Scheduling of daily activities: an optimization approach

Janody Pougala

Tim Hillel

Michel Bierlaire

Transport and Mobility Laboratory (EPFL) **May 2020**

STRC | 20th Swiss Transport Research Conference
Monte Verità / Ascona, May 13 – 15, 2020

Transport and Mobility Laboratory (EPFL)

Scheduling of daily activities: an optimization approach

Janody Pougala, Tim Hillel, Michel Bierlaire
Transport and Mobility Laboratory
Ecole Polytechnique Fédérale de Lausanne
Station 18
1015 Lausanne
phone:
fax:
{janody.pougala,tim.hillel,michel.bierlaire}@epfl.ch

May 2020

Abstract

Transport planners and operators have to face nowadays increasingly saturated networks and shifts in mobility behaviors driven by societal changes and emerging technologies. Traditional trip-based models become very limited in terms of behavioral realism when it comes to anticipating and accommodating the users' new, and often hard-to-capture needs. The shift towards activity-based approaches is thus natural, as this alternative is better equipped to deal with individual-level granularity (Castiglione *et al.,* 2014). The assumption behind these models is that all transport-related choices made by a person (e.g. number of trips, location and mode choice) are derived from the need to do activities (Bowman and Ben-Akiva*,* 2001), and their spatiotemporal sequence. We propose a modeling approach based on first principles: a traveler schedules their activities in order to maximize the total time-dependent utility they can derive out of them, thus solving a mixed integer optimization problem. The new model generates distributions of schedules for each individual, from which likely outcomes can be drawn. This approach enables simultaneous consideration of multiple choice dimensions (e.g. activity, location, mode choices). This allows for more flexibility than sequential approaches which tend to not appropriately represent scheduling trade-offs. The model is tested using trip diary data from the 2015 Swiss Mobility and Transport Microcensus and a dataset collected by the authors. The results show that the model can generate realistic activity schedules for a wide range of individuals.

Keywords

Activity-based modeling; Transport demand; Mixed-integer optimization; Simulation

1 Introduction

Trip-based models have been for decades the traditional approach to forecast travel demand. In trip-based analyses, trip purpose, origins and destinations are usually predicted independently, then paired and assigned to the transport network in subsequent steps. As the interrelations between these choices are not taken into account, trip-based models are limited when it comes to aid decision-makers to manage existing networks (Castiglione *et al.,* 2014) or deal with out-of-the-norm situations that often require an immediate but accurate response. This has become apparent in the scope of the recent COVID-19 outbreak: in order to enforce efficient and targeted measures to prevent and/or manage the spread of the virus, it has been essential for many governments around the world to grasp the complexities of their citizen's daily mobility habits, and understand how and why they move. It is in response to the lack of flexibility and behavioral realism of the trip-based approach that the activity-based stream of transport research has emerged. The activity-based approach is founded on the assumption that the demand for travel is induced by the need to perform activities. Contrary to trip-based models, activity-based models (ABM), consider the correlations between an individual's mobility choices. In addition, these models allow to understand mobility behaviors in wider social and environmental contexts, by taking into account interactions such as the influence of the household or the social circle, or model decisions that are driven by routine or subjective preferences rather than a hypothetical rationality. Ideally, the information obtained from activity-based models would represent a complete image of an individual's mobility and the multiple ways they are affected by and interact with their physical and social environment.

There have been many contributions focused on activity-based modeling in the transportation field, including examples of fully-functioning models that have been successfully implemented in practice. A review of these works can be found in Section 2. Authors have pointed out that current ABM do not quite fulfil all their promises yet (e.g. Axhausen (2000)), and several models fall back upon simplifying assumptions that significantly decrease the quality of the predictions. The approach presented in this paper attempts to reintroduce a high degree of behavioral realism by basing the modelling framework on first principles that are easily generalizable. The decision process is modelled as a mixed integer optimization problem. Individuals schedule their days by attempting to maximize the utility generated by the chosen sequence of activities, constrained by their available time and, more importantly, their preferences in terms of timing and locations. The framework and simulation process are presented in Section 3. Finally, results from an empirical investigation are shown and discussed in Section 4.

2 Literature review

Activity-based models aim to be behaviorally realistic by considering that the need to do activities drives the travel demand in space and time (Hägerstraand, 1970, Axhausen and Gärling, 1992, Kitamura, 1988). ABM consider mobility as multidimensional systems instead of single observations: models focus on overall behavioral patterns instead of discrete trips, decisions are analysed at the level of the household as opposed to isolated individuals, and dependences between events are taken into account (Timmermans, 2003, Pas, 1985).

Two main research streams have emerged within the scope of ABM. On one hand, utility-based models rely on the assumption that individuals maximize the utility they gain from performing activities. These models, such as Bowman and Ben-Akiva (2001), Adler and Ben-Akiva (1979), extend the traditional trip analysis by considering chains of trips (or tours). Mobility behaviour is thus modelled as a series of discrete choices, treated sequentially, solved with econometric methods such as advanced discrete choice models (Bowman and Ben-Akiva, 2001, Wang and Timmermans, 2000, Nurul Habib and Miller, 2009) or with micro-simulations (e.g. STARCHILD (Recker *et al.*, 1986), SMASH (Ettema *et al.*, 2000) CEMDAP (Bhat *et al.*, 2004), FAMOS (Pendyala *et al.*, 2005)). On the other hand, rule-based or computational process models such as Golledge *et al.* (1994), Arentze and Timmermans (2000) are based on the assumption that decision-makers are not driven by the desire to obtain optimal solutions, but rather consider context-dependent heuristics and conditional rules to make decisions (Timmermans, 2003).

Both approaches present a certain number of shortcomings. Rule-based models are mostly based on empirical analysis and lack a theoretical foundation to generalize them (Joh, 2004) and require extensive amounts of data to derive reliable *if-then* conditions. Econometric models tend to not consider behaviour itself, but rather assume it implicit to the full process (Timmermans, 2003), and they often need to be significantly simplified in order to be estimated.

The framework presented in this paper falls within the utility-based side of research. The novelty of the approach is that the output of the model is not a deterministic schedule, but rather a distribution of schedules for each given individual from which we are able to draw possible outcomes. A layer of randomness is thus included, allowing more flexibility in terms of modeling of behavior compared to the existing literature.

3 Integrated framework

We propose an utility-based mixed integer optimization model to generate individual schedules given a set of activities. We present in this section the core elements of our framework, and define all fundamental concepts and assumptions.

3.1 Definitions

We introduce the following definitions:

1. **Time:** we assume time to be discretized in δ time blocks of equal length, with T the time horizon (e.g. $T = 24$h),
2. **Space:** space is discretized in a finite set of locations $\mathcal{L}$. Each location is associated to at least one activity.
3. **Activity:** an activity i is uniquely defined as an action taking place in a physical location l, having a start time x_i and a duration τ_i. The sequence of activities $\{i, i+1\}$ generates a trip from location l_i to l_{i+1}, that can be performed using mode m. An activity than can be performed at multiple locations, or reached with different modes is modelled as multiple unique activities. For each individual n, we consider four possible sets of activities:

   a) *Feasible set $\mathcal{F}^n$:* all possible activities available to the individual n within a given time frame. This time frame may be larger than the time budget. For instance, for a daily scheduling process, the feasible set includes all activities that could be performed during the week, month, etc.,

   b) *Considered set C^n:* all activities that the individual n considers performing within their time budget. For example, given a list of activities to be performed in a week, the considered set includes activities that the individual plans to do in a given day. We let $C^n \subseteq \mathcal{F}^n$,

   c) *Scheduled set $\mathcal{S}^n$:* all activities the individual n schedules for a given day, based on the set (or agenda) they had previously considered. We let $\mathcal{S}^n \subseteq C^n$,

   d) *Realized set $\mathcal{R}^n$:* all activities actually performed by the individual n within their time budget. Given that the realized set is built from the scheduled set through external operations such as deletion, addition or substitution or activities, $\mathcal{R}^n$ and C^n could be distinct.

Only the considered and scheduled set are in the scope of our research.

3.2 First principles

The main assumption of the model is that scheduling a daily activity and mobility plan is a universal task. While the thought process involved in the generation of said schedule remains specific to the individual, we postulate that certain principles can be generalized to the whole population to explain the scheduling behavior. We consider the following axioms, introduced in the early works of Becker (1965) and Recker and Root (1981).

1. An individual n formulates an activity plan within a bounded timeframe (e.g. a day), referred to as *time budget*.

2. Each considered activity $i \in C^n$ is associated with a utility U_{in}, which quantifies the satisfaction derived by the individual when they perform the activity. Intuitively, we assume U_{in} to be *time-dependent*, as the gain of utility is not uniform across the distribution of durations and time-of-day decisions. For instance, one might enjoy working out for a given duration - but this satisfaction will decrease the longer the duration extends past this threshold, or, conversely, if they are unable to work out for a sufficient amount of time. Similarly, the same activity can yield different utilities depending on its time-of-day start (e.g. starting work on time vs. late, or running errands in the morning vs. in the evening). These observations rely on an important assumption: each individual n is *time-sensitive*, and they have quantifiable preferences for the timing of each activity (start, end, and/or duration).

3. The scheduling process itself is assumed to be driven by the desire to maximize the total satisfaction, or utility, provided by the activities subject to the given time budget constraint. Taking into account the timing preferences mentioned in item 2, we expect all schedule deviations (i.e. differences between what can be scheduled given the constraints and what they would rather do) to decrease the utility, with a rate that depends on the individual and their own flexibility towards activity i.

3.3 Utility functions

The central element of our framework is the utility U_{in} of performing activity i by the individual n. As expressed in Equation 1, the utility function is the sum of five main components:

$$U_{in} = U_{\text{const},in} + U_{\text{timing},in} + U_{\text{duration},in} + U_{\text{tt},in} + \varepsilon_{in} \tag{1}$$

- A constant utility of activity participation $U_{\text{const},in}$. Assuming this constant to be zero for all in-home activities, it represents the preference of performing the activity rather than staying at home, all other things being equal.
- Two terms $U_{\text{timing},in}$, $U_{\text{duration},in}$ which capture the (a priori negative) impact of schedule deviations on the total utility. Contrasting with Feil (2010), that only considers the disutility of being late to an activity, these terms express deviations in terms of both start time (early/late) and duration (too long/too short) respectively. They penalize divergences from the preferred schedule, to an intensity depending on the indivual's flexibility.
- A term $U_{\text{tt},in}$ which represents the utility of traveling to the location of the activity. Intuitively, this term has a negative impact on the total utility.
- A stochastic term $\varepsilon_{,in}$, defined as a sum of error components associated to specific values of the decision variables.

Utility of schedule deviations

1. Flexible: deviations from preferences for activity i are relatively unimportant, thus are less or not penalized.
2. Moderately flexible: deviations from preferences are moderately undesirable, and so are more penalized than in the flexible case.
3. Not flexible: deviations from preferences are not strongly undesirable, and are consequently highly penalized.

The impact of the three levels of flexibility on the penalization of schedule deviations is illustrated in Figure 1.These penalties may not be symmetrical, and it is necessary to define how one perceives a type schedule deviation in relation to the others. Intuitively, one can postulate that individuals do not penalize equally arriving early vs. late to the same activity, regardless of their flexibility. For start times, this observation is confirmed by studies on departure time preferences (Small, 1982, Arnott *et al.,* 1987). Fewer studies exist on the deviation from the optimal duration of an activity (usually the individual's preferred duration, but could be constrained by other factors such as the required daily working hours), however, several authors have identified possible *frustration* and *satiation* effects. (Ettema *et al.,* 2007)

In the early stages of the model, the values associated to each flexibility category are not specific to the individual. For example, all persons declaring to be flexible for any activity will penalize deviations in the same way. The model can thus account for priorities between different activities, which is analogous to the traditional classification of activities encountered in the literature (e.g. mandatory, maintenance and discretionary, and other similar categorizations) (Castiglione *et al.,* 2014), but this hierarchy is uniform across the population.

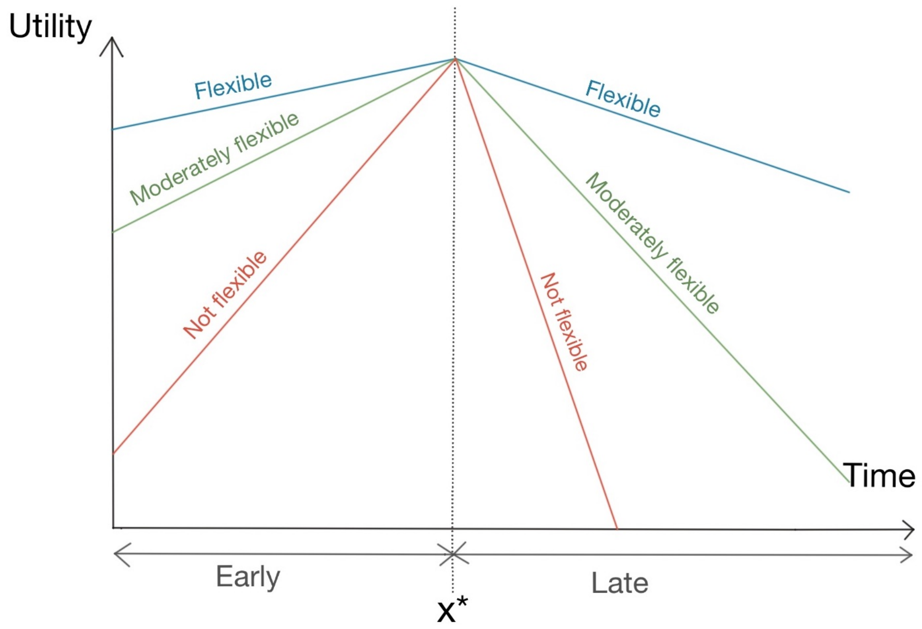


Figure 1: Impact of start time deviation on utility for different flexibilities

Equation 2 defines the impact on the utility of a deviation in regards to start time. When the activity is scheduled earlier than the desired time (i.e. $x^*_{in} - x_{in} > 0$), the deviation is penalized through the term $U_{\text{early},in}$, while $U_{\text{late},in} = 0$. On the other hand, if the activity is scheduled later than preferred (i.e. $x^*_{in} - x_{in} < 0$), the deviation is penalized through $U_{\text{late},in}$, while $U_{\text{early},in} = 0$. If the scheduled start time is equal to the desired time (i.e. $x^*_{in} - x_{in} = 0$), the utilty is not penalized. The same logic is applied to the scheduled duration, as defined by equation 3.

$$\begin{aligned} U_{\text{timing},in} &= U_{\text{early},in} + U_{\text{late},in} \\ &= \theta_{ek} \max(0;\ x^*_{in} - x_{in}) + \theta_{lk} \max(0;\ x_{in} - x^*_{in}) \end{aligned} \tag{2}$$

$$\begin{aligned} U_{\text{duration},in} &= U_{\text{short},in} + U_{\text{long},in} \\ &= \theta_{dsk} \max(0;\ \tau^*_{in} - \tau_{in}) + \theta_{dlk} \max(0;\ \tau_{in} - \tau^*_{in}) \end{aligned} \tag{3}$$

Utility of travel

At this stage of the model, we consider the simplest definition of the utility generated by traveling: a linear function of the travel time t. As defined in section 3.1, each activity i is defined by a unique location l_i, and can be reached from the location of the previous activity l_{i-1} by traveling with mode m_i. We assume that the impact of traveling on the utility of the activity is negative, meaning that an activity with a longer travel component will be regarded less favorably than an activity requiring a shorter travel time. We name θ_{ti} the penalty associated with traveling, and we consider it equal across individuals.

$$U_{\mathrm{tt}_{in}} = \theta_{ti} t\,(l_{i-1}, l_i, m_i) \tag{4}$$

This formulation can be easily extended to include additional features such as network-specific attributes (e.g. level of service).

Error components

We assume that each specific value k of a decision variable y in our optimization problem (see section 3.4) has a corresponding error component, and that the total stochastic term for an activity is their sum (Eq. 5). This specification defines a choice model equivalent to an error component mixed logit model (Train, 2003).

$$\varepsilon_{in} = \sum_{y \in Y} \sum_{k \in K_y} \delta^k_{iy} \varepsilon^y_k + \xi_{in} \tag{5}$$

with Y the set of decision variables, K_y the set of possible values for decision variable y, δ^k_{iy} an indicator variable equal to 1 if value k was chosen for decision variable y. We assume that the error components ε^y_k follow a multivariate normal distribution with a known correlation structure Σ. ξ_{in} is an i.i.d. error term specific to the individual n, such that $\xi_{in} \sim EV(0, \mu)$.

This error structure allows to take into account the very high correlations between different alternatives (e.g. in real life, there would likely be little perceived difference between a duration of 5 minutes and 6 minutes). Note that the number of unique components k for the errors of the timing decisions (start time and duration) is not necessarily equal to the chosen time discretization δ (as defined in Section 3.1). It could be beneficial from a computational perspective to choose low values of k, and increase them progressively to reach a tradeoff between quality of results and performance. The same conclusion can be made in the case where time is continuous, although the sums would be replaced by integrals.

Total utility function

Replacing (2), (3), (4) and (5) into (1) yields the total time-dependent utility function for activity i performed by individual n:

$$U_{in} = c_{in} + \theta_{ek} \max(0;\ x^*_{in} - x_{in}) + \theta_{lk} \max(0;\ x_{in} - x^*_{in}) + \theta_{dsk} \max(0;\ \tau^*_{in} - \tau_{in}) + \theta_{dlk} \max(0;\ \tau_{in} - \tau^*_{in}) + \theta_{ti} tt(l_{i-1}, l_i, m_i) + \sum_{y \in Y} \sum_{k \in K_y} \delta^k_{iy} \varepsilon^y_k + \xi_{in} \quad (6)$$

3.4 Mixed integer optimization framework

We model a person n with a set of activities C^n and a time budget T, who schedules all activities $i \in C^n$ by solving a mixed integer optimization problem. where the total utility of the schedule is maximized.

$$\Omega = \max \sum_i \omega_{i_n} U_{i_n} \quad (7)$$

The decision variables of the problem are the following:

- ω_{i_n}: a binary variable equal to 1 if activity i is scheduled and 0 otherwise,
- z_{ij_n}: a binary variable equal to 1 if activity i follows activity j in the schedule and 0 otherwise,
- x_{i_n}, τ_{i_n}: positive continuous variables representing respectively the start time and the duration of activity i.

The problem is subject to a set of constraints:

$$\sum_i \sum_j \left(\tau_{i_n} + z_{ij_n} t_{ijm_n}\right) = T \tag{8}$$

$$\omega_{\text{dawn}_n} = \omega_{\text{dusk}_n} = 1 \tag{9}$$

$$\omega_{i_n} \leq \tau_{i_n} \qquad \forall i \in C^n \tag{10}$$

$$\tau_{i_n} \leq \omega_{i_n} T \qquad \forall i \in C^n \tag{11}$$

$$z_{ij_n} + z_{ji_n} \leq 1 \qquad \forall i, j \in C^n, j \neq i \tag{12}$$

$$z_{ii_n} = 0 \qquad \forall i \in C^n \tag{13}$$

$$z_{i,\text{dawn}_n} = z_{\text{dusk},j_n} = 0 \qquad \forall i, j \in C^n \tag{14}$$

$$\sum_{i,i \neq j} z_{ij_n} = \omega_{j_n} \qquad \forall j \in C^n, j \neq \text{dawn} \tag{15}$$

$$\sum_{j,j \neq i} z_{ij_n} = \omega_{i_n} \qquad \forall i \in C^n, i \neq \text{dusk} \tag{16}$$

$$\left(z_{ij_n} - 1\right) T \leq x_{in} + \tau_{i_n} + z_{ij_n} t_{ijm_n} - x_{jn} \leq \left(1 - z_{ij_n}\right) T \qquad \forall i, j \in C^n \tag{17}$$

$$\sum_i \omega_i \leq 1 \qquad \forall i \in G \tag{18}$$

$$x_{i_n} \geq \gamma_i^- \qquad \forall i \in C^n \tag{19}$$

$$x_{i_n} + \tau_{i_n} \leq \gamma_i^+ \qquad \forall i \in C^n \tag{20}$$

Equation(8) constrains the total time assigned to the activities in the schedule (duration and travel time to successive activity with mode m) to be equal to the time budget. Equation (9) ensures that each schedule begins and end at home (*dawn* and *dusk* are respectively the first and last in-home activity of the day). Equations (10) and (11) enforce consistency with the activity durations, by requiring the activity to have a duration of at least one unit if it takes place (10), and for the activity to have zero duration if it does not take place. Equations (12)-(17) constrain the sequence of the activities: (12) ensures that two activities i and j can only follow each other once (thus can only be scheduled once) and (13) ensures that an activity cannot follow itself. (14), (15), (16) state that each activity but the first has only one predecessor, and each but the last only one successor. (17) enforces time consistency between two consecutive activities (with t_{ijm_n} the travel time between them using mode m). (18) ensures that only one activity within a group of duplicates G (as defined in section 3.1) is selected. Finally, (19) and (20) are time windows constraints. The outcome of the model is a feasible schedule S which includes activities from considered set C^n, and which complies with the constraints. As the utility functions of all activities depend on the error term, we expect different draws of ε_{i_n} to generate different solutions (see Section 3.5).

3.5 Simulation

The outputs of the model are defined conditionally on the distributions of the error terms ε_{i_n} and the parameters θ_n (which in this case are the penalty values). A simulation is thus performed in order to draw solutions (i.e. deterministic schedules) from the choice model. While traditional methods such as Metropolis-Hastings are appropriate to deal with combinatorial problems of this nature, we opt for a simpler approach: we draw first from the distribution of the error terms and the parameters, then solve the optimization problem for this instance. The resulting schedule is a single draw from the choice model.

4 Empirical investigation

4.1 Case study 1: Authors' dataset

The first case-study uses data collected from a survey on habitual daily activities and scheduling preferences for a sample of 10 individuals over the course of a 7-day week. On each day during the week, the respondents filled in their intended schedule for the following day (start time and duration for each activity), as well as all considered locations for each activity. In addition, they were asked to specify their scheduling flexibility for each activity, by giving a score $S = \{-1, 0, 1\}$, 1 indicating low flexibility, moderate flexibility, and high flexibility respectively. On the first day of the survey, the respondents were asked to fill the activities they intended to perform during the week, regardless of the day.

The activities were sorted in 11 categories (home, work[1], education[2], shopping[3], errands and services[4], escort, fitness, leisure, other). A travel time matrix was created for each respondent based on the locations they indicated having considered for the day. In the absence of information on mode choice (the respondents having only indicated which modes they had access to) the matrix was computed using the Google Directions API for all available modes (car, public transportation, bicycle and walking).

[1]This activity is only available to respondents who are regularly employed and have a usual workplace they have to travel to. It is worth noting that some of the respondents had a possibility to work from home. In this case, we categorized this as a *work* activity taking place at the location of the home, rather than including it in the *home* category. Nevertheless, we assume that working from home and working at the regular workplace do not necessarily share the same characteristics (e.g. flexibility, duration).

[2]Only available to respondents who are students

[3]Shopping for non-essential items.

[4]Shopping for essential items, e.g. food, and use of services e.g. medical appointments.

4.1.1 Results

Figure 2 illustrates the results obtained for one individual, and different draws of the error term ε_{in}. The chosen penalty values are summarized in Table 1[5], and the error term random variables for start time and duration were discretized as follows:

- Start time: 4 iid random variables $\varepsilon_k^x \sim \mathcal{N}(0, 1)$, for every 6h time block.
- Duration: 6 iid random variables $\varepsilon_k^\tau \sim \mathcal{N}(0, 1)$, with the following intervals: $\tau_{in} \in \{[0, 1h],\ ]1h, 3h],\ ]3h, 8h],\ ]8h, 12h],\ ]12h, 16h],\ ]16h, 24h[\}$

The individual of interest has the following considered set:

1. Working in the morning and the afternoon, either at their usual workplace or at home,
2. Having a lunch break, either to eat at a restaurant near their workplace, or to run errands.
3. A fitness session, preferably after work.
4. A leisure outing in the evening.

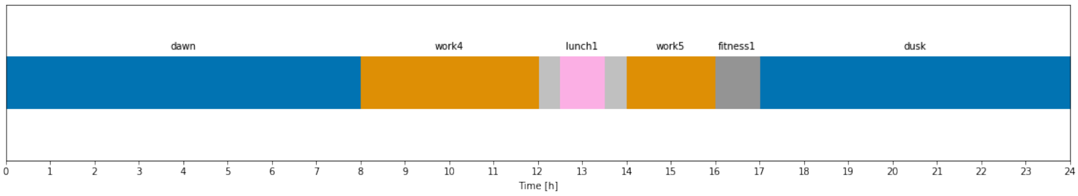
Table 1: Penalty values by flexibility, in units of utility

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

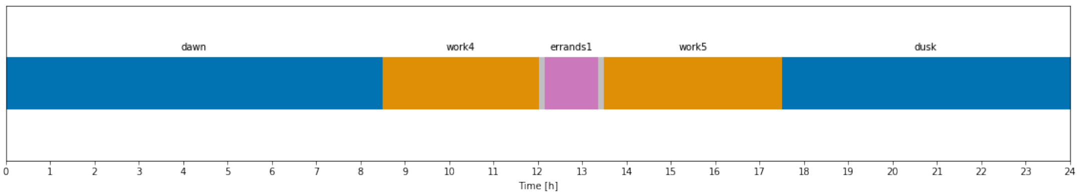
The first example (Fig. 2(a)) shows a schedule for a working day where both the morning and afternoon work blocks are scheduled at home, with a lunch break at a restaurant near the workplace, and a fitness session after work. In the example in Fig. 2(b) there is no fitness nor eating out planned, but the person is scheduled to run errands during lunchtime.The work

[5]The penalty values were arbitrarily assigned, using results from (Small, 1982)

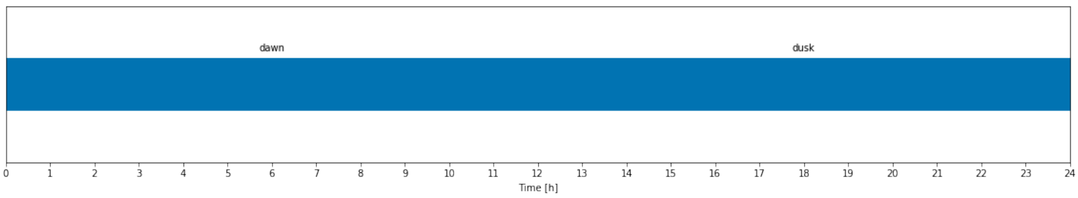
duration is 1h longer than the previous solution. While both schedules are different, they are both feasible in the sense that the mandatory activities (work) are scheduled and are relatively close to the stated timing preferences, while lower priority activities (lunch, errands, fitness), are only scheduled if possible and if they provide a substantial gain in utility. Given that the leisure outing is not scheduled in any of the 4 presented schedules, we can infer that neither condition was met. On the other hand Fig. 2(c) and 2(d) show examples that are sound in the context of the model, but unlikely in terms of behavior: 2(c) represents a day with no activity scheduled, and in 2(d), only half of the work time scheduled. Both these instances highlight the importance of choosing the right combination of error terms to obtain schedules that are not only feasible, but also realistic given context-specific conditions.



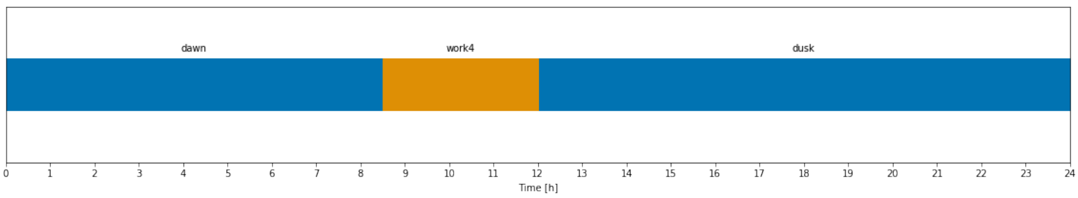
(a) Solution 1



(b) Solution 2



(c) Solution 3



(d) Solution 4

Figure 2: Example of results from the authors' dataset

4.2 Case study 2: Swiss Mobility and Transport microcensus

The second case study uses data from the Swiss Mobility and Transport microcensus (MTMC), a Swiss nationwide survey gathering insights on the mobility behaviours of local residents (OFS, 2015). Respondents provide their socio-economic characteristics (e.g. age, gender, income) and those of the other members of their household, and information on their daily mobility habits, and detailed records of their trips during a reference period (1 day). The 2015 edition of the MTMC contains 57'090 individuals, and 43'630 trip diaries. We test the model on Lausanne residents only, reducing the number of daily trip diaries to 2'227.

From the 13 trip purposes (and an additional 18 *leisure* subcategories) available in the travel diaries, we keep only 9: home, work, education, shopping, errands and use of services, business trips, leisure and escort. The start, end and durations of each activity are derived from the timings of the recorded trips. The latitude and longitude values are provided for each visited locations, and these measures are used to produce a travel time matrix using the Google Directions API. For the sake of simplicity, we ignore mode choice, and consider only the car mode.

A major limitation of this revealed preference dataset is the lack of subjective and qualitative information, such as the preferences of the individual in terms of start times and durations, their flexibility or simply whether the activities they have recorded for the day were freely chosen or constrained in any way. To estimate the model, we make the following simplifications:

- The desired start times and durations are assumed equal to timings recorded by the individual.
- The feasible time windows are obtained using the average values for start and end times for each activity in an out-of-sample distribution, obtained using 30% of the observations in the Lausanne sample.
- We assume that the individuals visits all locations of their considered set. This implies that there can be no duplicates of activities and therefore constraint (18) does not apply.
- Activities are classified in 3 categories (mandatory, maintenance and discretionary, to which a flexibility profile is assigned, uniformly across all population (Table 2). Deviation penalties are defined based on this classification (Table 1).

[6] Not including mandatory home stays *dawn* and *dusk*

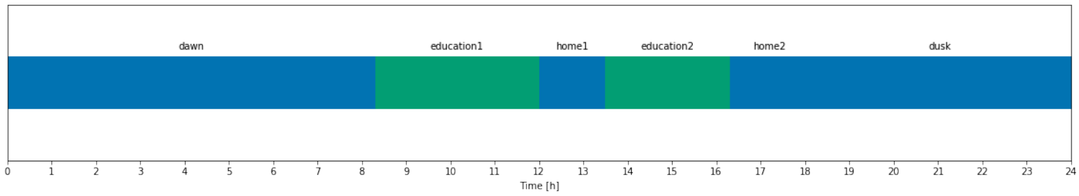
Table 2: Categories and flexibility profiles for activities in the MTMC
The following acronyms are used: F=flexible, MF=moderately flexible, NF=not flexible

Activity	Category	Flexibility profile	
Work		*Start*	*Duration*
Education	Mandatory	Early: NF	Short: NF
Business trip		Late: MF	Long: NF
Errands, use of services	Maintenance	Early: MF	Short: MF
Escort		Late: MF	Long: F
Home[6]			
Shopping	Discretionary	Early: F	Late: MF
Leisure		Short: F	Long: F

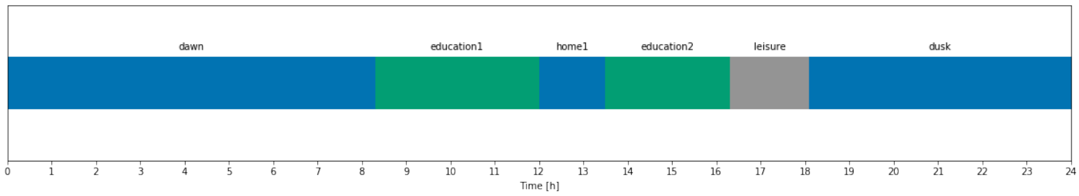
4.2.1 Results

We present one example from an individual in the MTMC. The set of considered activities contains two education activities (preferred in the morning, and in the afternoon), with a return at home during lunchtime. A leisure activity is also considered, to start at the end of the last education period, followed by a return home.

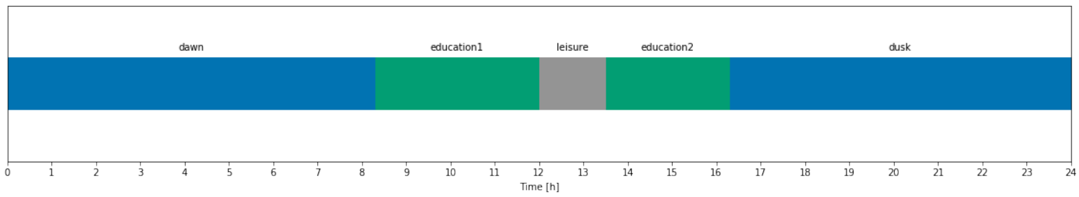
Figure 3 shows three unique outputs produced by the model, for different draws of ε_{in} (keeping the same discretization as in section 4.1). The first option (Fig. 3(a)) shows a sequence in which both *education* instances are scheduled, including the return home at noon. In the second option (Fig. 3(b)), the *leisure* activity is scheduled, relatively closely to the individual's preferences. The third solution (Fig. 3(c)) also includes the leisure activity, however the proposed timing has deviated substantially from the assumed preferences. Unsurprisingly, the changes in solution affect mainly the discretionary activity, for which deviations are far less penalized than its mandatory counterparts.



(a) Solution 1



(b) Solution 2



(c) Solution 3

Figure 3: Example of results from the MTMC

5 Conclusion and future work

This paper presents an integrated framework to solve the activity-based problem by using an optimization approach. The advantage of the proposed methodology is two-fold:

1. all choices pertaining to daily mobility (activity scheduling, mode choice, destination) can be considered simultaneously,
2. the probabilistic nature of the model brings more flexibility and realism to the results.

In its current stage, the model relies on a number of assumptions to produce results, as the required insights are not always available in traditional data sources such as travel diaries. One remaining challenge is thus to provide heuristics to obtain estimators for the missing attributes. Specifically, information such as the activities considered by the individual (as opposed to those they record in the travel diary), their preferences in terms of start time, duration or frequency of the activity, or their flexibility are difficult to derive from straightforward, factual surveys.

Table 3 summarizes data requirements and two possible solutions to overcome the lack of information. The *heuristic* column describes straightforward implementations, which can be used to obtain initial results in the absence of exhaustive data.

A crucial improvement for future iterations of the model is the enhancement of the utility functions, notably by integrating socio-economic attributes. A significant challenge will be to estimate the values of the parameters from the observations, which can be approached with Bayesian estimation methods.

Table 3: Data requirements for operational model

Requirements	Rigorous solution	Heuristic
Desired start times and durations	Habit analysis and identification of typical timings in multidays diaries	Out-of-sample distributions, with geographical sampling
Flexibility	Habit analysis in multidays diaires: flexibility would be the timing variability	Assign a flexibility profile to each activity based on literature classification
Penalty values	Calibrated on data; n-dependent	From literature, homogeneous across all population
Constant utility of activities	Calibrated on data	Captured by error term
Feasible time windows	Data collection	Minima and maxima values in out-of-sample distributions of start and end times for each activity, across the population
Travel time matrix	Build a set of considered locations within defined radius of home and work location (e.g. using geographical sampling), then use Google Maps distance matrix API	Use Google API between locations recorded by individual n in diary
Variance of error term	Calibrate from data	Trial and error, minimization of distance with optimal schedule

6 References

Adler, T. and M. Ben-Akiva (1979) A theoretical and empirical model of trip chaining behavior, *Transportation Research Part B*, ISSN 01912615.

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

Arnott, R., A. De Palma and R. Lindsey (1987) *Schedule delay and departure time decisions with heterogeneous commuters*, University of Alberta, Department of Economics.

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

Axhausen, K. W. and T. Gärling (1992) Activity-based approaches to travel analysis: Conceptual frameworks, models, and research problems: Foreign summaries, *Transport Reviews*, ISSN 14645327.

Becker, G. S. (1965) A Theory of the Allocation of Time, *The Economic Journal*, ISSN 00130133.

Bhat, C. R., J. Y. Guo, S. Srinivasan and A. Sivakumar (2004) Comprehensive Econometric Microsimulator for Daily Activity-Travel Patterns, *Transportation Research Record: Journal of the Transportation Research Board*, **1894** (1) 57–66, jan 2004, ISSN 0361-1981.

Bowman, J. and M. Ben-Akiva (2001) Activity-based disaggregate travel demand model system with activity schedules, *Transportation Research Part A: Policy and Practice*, **35** (1) 1–28, jan 2001, ISSN 0965-8564.

Castiglione, J., M. Bradley and J. Gliebe (2014) *Activity-Based Travel Demand Models: A Primer*, Transportation Research Board, Washington, D.C., mar 2014, ISBN 978-0-309-27399-2.

Ettema, D., F. Bastin, J. Polak and O. Ashiru (2007) Modelling the joint choice of activity timing and duration, *Transportation Research Part A: Policy and Practice*, ISSN 09658564.

Ettema, D., A. Borgers and H. Timmermans (2000) A simulation model of activity scheduling heuristics: An empirical test, *Geographical and Environmental Modelling*, **4** (2) 175–187, ISSN 13615939.

Feil, M. (2010) Choosing the Daily Schedule: Expanding Activity-Based Travel Demand Modeling, Ph.D. Thesis, ETH Zürich.

Golledge, R. G., M.-P. Kwan and T. Garling (1994) Computational-Process Modelling of Household Travel Decisions Using a Geographical Information System, *Technical Report*.

Hägerstraand, T. (1970) What About People in Regional Science?, *Papers in Regional Science*, ISSN 14355957.

Joh, C.-H. (2004) Measuring and predicting adaptation in multidimensional activity-travel patterns, Ph.D. Thesis, Citeseer.

Kitamura, R. (1988) An evaluation of activity-based travel analysis.

Nurul Habib, K. M. and E. J. Miller (2009) Modelling activity generation: A utility-based model for activity-agenda formation, *Transportmetrica*, ISSN 18128602.

OFS (2015) Comportements de la population en matiÃ¨re de transports, *Technical Report*, OFS.

Pas, E. I. (1985) State of the art and research opportunities in travel demand: Another perspective, *Transportation Research Part A: General*, **19** (5-6) 460–464, ISSN 01912607.

Pendyala, R., R. Kitamura, A. Kikuchi, T. Yamamoto and S. Fujii (2005) Florida activity mobility simulator overview and preliminary validation results, January 2005.

Recker, W. W., M. G. McNally and G. S. Root (1986) A model of complex travel behavior: Part I-Theoretical development, *Transportation Research Part A: General*, ISSN 01912607.

Recker, W. W. and G. S. Root (1981) Toward a dynamic model of individual activity pattern formulation.

Small, K. A. (1982) The scheduling of consumer activities: work trips., *American Economic Review*, ISSN 00028282.

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

Train, K. E. (2003) *Discrete choice methods with simulation*, ISBN 9780511753930.

Wang, D. and H. Timmermans (2000) Conjoint-based model of activity engagement, timing, scheduling, and stop pattern formation, *Transportation Research Record*, ISSN 03611981.