schedule and utility components capturing the activity-travel behaviour:

$$U_S = U + \sum_{a=0}^{A-1} (U_a^{\text{participation}} + U_a^{\text{start time}} + U_a^{\text{duration}} + \sum_{b=0}^{A-1} U_{a,b}^{\text{travel}}). \quad (1)$$

The components and the associated assumptions are defined as follows:

1. A generic utility U that captures aspects of the schedule that are not associated with any activity (e.g. resource availability at the level of the household).
2. The utility $U_a^{\text{participation}}$ associated with the participation of the activity a, irrespective of its starting time and duration.

$$U_a^{\text{participation}} = \gamma_a + \beta_{\text{cost}} c_a + \varepsilon_{\text{participation}}, \quad (2)$$

where γ_a and β_{cost} are unknown parameters to be estimated from data, and $\varepsilon_participation$ is an error term.

3. The utility $U_a^{\text{start time}}$, which captures the perceived penalty created by deviations from the preferred starting time.

$$U_a^{\text{start time}} = \theta_a^{\text{early}} \max(0, x_a^* - x_a) + \theta_a^{\text{late}} \max(0, x_a - x_a^*) + \varepsilon_{\text{start time}}, \quad (3)$$

where $\theta_a^{\text{early}} \leq 0$ and $\theta_a^{\text{late}} \leq 0$ are unknown parameters to be estimated from data, and $\varepsilon_{\text{start time}}$ is an error term.

The first (resp. second) term captures the disutility of starting the activity earlier (resp. later) than the preferred starting time.

4. The utility U_a^{duration} associated with duration. This term captures the perceived penalty created by deviations from the preferred duration.

$$U_a^{\text{duration}} = \theta_a^{\text{short}} \max(0, \tau_a^* - \tau_a) + \theta_a^{\text{long}} \max(0, \tau_a - \tau_a^*) + \varepsilon_{\text{duration}} \quad (4)$$

where $\theta_a^{\text{short}} \leq 0$ and $\theta_a^{\text{long}} \leq 0$ are unknown parameters to be estimated from data, and $\varepsilon_{\text{duration}}$ is an error term. Similarly to the specification of start time, the first (resp. second) term captures the disutility of performing the activity for a shorter (resp. longer) duration than the preferred one,

5. For each pair of locations (ℓ_a, ℓ_b), respectively, the locations of activities a and b with $a \neq b$, the utility $U_{\text{travel}}^{a,b}$ associated with the trip from ℓ_a to ℓ_b. irrespective of the travel time. This term is composed of the penalty associated with the travel time ρ_{ab}, and other travel variables (including variables such as cost, level of service, etc.) Here, we illustrate the framework with a specification involving travel cost. It also includes an error term, capturing the unobserved variables.

$$U_{a,b}^{\text{travel}} = \beta_{t,\text{time}} \rho_{ab} + \beta_{t,\text{cost}} c_t + \varepsilon_{\text{travel}} \quad (5)$$

where $\beta_{t,\text{time}}$ and $\beta_{t,\text{cost}}$ are unknown parameters to be estimated from data, and $\varepsilon_{\text{travel}}$ is an error term.

The schedules generated by the simulator must be *feasible*, according to a set of constraints defined at the level of the individual or the household by the modeller. For example, a schedule is feasible if:

- it does not exceed the maximum (time or cost) budget,
- each activity starts when the trip following the previous activity is finished,
- trips using mode m are only made if and when m is available,
- each activity meets its respective requirements (e.g. participation of other members of the household, feasible time windows, follows/precedes another activity)
- ...

The parameters involved in the utility function are summarised in Table 3. Indices S, a, and n denote respectively a schedule, an activity and an individual. The *Logit model* column indicates which parameters are estimated in the current study, with results presented in Section 4.

Parameter	Notation	Associated variable	Estimated
Alternative-specific constants	$\gamma_{S,n}$	-	
Activity-specific constant	$\gamma_{a,n}$	-	Yes
Cost of activity participation	β_{cost_a}	Cost c_a	
Penalty start time (early)	θ_a^{early}	Deviation start time δ_{e,x_a}	Yes
Penalty start time (late)	θ_a^{late}	Deviation start time δ_{ℓ,x_a}	Yes
Penalty duration (short)	θ_a^{short}	Deviation duration δ_{s,τ_a}	Yes
Penalty duration (long)	θ_a^{long}	Deviation duration δ_{ℓ,τ_a}	Yes
Travel cost	$\beta_{t,cost}$	Cost c_t	
Travel time	$\beta_{t,time}$	Time ρ_{ab}	

Table 2: Parameters of the utility function

3.2 Parameter estimation

The scheduling process can be defined as a discrete choice model where the alternatives are full daily schedules, each associated with a utility.

In principle, maximum likelihood estimation requires complete enumeration of the alternatives in the choice set. It is possible, though, to estimate the parameters using only a sample of alternatives. This is actually necessary in the activity-travel context, where the full choice set C_n of alternatives is combinatorial and characterised by complex constraints. For each individual n in the sample, we consider a sample of alternatives $\tilde{C}_n$. The maximisation of the likelihood function yields consistent parameter estimates if a correction term $\ln P_n(\tilde{C}_n|i)$ is introduced to take into account sampling biases (Ben-Akiva and Lerman, 1985):

$$P_{in} = P_n(i|\tilde{C}_n) = \frac{e^{\mu V_{in} + \ln P_n(\tilde{C}_n|i)}}{\sum_{j \in \tilde{C}_n} e^{\mu V_{jn} + \ln P_n(\tilde{C}_n|j)}} \quad (6)$$

Parameter	Notation	Associated variable	Logit model
Alternative-specific constants	$\gamma_{S,n}$	-	
Activity-specific constant	$\gamma_{a,n}$	-	Yes
Cost of activity participation	β_{cost_a}	Cost c_a	
Penalty start time (early)	θ_a^{early}	Deviation start time δ_{e,x_a}	Yes
Penalty start time (late)	θ_a^{late}	Deviation start time δ_{ℓ,x_a}	Yes
Penalty duration (short)	θ_a^{short}	Deviation duration δ_{s,τ_a}	Yes
Penalty duration (long)	θ_a^{long}	Deviation duration δ_{ℓ,τ_a}	Yes
Travel cost	$\beta_{t,\text{cost}}$	Cost c_t	
Travel time	$\beta_{t,\text{time}}$	Time ρ_{ab}	
Error term (participation)	$\varepsilon_{\text{participation}}$	-	
Error term (start time)	$\varepsilon_{\text{start time}}$	-	
Error term (duration)	$\varepsilon_{\text{duration}}$	-	
Error term (travel time)	$\varepsilon_{\text{travel}}$	-	

Table 3: Parameters of the utility function. The *Logit model* column indicates whether the parameter is estimated in the logit specification

The alternative-specific correction term $\ln P_n(\tilde{C}_n|i)$ is the logarithm of the conditional probability of sampling the choice set $\tilde{C}_n$ given that i is the alternative chosen by person n. This value depends on the protocol used to generate the choice set.

Each component of the utility function (Equations 2-5) is associated with a random term. This defines a mixed logit model with error components, by creating correlations between alternatives which share the same values for each dimension. The model reduces to a simple logit model if we assume the error terms to be i.i.d. and Extreme Value distributed, meaning that there is no correlation between alternatives. This assumption is adopted in the case study presented in Section 4.

3.3 Choice set generation

The estimation of parameters using maximum likelihood estimation requires an evaluation of the likelihood function for each alternative of the choice set $\tilde{C}_n$. If $\tilde{C}_n$ is a subset of the universal choice set of alternatives C_n, the likelihood function must be corrected with the probability of sampling the choice set $\tilde{C}_n$ given the chosen alternatives (6). This probability depends on the generation protocol for the sample. The procedure must therefore be able to produce tractable probabilities, while ensuring the generation of a pertinent choice set for the estimation of parameters.

More specifically, the choice set should contain alternatives with high probability of being chosen, to represent a choice set that the individual would actually consider. However, estimating a model with such a choice set would lead to biased model parameters, which would, in turn, decrease the accuracy and realism of the model predictions. On the other hand, the size of the solution space requires a strategic procedure to sample alternatives, to avoid only selecting non-informative, or low probability, schedules. The strategy to build the choice set must therefore generate an ensemble of high probability schedules, to estimate significant and meaningful parameters, while still containing low probability alternatives to decrease the model bias. (Bierlaire and Krueger, 2020).

The importance sampling of alternatives with the Metropolis-Hastings algorithm (Flötteröd and Bierlaire, 2013, Danalet and Bierlaire, 2015) is a good strategy to achieve this objective, while keeping tractable probabilities to derive the sample correction for the likelihood function.

The Metropolis-Hastings algorithm (Hastings, 1970) is a Markov Chain Monte-Carlo method used to generate samples from a multidimensional distribution, using a predefined acceptance/rejection rule. The procedure is summarised in algorithm 1.

Algorithm 1 Metropolis-Hastings algortihm (Gelman et al., 1995)

Choose starting point X_0 from starting distribution $p(X_0)$
for $t = 1, 2, ...$ **do**
    Sample a candidate point $X*$ from a transition distribution $q(X * | X_{t-1})$
    Compute acceptance probability $\alpha(X_{t-1}, X^*) = \min\left(\frac{p(X^*)q(X_{t-1}|X^*)}{p(X_{t-1})q(X^*|X_{t-1})}\right)$
    With probability $\alpha(X_{t-1}, X^*)$, $X_t \leftarrow X^*$, else $X_t \leftarrow X_{t-1}$

Each iteration of the random walk is therefore composed of two main steps:

1. Generation of a candidate point,
2. Acceptance or rejection of the candidate point.

In the context of the activity-based framework, each point (or state) is a schedule, and the target distribution is the schedule utility function (Equation (1)).

3.3.1 Generation of a candidate point

We define X_t the state(or point) at time t. X_t is a 24 hour schedule, discretised in *blocks* of duration $\tau \in [\tau_{min}, 24 - \tau_{min}]$ (with τ_{min} the minimum block duration). The new candidate point is a neighbouring schedule X^*, i.e. a schedule that only differs in one dimension (time, space, or activity participation - see fig. 2). We define heuristics (operators) $\omega \in \Omega$ to create X^* by modifying the current state. X^* is then accepted or rejected with a given acceptance probability.

Each operator ω can be selected with a probability P_ω, decided by the modeller.

Each schedule X_t is characterised by one or more *anchor* nodes ν, at the start of a block, indicating the position of the operator changes. In this context, each block corresponds to the temporal magnitude of the change.

Each operator must generate a feasible schedule, as defined in Section 3.1. In addition, the following conditions must be satisfied by the algorithm:

- Each iteration of the Metropolis-Hastings algorithm must be irreducible, meaning that each state of the chain can be reached in a single step:

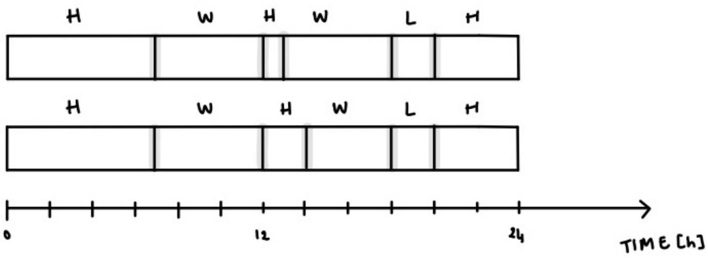


Figure 2: Example of neighbouring schedules. The schedules differ in the duration of the time spent at home during lunch time.

$$Q(X_t|X_{t-1}) > 0 \qquad \forall X_t, X_{t-1} \tag{7}$$

For this reason, each operator should apply single changes, or the combination of operators should lead to a state that can only be reached with this combination.

- Each iteration of the Metropolis-Hastings algorithm must be reversible, i.e. the forward probability (probability to do the change) and backward probability (probability to undo the change and go back to the previous state) must be strictly positive.

$$Q(X_t|X_{t-1}) > 0 \qquad \forall X_t, X_{t-1} \tag{8}$$
$$Q(X_{t-1}|X_t) > 0 \qquad \forall X_t, X_{t-1} \tag{9}$$

Defining single change operators enables to derive tractable probabilities.

The following list describes examples of operators that meet these requirements. Other operators can be created according to the modeller's needs and specifications. We illustrate their effect on an example schedule, shown in Figure 3. In its initial state, we assume time to be discretised in 24 blocks of length $\delta = 1h$. We consider two activities: *work* and *leisure*, each associated with a start time x_w and x_l, a duration τ_w and τ_l, and locations ℓ_w, ℓ_l. Considering that home is at location ℓ_h (and $\ell_h \neq \ell_w \neq \ell_l$), the individual travels to the other activities using modes m_w and m_l.

Anchor The anchor operator ω_{anchor} adds an anchor node v in the schedule. This change does not affect the activity sequence, but allows to change the position of the potential modifications of the other operators.

The transition probability associated with this change is the probability of selecting one of the existing blocks as anchor node.

Assign The assign operator ω_{assign} assigns an activity $j \in \mathcal{A}$ to a block of duration δ at position v, which was previously assigned to activity i. $\mathcal{A}$ is a set of N possible activities. The assignment is done with replacement, which means that $P(i = j) > 0$. To respect validity requirements, the resulting schedule must always start and end at home.

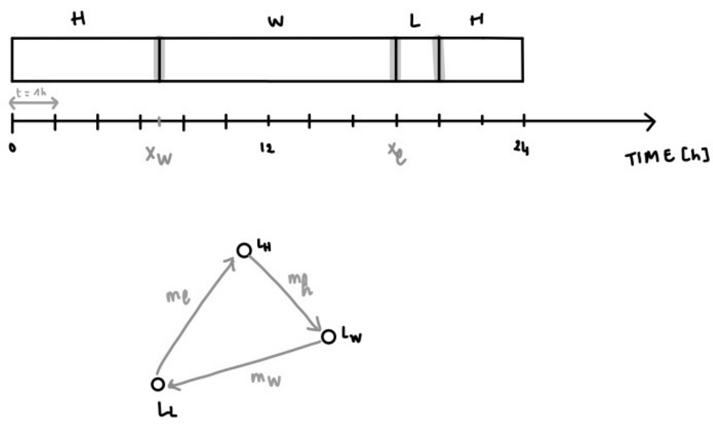


Figure 3: Initial schedule

Figure 4 illustrates an example of modification applied by the *assign* operator on the initial schedule.

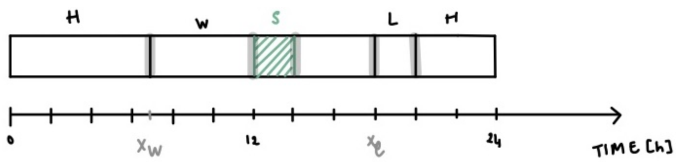


Figure 4: Change applied by the *assign* operator

Swap The operator ω_{swap} randomly swaps two adjacent blocks. A block at position ν, b_ν, is randomly selected, then is swapped with the following block. In order to respect the validity requirements, the resulting schedule must always start and end at home.

Figure 5 illustrates an example of modification applied by the *swap* operator on the initial schedule.

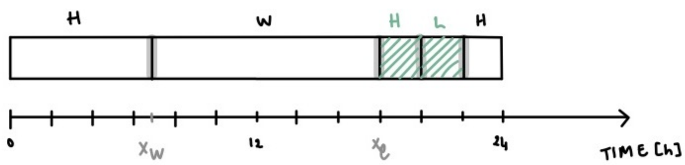


Figure 5: Change applied by the *swap* operator

Inflate/Deflate The *inflate/deflate* operator $\omega_{\text{inf/def}}$ allows to perform a shift of the schedule by randomly inflating the duration (i.e. adding one block of length δ) of the activity i at position ν and deflating the duration (i.e. removing one block of length δ) of an activity j of the schedule. The direction of the inflation and deflation (affecting the previous or following block of the selected one) is randomly chosen. If $i = j$, the operator only shifts the start time of the activity, while maintaining its duration. This operator modifies durations without generating infeasible schedules (e.g. schedules with a total duration that is different than the time budget). In order to ensure the validity constraint that the schedule must start and end at home, the first and last time block of the schedule cannot be modified.

Figure 6 illustrates an example of modification applied by the *inflate/deflate* operator on the initial schedule.

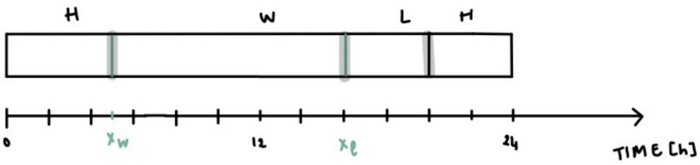


Figure 6: Change applied by the *inflate/deflate* operator

Location The *location* operator ω_{loc} changes the location ℓ_i of a randomly selected activity i at position ν, with probability P_{loc}. The new location is selected from a set of locations $\mathcal{L}$ that is considered known. The travel times following this change are recomputed, and any excess or shortage of time as compared to the available time budget is absorbed by the time at home. For this reason, and to remain compliant with validity constraints, the resulting change cannot go over the time budget by more than the minimum time at home (i.e. 2δ). In addition, the home location ℓ_h cannot be changed. The selection of a location must therefore be done according to a distribution $P_\ell(\rho)$ which is conditional on the travel times ρ. We assume that this distribution is exogenous to the choice-set generation algorithm.

Figure 7 illustrates an example of modification applied by the *location* operator on the initial schedule.

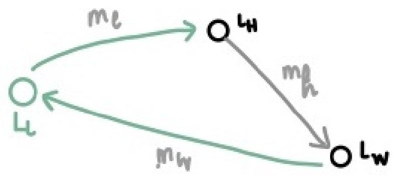


Figure 7: Change applied by the *location* operator

Mode Similarly to the *location* operator, the *mode* operator ω_{mode} changes the mode m of the outbound trip of a randomly selected activity i at position ν.. The new mode is selected from a set of modes $\mathcal{M}$ that is considered known. The travel times following this change are recomputed, and any excess or shortage of time as compared to the available time budget is absorbed by the time at home. For this reason, and to remain compliant with validity constraints, the resulting change cannot go over the time budget by more than the minimum time at home (i.e. 2δ). The selection of a mode must therefore be done according to a distribution $P_m(\rho)$ which is conditional on the travel times ρ. We assume that this distribution is exogenous to the choice-set generation algorithm. As the last home activity is not linked to an outbound trip, it cannot be selected for a mode change.

Figure 8 illustrates an example of modification applied by the *mode* operator on the initial schedule.

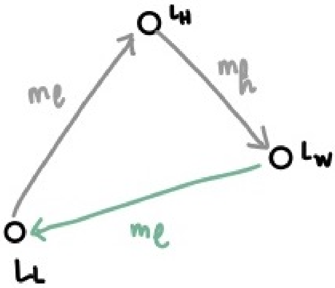


Figure 8: Change applied by the *mode* operator

Block The block operator ω_{block} modifies the time discretisation by changing the length δ of the schedule blocks (e.g. from $\delta = 30$ to $\delta = 15$ minutes). This change does not affect the activity sequence, but allows to change the scale of the potential modifications of the other operators.

The transition probability associated with this change is the probability of selecting one of the possible discretisations.

Figure 9 illustrates an example of modification applied by the *block* operator on the previously introduced initial schedule.

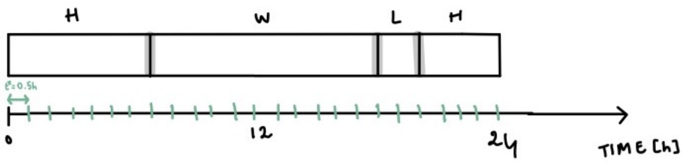


Figure 9: Change applied by the *block* operator

Combination This meta-operator ω_{meta} combines n distinct operators from the full set of operators Ω. n is an arbitrary number such that $n \in 2, ..., N_{op}$, with N_{op} the number of available operators. The transition probabilities of the change are the combined forward (resp. backward) probabilities of the selected operators. Combining operators through a meta-operator instead of randomly selecting them “on the fly" during the random walk process offers the advantage of making it easier for the modeller to track the behaviour of the process. Specifically, the impact of each operator, whether applied individually or in conjunction with others, can be evaluated.

We summarise the previous list in Table 4. As previously mentioned, this list is not exhaustive: other operators can be created or combined to fit the requirements of the intended applications, or simply to improve the performance of the MH algorithm.

Name	Choice dimension	Description	Probability
Anchor	-	Adds or deletes an anchor node	P_{anchor}
Assign	Activity participation	Assigns activity to a given block	P_{assign}
Swap	Activity participation, Time	Swaps the activities of two adjacent blocks	P_{swap}
Inflate, Deflate	Time	Inflates or deflates the duration of a given activity	$P_{inf, def}$
Mode	Mode of transportation	Changes the mode of transportation associated with activity	P_{mode}
Location	Location	Changes the location associated with activity	P_{loc}
Block	-	Modifies time discretisation of the schedule	P_{block}
Meta-operator	All	Combines two or more operators	P_{meta}

Table 4: Example of operators

More details on the operators and the derivations of the transition probabilities can be found in Pougala et al. (2021).

3.3.2 Acceptance of candidate points

The target distribution of the MH algorithm is the schedule utility function (Equation (1)), conditional on the distribution of the error terms, and with unknown parameters to be estimated. The acceptance probability is defined by:

$$\alpha(X_{t-1}, X^*) = \min\left(\frac{p(X^*)q(X_{t-1}, X^*)}{p(X_{t-1})q(X^*, X_{t-1})}\right) \quad (10)$$

where X^* is the candidate state, $p(i)$ is an unnormalised positive weight, proportional to the target probability (Flötteröd and Bierlaire, 2013) and $q(i, j)$ is the transition probability to go from state i to state j.

Similarly to Danalet and Bierlaire (2015), for each state X_t, the target weight $p(X_t)$ is defined by:

$$p(X_t) = \tilde{U_S}(X_t) \quad (11)$$

where $\tilde{U_S}$ is a schedule utility function with the same specification as the target (Equation (1)) but with parameters calibrated on a randomly generated choice set.

The transition distribution q is directly obtained from the working operator.

Therefore, the general algorithm (algorithm 1) can be adapted to the ABM context, as summarised in Algorithm 2.

Algorithm 2 Choice set generation for the ABM with Metropolis-Hastings

$t \leftarrow 0$, initialise state with random schedule $X_t \leftarrow S_0$
Initialise utility function with random parameters $\tilde{U_S}$
for $t = 1, 2, ...$ **do**
  Choose operator ω with probability P_ω
  $X^*, q(X_t, X^*) \leftarrow$ **ApplyChange**(ω, X_t)
  function APPLYCHANGE(ω, state X)
    return new state X', transition probability $q(X, X')$
  Compute target weight $p(X^*) = U_S\tilde{(}X^*)$
  Compute acceptance probability $\alpha(X_t, X^*) = \min\left(\frac{p(X^*)q(X_t|X^*)}{p(X_t)q(X^*|X_t)}\right)$
  With probability $\alpha(X_t, X^*)$, $X_{t+1} \leftarrow X^*$, else $X_{t+1} \leftarrow X_t$

Following Ben-Akiva and Lerman (1985), we define for each individual n the alternative specific corrective term for a choice set C_n of size $J+1$ with $\tilde{J}$ unique alternatives (Equation (12)). Each alternative j is sampled from the target distribution of the Metropolis-Hastings algorithm with probability q_{jn}, such that $q_{jn} = 0$ if $j \notin C_n$.

$$q(\mathcal{C}_n|i_n) = \frac{1}{q_{in}} \prod_{j \in C_n} \left(\sum_{j \in C_n} q_{jn} \right)^{J+1-\tilde{J}} \tag{12}$$

3.3.3 Implementation notes

Selection probabilities for operators The probabilities of selecting and applying an operator are arbitrary and to be defined by the modeller. An iterative approach to the choice set generation might highlight an imbalance in the rate of accepted schedules per generating operator. In this case, an equilibrium can be achieved by fine-tuning the operator choice probabilities, e.g. by selecting fewer times the operators that are more likely to produce accepted changes.

Schedule feasibility The states generated by the process must meet validity criteria such as starting and ending at home, or having consistent timings between consecutive activities. One risk when defining operators is that they change a current feasible schedule into an infeasible state. For example, changing the duration of an activity may lead to a total duration that differs from the available time budget. One solution, as done in this paper, is to define operators that do not inherently induce infeasibility. This provides the advantage of making the transition probabilities easier to compute, but limits the possible changes that can be applied. On the other hand, allowing for infeasibility in the operators results can lead to more varied results. An operator that restores feasibility at the end of the process (e.g. modifying the time spent at home to absorb timing gaps or excesses in the schedule). However, as these changes would be dependent on the current state, computing the associated transition probabilities would prove more difficult.

Target weights The target weights for each state X_t are defined as the utility function evaluated at X_t. However, the function evaluation is conditional on the values of its parameters, that we attempt to estimate with the random walk. Lemp and Kockelman (2012) proposes an iterative process to compute the weights in importance sampling, by updating the weights with models estimated at the previous iterations. For example, Danalet and Bierlaire (2015) use parameters calibrated on a randomly generated choice set as a starting point for their Metropolis-Hastings process.

Initial schedule The methodology requires the initialisation of starting point, which is arbitrarily chosen. A randomly generated schedule can be used for this task, but for the sake of model efficiency and realism of the resulting choice, starting with a known high-probability schedule (e.g. a daily schedule that was recorded in a travel survey) can be considered. This allows to select more efficiently alternatives that are likely to be considered by the individual. However, as discussed previously, one must be careful to also include lower probability alternatives. The parameters of the random walk (e.g. acceptance ratio) must thus be adjusted to avoid such biases.

4 Empirical investigation

The objective of the empirical investigation is to apply the methodology on a real-life case study to illustrate the parameter estimation procedure. We use the Mobility and Transport Microcensus (MTMC), a Swiss nationwide survey gathering insights on the mobility behaviours of local residents (Office fédéral de la statistique and Office fédéral du développement Territorial, 2017). Respondents provide their socio-economic characteristics (e.g. age, gender, income) and those of the other members of their household. Information on their daily mobility habits and detailed records of their trips during a reference period (1 day) are also available. The 2015 edition of the MTMC contains 57'090 individuals, and 43'630 trip diaries. In order to illustrate a real-life application of the simulator, we focus on the sample of full-time students residing in Lausanne (236 individuals).

We start by generating the choice sets of daily schedules for each individual in the sample. Each choice set is composed of 10 alternatives, including the chosen (recorded) schedule.

The second step, once the choice sets have been generated, is to estimate the parameters of the utility function for the sample. For each individual and each alternative in their respective choice sets, we evaluate the sample correction term (eq. (12)) to be added to the utility function.

4.1 Choice sets

For each person in the train dataset, we generate a choice set of 10 alternatives (including the observed schedules) following the methodology presented in Section 3.3. The initial state X_0 of the random walk is the observed schedule.

We implement 6 operators: *Block* and *Anchor*, which influence the impact of the other operators, and *Assign, Swap, Inflate/Deflate* which modify the schedule directly. A *Meta*-operator is implemented to combine the actions of two or more operators. Each operator can be chosen with a uniform probability $P_{\text{operators}}$.

The target distribution of the random walk is the utility function of the activity-based model (Equation (1)), with a set of parameters β_0 that were estimated using randomly sampled choice sets. The target weights are evaluations of this utility function for the current state.

The random walk (Algorithm 2) is performed for a number of iterations n_{iter}. We discard $n_{\text{warm-up}}$ of these iterations to sample from a stabilised distribution. To create the choice set, we draw 9 alternatives by only keeping 1 out of n_{skip} schedules.

Feature	Definition	Value
Ω	Set of operators	Block, Assign,Anchor, Swap, Inf/Def, Meta
$N_{\text{operators}}$	Number of operators	6
$P_{\text{operators}}$	Operator selection probability	$1/N_{\text{operators}}$
n_{iter}	Number of iterations	100'000
$n_{\text{warm-up}}$	Warm-up iterations	50'000
n_{skip}	Skipped iterations	20

Table 5: Experimental set up of the random walk

The algorithm was run on a server (2 Skylake processors at 2.3 GHz and 192GB RAM, with 18 CPUs each, running in parallel) for each of the 236 students in the sample, for a total runtime of 2.22 minutes.

4.2 Logit specification

In this paper, we test the parameters resulting from three model specifications:

1. **Model 0:** A generic utility function with parameters from the literature (not estimated).
2. **Model 1:** A generic utility function, where we classify activities according to two levels of flexibility, and estimate the corresponding parameters for both categories.
3. **Model 2:**. An activity-specific utility function, where we estimate all activity-specific parameters and constants.

Both Model 0 and Model 1 are used as benchmarks for the estimates of Model 2.

We consider 5 different activities: home, work, education, leisure and shopping.

Following the definition of Pougala et al. (2022), travel is not considered as a standalone activity, but is always associated with the origin activity of the trip, if applicable.

We make the following additional simplifications:

- We do not estimate travel parameters, and consider them null in eq. (1),
- The scheduling preferences (desired start time and durations) are derived from the dataset. For each activity, we fit a distribution (either normal or log normal) across the student population. The calibrated parameters are reported in table 6. For each person, we draw values of desired start times and durations from these distributions.

Activity	Start time	Duration
Home	-	$\mathcal{N}(17.4, 3.4)$
Work	Log-$\mathcal{N}(0.65, 4.2, 3.4)$	$\mathcal{N}(7.6, 3.7)$
Education	Log-$\mathcal{N}(0.4, 6.2, 1.7)$	$\mathcal{N}(6.7, 2.1)$
Leisure	$\mathcal{N}(14.3, 3.5)$	$\mathcal{N}(3.5, 2.7)$
Shopping	Log-$\mathcal{N}(0.3, 4.6, 9.0)$	Log-$\mathcal{N}(1.3, 0.15, 0.32)$

Table 6: Desired times distributions in sample

Therefore, the utility function defined in eq. (1) can be written as follows for model 1:

$$U_S^1 = \gamma_a + \sum_f \lambda_f^a [\theta_f^{\text{early}} \max(0, x_a^* - x_a) + \theta_f^{\text{late}} \max(0, x_a - x_a^*) \\ + \theta_f^{\text{short}} \max(0, \tau_a^* - \tau_a) + \theta_f^{\text{long}} \max(0, \tau_a - \tau_a^*)] + \varepsilon_S$$

with f a category of flexibility $f \in$ {Flexible, Not Flexible}. λ_f^a is an indicator variable that is 1 if activity a belongs to category f, and is an input to the model. In this case study, education and work are considered not flexible, while leisure, shopping and home are considered flexible.

For model 2, the utility function can be written:

$$U_S^2 = \gamma_a + \sum_a [\theta_a^{\text{early}} \max(0, x_a^* - x_a) + \theta_a^{\text{late}} \max(0, x_a - x_a^*) \\ + \theta_a^{\text{short}} \max(0, \tau_a^* - \tau_a) + \theta_a^{\text{long}} \max(0, \tau_a - \tau_a^*)] + \varepsilon_S$$

Both models are estimated with PandasBiogeme (Bierlaire, 2020). The estimation process is done using 70% of observations in the sample data, where one observation is the daily schedule of one individual.

Finally, we simulate daily schedules for the Lausanne sample. In order to visualise the behaviour of the simulator conditionally upon the input parameters, we test three different sets of coefficients: (i) parameters from the literature (ii) generic parameters estimated on 70% of Lausanne data (estimation model 1) (iii) and activity-specific parameters estimated on 70% of Lausanne data (estimation model 2). We compare the schedule distributions and distributions of start times and durations resulting from these 3 configurations with the observed distribution from the dataset.

4.2.1 Parameters

Literature parameters The parameters from the literature were used in the first implementation of the framework, as described in Pougala et al. (2022). Values from the departure time choice literature (e.g. ratios from Small (1982)) were used to derive the parameters defined in table 7. The penalty parameters are specific to each flexibility category (flexible (F) or non flexible (NF) activities). In this set of parameters, we do not consider activity-specific constants ($\gamma_a = 0 \; \forall a \in A$). The assumption behind this is that, all else being equal, there is no inherent preference to perform any activity (home included). Any effect of this nature is therefore fully included in the random term of the schedule ε_S.

	Parameter	Param. estimate
1	F early	0.0
1	F late	0.0
2	F long	-0.61
2	F short	-0.61
3	NF early	-2.4
4	NF late	-9.6
5	NF short	-9.6
6	NF long	-9.6

Table 7: Parameters from the literature

Model 1 The home activity is used as a reference, such that $\gamma_{\text{home}} = 0$. The magnitudes and signs of the other constants are relative to the baseline behaviour which is staying at home. The estimated parameters are summarised in table 8. For flexible activities, the parameters indicate a similar behaviour to what is found in the literature: being late is more penalised than being early (approximately by a factor of 2). The penalties associated with duration have comparable magnitudes, although they are not statistically significant ($p > 0.05$). On the other hand, for non flexible activities, being early seems to be more negatively perceived than being late. The duration penalties are symmetrical.

Model 2 We consider both activity-specific constants and schedule deviation penalties. For all parameters, the home activity is set as a reference, such that $\gamma_{\text{home}} = 0$. As for model 1, the magnitudes and signs of the other coefficients are therefore relative to the home baseline. We estimate 20 parameters for this model (5 per activity), which are summarised in table 9.

	Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
1	F: early	-0.175	0.12	-1.46	0.145*
2	F: late	-0.333	0.14	-2.38	0.0171
3	F: long	-0.105	0.0722	-1.45	0.146*
4	F: short	-0.114	0.194	-0.585	0.559*
5	NF: early	-1.14	0.367	-3.10	0.00191
6	NF: late	-0.829	0.229	-3.61	0.0003
7	NF: long	-1.20	0.393	-3.05	0.00231
8	NF: short	-1.19	0.468	-2.54	0.0011
9	Education: ASC	16.0	2.46	6.49	8.63e-11
10	Leisure: ASC	8.81	1.7	5.17	2.28e-07
11	Shopping: ASC	6.85	1.80	3.80	0.000146
12	Work: ASC	16.0	2.58	6.18	6.57e-10

Table 8: Estimation results for Model 1 on student population. The asterisk (*) indicates parameters that are not statistically significant based on their p-value.

For education, all of the parameters are statistically significant. Being early is slightly less penalised than being late, although the penalties are almost symmetrical. The same observation can be made for the penalties associated with duration. For work, the penalty for being late is not statistically significant (p-value > 0.05), while being early is significantly penalised. The penalties associated with duration have a more negative impact on the utility function; in particular, the activity running for longer than desired is highly penalised.

Interestingly, most of the parameters associated with leisure are not statistically significant. This could indicate that leisure is not a particularly time constraining activity for students, in the sense that it is less likely to trigger trade-offs in the scheduling process than the other activities.

On the other hand, shopping displays high penalties for scheduling deviations, especially with respect to start time. This behaviour does not support the assumption used in the previous model that shopping is a flexible activity.

Figure 10 illustrates some schedules generated with activity-specific parameters.

4.3 Mixed logit specification

We relax the IIA assumption of the logit model by estimating an error components specification of the model. We create correlations between the utilities of the alternatives by estimating the error terms included in each component of the utility function (eqs. (4) to (5)). For practicality, we consider the following assumptions for the distributions of these error terms:

- *Participation*: for each activity a, the error term $\varepsilon^{a}_{\text{participation}} \sim \mathcal{N}(\mu_{\text{part},a}, \sigma^2_{\text{part},a})$ is included in the utility term related to the participation to a (Equation (2)) and captures

	Parameter	Param. estimate	Rob. std err	Rob. t-stat	Rob. p-value
1	Education: ASC	18.7	3.17	5.89	3.79e-09
2	Education: early	-1.35	0.449	-3.01	0.00264
3	Education: late	-1.63	0.416	-3.91	9.05e-05
4	Education: long	-1.14	0.398	-2.86	0.00428
5	Education: short	-1.75	0.457	-3.84	0.000123
6	Leisure: ASC	8.74	1.94	4.50	6.79e-06
7	Leisure: early	-0.0996	0.119	-0.836	0.403*
8	Leisure: late	-0.239	0.115	-2.07	0.0385
9	Leisure: long	-0.08	0.0617	-1.30	0.195*
10	Leisure: short	-0.101	0.149	-0.682	0.495*
11	Shopping: ASC	10.5	2.20	4.78	1.74e-06
12	Shopping: early	-1.01	0.287	-3.51	0.000443
13	Shopping: late	-0.858	0.237	-3.63	0.000284
14	Shopping: long	-0.683	0.387	-1.76	0.0779*
15	Shopping: short	-1.81	1.73	-1.04	0.297*
16	Work: ASC	13.1	2.64	4.96	7.16e-07
17	Work: early	-0.619	0.217	-2.85	0.00438
18	Work: late	-0.338	0.168	-2.02	0.0438
19	Work: long	-1.22	0.348	-3.51	0.000441
20	Work: short	-0.932	0.213	-4.37	1.23e-05

Table 9: Estimation results for Model 2 on student population. The asterisk (*) indicates parameters that are not statistically significant based on their p-value.

the unobserved correlations between schedules which include a,

- *Start time*: for each activity a, we define the error term $\varepsilon^a_{\text{start time}} = \xi^a_{AM}\delta^a_{AM} + \xi^a_{PM}(1-\delta^a_{AM})$, with ξ^a_{AM} and ξ^a_{PM} two normally distributed random quantities, respectively associated with a start time in the first half of the day ($x_a \in$ [00:00, 12:59]) and in the second ($x_a \in$ [13:00, 23:59]). δ^a_{AM} is an indicator variable that is equal to 1 if a starts in the first half of the day.

- *Duration*: we consider the error term $\varepsilon_{\text{duration}} = \xi_{\text{edu}}\delta_{\text{edu}} + \xi_{\text{other}}(1-\delta_{\text{edu}})$, where ξ_{edu} is a normally distributed term that captures the correlations between alternatives for which a significant amount of the available duration is spent doing a primary activity (education or work). δ_{edu} is an indicator variable that is equal to 1 if $\tau_a \geq 4h\ \forall a \in$ [education, work].

Each error component ξ_X is assumed to follow a normal distribution $\mathcal{N}(\mu_X, \sigma^2_X)$. For practicality, we assume that all the components have a zero mean ($\mu_X = 0$), and therefore only estimate the standard deviations σ_X.

Equation (1) can therefore be rewritten:

$$
\begin{aligned}
U_S = &\gamma_a \\
&+ \sum_a [\theta^{\text{early}}_a \max(0, x^*_a - x_a) + \theta^{\text{late}}_a \max(0, x_a - x^*_a) \\
&+ \theta^{\text{short}}_a \max(0, \tau^*_a - \tau_a) + \theta^{\text{long}}_a \max(0, \tau_a - \tau^*_a)\varepsilon^a_{\text{participation}} + \varepsilon^a_{\text{start time}}] \\
&+ \varepsilon_{\text{duration}} + \nu_S
\end{aligned}
$$

where ν_S is iid Extreme Value distributed.

Similarly to the logit model specification, we consider a generic (**Model 3**) and an activity-specific (**Model 4**) specification. The assumptions for the other parameters are the same as for the simple logit specification.

4.4 Simulation results

Using the parameters described in the previous section, we simulate schedules for the test dataset. The simulation procedure is described in detail in Pougala et al. (2022): at each iteration $i \leq n_{\text{max}}$, we draw a random term ε_i from a known distribution. We solve the utility maximisation problem for this error instance to obtain a draw from the underlying schedule distribution. We draw $n_{\text{max}} = 100$ schedules for each individual in the sample.

To compare the results of each model with the original data, we analyse the simulated frequencies of activity participation per hour of the day, simulated durations and start times for each activity. We compute the Kolmogorov-Smirnov statistic between the original and simulated distributions for a quantitave evaluation of the goodness-of-fit of these distributions.

Figure 10: Examples of simulated schedules (Model 2)

4.4.1 Simulated statistics

We compare some descriptive statistics of the simulated sample with those observed in the dataset. These statistics are daily averages of: the time spent out-of-home (total and for each activity) and proportion of schedule containing each activity. These statistics are derived exclusively for schedules which contain at least one activity out-of-home. It is worth noting that all three models generate significantly more fully-at-home days (about 5 times more than what is observed in the MTMC data).

The results are summarised in Table 10 and Table 11 respectively. The two estimated models (generic and activity-specific) generate average durations that are closer to the observed ones than the model with parameters from the literature. They are especially accurate for the average total time, but the proportions across activities are not as well captured. For example, the average durations spent in education are underestimated by about 1h, while the time spent in leisure is overestimated (by 2h in the case of the activity-specific model).

Activity	Data	Literature	Generic	Activity-specific
Total	04:53	02:54	04:10	05:19
Education	03:32	01:11	02:25	02:29
Leisure	00:39	00:58	01:17	02:32
Shopping	00:08	00:22	00:21	00:10
Work	00:26	00:05	00:07	00:08

Table 10: Average out-of-home duration, in hh:min

Regarding the proportion of schedules containing each activity (table 11) the activity-specific model significantly underestimates the frequency of each activity. This is likely due to the approximation of the desired start times, which are computed for only one instance of the activity, and do not properly account for bimodality or asymmetry in timing preferences (e.g. different desired start times for doing work in the morning or in the afternoon). This point in discussed further in Section 4.5.1.

Activity	Data	Literature	Generic	Activity-specific
Education	0.86	0.56	0.60	0.37
Leisure	0.98	0.52	0.56	0.75
Shopping	0.28	0.21	0.19	0.11
Work	0.09	0.06	0.05	0.02

Table 11: Proportion of schedules containing each activity

4.4.2 Time of day participation

Figure 11 shows the typical distribution of activities in the course of a day, for schedules including at least one activity out of home. The height of each bar represents the proportion of the sample that is participating in each activity at a given moment of time. Before

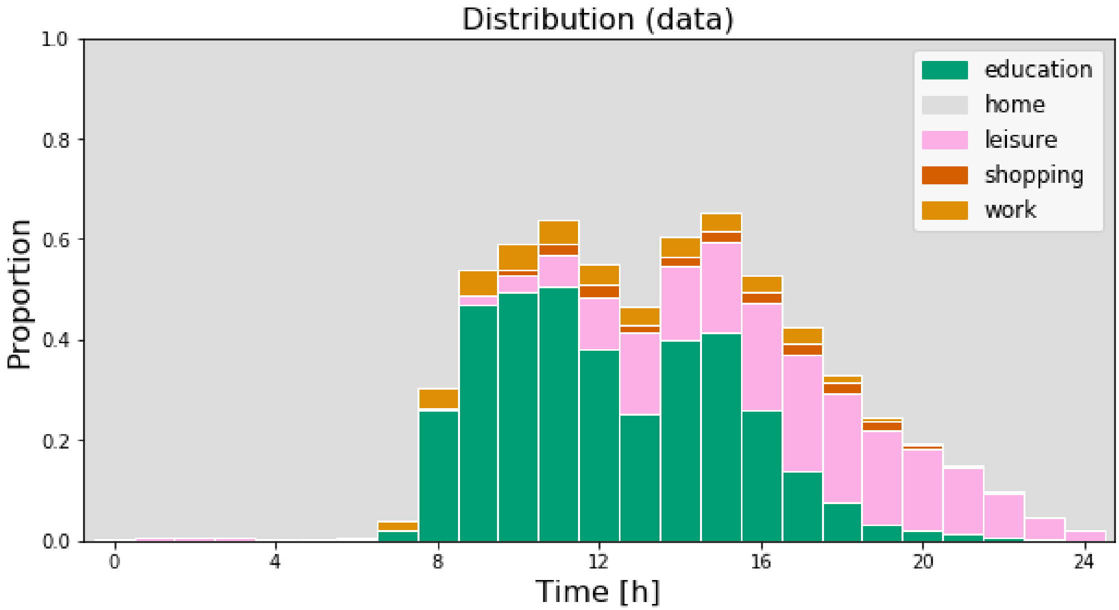


Figure 11: Observed time of day activity frequency. The height of the bars is the proportion of people participating in each activity at a given moment.

7:00, almost all of the individuals in the sample are home. The proportion of people undertaking their main *education* activity steadily increases during the morning, to reach a peak at 11h (50%). The proportion decreases at lunch time (40% to 25% between 12:00 and 13:00) and goes up again in the afternoon. The *leisure* activity is the second most frequent activity from 10:00 to 15:00. From 16:00 onward, it surpasses *education*. *Work* is the third most frequent activity, although in much smaller proportions than the previous two. Its profile is similar to *education*.

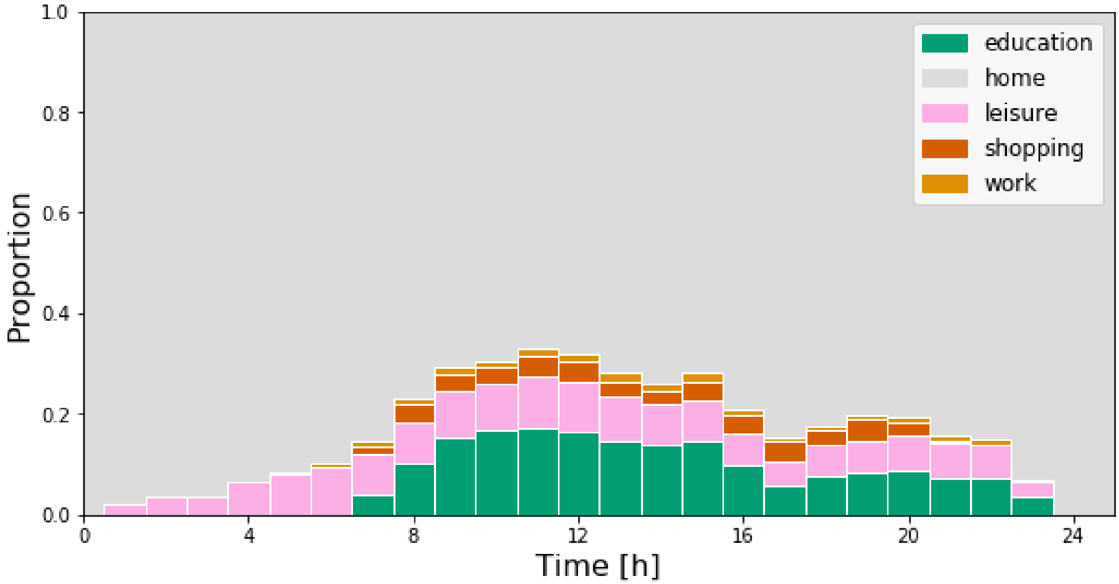
Figure 12 shows the distributions for out-of-home schedules[1] resulting from the simulator framework with the 3 mentioned configurations: with parameters from the literature (fig. 12a), generic schedule deviation penalties (fig. 12b), and activity-specific penalties and constants (fig. 12c). All three configurations are able to capture the importance of *education* relative to the other activities in the schedule. However, as mentioned in the previous section, for all models, the majority of generated schedules are full days at home (i.e. no out-of-home activity scheduled).

The original profile of the education activity, with a distinct peak period, is best captured with the estimated parameters, both generic and activity specific. In both cases, the peak is reached before 9:00, as opposed to the observed 11:00 peak. This discrepancy is likely due to the assumption of a unimodal desired start time; a multimodal distribution (closer to the observed one) would improve the fit of the simulated distribution.

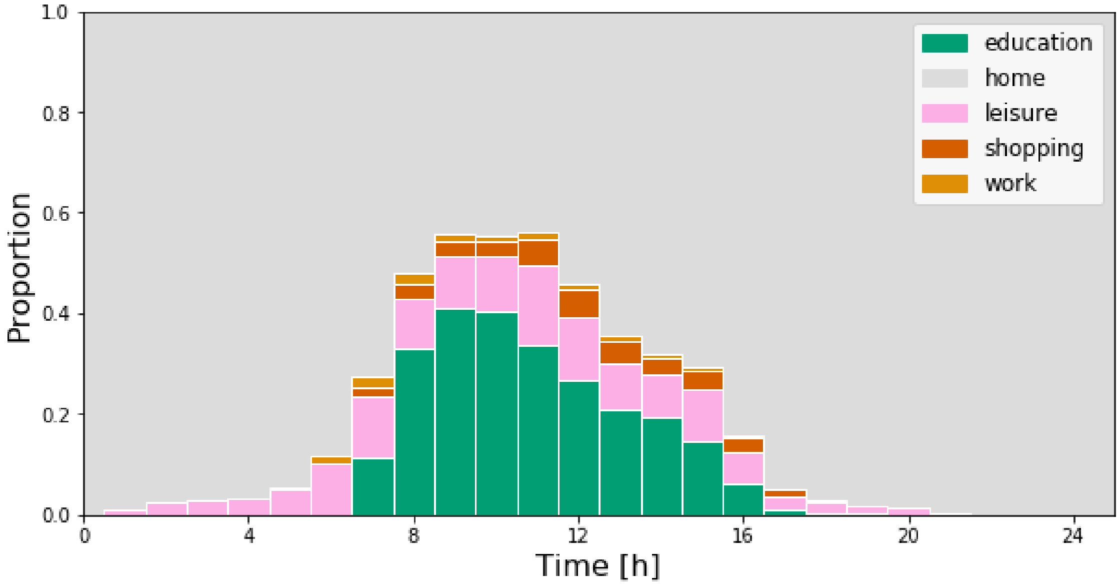
Interestingly, the *leisure* activity — and by extension, all activities previously defined as flexible — has very different simulated profiles from the observed one. With the literature parameters, the share of leisure is constant for most of the day, and comparable to the share of education. On the other hand, with the activity-specific parameters, the activity

[1]Out of the 20 simulated schedules for each individual in the sample.

is overrepresented during the night (midnight to 7:00), as compared to the other simulated activities, and the leisure observations in the data for this time period. The rest of day, the profile is similar to the real one.



(a) Parameters from literature

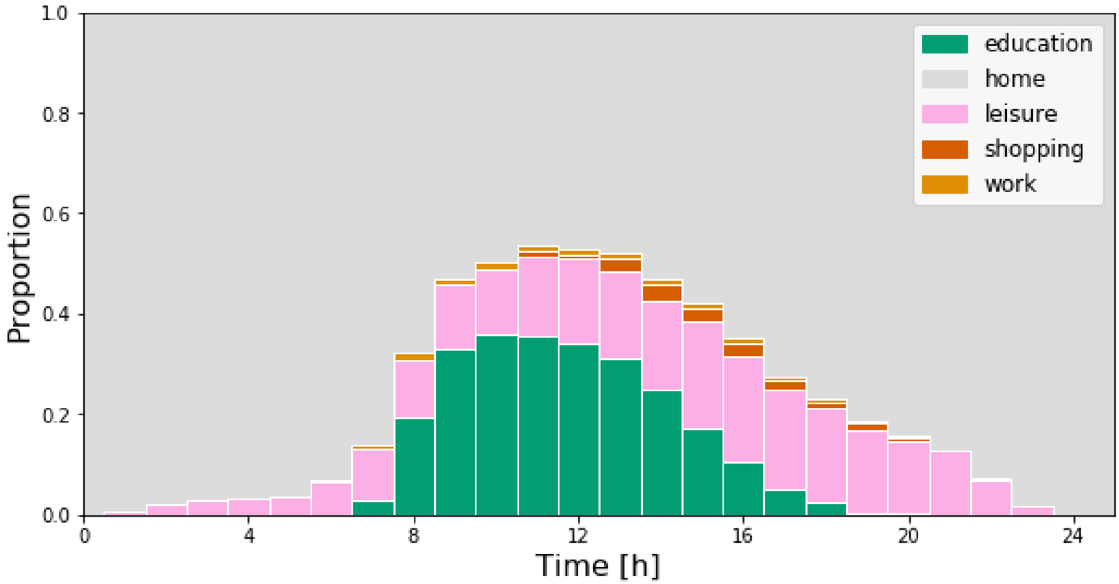


(b) Generic parameters (estimation model 1)

Figure 12: Simulated time of day activity frequency

4.4.3 Start time

We compare the simulated start times for each activity and each model, by visualising the kernel density estimations of the models with parameters from the literature, generic and activity-specific parameters (fig. 13), and respective Kolmogorov-Smirnov (KS) statistic compared to the observed dataset (a lower KS indicates a better fit).



(c) Activity-specific parameters (estimation model 2)

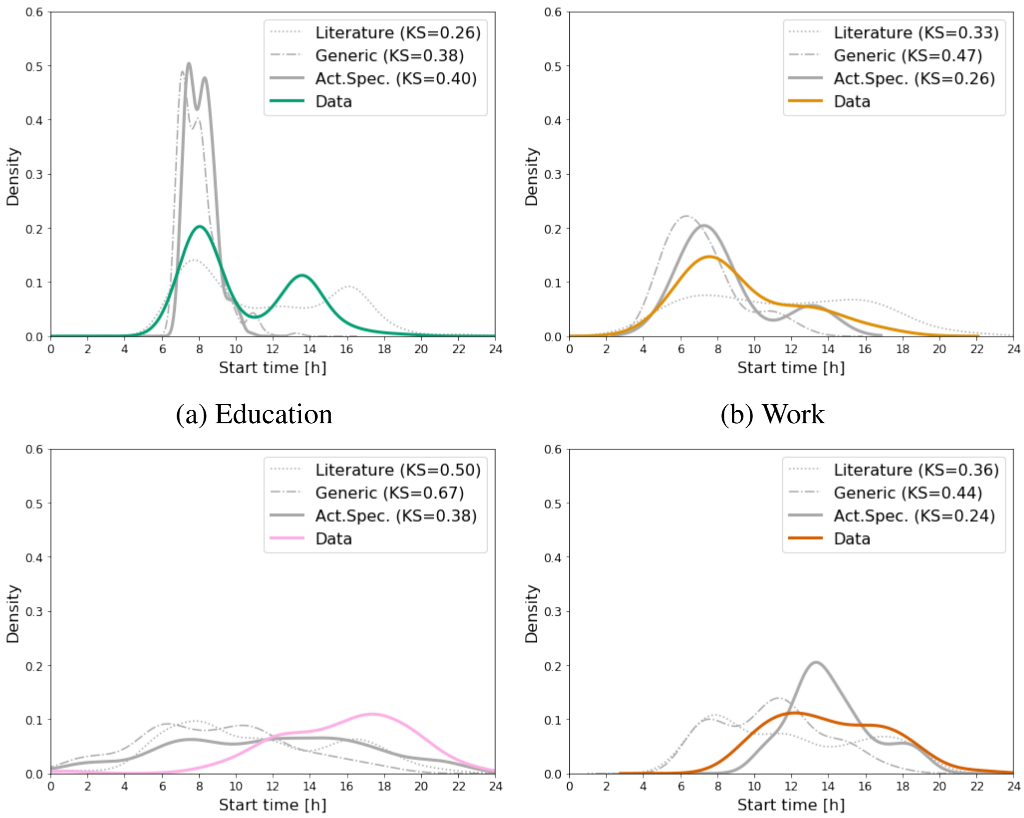
Figure 12: Simulated time of day activity frequency (cont.)

With the exception of *education* the activity-specific model is the model that better reproduces the distributions of start time (lowest KS). The observed distribution of *education* is truly bimodal, which is not properly captured by either of the estimated models. This is likely due to the approximation of desired times to a unimodal distribution. The model with parameters from the literature produces a relatively good fit, but this distribution varies very little from one activity to another.

4.4.4 Duration

Similarly, we compare the simulated durations for each activity and each model, by visualising the kernel density estimations of each model (fig. 14) and computing their respective KS statistic.

For all activities, the model with parameters from the literature tends to generate short activities ($\tau_a \leq 2$ hours) more frequently, and in smaller proportions activities with a duration of about 8 hours (for education, leisure and shopping). The two models with estimated parameters generate more diverse patterns with respect to duration: the generic model seems to capture well the bimodality of *education*. On the other hand, the activity-specific model generates better distributions for *work* and *leisure*. All three models tend to generate short instances of the *shopping* activity, although there is a non negligible number of schedules with very long shopping activities (8 hours), which is not close to what was observed nor particularly realistic. This limitation is also reflected by the high value of the KS statistic.



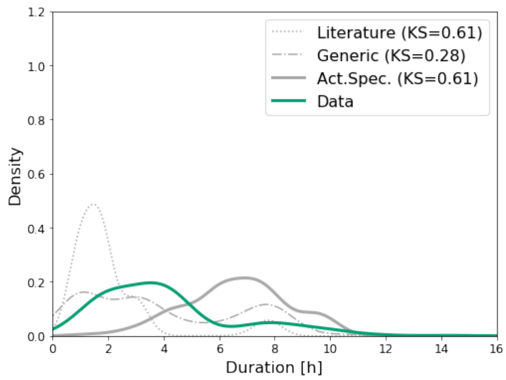
(a) Education

(b) Work

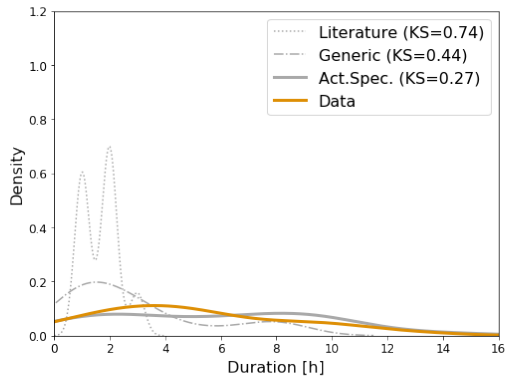
(c) Leisure

(d) Shopping

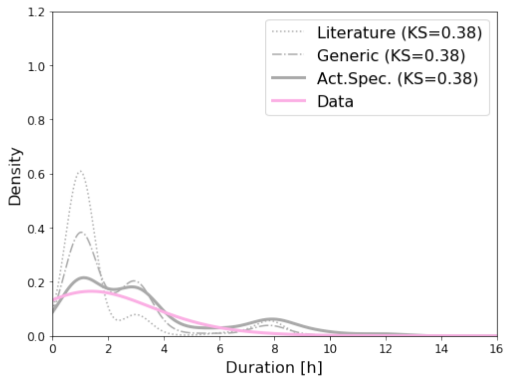
Figure 13: Simulated start times, per model and activity



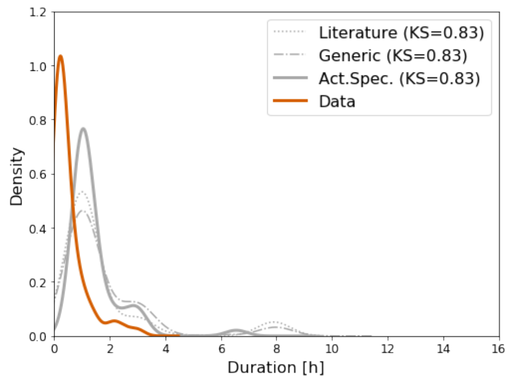
(a) Education



(b) Work



(c) Leisure



(d) Shopping

Figure 14: Simulated durations, per model and activity

4.5 Distance metrics

We quantify the performance of the three specifications by analysing the *distance* of the respective solutions (distributions of schedules as illustrated in fig. 12. The distributions are compared on the basis of several metrics:

1. Inspired by the work of Recker et al. (2008), we measure the error in the discrete components (activity participation/sequence) with the *edit distance* (the minimum number of changes required to make two sequences equal[2], where the changes can be the insertion, deletion or substitution of an element. Higher values for the edit distance indicate increasingly different sequences. Here, we consider a unidimensional edit distance and only measure the differences in the order of activities. To do so, we transform the schedule into a sentence where the words are activities, and compute the Levenshtein distance from the original sentence to each draw.

2. The continuous temporal components (start time, duration), are compared by deriving the average *time overlap* between the observed schedule and the simulated one, for a given activity. The time overlap is defined as the duration for which a given activity is taking place at the same time in both schedules. The average is computed over all simulated activities.

3. The Hellinger distance is computed to measure the similarity of the discretised time-of-day distributions of each activity. The Hellinger distance between two finite distributions P and Q is defined as:

$$h(P, Q) = \sqrt{\frac{1}{2}\sum_{i=1}^{k}(\sqrt{p} - \sqrt{q})^2} \tag{13}$$

   For the sake of illustration, we measure the similarity for the distributions of *home*, *work*, and *education*.

Table 12 gives the average values of the edit distance and the time overlap for the three configurations, as well as the bootstrap 95% confidence intervals. There are no significant differences in the edit distance, and the solutions drawn with all three models require on average 5 edit changes (considering equal costs for all possible operations) to retrieve the original sequence. In terms of time overlap, the activity-specific model appears to be the best performing specification.

Table 13 gives the Hellinger distance for the *home, education* and *work* time-of-day distributions and each model specification (parameters from the literature, generic flexibility parameters, and activity-specific parameters). The results are coherent with the visual analysis: none of the three models perfectly reproduces the observed distributions, especially for *education*. The generic model displays higher distances than the model with parameters from the literature. The chosen classification does not appear to be adequate, as there seems to be a loss of information. On the other hand, the distributions generated with the activity-specific parameters are overall (on average and total), closer to the data.

[2]Here, a *sequence* refers to the order in which the activities are scheduled. It does not include the temporal dimension.

Metric	Model		
	Literature	Generic	Act. specific
Edit distance	5.23 [4.77, 5.78]	5.42 [4.98, 5.89]	5.69 [5.12, 6.30]
Time overlap [h]	5.78 [5.48, 6.01]	4.92 [4.62, 5.21]	6.83 [6.36, 7.29]

Table 12: Average edit distance and time overlap per model, with 95% confidence intervals

The improvement over the parameters from the literature is encouraging given the simplicity of the specified model. More significant gains can be expected by a refinement of the specification.

Activity	Model		
	Literature	Generic	Act. specific
Home	0.718	0.745	0.716
Education	0.940	1.131	0.977
Work	0.356	0.427	0.238
Average	0.671	0.768	0.644
Total	2.014	2.303	1.931

Table 13: Hellinger distance per activity and model

4.5.1 Discussion

This empirical investigation using the MTMC has demonstrated the added value of estimating the parameters for the accuracy and realism of the simulated schedules, as opposed to using generic parameters. Removing a layer of abstraction by estimating activity-specific parameters instead of generic parameters aggregated over the set of activities has shown to provide results fitting the observed distribution better.

The application of the methodology has also highlighted some limitations: the simplifying assumptions formulated to estimate the problem have a significant impact on the quality of the solutions. For instance, the distributional assumptions of the desired times are too restrictive in this case. More specifically, multimodal distributions for the activity start times seem more appropriate and reflective of the observations. This change requires to reconsider the definition of activities, as it implies that the behaviour towards an activity of the same type (e.g. work) would differ depending on when it is scheduled.

Another finding is that, while the simulated profiles are close to the observed ones, all three tested models simulate significantly more schedules with no out-of-home activities than what is actually observed. The fact that this phenomenon is also observed with parameters from the literature suggests that the specification itself does not appropriately model the reality. Indeed, because of its restrictive assumptions on the independence of alternatives the logit model does not account for the correlations, interactions and unobserved behaviour who clearly impact the scheduling decisions (and specifically, the decision to travel out of home). More complex specifications must be investigated, starting with mixed logit models which relax the IIA assumption.

5 Conclusion and future work

We have presented a procedure to estimate the parameters of the OASIS framework, which includes the optimisation-based simulator introduced in Pougala et al. (2022). The estimation process includes: (i) the generation of a choice set for parameter estimation, with a sufficiently high variety of alternatives to ensure unbiased and stable parameter estimates, with tractable sample probabilities, and (ii) the discrete choice estimation of the parameters for different model specifications. We have applied our methodology on a simple case: a time-dependent and linear-in-parameters utility function, and a small dataset. The resulting parameters are mostly statistically significant and behaviourally interpretable, even with a relatively small number of alternatives in the choice set. Using the parameters as input for the activity-based simulator, we can demonstrate that the simulated distribution is closer to the observed one with the estimated parameters as opposed to a benchmark from the literature, with respect to the simulated activity participation and duration.

In this paper, we have focused on demonstrating the feasibility and added value of the methodology. This is a necessary foundation for the framework to be able to solve problems of higher complexity, including social interactions or multi-day behaviour. Methodological improvements such as relaxing the linear-in-parameter assumption for the utility specification, or choosing appropriate model structures to manage the high correlations (e.g mixtures of logit, latent class models) are expected to significantly improve the quality of the estimation and the associated simulation results.

Other simplifying assumptions, such as the distribution of desired times which is assumed unimodal in this paper, can significantly impact the quality of the estimations and must therefore be carefully investigated. For desired times specifically, they could be included as parameters to be estimated. Depending on the chosen model specification, this could require an iterative process to be solved.

Regarding validation, we will investigate in future work a multidimensional distance metric to compare observed and simulated schedules, similarly to the multidimensional sequence alignment technique used by Recker et al. (2008) or Joh et al. (2002). In addition, the calibration of parameters on a synthetic population would allow to evaluate the estimation quality against known control variables.

Regardless, the results of this paper open the way for significant contributions in activity-based modelling: the methodology to estimate the parameters allows researchers to explicitly consider behaviour in the activity-based analysis, which is usually a limiting factor in econometric models. An important contribution is that the methodology remains the same for any change of context-specific constraints and features (e.g. adding or removing a choice dimension, extending the analysis to include multiple people or multiple days,...) or change in utility specification. Modellers can therefore develop flexible and tailored models for a variety of applications to integrate in the framework in a straightforward way. The parameters can then be estimated, even with limited data, with positive impact on the realism the resulting simulations.

Acknowledgments

This work is supported by an Interdisciplinary Cluster Grant from EPFL's School of Architecture, Civil and Environmental Engineering (ENAC). This work was further supported by research funding from the Accenture Turing Strategic Partnership on Digital Twins.

References

Allahviranloo, M., Regue, R., and Recker, W. (2014). Pattern Clustering and Activity Inference.

Arentze, T., Ettema, D., and Timmermans, H. (2011). Estimating a model of dynamic activity generation based on one-day observations: Method and results. *Transportation Research Part B: Methodological*, 45(2):447–460.

Arentze, T. and Timmermans, H. (2000). *Albatross: a learning based transportation oriented simulation system*. Citeseer.

Arentze, T. A. and Timmermans, H. J. (2009). A need-based model of multi-day, multi-person activity generation. *Transportation Research Part B: Methodological*.

Axhausen, K. W. (2000). Activity-based modelling: Research directions and possibilities. *Arbeitsberichte Verkehrs- und Raumplanung*, 48.

Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*, volume 9. MIT press, Cambridge, MA, USA.

Bierlaire, M. (2020). A short introduction to pandasbiogeme. *A short introduction to PandasBiogeme*.

Bierlaire, M. and Krueger, R. (2020). Sampling and discrete choice. Technical Report TRANSP-OR 201109, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

Bowman, J. L. and Ben-Akiva, M. E. (2001). Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice*, 35(1):1–28.

Castiglione, J., Bradley, M., and Gliebe, J. (2014). *Activity-Based Travel Demand Models: A Primer*. Transportation Research Board, Washington, D.C.

Chapin, F. S. (1974). *Human activity patterns in the city: Things people do in time and in space*, volume 13. Wiley-Interscience.

Chen, S., Prakash, A. A., De Azevedo, C. L., and Ben-Akiva, M. (2020). Formulation and solution approach for calibrating activity-based travel demand model-system via microsimulation. *Transportation Research Part C: Emerging Technologies*, 119:102650.

Chow, J. Y. and Recker, W. W. (2012). Inverse optimization with endogenous arrival time constraints to calibrate the household activity pattern problem. *Transportation Research Part B: Methodological*, 46(3):463–479.

Danalet, A. and Bierlaire, M. (2015). Importance sampling for activity path choice. In *15th Swiss Transport Research Conference*.

Flötteröd, G. and Bierlaire, M. (2013). Metropolis–Hastings sampling of paths. *Transportation Research Part B: Methodological*, 48:53–66.

Frejinger, E., Bierlaire, M., and Ben-Akiva, M. (2009). Sampling of alternatives for route choice modeling. *Transportation Research Part B: Methodological*, 43(10):984–994.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Guevara, C. and Ben-Akiva, M. (2013). Sampling of alternatives in logit mixture models. *Transportation Research Part B: Methodological*, 58:185 – 198.

Hägerstraand, T. (1970). What About People in Regional Science? *Papers in Regional Science*.

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications.

Joh, C.-H., Arentze, T., Hofman, F., and Timmermans, H. (2002). Activity pattern similarity: a multidimensional sequence alignment method. *Transportation Research Part B: Methodological*, 36(5):385–403.

Lemp, J. D. and Kockelman, K. M. (2012). Strategic sampling for large choice sets in estimation and application. *Transportation Research Part A: Policy and Practice*, 46(3):602–613.

Nijland, L., Arentze, T., and Timmermans, H. (2009). Estimating the parameters of a dynamic need-based activity generation model. In *12th International Conference on Travel Behaviour Research (IATBR 2009), December 13-18, 2009, Jaipur, India*.

Office fédéral de la statistique and Office fédéral du développement Territorial (2017). Comportement de la population en matière de transports. Résultats du microrecensement mobilité et transports 2015. Technical report, Neuchâtel, Berne.

Pagliara, F. and Timmermans, H. (2009). Choice set generation in spatial contexts: a review. *Transportation Letters*, 1(3):181–196.

Pougala, J., Hillel, T., and Bierlaire, M. (2021). Choice set generation for activity-based models. In *Proceedings of the 21st Swiss Transport Research Conference*, Ascona, Switzerland.

Pougala, J., Hillel, T., and Bierlaire, M. (2022). Capturing trade-offs between daily scheduling choices. *Journal of Choice Modelling*, 43:100354.

Rasouli, S. and Timmermans, H. (2014). Activity-based models of travel demand: promises, progress and prospects. *International Journal of Urban Sciences*, 18(1):31–60.

Recker, W., Duan, J., and Wang, H. (2008). Development of an Estimation Procedure for an Activity-Based Travel Demand Model. *Computer-Aided Civil and Infrastructure Engineering*, 23(7):483–501.

Recker, W. W. (1995). The household activity pattern problem: general formulation and solution. *Transportation Research Part B: Methodological*, 29(1):61–77.

Small, K. A. (1982). The scheduling of consumer activities: work trips. *American Economic Review*.

Train, K. (2009). *Discrete choice methods with simulation*. Cambridge University Press, Cambridge, 2nd ed. edition.

Vovsha, P., Bradley, M., and Bowman, J. (2005). Activity-based travel forecasting models in the united states: progress since 1995 and prospects for the future. In *EIRASS Conference on Progress in Activity-Based Analysis*.

Xu, Z., Kang, J. E., and Chen, R. (2017). A Random Utility Based Estimation Framework for the Household Activity Pattern Problem. *Transportation Research Procedia*, 23:809–826.