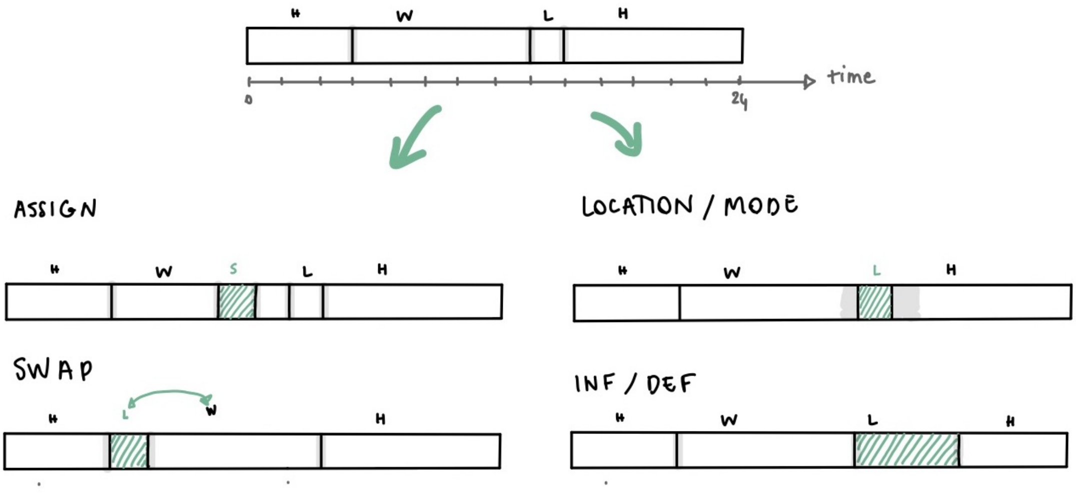


Choice set generation for activity-based models

Janody Pougala

Tim Hillel

Michel Bierlaire

Transport and Mobility Laboratory

September 2021

STRC | **21th Swiss Transport Research Conference**
Monte Verità / Ascona, September 12 – 14, 2021

Transport and Mobility Laboratory

Choice set generation for activity-based models

Janody Pougala, Tim Hillel, Michel Bierlaire
Transport and Mobility Laboratory
(TRANSP-OR)
EPFL

phone:
fax:
{janody.pougala,tim.hillel,
michel.bierlaire}@epfl.ch

September 2021

Abstract

Activity-based models have seen a significant increase in research focus in the past decade. Based on the fundamental assumption that travel demand is derived from the need to do activities and time and space constraints (Hägerstraand, 1970, Chapin, 1974). ABM offer a more flexible and behaviourally centred alternative to traditional trip-based approaches. Econometric — or utility-based — activity-based models (e.g., Adler and Ben-Akiva, 1979, Bowman and Ben-Akiva, 2001) postulate that the process of activity generation and scheduling can be modelled as discrete choices. Individuals derive a utility from performing activities, and they schedule them as to maximise the total utility of the schedule. In classical discrete choice model applications, the parameters of the utility functions can be estimated by deriving their maximum likelihood estimators. As the likelihood function is defined over a full enumeration of the alternatives in the choice set, this approach is limited for activity-based applications: the set of possible activities and their spatio-temporal sequence is combinatorial and not fully observed by either the decision-maker or the modeller. While discrete choice models can be estimated over samples of alternatives (e.g., Guevara and Ben-Akiva, 2013) an appropriate definition of such sample is as crucial as it is challenging. This paper presents a methodology to sample a choice set of full daily schedules for a given individual and a list of activities. The Metropolis-Hastings algorithm allows us to explore the space efficiently and draw both high and lower probability alternatives for consistent estimation of the parameters. The methodology is tested on a sample of individuals from the 2015 Swiss Mobility and Transport Microcensus (Office fédéral de la statistique and Office fédéral du développement Territorial, 2017).

Keywords

Activity-based modelling, Metropolis-Hastings, choice set, simulation, sampling

1 Introduction

Activity-based models stem from the fundamental assumption that travel demand is derived from the needs of individuals to perform activities (Bowman and Ben-Akiva, 2001) and that this need is influenced by space and time constraints (Chapin, 1974, Hägerstraand, 1970). There exist many examples of activity-based models, which fall within two main categories: rule-based models and econometric models. Rule-based, or computational process, models (e.g. Golledge *et al.*, 1994, Timmermans, 2003, Arentze and Timmermans, 2000) use decision rules to derive feasible solutions. Econometric models postulate that scheduling can be explained with econometric processes such as random utility maximisation. Advanced econometric techniques such as discrete choice models are used to explain and predict the activity-travel behaviour of individuals. Sequential discrete choice models (e.g. Adler and Ben-Akiva, 1979, Bowman and Ben-Akiva, 2001) consider a series of choices done consecutively with varying amounts of feedback between each step. Joint models (e.g. Ettema *et al.*, 2007, Nurul Habib, 2018, Charypar and Nagel, 2005) also integrate correlations between each aspect of the scheduling decision by evaluating them simultaneously.

Using discrete choice models and concepts such as random utility maximisation in an activity-based context implies the derivation of choice probabilities, and the calibration of maximum likelihood estimators of the parameters of the utility functions. This requires the modeller to assume a finite and enumerable choice set, and universal across the population. This assumption is difficult to justify in the context of activity-based models: if one considers as schedule to be a discrete alternative subject to choice, then the choice set comprised of all possible combinations of this schedule is huge. For example, a daily schedule discretised in t blocks of time (e.g. 5 minutes intervals), and a possibility to perform n different activities at each block, at l possible locations and a choice between m modes of transportation, then the full choice set contains $\binom{n \cdot l \cdot m+t-1}{t}$ combinations. For $n = 2$ activities, $l = 2$ possible locations, $m = 2$ modes of transportation and $t = 24$ blocks of 1h, we are dealing with more than 2 million possible schedules. It is evident that individuals do not consider the entirety of the choice set when they make decisions for their schedules of activities. They are only aware of a fraction of it, and out of these alternatives that they can think of, they would discard many as unfeasible because they do not fit an arbitrary and context-dependent set of requirements and constraints. The challenge for modellers is therefore to generate a choice set that is both realistic and appropriate to estimate parameters.

In this paper, we present a methodology based on the Metropolis-Hastings algorithm to strategically sample alternatives across time and space, to estimate the parameters of the utility functions of econometric activity-based models. This methodology is an extension of the process developed by Flötteröd and Bierlaire (2013) for sampling of alternatives in route choice models, and its application to activity networks by Danalet and Bierlaire (2015). Section 2 presents a brief review of current approaches to estimate choice sets, with an emphasis on applications in utility-based ABMs. We describe the methodology in Section 3, and a practical application in Section 4, using the optimisation framework for activity scheduling described by Pougala *et al.* (2021). Note that this paper is merely a proof of concept of the sampling methodology applied to an activity-based framework where scheduling and mobility choices are regarded simultaneously. As such, we discuss in Section 5 leads for future work and important considerations to take into account and to investigate to make the concept operational for implementation.

2 Literature review

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

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

The early works of Hägerstraand (1970) and Chapin (1974) established the fundamental assumption of activity-based models, that the need to do activities drives the travel demand in space and time. Consequently, mobility is modelled as a multidimensional system rather than a set of discrete observations. Unlike traditional trip-based models, ABMs 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.

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 scheduling decisions can be modelled as discrete choices: sequential discrete choice models consider a series of choices done consecutively with varying amounts of feedback between each step. On the other hand, joint models also integrate correlations between each aspect of the scheduling decision by evaluating them simultaneously. Other models do not consider the choice as fully discrete, but an hybrid consumption of discrete and continuous "goods".

Little work in the field of activity-based modelling specifically tackles choice set generation for estimation of model parameters. While precise choice generation methodologies have not been greatly explored in activity-based models, it is an issue that has seen more focus in spatial applications such as route, destination or residential choice modelling. Two main types of approaches can be found in the literature: deterministic and stochastic choice set generation models (Pagliara and Timmermans, 2009). Models who use a deterministic approach will typically include a choice set predefined by the modeller, or samples of alternatives obtained with decision rules reflecting the domain knowledge. On the other hand, stochastic approaches do not assume that the choice set is universal and known, but rather model the uncertainty associated with it.

In addition to the assumption that choice sets are not universal (i.e. homogeneous across the population) and fully known to the modeller, realistic choice set generation implies considering dynamic choice sets (i.e. that evolve with time and additional endogenous or exogenous information), that may not be fully known to the decision-makers themselves. This is especially true in spatio-temporal applications, which often involve combinatorial spaces. Shocker *et al.* (1991) distinguishes three different sets:

1. the *awareness set*, the set of alternatives within the universal set that the consumer knows of and which are appropriate to satisfy their goals,
2. the *consideration set*, which is the set of alternatives from the awareness set that are accessible at a particular point in time,
3. the *choice set*, which is the set that the consumer considers immediately before making a choice.

As the awareness and consideration sets are not always available in traditional data sources (e.g. time use surveys, travel diaries), it is important to develop a strategy to generate them in an efficient and realistic way.

In route choice modelling, Flötteröd and Bierlaire (2013) describe a methodology to sample paths from a given distribution in a network, which produces a choice set that meets these requirements. They use the Metropolis-Hastings algorithm (Hastings, 1970) to explore the solution space in an efficient way:

1. First, they propose an initial shortest path between an origin and a destination,
2. they perform random modifications on the path with a known probability, and accept or reject the change based on an acceptance probability defined by the modeller. The process is carried until the defined Markov chain reaches stationarity.

In the paper, the changes to the current state are applied using *operators*: the *splice* operator, the *shuffle* operator, and the combination of both. Splicing the path consists in randomly drawing an insertion node with a given probability, then recomputing the shortest path. In the shuffle operation, the order of the existing nodes in the path is changed with a given probability, and the shortest path is recomputed.

Danalet and Bierlaire (2015) have adapted and applied the methodology proposed by Flötteröd and Bierlaire (2013) to sample alternatives in an activity-based context. The alternatives are activity schedules, which are represented as paths in a defined network. The nodes of the network are activities potentially performed for a unit of time, and the edges connecting them represent successful performance and succession between activities. They therefore consider a network with $KT + 2$ nodes and $2K + K^2T - 1$ edges, where K is the number of activity types, and T the number of time units in the given temporal horizon. As they want to include attractive alternatives in their choice set, they define an attractivity measure for each node based on their frequency of observation and the frequency of the length of activity-episodes in the network. They validate the method on a synthetic network and on a real dataset describing pedestrian behaviour, and by calibrating the parameters of a discrete choice model with a utility associated to each activity path. They find that importance sampling with the Metropolis-Hastings algorithm provides a better model fit than randomly sampling the choice model.

In this paper, we extend the works of Flötteröd and Bierlaire (2013) and Danalet and Bierlaire (2015) by proposing an application of the model to estimate the parameters of the activity-based optimisation framework described in Pougala *et al.* (2021). The key difference with Danalet and Bierlaire's work is the representation of the schedule: while they consider activity paths which are only time dependent, we integrate additional choice dimensions such as location and mode choice. These dimensions are also considered simultaneously within the framework, with the aim to capture trade-offs and interrelations between choices.

Expanding on the idea of generating neighbouring states with small changes, we introduce new operators that can modify specific aspects of the schedule, over each of the choice dimensions.

3 Methodology

3.1 Definitions

Following the framework developed and presented in Pougala *et al.* (2021), we introduce several fundamental definitions to set up the context of the problem.

- **Time:** we assume time to be discretised in time blocks of equal length t, with T the time horizon (e.g. $T = 24$h),
- **Space:** space is discretised in a finite set of locations $\mathcal{L}$. Each location is associated to at least one activity.
- **Activity:** for each individual n, an activity a is uniquely defined as an action taking place in a physical location $\ell_a \in \mathcal{L}$, having a start time x_a and a duration τ_a. The sequence of activities $\{a, a+1\}$ generates a trip from location ℓ_a to ℓ_{a+1}, that can be performed using mode $m_a \in \mathcal{M}$. $\mathcal{M}$ is the set of modes of transportation that are available to the individual. Note that, if the next activity takes place at the same location, the duration of the trip is simply zero.
- **Schedule:** a schedule $\mathcal{S}$ is the outcome of the individual's decisions with respect to activity participation, activity location, activity scheduling, transportation mode choice, and any other dimension added at the discretion of the modeller (e.g. route choice). More specifically, a schedule $\mathcal{S}$ is a sequence of S activities $(a_0, \ldots, a_S)$, 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.
  A schedule is *valid* if
  - it spans the whole time horizon, that is if

$$\tau_{\text{dawn}} + \tau_{\text{dusk}} + \sum_{n=1}^{N-1} (\tau_{a_n} + \rho(\ell_{a_n}, \ell_{a_{n+1}}, m_{a_n})) = T, \tag{1}$$

  - it does not exceed the maximum budget, that is if

$$\sum_{n=1}^{N-1} (c_{a_n} + \kappa(\ell_{a_n}, \ell_{a_{n+1}}, m_{a_n})) \leq B, \tag{2}$$

– each activity starts when the trip following the previous activity is finished, that is

$$x_{a_{n+1}} = x_{a_n} + \tau_{a_n} + \rho(\ell_{a_n}, \ell_{a_{n+1}}, m_{a_n}), \forall n = 0, \dots, N-1, \tag{3}$$

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

$$\tau_{a_n} \geq \tau_a^{\min}. \tag{4}$$

– all mandatory activities are included,
– only one activity from a set of considered duplicates (i.e. same activity with different associated locations or modes) is included in the schedule.

Each valid schedule $\mathcal{S}$ is associated with a time-dependent utility function, which is the sum over the utilities of each activity $a \in \mathcal{S}$. These utilities include components such as the utility of participating to the activity, the (dis)utility of travelling, or of deviating from a preferred schedule. An example of specification of $U_{\mathcal{S}}$ is presented in section 3.2.2.

- **Choice set:** We adopt a definition similar to the one proposed by Shocker *et al.* (1991), discussed in section 2. An entire schedule (including activity participation, timings, locations and modes of transportation) is defined as one alternative, that an individual can choose. The choice set therefore contains several distinct alternatives. We call *feasible set* $\mathcal{F}^n$ the ensemble of valid schedules. This is the full choice set of the problem, which is combinatorial and therefore cannot be enumerated. Out of all possible schedule alternatives, the individual is only aware of a sample that defines the *considered set* C^n. This set, while finite, is not readily available to the modeller, which instead has to rely on the schedule that was actually chosen and recorded to infer behaviour. The *realised schedule* is the chosen alternative.

  Figure 1 illustrates the definition of the choice sets, and how they relate to each other.

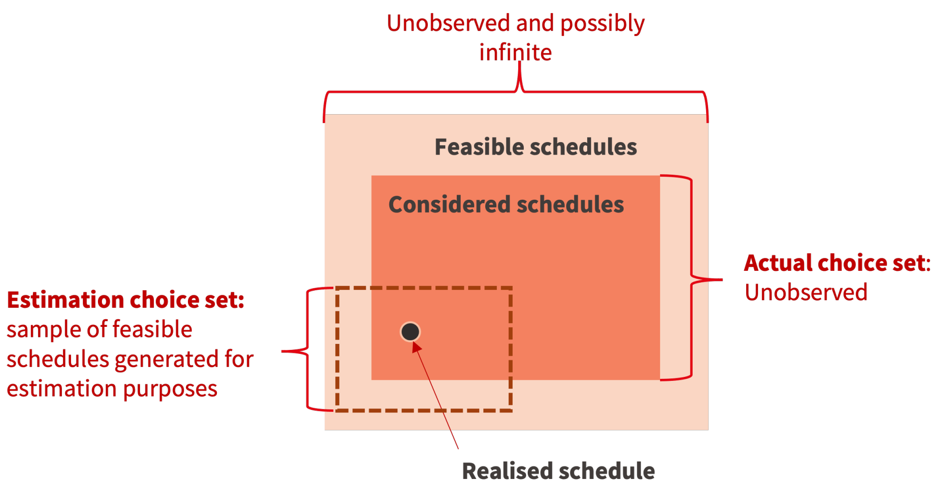


Figure 1: Definition of choice sets

3.2 Strategic generation with Metropolis-Hastings algorithm

Intuitively, we want the choice set that we generate for each individual to be as close as possible to what they would actually *consider*, meaning that it contains alternatives with high probability of being chosen. However, estimating a model with such a choice set would lead to very biased model parameters, which would, in turn, decrease the accuracy and realism of the model predictions. On the other hand, the solution space (i.e. the *feasible* set) is so large that by randomly sampling alternatives we risk selecting only meaningless schedules.

The strategy to build the choice set must therefore generate an ensemble of high probability schedules, to estimate significant and meaningful parameters, while still containing low probability alternatives to decrease the model bias. (Bierlaire and Krueger, 2020).

Extending the methodology developed by Flötteröd and Bierlaire (2013) and its application to activity-based contexts by Danalet and Bierlaire (2015), we model the choice set generation as a Markov process. The topology of the network in the activity-based context is very complex, and requires an appropriate methodology to efficiently explore the solution space.

We define X_t the *state* at time t. X_t is a 24 hour schedule, discretised in blocks of duration $\tau \in [\tau_{min}, 24 - \tau_{min}]$ (with τ_{min} the minimum block duration). The choice set is generated by exploring the *neighbouring* schedules of the state X_t. Two schedules are considered *neighbours* if they only differ in one element of one dimension (among time, space, or activity participation). Figure 2 shows an example of neighbouring schedules.

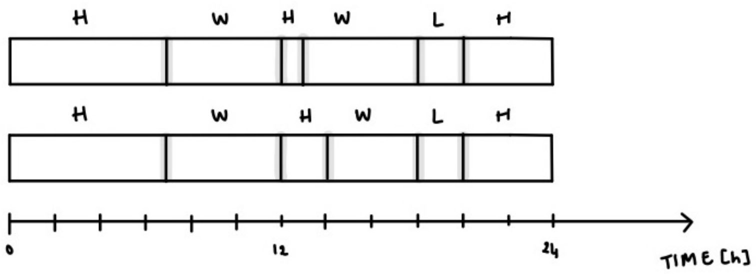


Figure 2: Example of neighbouring schedules. The schedules differ in the duration of the time spent at home during lunch time.

A new state X_{t+1} is proposed by using a set of operators Ω. These operators are heuristics that modify the current state to create a neighbour, with a probability P_ω. Examples of operators are described in section 3.2.3. A crucial requirement for each operator is that they must generate a state that meets the validity constraints described in 3.1.

The goal is to repeat this process until the chain reaches stationarity. At this point, we can assume that the generated schedules are draws from the distribution of schedules across the choice set.

3.2.1 Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm (Hastings, 1970) generates a random walk with an acceptance/rejection rule to converge to a specified target distribution (Gelman *et al.,* 1995). The procedure is summarised in algorithm 1.

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

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

3.2.2 Target distribution

The objective is to generate a choice set with attractive alternatives in order to estimate consistent parameters. Using the choice model to define the weights of the target distribution (i.e., setting $b(X_t) = U(X_t)$) is a convenient way to define the attractiveness of each drawn alternative. We use the same utility specification as Pougala *et al.* (2021) (Equation (5)), which is time-dependent and linear in parameters.

$$U_S = U + \sum_{a=0}^{A-1}(U_a^1 + U_a^2 + U_a^3 + \sum_{b=0}^{A-1}(U_{a,b}^4 + U_{a,b}^5)). \tag{5}$$

Its components and the associated assumptions are defined as follows:

1. A generic utility U that captures aspects of the schedule that are not associated with any activity. For instance, the agent may prefer that all shopping activities take place in the afternoon, or may dislike days with too many activities.

2. The utility U_a^1 associated with the participation of the activity a, irrespective of its starting time and duration. This term may include any variable such as level of service, cost, etc. It may also include an error term, capturing the unobserved variables.

$$U_a^1 = \beta_{\text{cost}} * c_a + \varepsilon_1 \tag{6}$$

3. the utility U_a^2 associated with starting time. This term captures the perceived penalty created by deviations from the preferred starting time. We define it as a deterministic (dis)utility:

$$U_a^2 = V_a^2 \tag{7}$$

with:

$$V_a^2 = \theta^e_{a_k} \max(0, x_a^- - x_a) + \theta^\ell_{a_k} \max(0, x_a - x_a^+), \tag{8}$$

where $\theta^e_{a_k} \leq 0$ and $\theta^\ell_{a_k} \leq 0$ are unknown parameters. The first (resp. second) term captures the disutility of starting the activity earlier (resp. later) than the preferred starting time. The index k captures the level of flexibility with respect to the scheduling of the activity, such that $k \in$ {Flexible, Moderately Flexible, Not Flexible}.

4. the utility U_a^3 associated with duration. This term captures the perceived penalty created by deviations from the preferred duration. We define it as a deterministic (dis)utility:

$$U_a^3 = V_a^3 \tag{9}$$

with:

$$V_a^3 = \beta^e_{a_{a,k}} \max(0, \tau^-_{a_k} - \tau_a) + \beta^\ell_{a_k} \max(0, \tau_a - \tau_a^+), \tag{10}$$

where $\beta^e_{a_k} \leq 0$ and $\beta^\ell_{a_k} \leq 0$ are unknown parameters. Similarly to the specification of start time, the first (resp. second) term captures the disutility of performing the activity for a shorter (resp. longer) duration than the preferred one,

5. For each pair of locations (ℓ_a, ℓ_b), respectively the locations of activities a and b with $a \neq b$, the utility $U^4_{a,b}$ associated with the trip from ℓ_a to ℓ_b, irrespective of the travel time. This term may include variables such as cost, level of service, etc. It may also include an error term, capturing the unobserved variables.

$$U^4_{a,b} = \beta_{t,\text{cost}} * c_t + \varepsilon_4 \tag{11}$$

6. For each pair of locations (ℓ_a, ℓ_b), the utility $U^5_{a,b}$, which captures the penalty associated with the travel time from ℓ_a to ℓ_b. We assume a deterministic specification:

$$U^5_{a,b} = V^5_{a,b} \tag{12}$$

with

$$V^5_{a,b} = \theta_t \rho_{ab}, \tag{13}$$

where θ_t is an unknown parameter, and ρ_{ab} is the travel time to the next location.

3.2.3 Operators

This list describes examples of operators that can be implemented in the defined problem set-up. Note that this list is not exhaustive, as many other operators can be created according to the modeller's needs and specifications.

The selected operators must meet the following requirements:

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

$$Q(X_t|X_{t-1}) > 0 \qquad \forall X_t, X_{t-1} \tag{14}$$

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

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

$$Q(X_t|X_{t-1}) > 0 \qquad \forall X_t, X_{t-1} \tag{15}$$

$$Q(X_{t-1}|X_t) > 0 \qquad \forall X_t, X_{t-1} \tag{16}$$

  Defining single change operators is interesting for the tractability of these quantities.

- The proposal state X_{t+1} resulting from the action of the operator or combination of operators should be *feasible*, i.e. it should meet the validity requirements described in section 3.1. As the target weights are defined based on the choice problem, invalid schedules cannot be accepted.

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

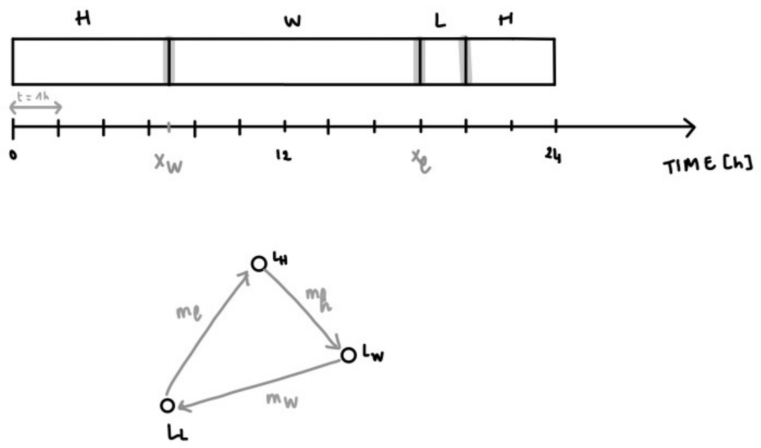


Figure 3: Initial schedule

Block The block operator ω_{block} modifies the time discretisation with a given probability P_{block}, by changing the length δ of the schedule blocks (e.g. from $\delta = 30$ to $\delta = 15$ minutes). The possible discretisations are set to: $\delta = \{1min, 5min, 15min, 30min, 60min\}$. This change does not affect the activity sequence, but allows to change the scale of the potential modifications of the other operators.

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

$$Q(X_t|X_{t-1}) = Q(X_{t-1}|X_t) = \frac{1}{N_\delta} \tag{17}$$

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

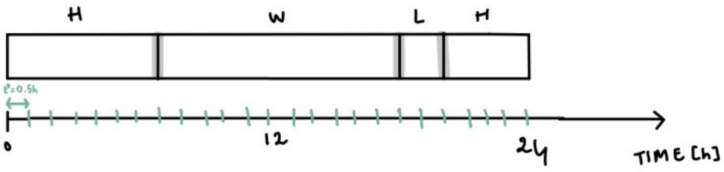


Figure 4: Change applied by the *block* operator

Assign With given probability P_{assign}, the assign operator ω_{assign} assigns an activity $j \in \mathcal{A}$ to a given block of duration δ, which was previously assigned to activity i. $\mathcal{A}$ is a set of N possible activities. The assignment is done with replacement, which means that $P(i = j) > 0$. To respect validity requirements, the resulting schedule must always start and end at home. Considering b_p the block at position p, with $p = (0, ..., T - \delta, T)$, the transition probabilities of the change can therefore be defined as the product of the probability of choosing a valid block multiplied by the probability of choosing one of the N activities:

$$Q(X_t|X_{t-1}) = Q(X_{t-1}|X_t) = \begin{cases} P_{\text{assign}} \frac{T-2\delta}{NT\delta}, & \text{if } b_i \notin \{b_0, b_T\} \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

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

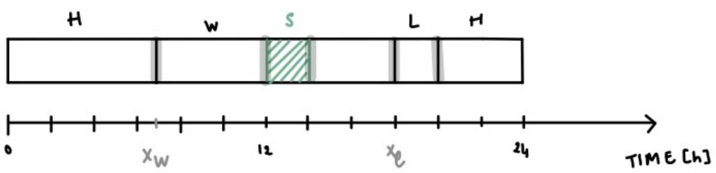


Figure 5: Change applied by the *assign* operator

Swap The operator ω_{swap} randomly swaps two adjacent blocks with probability P_{swap}. A block at position p b_p is randomly selected, then is swapped with the following block. In order to respect the validity requirements, the resulting schedule must always start and end at home. Considering b_p the block at position p, with $p = (0, ..., T - \delta, T)$, then $b_0, b_{T-\delta}$ and b_T being selected would lead to an infeasible schedule. The transition probabilities of the change can therefore be defined as :

$$Q(X_t|X_{t-1}) = Q(X_{t-1}|X_t) = \begin{cases} P_{\text{swap}} \frac{T-3\delta}{NT\delta}, & \text{if } b_i \notin \{b_0, b_{T-\delta}, b_T\} \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

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

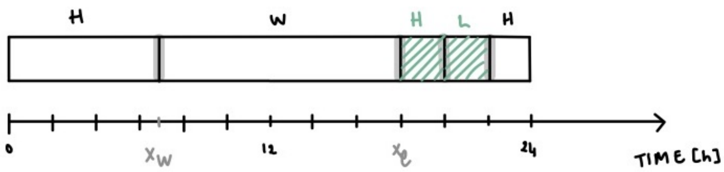


Figure 6: Change applied by the *swap* operator

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

$$Q(X_t|X_{t-1}) = Q(X_{t-1}|X_t) = \begin{cases} P_{\text{infdef}} P_{\text{direction}} \frac{2(T-2\delta)}{NT\delta}, & \text{if } b_i \notin \{b_0, b_T\} \\ 0, & \text{otherwise} \end{cases} \tag{20}$$

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

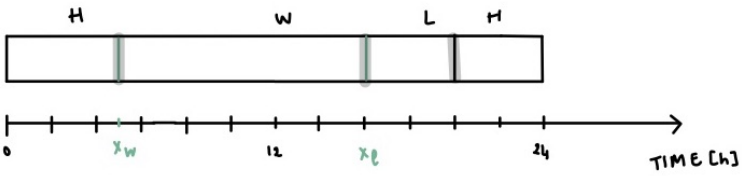


Figure 7: Change applied by the *translate* operator

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

$$Q(X_t|X_{t-1}) = Q(X_{t-1}|X_t) = \begin{cases} P_{\text{loc}} P_\ell \frac{T-2\delta}{NT\delta}, & \text{if } b_i \notin \{b_0, b_T\} \\ 0, & \text{otherwise} \end{cases} \tag{21}$$

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

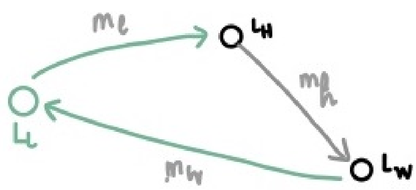


Figure 8: Change applied by the *location* operator

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

$$Q(X_t|X_{t-1}) = Q(X_{t-1}|X_t) = \begin{cases} P_{\text{mode}} P_m \frac{T-2\delta}{NT\delta}, & \text{if } b_i \notin \{b_0, b_T\} \\ 0, & \text{otherwise} \end{cases} \tag{22}$$

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

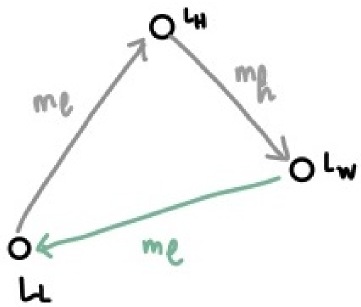


Figure 9: Change applied by the *mode* operator

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

3.2.4 Random walk

One step of the random walk moves from state X_t to X_{t+1} with acceptance probability $\alpha(X_t, X_{t+1})$. At each iteration, a candidate state X^* is drawn by applying changes from one or more arbitrarily chosen operators. The procedure is detailed in Algorithm 2.

Algorithm 2 Choice set generation

$n \leftarrow 0$, initialise state with random schedule $X_n \leftarrow S_0$
while $n \leq n_{iter}$ **do**
  Choose operator ω
  With probability P_ω, $X^* \leftarrow$ **Operator**(X_n)
  Compute acceptance probability $\alpha(X_n, X^*) = \min\left(\frac{b(X^*)q(X_n|X^*)}{b(X_n)q(X^*|X_n)}\right)$
  With probability $\alpha(X_n, X^*)$, $X_{n+1} \leftarrow X^*$, else $X_{n+1} \leftarrow X_n$
end while

3.2.5 Implementation notes

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

Schedule feasibility As described in section 3.1, the states generated by the process must meet validity criteria such as starting and ending at home, or having consistent timings between consecutive activities. One risk when defining operators is that they change a current feasible schedule into an infeasible state. For example, changing the duration of an activity may lead to a total duration that differs from the available time budget. One solution, as done in this paper, is to define operators that do not inherently induce infeasibility. This provides the advantage of making the transition probabilities easier to compute, but limits the possible changes that can be applied. On the other hand, allowing for infeasibility in the operators results can lead to more varied results. An operator that restores feasibility at the end of the process (e.g. modifying the time spent at home to absorb timing gaps or excesses in the schedule). However, as these changes would be dependent on the current state, computing the associated transition probabilities would prove more difficult.

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

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

4 Empirical investigation

The Mobility and Transport Microcensus 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 Microcensus contains 57’090 individuals, and 43’630 trip diaries.

We use the Metropolis-Hastings algorithm to generate a choice set for the *student* population of Lausanne (236 schedules). They are individuals who have declared being full-time students, which means that the *education* activity is expected to be one of the main out-of-home activities.

4.1 Initialisation

The initial parameters of the model, used to evaluate the weights characterising the target distribution, are calibrated on the full sample of Lausanne residents (students and non-students), with 1118 diaries. Table 1 gives the values of the significant estimated parameters. This calibration was performed by estimating the parameters using a choice set of size $N = 100$, with 99 randomly generated schedules, and the chosen schedule that was recorded in the survey. Note that *home* is chosen as the reference alternative. Consistently with random utility theory, the constants of the other activities can therefore be interpreted as the utility gained from performing out-of-home activities as opposed to staying at home, all else being equal.

Table 1: Parameters calibrated on randomly generated model.
NS: Not significant

Activity	Constant	Duration [h]	Start time[h]	Penalty: early	Penalty: late	Penalty: short	Penalty: long
Business trip	3.34	13.29	7.44	-2.65	-0.29	-0.25	-38.35
Education	5.78	5.97	6.00	-1.92	-0.22	-1.17	-0.22
Errands, services	2.61	NS	17.56	-0.0087	-1.15	NS	-0.75
Escort	3.90	NS	11.99	-0.32	-0.36	NS	-0.91
Home	-	23.98	-	-	-	-0.30	-266
Leisure	4.29	0.51	8.46	-1.55	-0.21	0	NS
Shopping	34.67	NS	8.42	-2.50	-0.24	0.12	-0.98
Work	7.33	11.22	6.49	-1.97	-0.54	-0.69	-1.25

We initialise the following operators for the random walk: Block, Assign, Swap, Inflate/Deflate and Combination. For the sake of simplicity, we omit the travel dimension and assume that each activity takes place at the same location. We therefore only focus on the scheduling aspect. We also assume equal probability of selecting the aforementioned operators.

There are 8 activities that can be scheduled: home (not including mandatory start and end of the schedule), work, education, shopping, errands or use of services, business trip (e.g. work activity outside of the typical workplace), leisure and escort (e.g. accompanying someone to an activity). These categories are a simplification made by the authors of the original classification reported in the dataset. We have assumed that the probability of scheduling a type of activity (specifically for the *assign* operator) was not equal across types. Instead, we have used the frequency of each activity type $a \in \mathcal{A}$ in the sample of 236 students as a proxy for the probability of choosing to perform it at a given time. The frequency is defined as the number of schedules in the sample in which the activity is present. The values are reported in table 2.

Table 2: Frequency values in sample, per activity

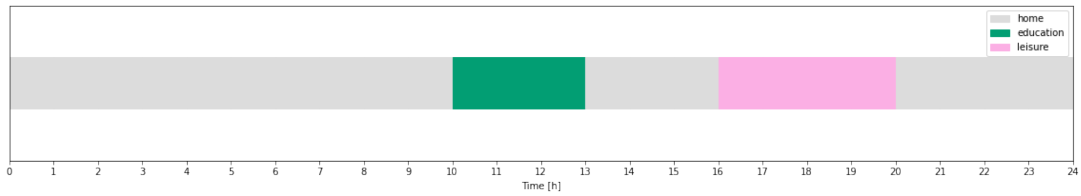
	Home	Leisure	Education	Shopping	Escort	Errands, services	Work	Business trip
Frequency [%]	38.8	23.0	20.6	8.7	2.8	2.8	2.5	0.82

4.2 Examples

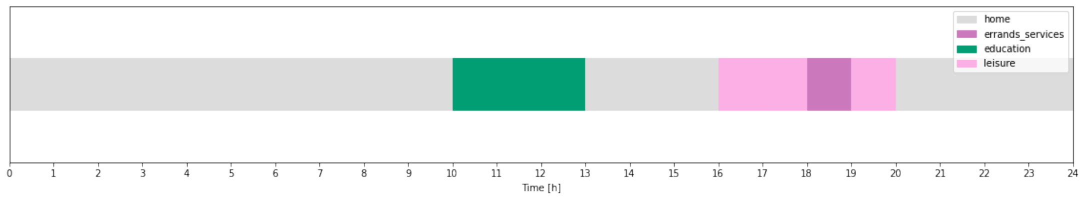
The first example is a choice set generated for one individual, randomly selected in the sample. We run 10'000 iterations of the algorithm, and keep 20 accepted schedules after a warm-up period. Figure 10 shows the initial schedule used as input of the procedure, and 4 generated schedules. As the selected schedules were consecutive, we can visualise the result of each accepted change. Visually, most of the accepted moves seem to be *assigning* new activities. Schedule 4 (Figure 10(e)) is interesting: it does not appear to be a reasonable schedule, with many splits of activities and short durations. This result indicates that the process is able to generate both attractive schedules (with respect to the utility function) such as Fig 10(b)-10(c), and alternatives with lower choice probabilities.

We repeated the procedure for all the individuals in the sample. We ran 10'000 iterations of the algorithm, and sampled accepted schedules after a warm-up period. We generated 5'000 schedules across the population. Figure 11 shows the distribution of start times for the different activity types across all the generated alternatives. For most activities, the start times appear to have been uniformly scheduled during day. On the other hand, *education* and *leisure* have more defined peaks. Looking at the frequency of scheduling of each activity (Figure 12), we notice that they are the two activities who are present in every generated schedule. This result is not surprising, as they are the most frequent out-of-home activities among Lausanne students (Table 2), and the observed schedules were used as starting points of the random walk.

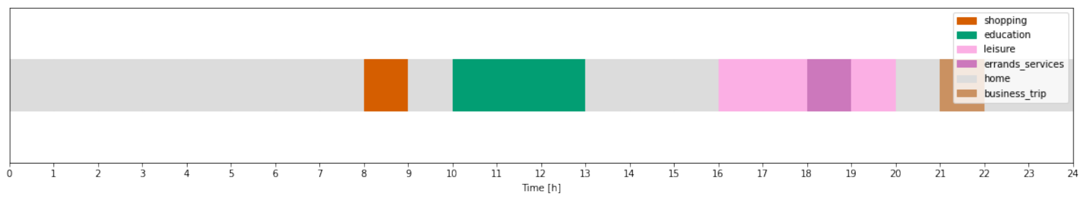
We are interested in understanding the impact of each operator on the acceptance of a generated schedule. Figure 13 illustrates the proportion of each operators among accepted moves. In the current set-up, and considering an equal probability of selecting one of the operators at each iteration, combining multiple changes (meta-operator) seems the most promising to achieve a schedule that will be accepted, followed by the *assign* operator. This makes sense: because the constants for participating in each activity type are positive and often larger in scale than the penalties for schedule deviations (Table 1), adding activities is more favourable in terms of utility gain than the other operations. The *block* operator applied alone does not produce any accepted schedule. This also makes sense, as this operator does not fundamentally change the current state, and therefore needs to be applied in conjunction with other operators.



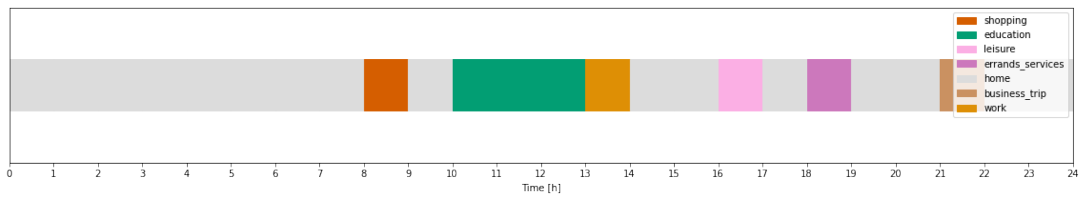
(a) Initial schedule



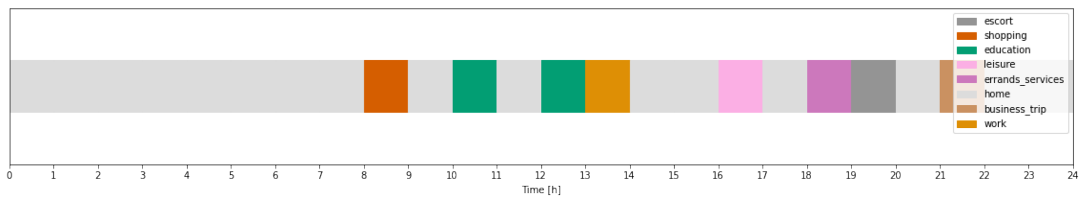
(b) Schedule 1



(c) Schedule 2



(d) Schedule3



(e) Schedule 4

Figure 10: Example alternatives from choice set

Taking a closer look to the combinations of operators (meta-operators), we can note that longer combinations (up to 4 operators) are more likely to produce accepted schedules (Figure 14). Figure 15 shows the prevalence of each operator in the accepted meta-operator combinations. The *assign* operator is the most frequent, especially when drawn multiple times. The *swap* operator is the second most combined operator, specifically in combination with *InflateDeflate* or applied multiple times in a row. Note that the map is not symmetric: for instance, applying the *Block* operator after *InflateDeflate* is a combination that is less present among accepted schedules than the other way around.

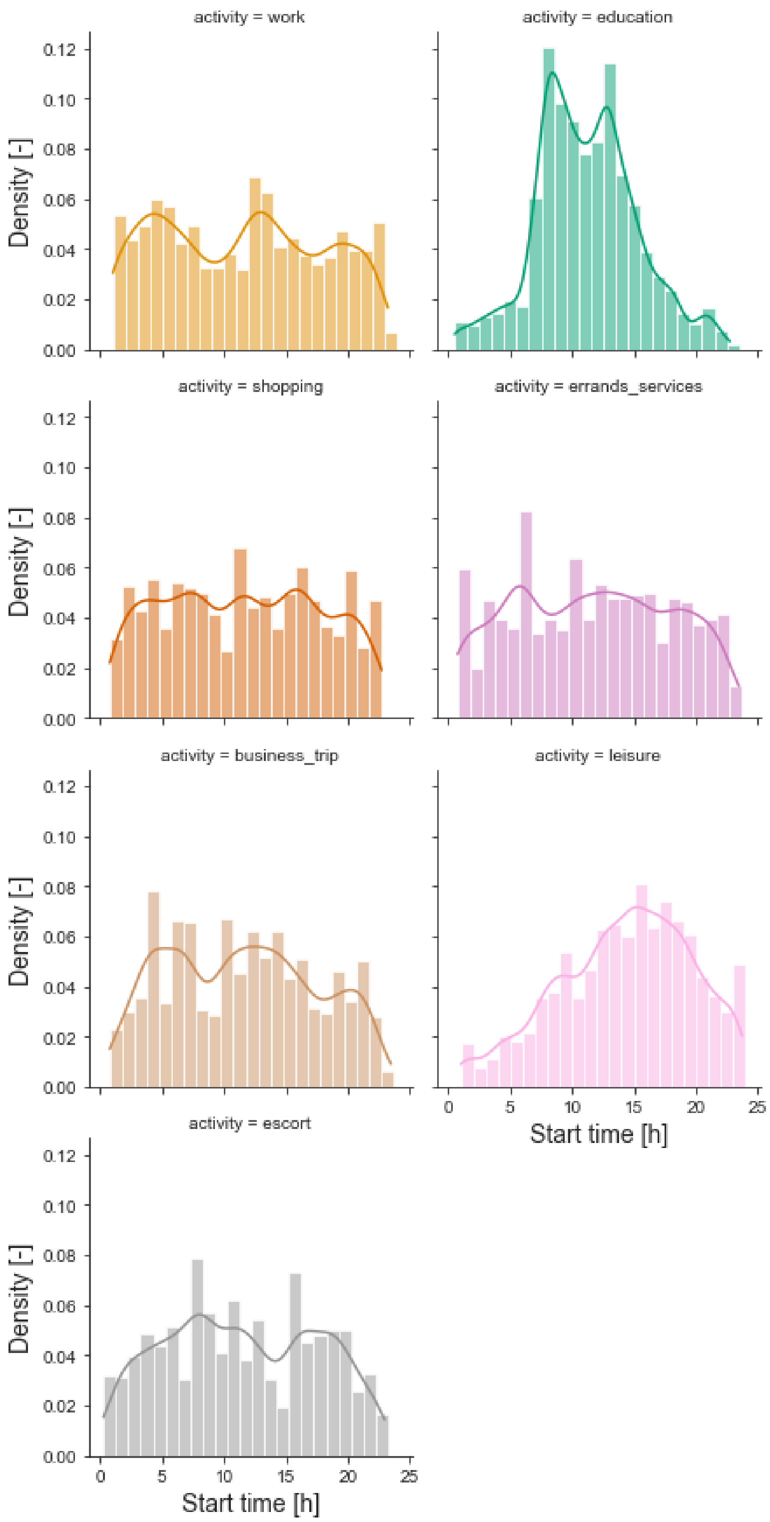


Figure 11: Distribution of start times in the generated choice sets

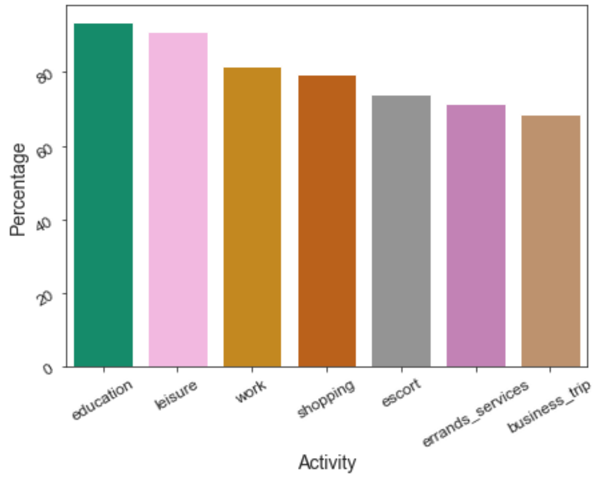


Figure 12: Frequency of activity types in the generated choice sets

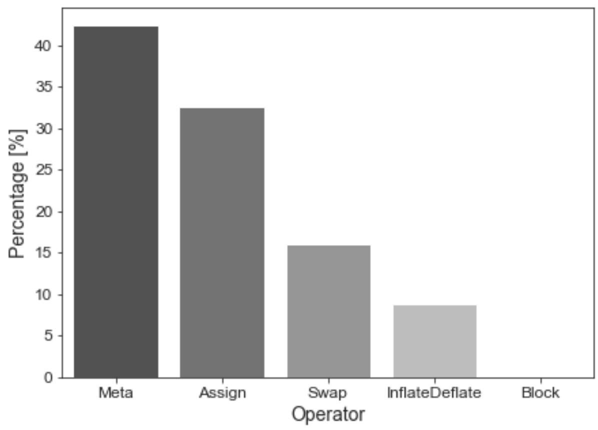


Figure 13: Frequency of operator types in accepted schedules

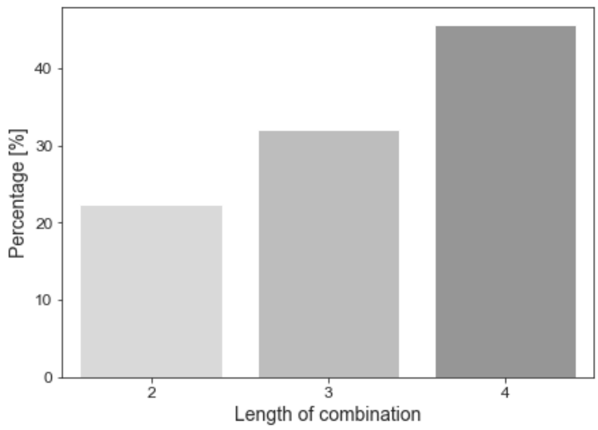


Figure 14: Typical lengths of combinations for accepted meta-operators

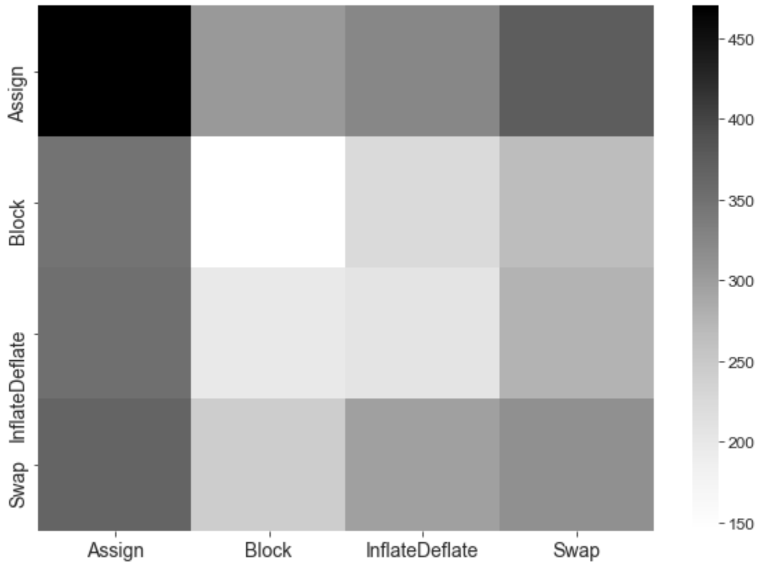


Figure 15: Frequency of pairs in accepted meta-operators. *y-axis: first operator, x-axis: second operator*

5 Conclusion and future work

In this paper, we have extended the alternative sampling methodology developed by Flötteröd and Bierlaire (2013) and Danalet and Bierlaire (2015) to generate choice sets for activity-based models where the alternatives are full activity schedules. We include choice dimensions such as activity participation, timings, location and mode choice. The Metropolis-Hastings algorithm is used to explore the combinatorial solution space. At each iteration, a new state (or neighbour schedule) is constructed with the use of operators which induce a single change in one of the dimensions of the schedule.

We have shown an example of application on an observed schedule as a proof of concept. Future work will be focused on validating the approach, and applying it on larger-scale samples. The validation can be performed by estimating parameters with the sampled choice set, and measure their bias with metrics such as the mean absolute error (as proposed by Lemp and Kockelman (2012)). These statistics can be compared to those obtained with randomly generated choice sets to evaluate the gain of information and accuracy. A synthetic population can be used to validate the parameter estimates against control values. The performance of the model can also be evaluated by considering metrics of quality of the choice set: for instance, indicators such as the degree of similarity (Flötteröd and Bierlaire, 2013) can be used to control the generation process.

A sensitivity analysis must be performed to understand the impact of different features and parameters of the algorithm. For example, parameters such as the probabilities to select certain operators or combinations of operators can be adjusted depending on their influence on the acceptance ratios. Similarly, the target weights guide the algorithm through the space; it is therefore crucial that they penalise or reward specific behaviours. An investigation on the effect of different utility functions (i.e. different proposals for target distributions) on the random walk will be performed. For instance, the examples presented in section 4 seem to indicate that a different relationship between constants and penalty parameters would lead to differences in accepted moves or operators.

Finally, we will focus on improving the performance of the algorithm by considering the convergence of the Markov chain. This will require the development of convergence metrics (Flötteröd and Bierlaire (2013) propose a measure of distance between generated paths) to monitor mixing and stationarity (Gelman *et al.,* 1995).

Acknowledgement

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

6 References

Adler, T. and M. Ben-Akiva (1979) A theoretical and empirical model of trip chaining behavior, *Transportation Research Part B: Methodological*, **13** (3) 243 – 257, ISSN 0191-2615.

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

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

Bowman, J. L. and M. E. 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.

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

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

Charypar, D. and K. Nagel (2005) Generating complete all-day activity plans with genetic algorithms, *Transportation*, **32** (4) 369–397.

Danalet, A. (2015) Activity choice modeling for pedestrian facilities, *Technical Report*, EPFL.

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

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*.

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

Gelman, A., J. B. Carlin, H. S. Stern and D. B. Rubin (1995) *Bayesian data analysis*, Chapman and Hall/CRC, ISBN 0429258410.

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

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

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

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

Lemp, J. D. and K. M. Kockelman (2012) Strategic sampling for large choice sets in estimation and application, *Transportation Research Part A: Policy and Practice*, **46** (3) 602–613, ISSN 0965-8564.

Nurul Habib, K. (2018) A comprehensive utility-based system of activity-travel scheduling options modelling (CUSTOM) for worker's daily activity scheduling processes, *Transportmetrica A: Transport Science*, **14** (4) 292–315, apr 2018, ISSN 23249943.

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

Pagliara, F. and H. Timmermans (2009) Choice set generation in spatial contexts: a review, *Transportation Letters*, **1** (3) 181–196, jul 2009, ISSN 1942-7867.

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.

Pougala, J., T. Hillel and M. Bierlaire (2021) Capturing trade-offs between daily scheduling choices, *Technical Report*, **TRANSP-OR 210101**, Transport and Mobility Laboratory, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

Shocker, A. D., M. Ben-Akiva, B. Boccara and P. Nedungadi (1991) Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions, *Marketing Letters 1991 2:3*, **2** (3) 181–197, aug 1991, ISSN 1573-059X.

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

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