Capturing trade-offs between daily scheduling choices

Janody Pougala [*] [†] Tim Hillel [*] Michel Bierlaire [*]

January 1, 2021

Report TRANSP-OR 210101
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, {janody.pougala,tim.hillel,michel.bierlaire}@epfl.ch

[†]Corresponding author

Abstract

We propose a new modelling approach for daily activity scheduling which integrates the different daily scheduling choice dimensions (activity participation, location, schedule, duration, and transportation mode) into a single optimisation problem. The fundamental behavioural principle behind our approach is that individuals schedule their day to maximise their overall derived utility from the activities they complete, according to their individual needs, constraints and preferences. By combining multiple choices into a single optimisation problem, our framework is able to capture the complex trade-offs between scheduling decisions for multiple activities, such as how spending longer in one activity will reduce the time-availability for other activities or how the order of activities changes the travel-times. The implemented framework takes as input a set of considered activities, with associated location and mode-of-travel, and uses these to produce empirical distributions of individual schedules given a set of considered activities, from which different daily schedules can be drawn. The model is illustrated using historic trip diary data from the Swiss Mobility and Transport Microcensus. The results demonstrate the ability of the proposed framework to generate complex distributions of starting time and duration for different activities within the tight time constraints.

Keywords: activity-based modeling, daily scheduling behaviour, mobility, mixed-integer optimization, random utility maximization

1 Introduction

The scheduling of daily activities is a complex process that combines multiple interconnected choices, including deciding which activities to perform in a day, as well as the timings, location and mode-of-travel for each performed activity. The daily scheduling process is a critical component of Activity-Based Models (ABMs) of transport demand, which assume that travel demand can be derived from the needs of individuals to perform activities (Bowman and Ben-Akiva, 2001). The behavioural realism of activity-based models is therefore dependent on the accurate modelling of the scheduling process. In this paper, we introduce a new approach to modelling individual activity scheduling, based on mixed-integer optimisation.

Real-world daily activity schedules represent more than a sequence of independent activities and locations, and are instead the result of unobserved dynamics, reasoning, and trade-offs. For example, an individual might leave work earlier than usual to be on time for a concert, or skip a regular exercise session entirely due to a high-workload. Being “in a rush”, having “plenty of time” or being able to “squeeze in” additional activities in otherwise packed schedules are universal experiences that indicate the flexibility in daily scheduling.

There exist many examples of activity-based models (as seen in section 2), and about as many methods to deal with scheduling trade-offs and their behavioural implications. Existing approaches can be grouped into two major paradigms: econometric models and rule-based models. Econometric models postulate that the scheduling of activities can be explained with econometric processes such as random utility maximisation. They do not explicitly model behaviour but rather consider it a consequence of the utility maximisation. The different choice dimensions are often modelled sequentially (e.g. Bowman and Ben-Akiva (2001), Hilgert et al. (2017), Bradley and Bowman (2008)). On the other hand, rule-based or computational process models use decision rules to derive feasible solutions. This makes them easier to implement in practice. But these rules are hard-coded and often arbitrary, which limits their generalisation.

Both approaches tend to come short of fully integrating a dynamic behaviour within a framework that is flexible enough to be interpreted in a wider context. This can be problematic when models are applied to help decision makers enforce efficient and targeted measures. As made evident by the recent global pandemic and the consequent strategies to manage the spread of the virus, this task requires a deep understanding of the motives behind the mobility choices of a person, how they interact with their environment (physical, social and cultural) and how they react to events or perturbation that might more or less significantly constrain their field of possibilities.

The approach introduced in this paper presents three key advantages compared to existing scheduling models in the field: (i) all choice dimensions (activity participation, activity location, activity schedule, activity duration, and transportation mode choice to travel to the next activity) are modelled simultaneously; (ii) the model produces an empirical distribution of individual schedules that can be investigated with simulation; and (iii) the framework is built on first behavioural principles and can be generalised to complex mobility situations.

The rest of the paper is laid out as follows: Section 2 presents a brief review of the literature, with an emphasis on utility-based models and simulators. The framework is detailed in Section 3, with an overview of the key components of the model and the simulation methodology. We illustrate the flexibility, operationality and realism of the framework on the Swiss Mobility and Transport Microcensus (MTMC) in Section 4. We then conclude with a discussion on current and future challenges.

2 Relevant literature

Activity-based models originally emerged in the 1970s as a response to the shortcomings of the traditional 4-step models (Vovsha et al., 2005, Castiglione et al., 2014), namely: (i) 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; (ii) models tend to suffer from biases due to unrealistic aggregations in time, space, and within the population; and (iii) 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. Unlike traditional trip-based models, ABM focus on overall behavioural patterns: decisions are analysed at the level of the household as opposed to seemingly independent individuals, and dependencies between events are taken into account (Timmermans, 2003, Pas, 1985). Specifically, modellers are interested in the link between activities and travel, often considered within a given timeframe. Typically, a single day is used as the unit of analysis. The resulting goal of studies in the literature is therefore to replicate as accurately as possible the interactions and considerations involved in the development of a daily schedule by an individual. Several authors have pointed that focusing only on one-day scheduling ignores day-to-day correlations and dynamics (e.g. Arentze et al. (2011)). However, proper implementation and validation of single-day models remains complex. In addition, the information required to go beyond single-day travel patterns is usually not readily available, and might require additional data collection and processing steps, including fusion of multiple data sources (e.g. Aschauer et al. (2019))

While the scheduling process is central to the activity-based research, there is no clear consensus on the representation and modelling of the daily scheduling process in utility-based frameworks. Typically, individuals are assumed to schedule activities by maximising the utility they can expect to gain. The timeframe is often introduced as a time budget that constrains the overall time expenditure. The earliest functional utility-based models is the logit model for household daily travel patterns developed by Adler and Ben-Akiva (1979), which assumes that households choose from a set of possible daily patterns, and uses a logit model to compute the choice probabilities for each alternative. This was followed by the disaggregate travel demand model developed by Bowman and Ben-Akiva (2001) that models a series of sequential decisions to generate an activity pattern and tours for the day. These decisions are, in order, the choice of activity pattern (staying at home or travelling), the primary tour time of day, the primary tour destination and mode, and finally the secondary tours times of day, destination and modes. The choice activity pattern is modelled using a nested logit model, the tour times of day are generated using a logit model, and the destination and mode with a logit model with alternative sampling. A set of rules is used to define a hierarchy among activities (primary vs. secondary). The models developed by Adler and Ben-Akiva, Bowman and Ben-Akiva are travel-centric: while both assume an interdependence among choices, they mostly focus on trip characteristics (e.g. tour frequency, number of stops, mode choice...). Behavioural mechanisms explaining the actual choice of activities and their sequence are examined less closely. In the context of these models, activity schedules and emerging behaviour are implicit and rather consequential to the predicted travel decisions. In addition, the discrete choices resulting in the final activity-travel pattern are treated sequentially. This allows for simple, clearly defined modelling assumptions, but limits the ability of the framework to capture trade-offs that individuals could make between different choice dimensions. Nonetheless, sequential estimation remains popular in the literature, especially for microsimulators (e.g. Recker et al. (1986), Pendyala et al. (2005), Smith et al. (1995), Ettema et al. (2000), Axhausen et al. (2016))

More recent works have focused on joint estimation of mobility choices, with a more explicit integration of emerging behaviour in the scheduling process. For instance, Nurul Habib and Miller (2009) use an utility-based approach to model the generation of activities (i.e. which activities are considered in the first place). In this case, the utility function is defined for an agenda (a set of activities to be scheduled), aiming to capture the trade-off between planned and unplanned activities. The choice probabilities are estimated with the Kuhn-Tucker optimality conditions in place of discrete choice models. The resulting agenda is then used as input for a discrete-continuous scheduling model that predicts sequentially the choice of activity (discrete choice) and the time expenditure for the chosen activity (continuous choice) (Nurul Habib, 2011). The discrete-continuous representation of activity schedules has been investigated extensively by Bhat et al. (2004) (see also Bhat (2005, 2018)). In their Multiple Discrete-Continuous Extreme Value (MDCEV) model, the scheduling process is modelled as a combination of a discrete choice (activity participation) and continuous choice (activity duration). Behaviour is explicitly considered with a non-linear utility function and satiation effects (decreasing marginal utility). The discrete-continuous approaches are a flexible solution to simultaneously consider multiple choice dimensions. However, they become limited when it comes to integrate time-of-day decisions, which are heavily influenced by external factors (e.g. shop opening times, working hours, commitments, etc.).

Joint estimation of multiple choice dimensions, including time-of-day, has been explored in other works. Ettema and Timmermans (2003) formulate an error-component discrete choice model to jointly estimate duration, time-of-day preference and effect of schedule delays on the utility function of the alternatives. They consider that individuals maximise the sum of the utility gained from travelling and from performing the activities, the latter composed of three elements: a time-of-day dependent utility, a duration utility, and a schedule delay utility dependent on the start time. Their model is thus able to accommodate more explicitly the discontinuities in utility introduced by the presence of these external constraints and preferences. However, the mainly focuses on time allocation for a given set of activities, and schedule dynamics linked to activity participation (e.g. dropping an activity if the timings are not convenient for the individual) cannot easily be taken into account.

Furthermore, time trade-offs between activities are not clearly defined in the aforementioned works. It is common in the econometric representation of activity-based models to treat time as one would goods, in terms of consumption, meaning that a marginal change in time is defined as a derivative of the utility function. English (2020) argues that this representation is problematic, as the marginal change in time cannot be interpreted as such. It depends on both the time change and the time replaced. In the context of activity-based models, the impact of an change in time on the utility depends

The framework presented in this paper follows the modelling philosophy presented by Ettema et al. (2000) by considering schedule deviations and simultaneously estimating choices of timing, duration, and trip characteristics (mode, location). Similarly, our framework is founded on first behavioural principles, which allows for flexible and generalisable modelling. We aim to provide a holistic point of view: by explicitly including the choice of activity participation, activity sequence, location and transport mode in the optimisation problem, we allow for a complete integration and interactions of schedule-related choice dimensions. In addition, we put a specific emphasis on timing preferences to capture and interpret scheduling trade-offs.

3 Modelling framework

The framework presented in this paper captures the choice of a valid schedule for a given time horizon (typically, a day) made by a single individual, called the agent. The central theory behind our approach is that individuals schedule their day to maximise their overall derived utility from the activities they complete, according to their individual needs, constraints and preferences. We therefore define a general utility function which captures the derived utility from an individual completing a considered activity, according to (i) the preference towards *participating* in that the type of activity, (ii) the desired and scheduled *duration* of the activity, (iii) the desired and scheduled *start-time* of the activity, (iv) the flexibility of the individual towards *schedule deviation* in start-time (early/late) and duration (long/short) for the activity, and (v) the required *travel-time* to arrive at the activity location from the previous location. We then define a mixed-integer optimisation problem for each individual which maximises the sum of the utilities of each completed activity in a schedule over a fixed time budget. This optimisation problem can therefore capture the *trade-offs* between scheduling decisions for multiple activities, such as how spending longer in one activity will reduce the time-availability for other activities or how the order of activities changes the travel-times. The overall framework takes as input a set of *considered activities*, with associated location and mode-of-travel, and uses this to define a distribution over possible schedules, from which likely scheduling choices can be stochastically drawn.

In this section, we introduce the modelling elements of the proposed framework. For the sake of clarity of the notations, no index is associated with the agent in the following analysis.

Time can be either continuous or discrete. The time horizon starts at $t = 0$ and finishes at $t = T$. Space is characterised by a discrete and finite list of L locations, indexed by ℓ. The location $\ell = 0$ is called "home", and is assumed to be the location of the agent at time $t = 0$ and time $t = T$.

The agent considers M transportation modes, indexed by m. The travel time between two locations ℓ_o and ℓ_d using mode m is denoted $\rho(\ell_o, \ell_d, m)$ and is exogenous. If ℓ_d cannot be reached from ℓ_o using mode m, then $\rho(\ell_o, \ell_d, m) = +\infty$.

The agent considers a set of A activities, indexed by a. Each activity a is associated with:

- a list L_a of possible locations where the activity could be performed,
- an indicator μ_a that is 1 if the activity is mandatory and 0 if it is optional,
- a time interval when the agent prefers to start the activity[1]: $[x_a^-, x_a^+]$, where $x_a^- \leq x_a^+$,
- a minimum duration $\tau_a^{\min}$,
- a range of desired durations $[\tau_a^-, \tau_a^+]$, where $\tau_a^{\min} \leq \tau_a^- \leq \tau_a^+$,

We define a binary indicator δ_{as} which equals 1 if activity a can be performed at location ℓ, and 0 otherwise. Each relevant pair activity/location is associated with a feasible time interval $[\gamma_{a\ell}^-, \gamma_{a\ell}^+]$. It stipulates that the activity can take place only during that time interval. For example, shopping can typically only happen during the opening hours of the

[1]Note that the assumption that the preferences in starting time and duration are captured by a unique time interval mathematically convenient, but may not be realistic. For instance, a student may prefer to sit an exam either early in he morning, or late in the afternoon. In that case, it would be modelled using two different activities.

selected shop. Note that the agent may consider a location for an activity even if there is no overlap between $[\gamma^-_{a\ell}, \gamma^+_{a\ell}]$ and $[x^-_a, x^+_a]$. While the former represents a hard constraint, the latter represents a preference.

3.1 Valid schedules

Given the above information, the agent considers valid schedules. A schedule is the outcome of the agent's decisions with respect to activity participation, activity location, activity scheduling and transportation mode choice. More specifically, a schedule $\mathcal{S}$ is a sequence of S activities $(a_0, \dots, a_S)$, starting with a dummy activity a_0 called "dawn", and finishing with a dummy activity a_S called "dusk", both of which take place at home. Each activity a is associated with an actual location ℓ_a, an actual starting time x_a and an actual duration τ_a. With the exception of the last activity "dusk", a trip is performed immediately after each scheduled activity a, using an actual mode of transportation m_a. Note that, if the next activity takes place at the same location, the duration of the trip is simply zero.

A schedule is *valid* if

- it spans the whole time horizon, that is if

$$\tau_{\text{dawn}} + \tau_{\text{dusk}} + \sum_{s=1}^{S-1} \big(\tau_a + \rho(\ell_{a_s}, \ell_{a_{s+1}}, m_{a_s})\big) = T, \tag{1}$$

- all mandatory activities are included,
- each activity starts when the trip following the previous activity is finished, that is

$$x_{a_{s+1}} = x_{a_s} + \tau_{a_s} + \rho(\ell_{a_s}, \ell_{a_{s+1}}, m_{a_s}), \forall s = 0, \dots, S-1, \tag{2}$$

- the duration of each activity is valid, that is if

$$\tau_a \geq \tau_a^{\text{min}}. \tag{3}$$

3.2 Preferences

The agent is assumed to be rational, and to select the preferred schedule among all possible valid schedules. The preferences of the agent are captured by a utility function $U_{\mathcal{S}}$[2] associated with each schedule $\mathcal{S}$.

From the point of view of the analyst, the main challenge is that the choice set cannot be enumerated, due to the combinatorial structure of the set of valid schedules. We propose to address this challenge by performing an explicit enumeration for decisions related to activity location and transportation mode; and an implicit enumeration for decisions related to activity participation and activity scheduling.

For each activity considered by the agent, we explicitly enumerate all possible combinations of the locations and modes associated with it. Each of these combinations is

[2]We use U to define random utilities, and V to define deterministic utilities.

considered as a separate activity in the model. Therefore, each activity a considered by the agent is modelled by the analyst using ML_a mutually exclusive activities, each associated with a unique location ℓ_a and a unique mode of transportation m_a. In addition, we impose the constraint that at most one of these duplicate activities can be selected in a given schedule. This explicit enumeration leads to K groups G_k of activities that are mutually exclusive. We can therefore simplify some notations. The feasible time interval of activity a can be denoted $[\gamma_a^-, \gamma_a^+]$ without ambiguity, as well as the travel time between two activities:

$$\rho_{ab} = \rho(\ell_a, \ell_b, m_a). \tag{4}$$

The implicit enumeration consists of solving the scheduling problem considered by the agent using a standard optimisation algorithm, that identifies the optimal solution without complete enumeration.

Before describing the scheduling problem, we introduce the model of the utility $U_{\mathcal{S}}$ associated by the agent with the schedule. We define it as the sum of a generic utility U associated with the whole schedule and, for each activity, (i) the utility U_s^1 associated with the participation of the activity a_s, irrespective of its starting time and duration; (ii) the utility V_s^2 associated with a starting time different from the preferred one; (iii) the utility V_s^3 associated with a duration different from the preferred one; (iv) the utility U_s^4 associated with the trip towards the next activity, irrespective of the travel time; and (v) the utility V_s^5 associated with the travel time to the next activity:

$$U_{\mathcal{S}} = U + \sum_{s=0}^{S-1} (U_s^1 + V_s^2 + V_s^3 + U_s^4 + V_s^5). \tag{5}$$

Note that no utility is associated with the dummy activity "dusk". We also assume that

$$U_0^1 = V_0^2 = V_0^3 = 0. \tag{6}$$

Indeed, as only differences of utility matter, the two dummy activities serve as reference and their utility is set to zero.

This specification provides a great deal of flexibility. There are no specific assumptions about U_s^1 and U_s^4, except that they must be independent on starting time and duration decisions. They can either be treated as random variables or not, and can involve any variable. The generic utility U captures aspects of the schedule that are not associated with a specific activity. For instance, the agent may prefer that all shopping activities take place in the afternoon, or may dislike days with too many activities. We discuss the assumptions associated with the generic utility below.

The term V_s^2 is defined as

$$V_s^2 = \theta^e_{a_{s,k}} \max(0, x_{a_s}^- - x_{a_s}) + \theta^\ell_{a_{s,k}} \max(0, x_{a_s} - x_{a_s}^+), \tag{7}$$

where $\theta^e_{a_{s,k}} \leq 0$ and $\theta^\ell_{a_{s,k}} \leq 0$ are unknown parameters to be estimated from data. The first (resp. second) term captures the disutility of starting the activity earlier (resp. later) than the preferred starting time, as illustrated in Figure 1. Note that the amplitude of the

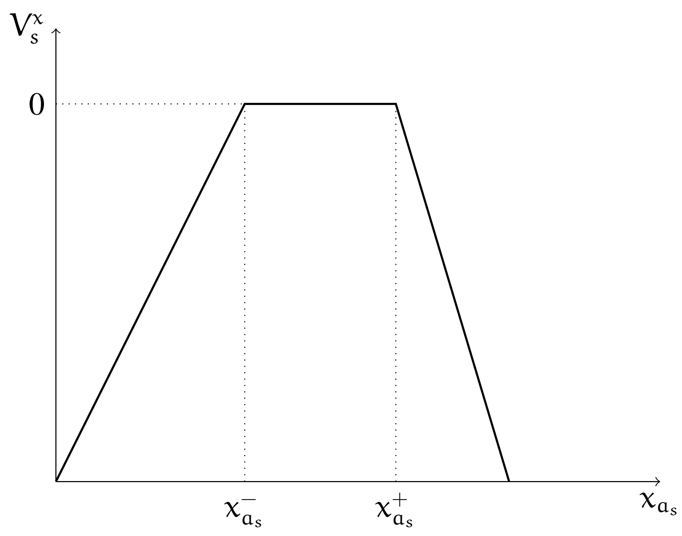


Figure 1: Utility associated with deviations from the preferred starting time of an activity

penalty, captured by the parameters θ, may vary across groups of activities. The index k captures the level of flexibility with respect to the scheduling of the activity.

The term V_s^3 is similarly defined:

$$V_s^3 = \beta^e_{a_{s,k}} \max(0, \tau^-_{a_{s,k}} - \tau_{a_s}) + \beta^\ell_{a_{s,k}} \max(0, \tau_{a_s} - \tau^+_{a_s}), \tag{8}$$

where $\beta^e_{a_{s,k}} \leq 0$ and $\beta^\ell_{a_{s,k}} \leq 0$ are unknown parameters to be estimated from data.

The term V_s^5 is the disutility of travel time:

$$V_s^5 = \theta_t \rho_{a_s, a_{s+1}}, \tag{9}$$

where θ_t is an unknown parameter to be estimated from data, and $\rho_{a_s, a_{s+1}}$ is the travel time to the next location.

3.3 Choice

The agent is assumed to select the valid schedule with the highest utility. She therefore solves an optimisation problem to maximise the utility function under the validity constraints. However, from the point of view of the analyst, the utility function (5) is captured by a random variable, and the model associates a choice probability with each valid schedule. In order to deal with this uncertainty, we propose a simulation approach, where the optimisation problem is explicitly solved for several realisations of the random utility. The resulting schedule is a realisation from the choice model. The advantage of this approach is that each generated schedule is valid by design, explicitly capturing the trade-offs made by the agent.

For each activity a, we first generate realisations of U^1, that we denote by V_a^1. For each pair (a, b) of locations, we also generate realisation of U^4, that we denote by V_{ab}^4. We characterise the decision of the agent using the following decision variables:

- ω_a: binary variable that is 1 if activity a is selected in the schedule, and 0 otherwise,

- z_{ab}: binary variable that is 1 if activity b is scheduled immediately after activity a, where $a \neq b$,
- x_a: starting time of activity a,
- τ_a: duration of activity a,

and we denote the corresponding vectors by ω, z, x, τ. We consider a realisation of the generic utility U, denoted by $V(\omega, z, x, \tau)$, to emphasise that it depends on the decision variables.

The objective function is derived from (5):

$$\max_{\omega,z,x,\tau} V(\omega, z, x, \tau) + \sum_{a=0}^{A} \omega_a (V_a^1 + V_a^2 + V_a^3) + \sum_{a=0}^{A} \sum_{b=0}^{A} z_{ab}(V_{ab}^4 + V_{ab}^5). \tag{10}$$

Note that the term $\sum_{b=0}^{A} z_{ab} V_{ab}^5$ corresponds to V_s^5 in (9). The constraints are

$$\sum_a \sum_b (\tau_a + z_{ab}\rho_{ab}) = T, \tag{11}$$

$$\omega_{\text{dawn}} = \omega_{\text{dusk}} = 1, \tag{12}$$

$$\tau_a \geq \omega_a \tau_a^{\text{min}}, \quad \forall a \in A, \tag{13}$$

$$\tau_a \leq \omega_a T, \quad \forall a \in A, \tag{14}$$

$$z_{ab} + z_{ba} \leq 1, \quad \forall a, b \in A, i \neq b, \tag{15}$$

$$z_{a,\text{dawn}} = z_{\text{dusk},a} = 0, \quad \forall a \in A, \tag{16}$$

$$\sum_a z_{ab} = \omega_b, \quad \forall b \in A, b \neq \text{dawn}, \tag{17}$$

$$\sum_b z_{ab} = \omega_a, \quad \forall a \in A, a \neq \text{dusk}, \tag{18}$$

$$(z_{ab} - 1)T \leq x_a + \tau_a + z_{ab}\rho_{ab} - x_b, \quad \forall a, b \in A, i \neq b, \tag{19}$$

$$(1 - z_{ab})T \geq x_a + \tau_a + z_{ab}\rho_{ab} - x_b, \quad \forall a, b \in A, i \neq b, \tag{20}$$

$$\sum_{a \in G_k} \omega_a, \leq 1 \quad k = 1, \dots, K, \tag{21}$$

$$x_a \geq \gamma_a^-, \quad \forall a \in A, \tag{22}$$

$$x_a + \tau_a \leq \gamma_a^+, \quad \forall a \in A. \tag{23}$$

Equation (11) constrains the total time assigned to the activities in the schedule (sums of durations and travel times) to be equal to the time budget. Equation (12) ensures that each schedule begins and ends with the dummy activities *dawn* and *dusk*. Equations (13) and (14) enforce consistency with the activity duration by requiring the activity to have a duration greater or equal than the minimal duration (3) and for the activity to have zero duration if it does not take place. Equations (15)-(19) constrain the sequence of the activities: (15) ensures that two activities a and b can only follow each other once (thus can only be scheduled once). As it is defined for distinct activities only, it also ensures that an activity cannot follow itself. Equations (16)-(18) state that each activity has only one predecessor (excluding the first activity), and each activity only one successor (excluding the last activity). Equation (19) enforces time consistency between two consecutive activities (with travel time ρ_{ab}). Equation (21) ensures that only one activity within a group of duplicates G is selected. Finally, (22) and (23) are time-window constraints.

Note that all the constraints in this formulation are linear in the decision variables. Therefore, if the objective function is also expressed as a linear function of the constraints, we obtain a linear integer optimisation problem[3], that can be solved by standard mathematical programming algorithms (e.g. branch-and-bound, branch-and-cut, constraint programming, etc.).

Hence, we add the assumption that the generic utility $V(\omega, z, x, \tau)$ must be specified as a linear function of the decision variables. This assumption is common in the mathematical programming literature, and is not overly restrictive.

4 Empirical investigation

In order to illustrate the optimisation-based simulation concept introduced in Section 3, we rely on a real dataset to generate the inputs. The objective is to show that, given sets of possible activities, locations, modes and timing preferences, the model is able to generate realisations of chosen daily schedules.

The MTMC is 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. We use only the data corresponding to the residents of Lausanne, for a total of 2'227 diaries.

4.1 Inputs

The required inputs (as defined in Section 3) are not necessarily available in traditional travel surveys, including the MTMC. The challenge is thus to provide heuristics to obtain estimators for the missing attributes.

Table 1 summarises the data requirements for the operational model, as well as two possible solutions to overcome the lack of information for each requirement. The *heuristic* column describes methods that have been applied in this paper, with results presented in section 4.3.

4.2 Utility specification

Allowing for the available inputs for this case study, the schedule utility function expressed in (5) has been simplified as follows:

[3] If time is modelled using a continuous variable, we solve a linear mixed integer optimisation problem.

Table 1: Data requirements for operational model

Requirements	Rigorous solution	Heuristic
Considered activities A	Activity choice set generation algorithm for each individual	Description of actual schedule from dataset
Considered modes M	Mode choice set generation algorithm for each individual	Consider all 5 main modes (driving, passenger, public transport, walk, cycle)
Considered locations L_a	Location set generation algorithm for each individual	Description of actual schedule from dataset
Desired start time and duration ranges $[x_a^-, x_a^+]$ and $[\tau_a^-, \tau_a^+]$,	Habit analysis and identification of typical timings in multi-day diaries	Ranges replaced by recorded values in dataset
Flexibility k	Habit analysis in multi-day diaries — flexibility would be the timing variability	Assign a discrete flexibility profile to each activity based on literature classification.
Penalty values (θ, β)	Calibrated on data — n-dependent	From literature, homogeneous across all population
Feasible time windows $[\gamma_{a\ell}^-, \gamma_{a\ell}^+]$	Data collection	Out-of-sample distributions of start and end times for each activity, across the population
Minimum duration τ_a^{min}	Habit analysis in multi-day diaries	Set to 0

1. The time-independent utilities (generic schedule utility U, activity participation U_s^1 and trip utility U_s^4) are set to 0.
2. The randomness is introduced by adding a random error term $\varepsilon_s \sim \mathcal{N}\left(0, \sigma^2\right)$, with variance σ^2 set to 1.
3. The ranges of start time preferences $[x_a^-, x_a^+]$ are replaced by a punctual desired time (i.e. $x_a^- = x_a^+ = x^*$), and the associated utility V_s^2 is therefore defined as:

$$V_s^2 = \theta^e_{a_{s,k}} \max(0, x^*_{a_s} - x_{a_s}) + \theta^\ell_{a_{s,k}} \max(0, x_{a_s} - x^*_{a_s}), \tag{24}$$

   The same assumption is made for the preferred durations and their associated utility V_s^3, similarly defined as:

$$V_s^3 = \beta^e_{a_{s,k}} \max(0, \tau^*_{a_{s,k}} - \tau_{a_s}) + \beta^\ell_{a_{s,k}} \max(0, \tau_{a_s} - \tau^*_{a_s}), \tag{25}$$

4. The flexibility in time k is modelled using a discrete indicator that can describe 3 possible behaviours (Figure 2):
   (a) Flexible (F): deviations from preferences for activity a are relatively unimportant, thus are less penalised.
   (b) Moderately flexible (MF): deviations from preferences are moderately undesirable, and so are penalised more than in the flexible case.
   (c) Not flexible (NF): deviations from preferences are strongly undesirable, and are consequently highly penalised.

Each activity is associated with one level of flexibility, and each level is characterised by specific values of the penalty parameters. The flexibility assignments for each activity are summarised in Tables 2 and 3. For the sake of simplicity, we consider that the parameters are deterministic instead of randomly distributed across the population. We have chosen values based on results from the departure time choice literature (Small, 1982).

4.3 Results

We present four examples from the MTMC: two students, identified as Alice and Bryan; a worker, Claire, and an unemployed person, Dylan. The set of considered activities, timing preferences, activity locations and modes for each individual are reported in Table 4. Certain activities were duplicated to offer different mode and location options. Figures 3-6 show unique outputs produced by the model, for different draws of ε_s.

For Alice, all solutions show sequences where both of the *education* instances are scheduled. Regarding the *leisure* activity, only the second schedule (Fig. 3b) includes it with timings consistent with her preferences. The other two solutions (Fig. 3a and 3c), this activity is scheduled at a different time of day than the desired times (in the morning and at lunch time, respectively).

For Bryan, the first two solutions both include shopping, but at different locations. In the third solution (Fig. 4c), the shopping activity does not appear in the schedule, indicating that staying at home has a higher overall utility.

Table 2: Categories and flexibility profiles for activities in the MTMC.

Activity	Category	Flexibility profile[a] *Start*	*Duration*
Work Education Business trip	Mandatory[b]	Early: NF Late: MF	Short: NF Long: NF
Errands, use of services Escort	Maintenance	Early: MF Late: MF	Short: MF Long: F
Home[c] Shopping Leisure	Discretionary	Early: F Late: MF	Short: F Long: F

[a] F = Flexible, MF = Moderately flexible, NF = Not flexible.
[b] In this example, we use the term *mandatory* to refer to non-flexible activities with high utilities.
[c] Not including mandatory home stays *dawn* and *dusk*.

Table 3: Penalty values by flexibility, in units of utility

Deviation	Flexibility	Penalty θ
Early start	Flexible (F) Moderately flexible (MF) Not flexible (NF)	0 -0.61 -2.4
Late start	F MF NF	0 -2.4 -9.6
Short duration	F MF NF	-0.61 -2.4 -9.6
Long duration	F MF NF	-0.61 -2.4 -9.6

Table 4: Considered activities and preferences for each individual.

Person	Activity	x_a^* (hh:mm)	τ_a^*(hh:mm)	Location[a]	Mode
Alice	Education (AM)	8:20	3:40	Campus	Car
	Education (PM)	13:30	2:45	Campus	Car
	Education (AM)	8:20	3:40	Campus	PT
	Education (PM)	13:30	2:45	Campus	PT
	Leisure	17:10	0:50	Campus	Car
Bryan	Education	7:30	4:40	Campus	Car
	Shopping	16:30	2:00	Downtown	Car
	Shopping	16:30	2:00	Campus	Car
Claire	Work (A)	14:25	4:25	Office	Car
	Work (B)	14:25	4:25	Office	PT
	Work (C)	14:25	4:25	Library	Car
	Errands	9:45	0:15	Chemist	Car
	Escort	14:10	0:01	Downtown	Car
	Leisure	8:00	1:00	Downtown	Car
	Shopping	13:00	2:00	Shop	Car
Dylan	Escort (Afternoon)	15:10	0:50	School	Car
	Errands	16:40	1:50	Shop	Car
	Escort (Evening)	18:50	0:03	School	Car
	Leisure	19:20	1:30	Gym	Car
	Leisure	19:20	1:30	Gym	Cycling

[a] Each location is assigned unique coordinates for which travel times are estimated.

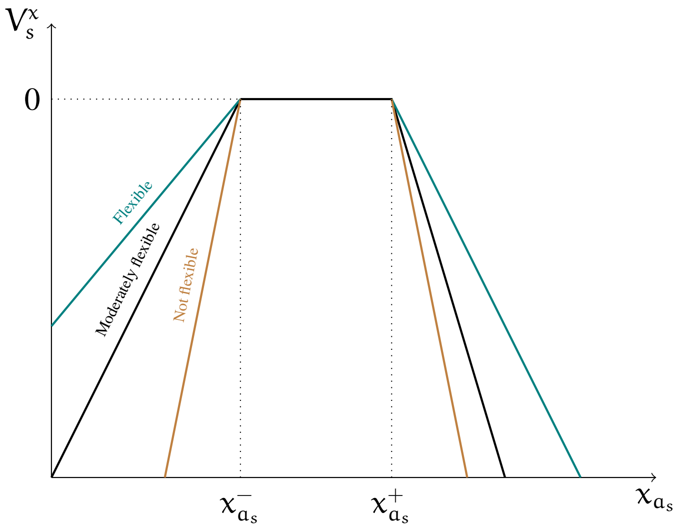
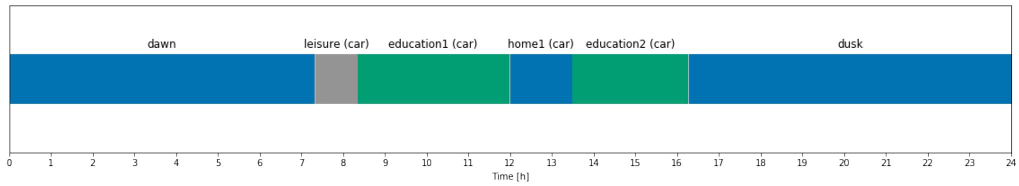
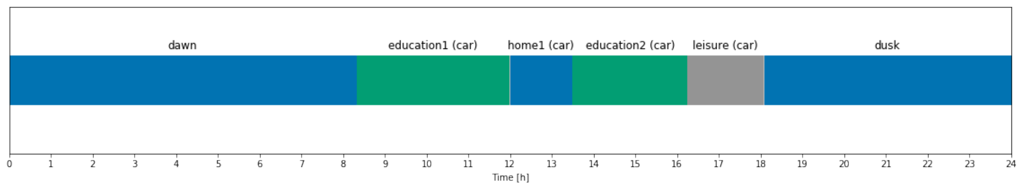


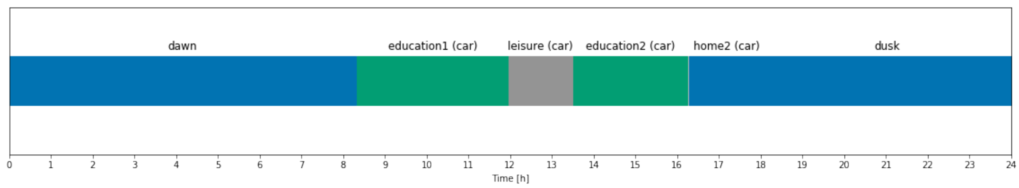
Figure 2: Utility associated with deviations from the preferred starting time of an activity, and levels of flexibility



(a) Solution 1



(b) Solution 2

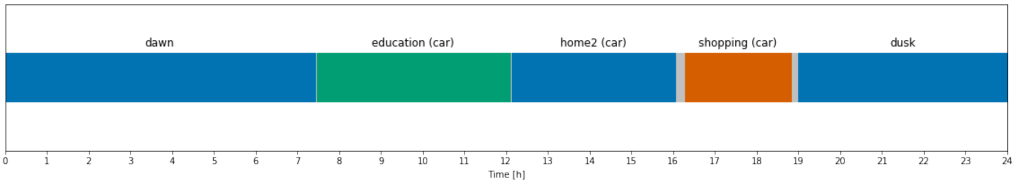


(c) Solution 3

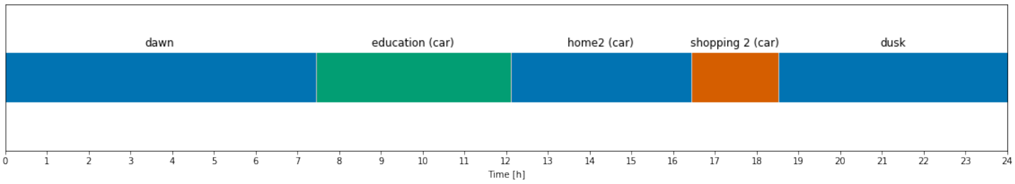
Figure 3: Generated schedules for Alice

The solutions for Claire (shown in Fig. 5) are similar in that all include *work*, with timings that do not diverge substantially from the preferences. On the other hand, the discretionary activities (here, errands, escort, leisure and shopping) are not always scheduled, and when they are, the scheduled timings can be far from the preferences (e.g. Fig. 5d). Figures 5b and 5d show schedules where the second location for work is chosen.

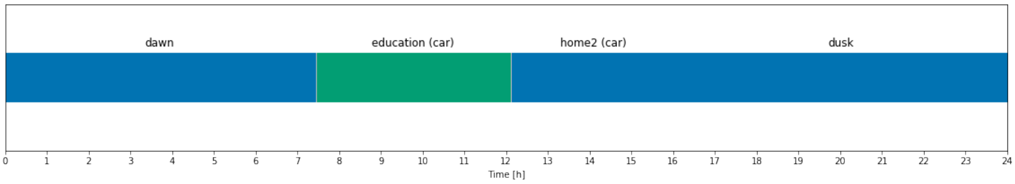
Dylan differs from the other selected individuals in that his set of considered activities



(a) Solution 1



(b) Solution 2



(c) Solution 3

Figure 4: Generated schedules for Bryan

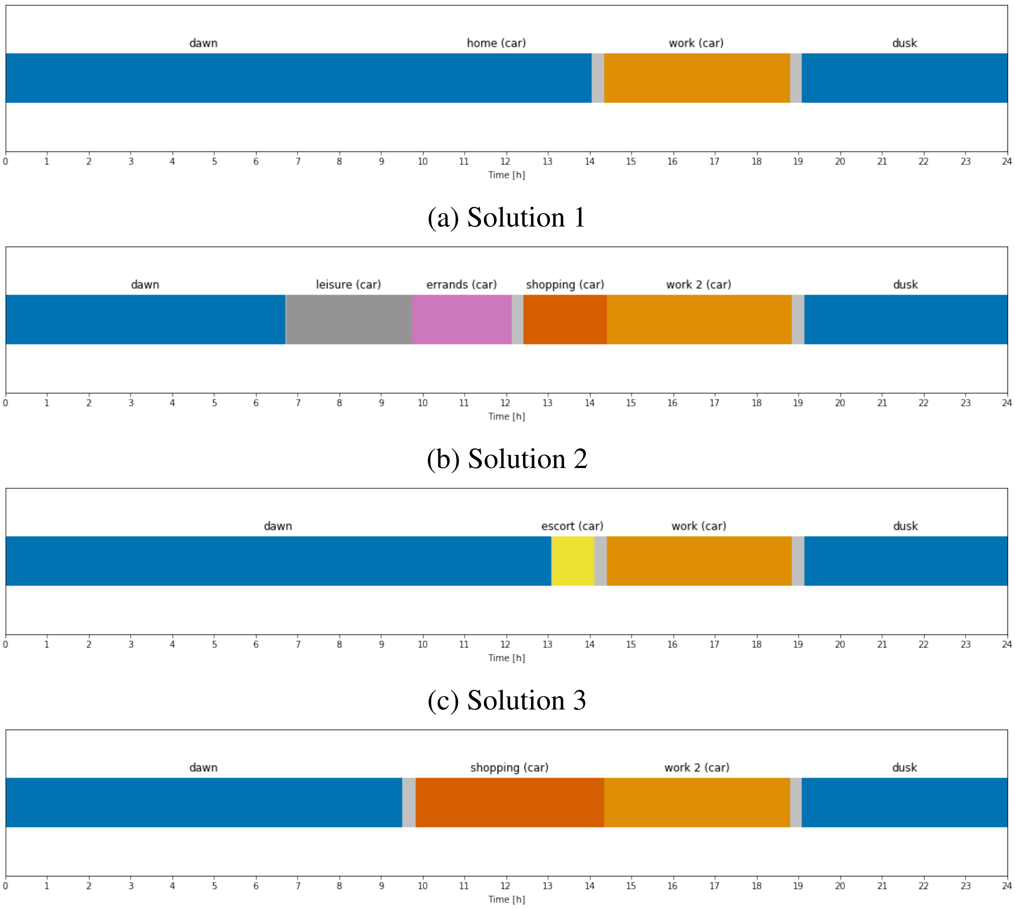
does not contain any highly constrained activity such as work or education. The leisure activity is included in the three options, but with varying durations. When included, the escort and errands activities stay relatively close to the preferences.

For the choice of transportation mode, none of the solutions include the public transport option. This indicates a consistently higher attractiveness of the car mode for the given parameters.

These results show that the variations in solutions affect mainly the discretionary activities, which have lower penalties for schedule deviations than less flexible activities. Note that we have selected only a small number of unique solutions out of all the generated solutions. The heterogeneity of the solution space (i.e. the distribution from which schedules are drawn) is driven by the relative values of the parameters and the error terms. More specifically, very high penalties (compared to the error variances) lead to semi deterministic problems where the scheduler will consistently output very similar (or the same) schedules. On the other hand, error terms with very high variance (compared to the penalties) will lead to a diverse set of solutions. An appropriate scale for the error terms must therefore be determined such that the model can generate solutions that are both varied and meaningful.

4.4 Schedule distribution

As mentioned in Section 3, the outcome of the framework is a series of realisations of schedules. To illustrate this concept, we return to the example of Claire, presented in the previous section.



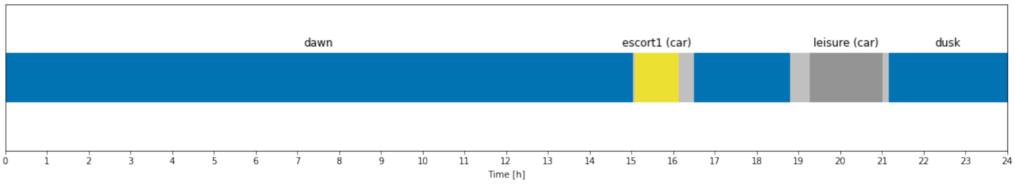
(a) Solution 1

(b) Solution 2

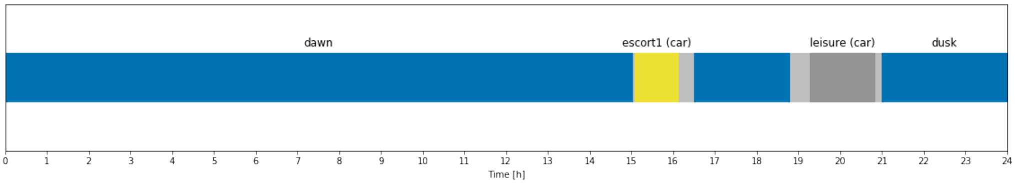
(c) Solution 3

(d) Solution 4

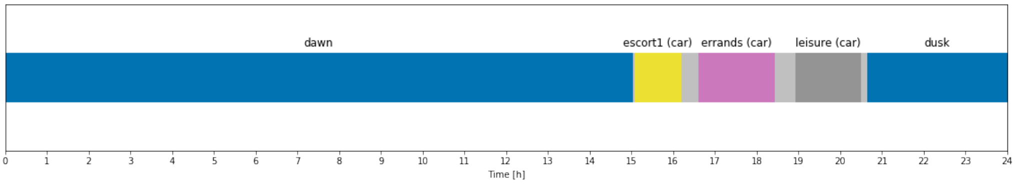
Figure 5: Generated schedules for Claire



(a) Solution 1



(b) Solution 2



(c) Solution 3

Figure 6: Generated schedules for Dylan

The framework is designed to capture the interactions of the activities in the schedule, leading to complex distributions of activity participation, start-time, and duration for each considered activity. This is illustrated in Figures 7-11, that show the distribution of these quantities across 1000 realisations of chosen schedules.

Regarding activity participation, we can note that 55.7% of the generated schedules contain out-of-home activities (as opposed to a full day spent at home). Out of these solutions, *Work* is among the most scheduled activities, and more than half the out-of home schedules contain the discretionary activity *shopping*. The *errands* activity is the third most scheduled activity, followed by *leisure* and *escort*. The latter has conflicting timings with both work and shopping, which may offer higher utility gains depending on the value error terms.

The simulation results are driven by the random quantities in the utility function, such as the utility of participation U_s^1 and the error term ε_S. Figure 8 shows the influence of different values of the utility of participation of the work activity ($U_s^1 = \{0, 10, 50, 100\}$, on the overall activity participation. The utility of participation for the other activities is still set to 0. For higher utilities, the work activity is always scheduled. However, this change has a limited effect on other activities, which are scheduled in similar proportions in all cases.

Figure 9 illustrates the effect of the variance of the error term on the activity participation. For very high variances, there is no clear difference between activities in terms of participation. This suggests that the magnitude of the error term overpowers the schedule deviations terms, thus limiting the hierarchy between activities.

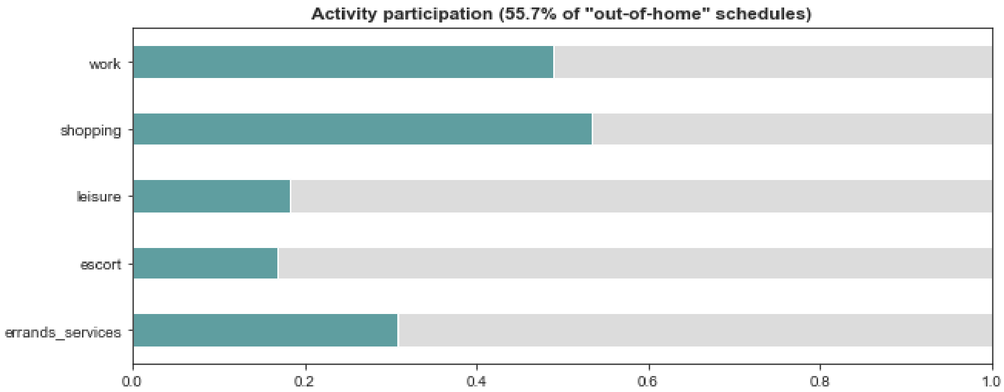


Figure 7: Distribution of activity participation (1000 runs)

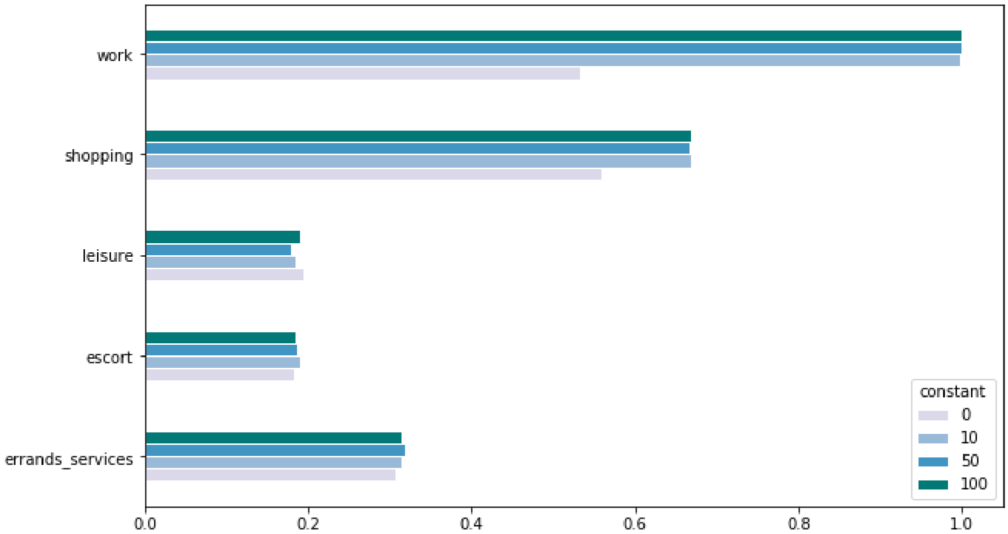


Figure 8: Distribution of activity participation, for different values of the participation utility

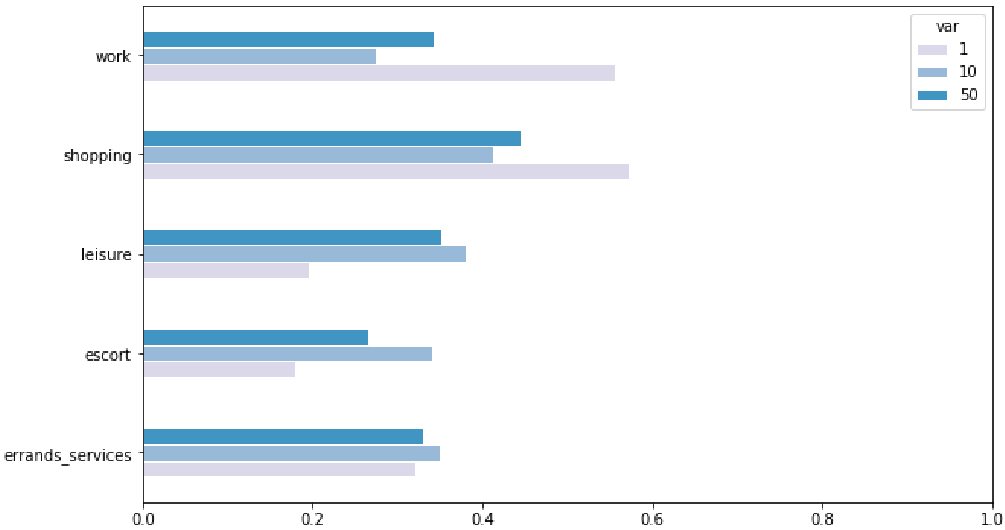


Figure 9: Distribution of activity participation, for different variances of the random term

The distributions of start times (Fig. 10) appear to be complex, and most of the time multimodal, with the noticeable exception of the “home" activity[4]. The distribution of *work* is seemingly unimodal, centred around the desired start time with very low variance. This is due to the high penalties associated with schedule deviations for this activity (cf. §3). It is worth noting that the *escort* and *shopping* activities are not centred around the desired time. Given that these are the two activities that had conflicting timings, this result shows the trade-off made during the optimisation process: in most schedules, these activities are started earlier to accommodate other activities for which the penalties for schedule deviations are higher.

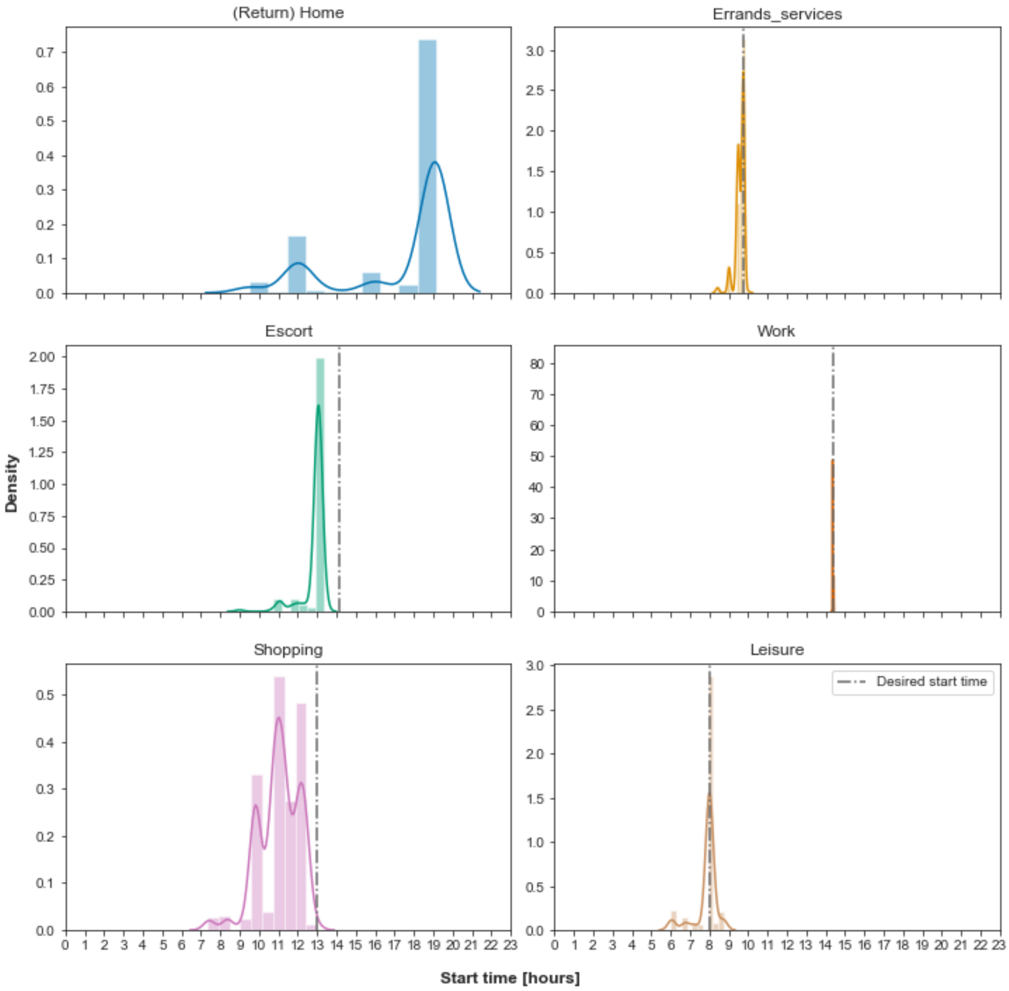


Figure 10: Distribution of start times per activity (1000 runs)
NB: The scale of the y-axis has been chosen for visibility.

Similar observations can be made for the distributions of durations (Fig. 11): the duration assigned to *work* is almost deterministic, and centred around the desired duration, while the durations allocated to flexible activities are multimodal. Again, when the desired durations involve schedule conflicts, the distributions are not centred around around the desired duration, and tend to have large spread (e.g. shopping, errands).

[4]Regarding the *home* activity, given the constraint that the day must start and end at home, we only show the time of the last return home.

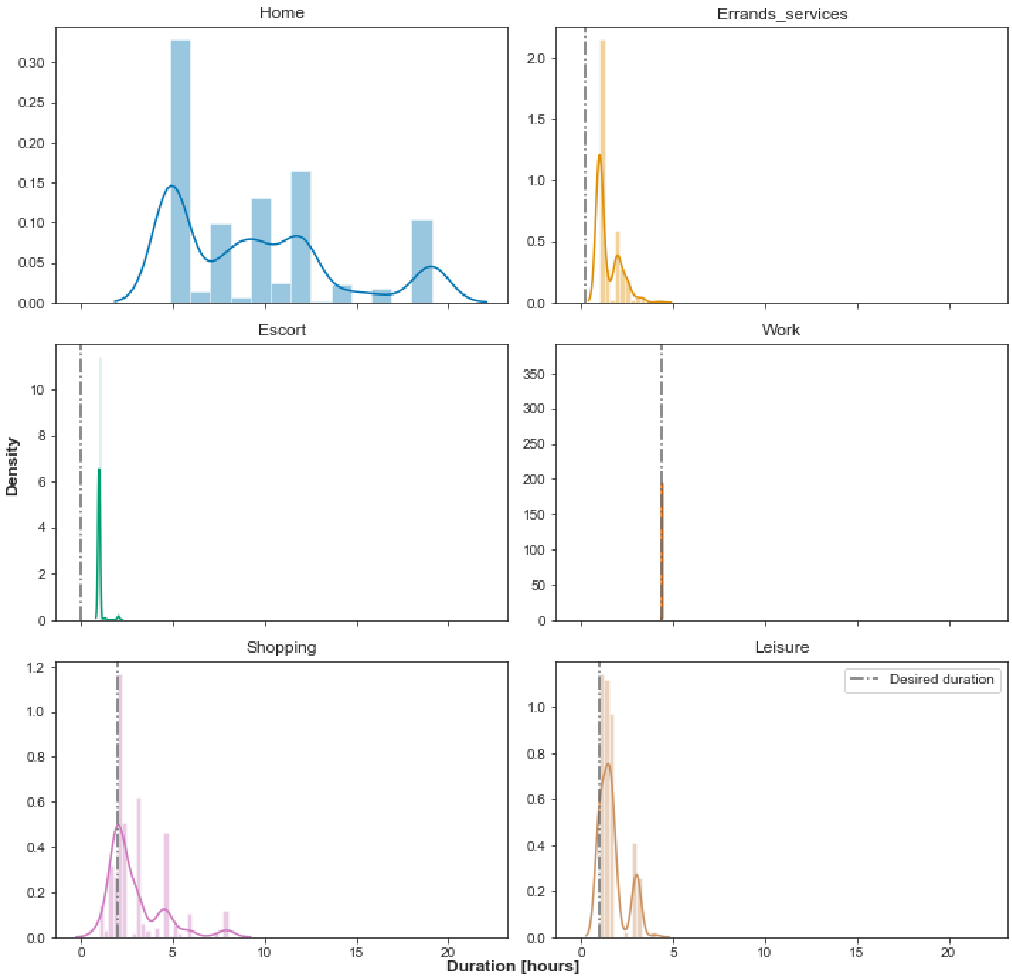


Figure 11: Distribution of duration per activity (1000 runs)

These distributions are also affected by the random terms. For instance, Figure 12 shows the distribution of start times obtained by increasing the variance of the random term to 10. While the change does not significantly impact the distribution of start times for the *work* activity, the *escort*, *leisure* and *shopping* activities are more spread in time, leading to more variety in the generated solutions. However, all distributions still seem to have a mode relatively close to, or centred around, the desired start time.

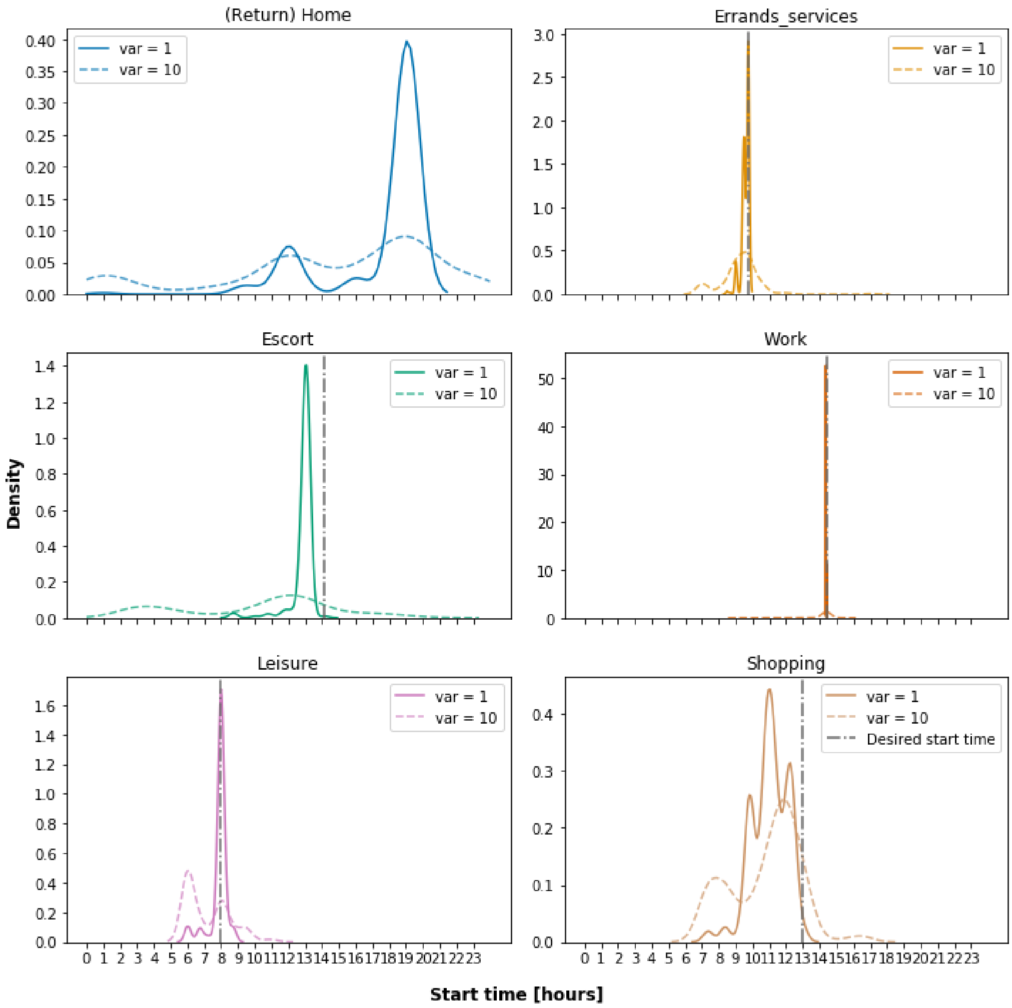


Figure 12: Distribution of start times per activity and variance of the random term

The experimental results show that the framework is able to generate different realisations of chosen schedules for given sets of considered activities, locations, modes, and timing preferences.

The multimodal distributions of the decision variables (start times and durations) highlight the scheduling trade-offs that are made during the optimisation process. These variations impact "flexible" activities in particular, which are characterised by lower penalties for schedule deviations.

Furthermore, the distributions emphasise the influence of the parameters of the model on its outputs, and consequently, the importance of selecting ranges of values that ensure both varied and stable solutions.

5 Conclusion and future work

This paper presents an integrated framework to model the trade-offs made by individuals when scheduling activities. The main characteristics of our methodology are as follows:

- All choices pertaining to daily mobility (activity scheduling, mode choice, activity location) can be considered simultaneously, and trade-offs between these choices are easily modelled.
- The choice of a schedule is explicitly modelled as a mixed integer optimisation problem solved by the decision maker.
- Due to the complexity of the choice model, there is no close form probability formulation. Instead, the framework allows the empirical distribution of the choice model to be estimated using simulation.

The proposed framework has been designed to deal with the complex interactions of all the dimensions involved in activity-based models. Clearly, it comes with a cost. In particular, there are several challenges that will have to be addressed by future research, including:

- Preparation of the input data,
- Estimation of the unknown parameters from data,
- Interactions among agents, and
- Decisions made at the household level.

In the above case study, we have assumed that the parameters of the model were not known. It is usually not the case in practice: they must be estimated from data, using, for instance, maximum likelihood estimation. One significant challenge for the application of maximum likelihood estimation to the activity-based context is the combinatorial nature of the choice set. As the alternatives, or possible schedules, cannot be enumerated, it is necessary to rely on samples of alternatives to estimate the model. This method is well documented in the literature (McFadden, 1978, Guevara and Ben-Akiva, 2013). The choice set generation itself can be performed with Markov Chain Monte Carlo methods, similarly to the method proposed by Flötteröd and Bierlaire (2013) in the context of route choice.

Furthermore, the combination of the activity scheduling decisions of many travellers in an area has an impact on the travel patterns of this area. As such, travel times are not exogenous, in contrast to what is assumed by our framework. The framework should therefore be coupled with mobility simulation tools, that take activity schedules as input, and generate indicators such as travel time and level of congestion as output (e.g. MATSIM (Axhausen et al., 2016), SUMO (Lopez et al., 2018), SimMobility (Adnan et al., 2016)), etc.)

Finally, the coordination of the activity scheduling decisions among all members of the same household is currently ignored by our framework. An extension that accounts for intra-household interactions is also an interesting topic for future research.

Acknowledgments

This work is part of a collaborative project with the Swiss Federal Railways (SBB), supported by the Swiss Innovation Agency (Innosuisse) (Project number: 2155006680).

Michel Bierlaire dedicates this article to John Polak, who introduced him to the fascinating world of travel behaviour in the early days of his PhD.

References

Adler, T. and Ben-Akiva, M. (1979). A theoretical and empirical model of trip chaining behavior. *Transportation Research Part B: Methodological*, 13(3):243 – 257.

Adnan, M., Pereira, F. C., Azevedo, C. M. L., Basak, K., Lovric, M., Raveau, S., Zhu, Y., Ferreira, J., Zegras, C., and Ben-Akiva, M. (2016). SimMobility: A multi-scale integrated agent-based simulation platform. In *95th Annual Meeting of the Transportation Research Board Forthcoming in Transportation Research Record.*

Arentze, T. A., Ettema, D., and Timmermans, H. J. (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.

Aschauer, F., Rösel, I., Hössinger, R., Kreis, H. B., and Gerike, R. (2019). Time use, mobility and expenditure: an innovative survey design for understanding individuals' trade-off processes. *Transportation*, 46(2):307–339.

Axhausen, K., Horni, A., and Nagel, K. (2016). *The multi-agent transport simulation MATSim*. Ubiquity Press.

Bhat, C. R. (2005). A multiple discrete-continuous extreme value model: Formulation and application to discretionary time-use decisions. *Transportation Research Part B: Methodological*, 39(8):679–707.

Bhat, C. R. (2018). A new flexible multiple discrete–continuous extreme value (MDCEV) choice model. *Transportation Research Part B: Methodological*, 110:261–279.

Bhat, C. R., Guo, J. Y., Srinivasan, S., and Sivakumar, A. (2004). Comprehensive Econometric Microsimulator for Daily Activity-Travel Patterns. *Transportation Research Record: Journal of the Transportation Research Board*, 1894(1):57–66.

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.

Bradley, M. and Bowman, J. (2008). Design Features of Activity-Based Microsimulation Models for U.S. Metropolitan Planning Organizations: A Summary. In *Transportation Research Board Conference Proceedings*, pages 11–20.

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.

English, E. (2020). Time in Consumer Theory.

Ettema, D., Borgers, A., and Timmermans, H. (2000). A simulation model of activity scheduling heuristics: An empirical test. *Geographical and Environmental Modelling*, 4(2):175–187.

Ettema, D. and Timmermans, H. (2003). Modeling Departure Time Choice in the Context of Activity Scheduling Behavior. In *Transportation Research Record*, number 1831, pages 39–46. National Research Council.

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

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*.

Hilgert, T., Heilig, M., Kagerbauer, M., and Vortisch, P. (2017). Modeling Week Activity Schedules for Travel Demand Models. *Transportation Research Record: Journal of the Transportation Research Board*, 2666(1):69–77.

Lopez, P. A., Behrisch, M., Bieker-Walz, L., Erdmann, J., Flötteröd, Y.-P., Hilbrich, R., Lücken, L., Rummel, J., Wagner, P., and Wießner, E. (2018). Microscopic traffic simulation using sumo. In *The 21st IEEE International Conference on Intelligent Transportation Systems*, pages 2575–2582. IEEE.

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673).

Nurul Habib, K. M. (2011). A random utility maximization (RUM) based dynamic activity scheduling model: Application in weekend activity scheduling. *Transportation*, 38(1):123–151.

Nurul Habib, K. M. and Miller, E. (2009). Modelling activity generation: A utility-based model for activity-agenda formation. *Transportmetrica*.

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.

Pas, E. (1985). State of the art and research opportunities in travel demand: Another perspective. *Transportation Research Part A: General*, 19(5-6):460–464.

Pendyala, R., Ryuichi, K., Akira, K., Toshiyuki, Y., and Satoshi, F. (2005). Florida Activity Mobility Simulator: Overview and preliminary validation results. *Transportation Research Record*, (1921):123–130.

Recker, W., McNally, M., and Root, G. (1986). A model of complex travel behavior: Part I—Theoretical development. *Transportation Research Part A: General*, 20(4):307 – 318.

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

Smith, L., Beckman, R., and Baggerly, K. (1995). TRANSIMS: Transportation analysis and simulation system. Technical report, Los Alamos National Lab., NM (United States).

Timmermans, H. (2003). Recent and expected directions in modelling transportation. *SATC 2003*.

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*.