EPFL

Resolving temporal scheduling conflicts in activity-based modelling

Patrick Manser [*]   Tom Haering [†]   Tim Hillel [‡] [†]

Janody Pougala [†]   Rico Krueger [†]   Michel Bierlaire [†]

9 December 2021

[*]Swiss Federal Railways (SBB), Hilfikerstrasse 1, 3000 Bern 65, Switzerland, {patrick.manser}@sbb.ch

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

[‡]University College London (UCL), London, UK

Abstract

This paper presents a novel activity-based demand model that combines an optimisation framework for continuous temporal scheduling decisions (i.e. activity timings and durations) with traditional discrete choice models for non-temporal choice dimensions (i.e. activity participation, number and type of tours, and destinations). The central concept of our approach is that individuals resolve time conflicts that arise from overlapping activities, e.g. needing to work and desiring to shop at the same time, in order to maximise their daily utility. Flexibility parameters indicate behavioural preferences to penalise deviations from desired timings. This framework has three advantages over existing activity-based modelling approaches: (i) the time conflicts between different temporal scheduling decisions including the activity sequence are treated jointly; (ii) flexibility parameters follow an utility-maximisation approach; and (iii) the framework can be used to estimate and simulate a city-scale case study in reasonable time. We introduce an estimation routine that allows flexibility parameters to be calibrated using real-world observations, as well as a simulation routine to efficiently resolve temporal conflicts using an optimisation model. The framework is applied to the full-time workers of a synthetic population for the city of Lausanne, Switzerland. We validate the model results against reported schedules. The results demonstrate the capabilities of our approach to reproduce empirical observations in a real-world case study.

Keywords: activity-based model; discrete choice; mathematical optimisation; maximum likelihood estimation; mixed-integer linear program

1 Introduction

Planning for future transport infrastructure and efficient service concepts in light of possible changes in population (e.g. demographic shifts), policy or technology relies on quantitative forecasts of travel demand. To that end, simulation models which aim to closely represent mobility behaviour are applied. The majority of models used in practice adopt an aggregated (i.e. macroscopic) approach. In macroscopic models, mobility is simulated as a set of trips or tours that are not connected within daily activity schedules. Furthermore, behaviour is typically clustered into segments without considering heterogeneous individual preferences (Boyce and Williams, 2015). The time distribution over the day is usually divided into broad categorical groups, e.g. peak hours and off-peak hours. Individual preferences and spatial-temporal constraints of travel are hence largely ignored.

Activity-based (i.e. microscopic) travel models treat mobility as individual decisions across many interwoven choice dimensions. These models afford a higher resolution representation of travel behaviour than traditional macroscopic models because they simulate each traveller as an autonomous decision-making unit and consider full consistency in time and space over a time period (e.g. a 24h-day) for each individual (Rasouli and Timmermans, 2014). In most activity-based models, the different choice dimensions – e.g. activity participation, destinations, activity duration or start times – are simulated sequentially (Castiglione et al., 2015, Davidson et al., 2011). Different modelling approaches can be used for each choice dimension. For example, activity participation and destination choice is well suited to discrete choice models, whilst other dimensions like activity duration and start time typically use rule-based or data-driven approaches (Scherr et al., 2020b). Sequential and rule-based approaches to generate activity-based schedules can be improved to model major system disruptions since they only partially consider behavioural preferences and feedback between choice dimensions.

In this paper, we apply a novel activity-based scheduling model for the city of Lausanne, Switzerland, in which temporal scheduling decisions (i.e. activity timings and durations) and travel times are considered jointly. The model targets to resolve activity-scheduling conflicts for each individual that arise from overlapping activities. In the case of a scheduling conflict, flexibility parameters indicate the behavioural preferences to shift timings of certain activity types compared to others. We adapt the framework introduced by Pougala et al. (2021a) that simulates desired choice dimensions simultaneously using a mixed-integer linear program (MILP) and that identifies behavioural flexibility preferences based on maximum likelihood estimation. The framework proposed by Pougala et al. has two major advantages: (i) it allows for interaction between multiple choice dimensions; and (ii) conflicts in time and space among different activities are resolved based on behavioural preferences. In this work, we introduce another contribution to the framework by implementing a computationally efficient version to demonstrate the application capabilities to a synthetic population for a city-scale case-study, with a detailed calibration and validation against empirically available observations. This is achieved by simulating the non-temporal choice dimensions (i.e. activity participation, number of tours, and considered locations) using existing discrete choice models prior to the optimisation model to reduce complexity within the MILP. The output of the framework is a realistic daily schedule containing a sequence of activities for each simulated individual. Each activity in this

schedule has a certain type, start time, duration, location as well as a mode to travel to the following activity. These schedules are an essential input for a traffic assignment – e.g. using the software MATSim (Horni et al., 2016) – which can be used to derive network loads to support decisions about future transport infrastructure investments.

The paper is structured as follows: First, a brief review of recent advances in activity-based demand modelling is provided. Next, the framework is introduced, including the parameter estimation routine and the MILP. This framework is then applied to the group of full-time workers of a synthetic population in the case study for the city of Lausanne, Switzerland. Finally, the results are presented, validated against empirical data from a survey and discussed.

2   Background

Early activity-based approach to travel demand modelling were proposed in the 1990s (Axhausen and Gärling, 1992). The main motivation driving the transition from traditional aggregated models to activity-based models is the lack of behavioural realism in the traditional approach, which does not allow for forecasting new policies such as congesting pricing, teleworking and ride-sharing incentives (Rasouli and Timmermans, 2014). For a more detailed overview of the development of the activity-based approach, we direct the reader to reviews provided by Bowman (2009) and Castiglione et al. (2015).

The first microscopic activity-based models were developed for several Northern American cities (Bowman and Ben-Akiva, 2001, Vovsha et al., 2005, Bhat et al., 2004). They follow an econometric, utility-maximising approach to simulate the choice behaviours of households and individuals. Typically, the activity schedules are built using a set of discrete choice models for mode ownership, tour bundling, activity selection, mode choice, and location choice. Multinomial and nested logit models are the most commonly applied model forms used in practice to represent the interactions of the various dimensions and to link the separate model steps. As Bhat et al. (2004) highlight, several important structural issues are not addressed within these early models, one of them being the relation of the time-of-day decision to the mode and destination choices.

Other early implementations of activity-based models use a rule-based approach. For example, the ALBATROSS framework (Arentze and Timmermans, 2004) uses decision trees to represent choice heuristics of individuals and derive these heuristics from activity travel data. Other rule-based frameworks that use decision trees for the scheduling procedure are FEATHERS (Bellemans et al., 2010) and TASHA (Miller and Roorda, 2003, Roorda et al., 2008). As stated in Auld et al. (2009), the issue with rule-based models is that they cannot predict modification choices, i.e. if an activity is modified, is it moved or shortened and by how much.

In Europe, the development of activity-based models was arguably driven by the increasingly popular agent-based transport simulation software MATSim (Horni et al., 2016). Most activity-based models are developed to generate schedules which are then fed into the dynamic traffic assignment of MATSim. The software MATSim itself provides a well-elaborated network simulation and route searching algorithms. However, the standard version of MATSim does not provide the functionality to generate activity-based demand including the choice dimensions of activity participation, sequence, locations and initial timings. For this purpose, Ziemke et al. (2015) use the econometric framework

CEMDAP (Bhat et al., 2004) to couple an existing activity-based model with MATSim. They take the original discrete choice parameters as estimated for Los Angeles, and apply them to a synthetic population of Berlin. Hilgert et al. (2017) propose the framework mobiTopp that covers the time period of one week. It is also inspired by the approach of Bowman and Ben-Akiva (2001) and applies a set of discrete choice models. Activity durations and start times are simulated as combination of discrete choice of an aggregated category and a weighted random draws within the chosen category. The resulting activity-based schedules of the mobiTopp model are then integrated into MATSim (Briem et al., 2019). A very similar approach is presented in Scherr et al. (2020b). They show a comprehensively validated microscopic model called SIMBA MOBi in which agents react to transport supply across all mobility choices. For the choices in the temporal dimension, it uses a rule-based approach that is based on weighted random draws from empirical distributions. Moeckel et al. (2020) propose a framework called MITO containing a simplified activity-based model. Destination choice is influenced by a travel time budget for every household, i.e. people who spent a lot of time commuting are less likely to do much other travel. A recent study from Hörl and Balac (2021) introduces a standardised process for generating activity-based travel demand based on open data and open software that is fully replicable by any user. Their approach is predominantly data-driven and does not focus on behavioural parameters. Drchal et al. (2019) implement another fully data-driven approach.

A newer generation of activity-based model tries to solve the scheduling problem with techniques like Hidden Markov Models or Bayesian Networks. Both Liu et al. (2015) and Saadi et al. (2016) introduce a Hidden Markov Model. The advantage of this method is that it considers trends of their activity sequencing from a temporal perspective. However, Saadi et al. (2016) point out that the approach presents a limitation at the time dimension. Joubert and De Waal (2020) present a Bayesian Network approach and highlight the benefits of a behaviourally rich travel demand model which allows for causal interpretation. This method can also account for temporal variables like activity duration.

There have been few attempts to combine activity-based modelling and mathematical optimisation techniques reported in the literature. An early study by Recker (2001) presents a theoretical mathematical formulation targeting to facilitate the practicality of activity-based modelling approaches. It unifies the complex interactions among the scheduling conflicts solved by households in performing their daily activities, while preserving the utility-maximising principles. Building from this work, Recker et al. (2008) introduce an estimation procedure for this optimisation framework. They estimate the relative importance of factors associated with spatial and temporal interactions among the activities in a schedule, however they conclude that the formulation for the time sequence is rather simplistic. Another framework called ADAPTS is demonstrated by Javanmardi et al. (2016), which implements a flexible non-linear optimisation model. The model is applied to a synthetic population for the Chicago region. The objective function aims to minimise the amount of changes in timing and duration of involved activities in a conflict situation. The weights used to account for individual activity preferences are constant weights for all activity types. Rizopoulos and Esztergár-Kiss (2020) develop an optimisation model for the interaction of activity scheduling and charging electric vehicles. For this purpose, the authors use a Genetic Algorithm that considers temporal flexibility which is defined based on heuristic rules and priority labels per activity. Esztergár-Kiss et al. (2020) sug-

gest an activity-based schedule optimisation method that includes temporal and spatial flexibility of the activities using a modified version of the Traveling Salesman Problem with Time Window constraints. Similar to Rizopoulos and Esztergár-Kiss (2020), they define heuristic flexibility priorities. Also, they focus on the travel episodes with a rather simple utility specification. Ballis and Dimitriou (2020) aim to convert multi-period and purpose-dependant origin-destination matrices into sets of activity schedules. They show a comprehensively validated framework that is mainly established on advanced graph-theoretical and combinatorial optimisation concepts. However, they only use simplified activity types and the activity schedules are not accompanied by socio-demographic information.

Many of the discussed issues are addressed by the activity-scheduling framework proposed by Pougala et al. (2021a). They provide a theoretical framework that combines multiple choice dimensions (activity participation, location, start time, duration and mode choice) into a single optimisation problem and that captures the complex trade-offs between scheduling decisions for multiple activities. The framework is based on the behavioural principle that individuals maximise their overall schedule utility according to their preferences and constraints for performing desired activities during one day. It represents all choices in the time dimension as continuous variables. The utility formulation is modular and preferences can be specified for each individual activity. The optimisation model is implemented as a MILP that can easily be extended by additional custom constraints. In their work, Pougala et al. take flexibility parameters from the literature without further calibration. Also, they demonstrate the framework for small samples of individuals only and do not proof the applicability in the large-scale context with a detailed validation against empirical observations.

This work builds on the theoretical framework of Pougala et al. and focuses on resolving scheduling conflicts in the temporal dimension, on calibrating the individual flexibility parameters and on computational efficiency to scale the model to a real-world large-scale simulation. The temporal dimension of the scheduling problem remains under-explored in literature compared to other dimensions such as location and mode choice, which have been demonstrated to be well calibrated in a microscopic models for various applications (Hörl et al., 2019, Scherr et al., 2020a). We aim at identifying a well calibrated utility specification based on empirically reported observations for temporal scheduling decisions, including the choices of activity start times and durations, while considering trip travel times between them. Also, the work aims to apply the optimisation framework to a synthetic population and reproduce observations as reported in the survey.

Stated succinctly, the contributions of our work are as follows:

- *Flexibility parameters*: An estimation routine is demonstrated that considers the temporal decision within the scheduling problem as a whole. In contrast to existing approaches (Bowman and Ben-Akiva, 2001, Bhat et al., 2004, Hilgert et al., 2017, Scherr et al., 2020b) which consider trip- or tour-based parameters, this approach estimates schedule-based parameters (total daily utility). Our formulation considers time as a continuous variable and the resulting flexibility parameters are based on behavioural principles. In the literature, time has either been modelled as a discrete variable in combination with utility-based principles (Bowman and Ben-Akiva, 2001, Castiglione et al., 2015) or as a continuous variable in combination with heuristic flexibility parameters (Javanmardi et al., 2016, Esztergár-Kiss et al.,

2020, Pougala et al., 2021a).

- *Real-world Application*: The implemented approach of coupling traditional discrete choice models with an optimisation model to reduce the number of decision variables targets to proof the applicability of the framework in a real-world large-scale case study, which is not done in Pougala et al. (2021a). This work applies the scheduling model to all 50'000 full-time workers of a synthetic population of Lausanne, Switzerland. The size is comparable to studies in the literature, which mostly simulate a certain percentage (e.g. 10 %) of the synthetic population for bigger cities (Ziemke et al., 2015, Briem et al., 2019).

3 Methodology

The presented work implements an approach that makes usage of existing traditional discrete choice models and combines them with an adapted version of the mathematical model as proposed by Pougala et al. (2021a). This approach targets to reduce the complexity of the problem and puts focus on parameter calibration and real-world application. To calibrate the individual flexibility parameters for a case study, we introduce an estimation routine that quantifies preferences based on empirically reported schedules in the Swiss mobility and transport microcensus (MTMC, BfS and ARE (2017)). The choice set for the estimation routine is generated using existing sequential discrete choice models (Scherr et al., 2020b). Based on the flexibility parameters, an optimisation model solves all temporal scheduling decisions (start times and durations of each activity) simultaneously while considering travel times. The non-temporal elements of the scheduling problem (i.e. activity participation, number and type of tours, and considered destinations) are simulated using the traditional discrete choice models as introduced by Scherr et al. (2020b) outside of the optimisation. They are fed as static activity sets into the optimisation framework. The proposed framework can be summarised with the two following streams, as illustrated in Figure 1:

1. *Parameter estimation:* The first stream (top row in Figure 1) aims to estimate individual preferences for a specific person group based on schedules reported in the MTMC (BfS and ARE, 2017). In this work, we focus on the group of full-time workers. Individual preferences are flexibility parameters that express the loss of utility when either deviating from a desired starting time or duration for each given activity type. The procedure is described in more detail in Section 3.2.

2. *Schedule simulation:* The second stream (bottom row) optimises the utility of the daily activity schedule for each individual in the synthetic population. The methodology to generate a synthetic population is provided in Bodenmann et al. (2019) and goes beyond the scope of this paper. Each individual is assumed to choose an activity set that they are going to perform during a 24h-period. Based on the estimated flexibility parameters, the optimisation approach then resolves time conflicts that arise from overlapping desired timings in the activity set. More insights into the methodology are given in Section 3.3.
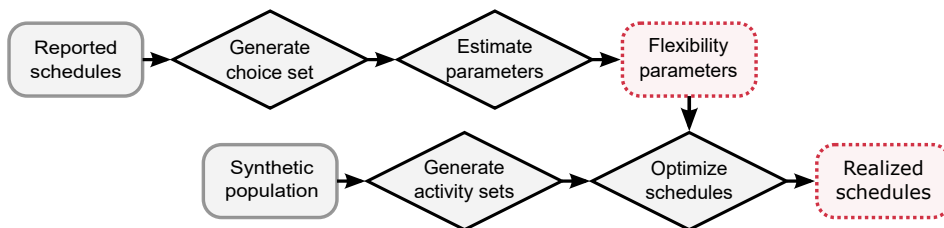
Figure 1: The two streams of the scheduling framework: calibration (top row) and simulation (bottom row).

3.1 General definitions

The following definitions are important in the context of this work:

Activity: Each $a \in \mathcal{A}$ is defined by a specific type, a desired start time x^*, a desired duration τ^*, and a set considered locations $\mathcal{L}$. Also, each activity is assigned to a tour type and an activity can be defined to be a sub-tour activity.

Activity set: An activity set $\mathcal{A}$ contains all activities an individual needs or wishes to perform during a time period (e.g. a 24-hour day) in no sequential order. It can contain multiple instances of activities of the same type (e.g. two leisure activities). An example for an activity set is $\mathcal{A} = \{$*dawn*, *dusk*, *home*, *work*, *leisure_1*, *leisure_2*$\}$. In this case, *leisure_1* and *leisure_2* are both leisure activities, but may have different desired start times, desired durations and considered location sets. We differ between three subsets of $\mathcal{A}$:

> *Home activity set*: The set $\mathcal{H} \subseteq \mathcal{A}$ contains the number of home activities an individual is going to perform. It includes activity types *home*, *dusk* and *dawn*. $\mathcal{H}$ must contain exactly one dawn and one dusk activity, whereas the number home activities is not constraint. We assume that each person wants to spend a fraction of their daily time budget at home. Therefore, we define $\tau^*_{\mathcal{H}}$ as the desired *home time budget*, which is the sum of the durations of all in-home activities: $\sum_{a \in \mathcal{H}} \tau_a = \tau_{\mathcal{H}}$. Each $h \in \mathcal{H}$ take place at the same location ($|\mathcal{L}| = 1$), which is given in the synthetic population data.

> *Primary activity set*: The set $\mathcal{P} \subset \mathcal{A}$ includes activity types *work* and *education*. Primary activities are defined as activities that take place at one long-term location (e.g. at the workplace). We assume that individuals are required to undertake primary activities for a total daily duration $\tau^*_{\mathcal{P}}$, such that the sum of the duration of each scheduled primary activity is equal to the *primary time budget*: $\sum_{a \in \mathcal{P}} \tau_a = \tau_{\mathcal{P}}$. The set $\mathcal{P}$ may be empty, if an individual does not need to spend time on any primary activity (i.e. $\tau^*_{\mathcal{P}}$=0).

> *Secondary activity set*: The set $\mathcal{S} \subset \mathcal{A}$ includes activity types *leisure*, *shopping*, *escort*, *business*, *further training* and *other*. Secondary activities are defined as activities with flexible locations. Hence, they may contain multiple considered locations, one of which is chosen in the optimisation model. In contrast to primary activities, the durations of secondary activities are treated individually rather than their daily sum. This comes with the assumptions that

7

people have individual activity duration budgets for secondary activity, but see the duration of primary activities as one daily objective. The set $\mathcal{S}$ may be empty, if an individual does not does not desire to take part at any secondary activity.

Schedule: Each schedule consists of a set of activities $\mathcal{A}$. To form a daily schedule, the activities must be in a ordered sequence without overlapping in time, including travel episodes. For this purpose, an individual needs to make the decision about a realised start time x, a realised duration τ, as well as a chosen location l for each $a \in \mathcal{A}$. Locations and sequence of activities result in travel episodes with a mode μ and a corresponding travel time tt. Activities of the same type must not take place consecutively if there is no travel episode in between (i.e. both are at the same location). A schedule always starts with a *dawn* activity at midnight and ends with a *dusk* activity at the defined time period. An example for a schedule is *dawn* $\rightarrow$ *work* $\rightarrow$ *leisure_2* $\rightarrow$ *home* $\rightarrow$ *leisure_1* $\rightarrow$ *dusk*. We consider two types of schedules:

> *Realised schedule*: In the choice set of the parameter estimation, it is the reported schedule as reported in a survey. In the optimisation context, it is the output of the optimisation model.

> *Feasible schedule*: Any possible deviation from a realised schedule can be feasible as long as it fits into the time period constraint and does not contain overlapping activities.

Tours: The set of tours $\mathcal{T}$ contains all out-of-home tours in the daily schedule of an individual. Each tour $t \in \mathcal{T}$ is defined to contain the subset of activities that take place between two home activities (e.g. *home* $\rightarrow$ *work* $\rightarrow$ *shopping* $\rightarrow$ *home*). The number of tours in a schedule hence equals the number of home activities in minus one ($|\mathcal{T}| = |\mathcal{H}| - 1$). $\mathcal{T}$ includes tour types *work*, *education* and *secondary*. Work and education tours include at least one and at most two primary activities of this type. Tours that include secondary activities only are referred to as secondary tours. The tour type is a static input in the optimisation context of this work.

Sub-tours: A sub-tour is a sequence of secondary activities between two primary activities without any home activity in between (e.g. *work – lunch – work*).

3.2 Estimation of individual flexibility preferences

This section explains the estimation of activity-specific flexibility parameters. Section 3.2.1 gives an insight into the methodology that is used to generate a competitive choice set. The model estimation uses the utility specification as given in Section 3.2.2 and is described in Section 3.2.3.

3.2.1 Choice set generation

In discrete choice modelling, individuals evaluate, compare and select alternatives from a choice set of mutually exclusive alternatives (Manski, 1977). In many applications, it is assumed that choice sets are fully known by the decision-makers and are fully or partially

observed by the modellers. This assumption is unrealistic for applications in which the choice set is large such as location or route modelling.

The way we define the choice sets in this paper is similar to the definition proposed by Shocker et al. (1991) in the context of marketing. We model the temporal dimension in the activity-scheduling process as the discrete choice between different combinations of activity timings and durations while also considering travel times between them. We assume that the decision-maker only possesses a partial knowledge of the available opportunities. The set of all available combinations is the *feasible schedule* set. This sample is finite but potentially very large. In addition, the sample of alternatives that the individual actually considers may neither be fully known by the modeller nor readily accessible from traditional data sources.

The utility-based framework described in Pougala et al. (2021a) requires the estimation of the parameters of the utility function, namely the penalties for schedule deviations from desired activity start times and durations, and the cost for travelling. The estimation requires a sample of alternatives with a formal description of the sampling protocol. Given the difficulty of this task in this context, we propose a heuristic approach in which choice sets are generated using a combination of the following two methods:

1. *Random alternatives:* We start with generating random schedule alternatives by randomly modifying the activity start times and durations in a given set of activities $\mathcal{A}$ as observed in the realised schedule of the MTMC. As the outcomes are random and not based on any empirical distribution, many alternatives are likely to have low levels of utility compared to the observed schedule. Relatively low levels of utility imply that the generated alternatives differ substantially from the observed schedule.

2. *Likely considered alternatives:* We generate activity schedules with variations of activity start times, durations and travel times that are likely to be considered by the decision-maker and hence biased towards high probability schedules. By likely considered activities we mean schedules that are observed by other individuals with similar characteristics (e.g. employment rate, age, urban home, etc.). For this purpose, the sequential activity-based demand model MOBi.plans as presented in Scherr et al. (2020b) is applied to draw distributions from the activities sets in the schedules as reported in the MTMC.

3.2.2 Utility specification

As the adapted version of the scheduling framework in this paper focuses on temporal scheduling decisions, it only uses the temporal components of the comprehensive utility specification as proposed in Pougala et al. (2021a) including travel times that arise from travelling between activities. In our work, the intrinsic attraction of an activity is not modelled with a constant inside the optimisation, but as a discrete choice based on the utilities of all available alternatives prior to the optimisation (see Section 3.3.1). Specifically, we include the utility terms relating to schedule deviations from timing and duration preferences for each activity *a* and the travel times between *a* and the consecutive activity in our

optimisation model for the utility specification of schedule alternative i:

$$U_i = \sum_{a \in \mathcal{A}_i} U_{\text{timing}}(x_a) + \sum_{a \in \mathcal{S}_i} U_{\text{duration}}(\tau_a) + \sum_{O \in \{\mathcal{P}_i, \mathcal{H}_i\}} U_{\text{duration}}(\textstyle\sum_{a \in O} \tau_a) + \sum_{a \in \mathcal{A}_i \setminus \{\text{dusk}\}} U_{\text{tt},a}(tt_a)$$

(1)

The components and the associated assumptions are defined as follows:

- $U_{\text{timing}}(x_a)$ indicates the impact for deviating from desired timings for each activity $a \in \mathcal{A}_i$. The desired start time is defined as x_a^* and x_a is the start time of activity a as stated in the choice set for alternative i. We introduce the two flexibility parameters β_a^{early} and β_a^{late} which penalise the difference $|x_a^* - x_a|$.

$$U_{\text{timing}}(x_a) = \beta_a^{\text{early}} \max(0; x_a^* - x_a) + \beta_a^{\text{late}} \max(0; x_a - x_a^*)$$

(2)

- $U_{\text{duration}}(\tau_a)$ captures the impact of deviating from desired durations. In the case of desired duration budgets, we differentiate between home, primary and secondary activities (see definitions in Section 3.1). For primary activities $a \in \mathcal{P}$, we compare the sum of all scheduled primary activities (i.e. $\tau_a = \sum_{p \in \mathcal{P}} \tau_p$) to the desired daily duration, or primary time budget ($\tau_a^* = \tau_{\mathcal{P}}^*$). The same assumption applies to home activities. For secondary activities $a \in \mathcal{S}$, we use the individual desired durations (i.e., $\tau_a^* \neq \tau_b^* \forall a, b \in \mathcal{S}, a \neq b$). β_s^{short} and β_s^{long} represent the loss in utility for deviating from a desired duration.

$$U_{\text{duration}}(\tau_a) = \beta_a^{\text{short}} \max(0; \tau_a^* - \tau_a) + \beta_a^{\text{long}} \max(0; \tau_a - \tau_a^*)$$

(3)

- $U_{\text{tt},a}(tt_a)$ is a disutility for the time spent travelling. Since the focus of this work lies on finding flexibility parameters, we fix β_{travel} to be -1. Different travel times for different alternatives are generated using an existing destination and mode choice model (Scherr et al., 2020b).

$$U_{\text{tt},a}(tt_a) = \beta_{\text{travel}} \, tt_a$$

(4)

3.2.3 Model estimation

Given the proposed utility specification, there are six types of parameters for each activity a to be estimated: (i) the desired start time x_a^*; (ii) the desired duration τ_a^*; and the four penalty terms for deviations from the desired start time and duration (iii) β_a^{early}, (iv) β_a^{late}, (v) β_a^{short}, and (vi) β_a^{long}.

We use two different approaches to determine the parameter values for these parameters:

1. For the desired start times and durations (x_a^* and τ_a^*), we use the mean values from empirical distributions for the activity types as found historic data as found in the MTMC. All activity types are grouped to have uni-modal distributions. Since persons in the MTMC only report on realised start times and durations (x_a and τ_a), this requires the assumption that the desired timings can be approached by the empirical mean of the realised timings.

10

2. Once the desired start times and durations are defined, the penalty terms (β_a^{early}, β_a^{late}, β_a^{short}, and β_a^{long}) are estimated from historic schedules using maximum likelihood estimation.

3.3   Schedule optimisation

Having defined the activity-specific flexibility parameters, we now introduce the optimisation model that targets to solve scheduling conflicts in the temporal dimension. Section 3.3.1 gives an overview over the generated activity sets that are treated as static inputs in the optimisation model. Then, Section 3.3.2 and Section 3.3.3 explain the implementation of the MILP.

3.3.1   Activity set generation

The optimisation framework used in this work takes as an input an activity set $\mathcal{A}$ for each individual as part of a synthetic population (Figure 1). This activity set $\mathcal{A}$ contains the long-term decisions (i.e. primary locations) and non-temporal decisions (i.e. activity participation, number and type of tours, and considered destinations). This requires the assumption that an individual knows the number and type of the activities it desires or needs to participate during the day prior to the time-conflict resolution. In this work, we apply the following sequence of discrete choice models[1] with the parameters as introduced and explained in more detail in Scherr et al. (2020b):

1. Long-term decisions about primary locations such as a work or school place, which depend on the choice of possessing a mobility tool such as a driving license and public transport subscription (Hillel et al., 2020).

2. Choice of number and type of primary out-of-home tours (i.e. work and education tours) depending on the long-term decisions as well as number and type of secondary activities within the primary tours. Also, the number and type sub-tour activities are generated during this choice step.

3. Decision about number of secondary tours as well as the number and type of secondary activities within each secondary tour.

4. Generation of a set of considered locations for each secondary activity based on draws from given distributions.

Table 1 gives an example of an activity set $\mathcal{A}$. The cells with sets in italic letters mean that these decisions are made in the optimisation model. Under consideration of these possible choice options, the activity sequence will be optimised as well. All other cells are static input for the optimisation model. It contains four home activities, which equals three possible out-of-home tours. One tour is primary, including one sub-tour activity. Since the sub-tour activity is static, this lunch activity must always take place between the two work activities. However, the two work activities can still swap their sequential order. Also, it is not decided if the accompany activity takes place before or after the

---

[1]The discrete choice models include socio-demographic attributes as well as level of service indicators (e.g. accessibility or parking costs).

work activity at this point. Three activities take place within secondary tours and cannot be moved to the work tour since the tour type is static. Their sequence is flexible, one shopping activity might end up in the same tour as the leisure activity if it fits best into in time and space preferences. Also, some of them have multiple considered locations, which will be decided later in the optimisation model.

Table 1: Example for an activity set $\mathcal{A}$ with the possible decision options.

Activity	Tour number	Tour type	Primary activity	Sub-tour activity	Considered locations
dawn	{1, 2, 3}	{work, secondary}	false	false	home place
work	1	work	true	{true, false}	office
lunch	1	work	false	true	office
work	1	work	true	{true, false}	office
escort	1	work	false	false	kindergarten
home	{1, 2, 3}	{work, secondary}	false	false	home place
shopping	{2, 3}	secondary	false	false	{shop 1, shop 2}
shopping	{2, 3}	secondary	false	false	{shop 3, shop 4}
home	{1, 2, 3}	{work, secondary}	false	false	home place
leisure	{2, 3}	secondary	false	false	gym
dusk	{1, 2, 3}	{work, secondary}	false	false	home place

The generation of the desired start times and durations for each activity in the context of this case study will be explained later in Section 4.2.

3.3.2 Optimisation model

This section summarises our adapted implementation of the MILP which is based on the framework introduced by Pougala et al. (2021a). For each schedule alternative i, an individual has a predefined set of activities $\mathcal{A}_i$ containing all activities a that are wished or needed to be performed by that individual within a bounded time period ξ (e.g. a 24-h day). In order to maximise the total schedule utility for each alternative i, each individual makes the decisions for each activity $a \in \mathcal{A}_i$ about: (i) the duration τ_a and the start time x_a, which implies the activity sequence; (ii) the tour number ϑ_a; (iii) the location λ_a; (iv) the mode μ_a; and (v) the travel time tt_a, which is an implication of the sequence, location and mode.

The objective for each individual is to maximise its total schedule utility which is defined in Equation (1):

$$\Omega = \max\ U_i(x, \tau, tt)$$

To this end, we introduce the following decision variables of the optimisation model:

$$
\begin{aligned}
x_a &\in \mathbb{R}_+ & \forall a \in \mathcal{A}_i \\
\tau_a &\in \mathbb{R}_+ & \forall a \in \mathcal{A}_i \\
tt_a &\in \mathbb{R}_+ & \forall a \in \mathcal{A}_i
\end{aligned}
\tag{5}
$$

where x_a and τ_a represent start time and duration of activity a, respectively, and tt_a represents the travel time from a to the following activity.

Another important set of decisions variables are the activity sequence indicators:

$$z_{ab} \in \{0, 1\} \qquad \forall a, b \in \mathcal{A}_i \tag{6}$$

where $z_{ab} = 1$ is equal to one if activity a takes place right before b. The travel time from a to b heavily depends on both trip origin and destination locations as well as the mode used for this trip. For this purpose, we introduce the location and mode choice variables:

$$
\begin{array}{llll}
\lambda_{al} & \in \ \{0, 1\} & \forall a \in \mathcal{A}_i \quad \forall l \in \{1, \ldots, |L_a|\} & =: \mathcal{L}_a \\
\mu_{am} & \in \ \{0, 1\} & \forall a \in \mathcal{A}_i \quad \forall m \in \{1, \ldots, |M|\} & =: \mathcal{M}
\end{array}
\tag{7}
$$

where $\mathcal{L}_a$ is the set of all considered locations for activity a and M is the set of all available modes. The location and mode choice variables have a direct impact on the travel times. The travel time variable can now be set as follows:

$$
tt_a \ = \ \sum_{b \in A, \ l_a \in \mathcal{L}_a, \ l_b \in \mathcal{L}_b, \ m \in \mathcal{M}} \Theta(l_a, l_b, m) \, z_{ab} \lambda_{al_a} \lambda_{bl_b} \mu_{abm} \qquad \forall a \in \mathcal{A}_i
\tag{8}
$$

where the travel time matrix Θ is exogenous and contains the travel times for all possible combinations of locations and modes within $\bigcup_{a \in \mathcal{A}_i} \mathcal{L}_a$ and $\mathcal{M}$. In the implementation of the optimisation model, we use a linearised form of this multiplication by introducing auxiliary variables.

The optimisation problem implemented in this work is now subject to the constraints:

$$
\sum_{a,b \in \mathcal{A}_i} (\tau_a + z_{ab} tt_a) \ = \ \xi
\tag{9}
$$

$$
\delta \ \leq \ \tau_a \ \leq \ \xi \qquad\qquad\qquad \forall a \in \mathcal{A}_i \quad (10)
$$

$$
z_{ab} + z_{ba} \ \leq \ 1 \qquad\qquad\qquad \forall a,b \in \mathcal{A}_i \quad (11)
$$

$$
z_{a,\text{dawn}} \ = \ z_{\text{dusk},a} \ = \ z_{aa} \ = \ 0 \qquad\qquad\qquad \forall a \in \mathcal{A}_i \quad (12)
$$

$$
z_{ab} \ = \ 0 \qquad\qquad\qquad \forall a,b \in \mathcal{P}_i \text{ or } \mathcal{H}_i \quad (13)
$$

$$
\sum_{b \neq a} z_{ab} \ = \ 1 \qquad\qquad\qquad \forall a \in \mathcal{A}_i \setminus \{\text{dusk}\} \quad (14)
$$

$$
\sum_{b \neq a} z_{ba} \ = \ 1 \qquad\qquad\qquad \forall a \in \mathcal{A}_i \setminus \{\text{dawn}\} \quad (15)
$$

$$
\sum_{l \in \mathcal{L}_a} \lambda_{al} \ = \ \sum_{m \in \mathcal{M}} \mu_{am} \ = \ 1 \qquad\qquad\qquad \forall a \in \mathcal{A}_i \quad (16)
$$

$$
(z_{ab} - 1)\xi \ \leq \ x_a + \tau_a + z_{ab} tt_a - x_b \ \leq \ (1 - z_{ab})\xi \qquad \forall a,b \in \mathcal{A}_i \quad (17)
$$

$$
\gamma_a^- \ \leq \ x_a \ \leq \ \gamma_a^+ - \tau_a \qquad\qquad\qquad \forall a \in \mathcal{A}_i \quad (18)
$$

Constraint (9) ensures that the total sum of all activity durations and travel times is exactly equal to the predefined time period ξ the schedule should fill out. The next constraint (10) guarantees that each activity's duration is between the minimal duration δ and maximal duration ξ. Constraint (11) imposes that an activity takes place either before or after another activity, possibly neither but never both. The bundled constraint (12) assures that no activity takes place before dawn or after dusk and that an activity cannot follow itself. Constraint (13) enforces that two primary activities or home activities cannot take place consecutively since this would break the general schedule structure as described in Section 3.1. The two constraints (14) and (15) ensure that every activity is preceded by another activity (except for dawn) and is also followed by some other activity (except for dusk). To ensure that only one location and mode is chosen per activity, constraint (16)

13

is set in place. Constraint (17) makes sure that if activity b follows activity a then a's start time plus duration plus the travel time from a to b is exactly equal to the start time of b. The last constraint (18) guarantees that every activity can only take place within its designated feasible time window.

3.3.3 Tour-based indicators and constraints

We extend the optimisation model by several indicator variables and constraints that capture a general schedule structure (e.g. number of primary activities per tour). These indicator variables are later used to constrain the decision options for certain variables such as tour type or sub-tour activity to the sets as introduced in Table 1.

The first introduced tour-related variable is a tour indicator:

$$\vartheta_{at} \in \{0, 1\} \qquad \forall a \in \mathcal{A}_i \quad \forall t \in \mathcal{T}_i \tag{19}$$

where $\mathcal{T}_i$ is the set of all possible out-of-home tours that can occur in a given schedule alternative i. If activity a takes place in tour t, ϑ_{at} equals one. The corresponding tour constraints are:

$$\begin{aligned}
\vartheta_{at} &\leq \vartheta_{bt} - z_{ab} + 1 & \forall a \in \mathcal{A}_i \setminus \{\text{dusk}\} \quad \forall b \in \mathcal{A}_i \setminus \mathcal{H}_i \quad \forall t \in \mathcal{T}_i \\
\vartheta_{at} &\geq \vartheta_{bt} + z_{ab} - 1 & \forall a \in \mathcal{A}_i \setminus \{\text{dusk}\} \quad \forall b \in \mathcal{A}_i \setminus \mathcal{H}_i \quad \forall t \in \mathcal{T}_i
\end{aligned} \tag{20}$$

where $\mathcal{H}_i$ is the set of all home activities. If $z_{ab} = 1$, the two constraints make sure that $\vartheta_{at} = \vartheta_{bt}$. Or in other words, if activities a and b are executed in sequence, both must be in the same tour. An exception is made for home activities, as these functions are break points between out-of-home tours. The break points make sure that the tour indicator ϑ_{at} always changes at a home activity h.

We use this tour-indicator variable to assign a tour type to every tour, which is either *work*, *education* or *secondary* (see Section 3.1). The type of the tour depends on the presence of a primary activity of a certain type within that tour. For the scope of this work, we keep the tour type fixed for each activity. Therefore, we add constraints that ensure that every activity a takes place in a tour with a type corresponding to the predefined activity tour type:

$$\sum_{t \in \mathcal{T}_i(a)} \vartheta_{at} \geq 1 \qquad \forall a \in \mathcal{A}_i \tag{21}$$

where $\mathcal{T}_i(a)$ is the set of tours with type equal to the tour type of activity a. With this constraint, it is ensured that a takes place in at least one tour, and this tour has the type as defined in the activity set $\mathcal{A}_i$.

Furthermore, we use the tour indicator to fix the number of primary activities per tour. This is achieved with the following constraint:

$$\sum_{a \in \mathcal{P}_i} \vartheta_{at} = p_t \qquad \forall t \in \mathcal{T}_i \tag{22}$$

where p_t is the number of primary activities in tour t as predefined in the activity set $\mathcal{A}_i$. To add additional modelling capability depending on the position of activity a in the schedule, we next introduce a variable that functions as a sub-tour indicator:

$$\psi_a \in \{0, 1\} \qquad \forall a \in \mathcal{A}_i \tag{23}$$

14

together with the following constraints

$$
\begin{aligned}
\psi_a &\leq \psi_b - z_{ab} + 1 & \forall a \in \mathcal{A}_i \quad \forall b \in \mathcal{A}_i \setminus \mathcal{P}_i \\
\psi_a &\geq \psi_b + z_{ab} - 1 & \forall a \in \mathcal{A}_i \quad \forall b \in \mathcal{A}_i \setminus \mathcal{P}_i
\end{aligned}
\tag{24}
$$

where $\mathcal{P}_i$ is the set of all primary activities. Per definition, a sub-tour is not allowed to start or end at home. For this reason, we fix the sub-tour indicator for activities in the set $\mathcal{H}_i$ to be zero:

$$
\psi_a = 0 \quad \forall a \in \mathcal{H}_i
\tag{25}
$$

Sub-tours are constructed in a way that the first instance of the primary activity is part of the sub-tour. The second instance after the sub-tour is not part of the sub-tour anymore, since the following trip is going home-wards (possibly with another activity in between). We use the information about the number of primary activities in a tour p_t as given in constraint (22) to introduce a minimum a amount of sub-tour activities:

$$
p_t \leq \sum_{a \in P_t} \psi_a + 1 \quad \forall t \in \mathcal{T}_i
\tag{26}
$$

which means that ψ_a has to be at least one for primary activities in tour t if two primary activities are present ($p_t = 2$). This will always be the first instance of the primary activities since the second instance is directed towards a home activity at some point and hence forced to be zero because of constraint (25). Lastly, at least one secondary sub-tour activity must happen between them since the two primary activities are not allowed to take place directly after each other (13).

4 Case study

This section shows the application of the proposed framework to the full-time workers in a synthetic population on a city-scale for Lausanne, Switzerland. Section 4.1 gives a general overview of the case study with the external inputs and general assumptions. Section 4.2 defines the desired timings and durations for all activity types which are used for both parameter estimation and schedule optimisation based on empirical data. In Section 4.3, the estimation of the flexibility parameters is demonstrated and discussed. Finally, Section 4.4 gives an overview over the implementation of the optimisation model and shows the results of applying the flexibility parameters to a synthetic population using the proposed mathematical framework.

4.1 Overview

The proposed framework is applied to the full-time workers living in the city of Lausanne, Switzerland. With more than 140'000 inhabitants (as of 2017), it is one of the biggest cities in Switzerland. Since behaviour depends on specific person attributes – for example, a student is likely to have a different behaviour compared to a full-time worker –, the population must be divided into different groups and the flexibility parameters calibrated separately for each group. For demonstration purposes in this work, we focus on the calibration of the parameters for full-time workers (employment rate higher than 80 % and not primarily in education). As shown in Figure 1, the framework needs two static inputs:

1. *Reported schedules*: We use the reported schedules from the Swiss MTMC (BfS and ARE, 2017). The computer-assisted telephone survey takes place every 5 years, most recently in 2015. It contains a sample of 57'090 persons from all over Switzerland. Each person reports on their conducted mobility schedule for a full day. A cleaning procedure is applied to remove persons who have a reporting date on the weekend and persons with non-valid schedules. Non-valid schedules include schedules that are not fully reported and schedules which do not comply with the constraints of the model (e.g. a person must always return home in the evening or primary and home activities must not take place consecutively). After the cleaning procedure, around 40'000 reported schedules remain in the observation set. For full-time workers in the whole country of Switzerland, the cleaned set in the MTMC contains 10'110 reported schedules.

2. *Synthetic population*: The nation-wide synthetic population of Switzerland for the year 2017 is generated by Bodenmann et al. (2019). It contains a very detailed person database with attributes such as age, employment rate, etc. as well as household structures and residences. In the synthetic population, almost 50'000 full-time workers are present in the city of Lausanne.

An additional third input for the model are travel times between different locations for all modes (see Section 3.3.2). We use static travel times which are derived from a network assignment using MATSim (Horni et al., 2016) and the Switzerland scenario from Scherr et al. (2020a). As a simplified assumption, we use constant travel times over the day. The public transport travel times are averaged over one hour (i.e. 7:00-8:00 h). Since the service frequency and the travel times of public transport in Switzerland are fairly constant throughout the day, constant travel times over the day are assumed to be valid. Car travel times are defined as the maximum between the congested state of the morning peak and the evening peak, which is a rather pessimistic assumption for car traffic. Travel times for the modes walk and bike are calculated based on the beeline distance between two locations, a detour factor and an average speed.

For both choice set generation and schedule optimisation, the following model steps are generated based on existing discrete choice models (see Section 3.3.1) as proposed by Scherr et al. (2020b) and assumed to be given in this case study: (i) The set of tours $\mathcal{T}$; (ii) the tour type of each activity $\vartheta_{at} \ \forall t \in \mathcal{T} \ \forall a \in \mathcal{A}$; (iii) the number of home activities $|\mathcal{H}| = |\mathcal{T}| + 1$; (iv) the set of primary activities $\mathcal{P}$ with a type and a number of primary activities per tour p_t; (v) the set of secondary activities $\mathcal{S}$ with a given activity type; (vi) the set of secondary sub-tour activities $\psi_a = 1$; and (vii) the considered destinations $\mathcal{L}_a \ \forall a \in \mathcal{A}$.

4.2 Definition of a desired timing and duration per activity type based on empirical distributions

A crucial element of the proposed framework is the definition of activity types for a specific person group with corresponding desired timings x_a^* and durations τ_a^* for each $a \in \mathcal{A}$. As these terms are defined in the utility specification in Equation (1), they are a common requirement for both parameter estimation and schedule optimisation. We use the main activity types for home, primary and secondary activities as defined in Section 3.1.

However, the flexibility of activity start times and durations might vary throughout the day (e.g. the flexibility of starting with a work activity might be different in the morning compared to starting with the afternoon work activity after lunch). Therefore, we split the main activity types into different behaviourally homogeneous sub-activities. By behaviourally homogeneous, we mean activities that have uni-modal distributions in the dimensions of both x_a^* and τ_a^* in the reported schedules. This is based on the assumption that activities with reported uni-modal distributions actually belong together and have the same meaning in terms of flexibility for specific desired timings and durations across all the individuals in a specific person group. As explained in Section 3.2.3, x_a^* and τ_a^* are then defined based on the mean values from empirical distributions in the MTMC. To find uni-modal distributions, the available information in the activity set $\mathcal{A}$ is used (e.g. number of primary activities or the number and type of tours).

Table 2 shows the empirical analysis of the activity types as observed in the MTMC. We divide activities into sub-activities by applying a manual cluster analysis based on visually observing the distributions for activity timings and durations. Our goal is to find uni-modal distributions while not splitting the activity type into too small clusters. In this work, all observations are assumed to be normally distributed. For each sub-activity distribution, we calculate the mean and the standard deviation for both x_a^* and τ_a^*.

A straightforward example is the activity type *work* in Table 2. It is divided into the two clusters of the first occurrence in the activity set $\mathcal{A}$ and – if present – following occurrences of work activities in $\mathcal{A}$. Our analysis shows that $\mathcal{A}$ contains at least one work activity in 80 % of the schedules (average occurrence 0.80). This means that on an average work day in Switzerland, 80 % of all full-time workers go to their work place at least once. In that case, workers are observed to begin their work activity with a mean timing of about 7:30 h. The observed distribution has a standard deviation of 0.9. The average occurrence of having more than one work activity in the activity set is 0.31, meaning that 31 % of the all individuals participate in at least one activity between the two work activities (e.g. returning home, or going out for lunch). The average desired timing x_a^* for returning back to the work place is at 13:15 h and follows more pronounced peak (standard deviation of 0.3). The total duration budget τ_a^* (sum over all work activity durations in $\mathcal{A}$) for full-time workers is distributed around a mean 9.5 h as observed in the survey.

Another notable example is the activity *home*. In the case of two work tours ($|\mathcal{T}_{work}| = 2$), exactly one home activity must take place between the two out-of-home tours per definition. This home activity between two work tours has a mean start time at 12:00 h and shows a very strong peak. Additionally, full-time works are returning back home in average between 17:00 and 18:00 h in the case of at least one work activity in $\mathcal{A}$. The observed total home duration budget is distributed around a mean of around 13 h (sum over all home activity durations in $\mathcal{A}$), meaning that the group of the full-time workers in average spends more than half of the day at home.

The only exception from the assumption of uni-modal peaks is the activity *escort*. It has two pronounced peaks which cannot be split into sub-activities with uni-modal peaks based on the information available in $\mathcal{A}$. Both morning and evening peak are equally likely and hence, we take a random draw from this set to assign a desired start time to an escort activity.

Table 2: Empirical analysis of activity types for full-time workers in the MTMC. Total number of observed schedules is 10'110.

Activity	Sub-activity indicator	Description	Avg. occurrence per schedule	Des. timing x^* [h] mean	std	Des. duration τ^* [h] mean	std
work	p_1	first in set	0.80	7.4	0.9		
	$p_2, ..., p_n$	following	0.31	13.2	0.3		
	$\sum_{a \in \mathcal{P}} \tau_a$	duration budget				9.5	0.5
leisure	$\psi = 1$	lunch	0.10	12.1	0.1	0.9	0.1
	$\vartheta_{work} = 1$	work tour	0.11	18.0	0.5	0.8	0.3
	$\vartheta_{secondary} = 1$	secondary tour	0.17	19.2	0.8	1.8	0.6
	$|\mathcal{P}| = 0$	no primary activity	0.16	13.0	4.0	1.5	0.7
shopping	$\psi = 1$	between work	0.06	12.1	0.1	0.4	0.2
	$\vartheta_{work} = 1$	work tour	0.16	17.2	0.5	0.4	0.2
	$\vartheta_{secondary} = 1$	secondary tour	0.22	14.0	1.5	0.4	0.2
escort		all day	0.16	{7.5, 18.0}	0.3	0.3	0.1
business		all day	0.03	10.0	1.0	4.0	2.0
education		all day	0.02	18.0	2.0	3.5	0.5
other		all day	0.19	10.0	1.0	4.0	2.0
dawn			1.00	0.0			
dusk	$|\mathcal{T}| > 1$	multi tours	0.41	15.4	1.0		
	$|\mathcal{T}_{work}| = 1$	one work tour	0.12	18.0	0.5		
	$|\mathcal{T}_{secondary}| = 1$	one secondary tour	0.47	16.4	0.7		
home	$h_1 \wedge |\mathcal{T}_{work}| = 2$	lunch	0.14	12.0	0.1		
	$|\mathcal{T}_{work}| > 0$	work tour present	0.25	17.2	0.2		
	$|\mathcal{T}_{work}| = 0$	no work tour	0.10	17.3	0.4		
	$\sum_{a \in \mathcal{H}} \tau_a$	duration budget				12.9	2.0

4.3 Flexibility parameter calibration

Besides the desired timings x_a^* and duration τ_a^* for each activity type, the estimation of flexibility parameters relies on strategic choice sets containing alternatives that decision-makers are likely to consider, as explained in Section 3.2.1. In this case study, a choice set is generated consisting of the realised schedule as well as a sample of 100 *random alternatives* and 100 *likely considered alternatives*. The alternative schedules are generated based on the realised schedules as reported in the MTMC of all full-time workers under consideration of the general schedule structure as explained in Section 3.3.1. The choice set size has been chosen such that computational time of both choice generation and parameter estimation remain manageable (i.e. under one day). A sensitivity analysis to investigate the impact of the size must be conducted in future work.

Having generated a choice set and defined the desired timings x_a^* and durations τ_a^* for each activity type (see Table 2), we can now apply Equation (1) to each alternative in the choice set to estimate the flexibility parameters β_a^{early}, β_a^{late}, β_a^{short}, and β_a^{long}. The aim of the flexibility parameters is to express behavioural preferences when it comes to resolving time conflicts between different types of activities. The parameters penalise deviations from desired timings and duration for each activity $a \in \mathcal{A}$ depending on the type of the activity. The higher the penalty compared to another activity, the less flexible or the more rigid the individual is when deviating from its desired timing or duration and the more likely it tries to reschedule another activity in the case of a conflict. To calibrate these parameters, we apply the estimation routine as explained in Section 3.2.3 and use the

software Biogeme (Bierlaire, 2020).

Table 3 shows the resulting flexibility parameters after applying the proposed estimation routine to the generated choice set. Empty cells mean that these parameters are not part of the utility specification. Cells with a β that equals 0 are not significant at a 5.0 % level. In terms of being earlier or later than the desired timing x^*, the lunch activities (either at home or in a restaurant) show the highest penalties (or loss in utility). This means that people are less willing to deviate from the desired timing for lunch activities compared to other activity types and try to move other activity types in the case of a conflict. Therefore, the desired timing for lunch activities (around 12:00 h as shown in Table 2) is a rigid point in time in the daily schedule of a full-time worker. The parameter for being early (β^{early}) at lunch is penalised by a factor of around two compared to being late (β^{late}) at lunch. This means that full-time workers weight – in terms of utility – being early at lunch by 5 min equal to being late by 10 min.

Another interesting result is that full-time workers get a relatively high penalty if they do not start their first work activity at the desired time (around 7:30 h). This means that they are not flexible with starting the first work activity, which highly affects the timing when people leave their homes in the morning. The home activity shows a high penalty for being late in the case of a work activity being present in the activity set. This means that workers are also not flexible with returning back home in the evening when they have to work on that day. Being more rigid with both starting and ending work activities compared to other activity types like leisure results in a constraint time frame for work activities between the desired timing of the first work activity and the desired timing of the returning-home activity. The secondary activities show in general low penalties (like *leisure* and *escort*) and are more likely to deviate from the desired timing, with some exceptions of activity types like the activity types *other* or also *shopping* that show higher penalties and therefore have more rigid timings.

In terms of deviating from the desired duration τ^* of an activity, some activity types show strong penalties β^{late} and β^{short}. For example, the full-time workers get a high penalty β^{short} for a short lunch activity. This means the they are not willing to take a lunch break shorter than the average desired duration (0.9 h, see Table 2). Also, it is an interesting result that full-time workers do only apply a small penalisation to a short work duration and do not penalise a long work duration. A possible interpretation is that they are in general flexible with their work duration, but the duration is determined by other factors like specific office hours. The flexibility parameter for a longer total time spent at home compared to the desired duration (12.9 h) indicates that the group of the full-time workers penalise being too long at home during the day. This means that they actively have the need or the desire to participate in activities outside of the home for a certain time (around half) of the day.

4.4 Optimised schedules

For the simulation of the schedules in this case study, we implement the MILP as described in Sections 3.3.2 and 3.3.3 in the framework OR-Tools[2] for Python. This general-purpose library allows for applying multiple solvers. To forestall the license burden for practitioners, the non-commercial SCIP-solver (Achterberg et al., 2008) is applied. Addi-

---

[2]`https://developers.google.com/optimization`, last accessed: 29.08.2021

Table 3: Flexibility parameter values and summary statistics for full-time workers. All parameters are significant at 5.0 % level.

Activity type	Parameter values			
	β_a^{early}	β_a^{late}	β_a^{short}	β_a^{long}
work:first in set	-0.615	-0.436		
work:following	-0.406	0		
work:duration budget			-0.022	0
leisure:lunch	-1.610	-0.821	-7.550	-1.360
leisure:work tour	-0.195	0	0	0
leisure:secondary tour	-0.076	0	-3.060	-0.692
leisure:no primary activity	-0.053	0	0	-0.588
shopping:between work	-0.545	-0.531	0	-2.240
shopping:work tour	-0.150	-0.316	0	-0.362
shopping:secondary tour	-0.239	-0.486	-5.910	-0.721
escort:all day	-0.175	0	0	-2.120
business:all day	-0.135	-0.267	-0.361	-0.863
education:all day	-0.246	-0.376	-0.376	-0.939
other:all day	-0.712	-0.548	0	0
dawn	0	0		
dusk:multi tours	-0.662	0		
dusk:one work tour	0	0		
dusk:one secondary tour	0	0		
home:lunch	-2.040	-0.929		
home:work tour present	-0.073	-0.596		
home:no work tour	0	-0.198		
home:duration budget			0	-0.354

Summary statistics

Number of parameters	64
Sample size	10'110
Initial log-likelihood	-57'295.58
Final log-likelihood	-47'847.37
$\bar{\rho}^2$	0.165
Estimation time (h)	13.00

tionally, the Ray[3] library is implemented to provide an efficient way for parallel computing. It reduces the computation time linearly by the amount of cores.

With this implementation of the optimisation model as well as all the inputs including

---

[3] `https://ray.io/`, last accessed: 29.08.2021

flexibility parameters from Table 3, the framework is applied to each full-time worker as of the synthetic population of Lausanne, Switzerland. Firstly, we generate an activity set $\mathcal{A}$ for each individual, as described in Section 3.3.1. For each activity $a \in \mathcal{A}$, we then take a random draw from a normal distribution to assign a desired timing x_a^* and duration τ_a^* based on its activity type as well as the corresponding mean and standard deviation as given in Table 2. The output of the optimisation model is a realised schedule with a final choice for each decision variable as given in Section 3.3.2 for the individual of the synthetic population.

A minority of the full-time workers (around 7 %) make the decision of not participating in any activity outside of the home during the day ($|\mathcal{A}| = 0$). After removing them from the optimisation model, around 45k schedules are simulated. We set the optimality gap in the MILP solver to be 0%, which guarantees that the found solution is optimal. In the optimisation model, the individuals may choose among the transport modes *walk*, *bike*, *public transport* (buses, trains, and all other means) and *car as a passenger*. If the individual owns a driving license and there is a car available in the household, the mode alternative *car as a driver* is also listed in the set of modes $\mathcal{M}$. The mode has to remain the same throughout an out-of-home tour. The time period ξ is 24 hours, meaning that every person must be back home at midnight. The feasible time window γ_a^- and γ_a^+ for each activity type is 4:00 h to 23:00 h, respectively. The only exception is the activity type leisure, with a γ_a^+ of 23:50 h.

Figure 2 depicts the final aggregated results of the application of the proposed framework to each full-time worker in Lausanne. It shows the relative frequency of individuals who are participating at certain activity type at a given time of the day (i.e. activity profiles). The figure compares the results of the framework (left Figure 2a) to the observed activity profiles in the MTMC (right Figure 2b). Overall, the results of our framework fit the curve of the MTMC nicely. Everyone starts the day at home (light grey area). Between 7:00 h and 17:00 h, the work activity (red) is dominating. The work shows a steep peak in the morning and at 10:00 h, between 60 and 80 % of all full-time workers are staying at their work place. At 12:00 h, there is a sudden discontinuity, when people are going to have lunch. This is also indicated by an increasing participation in lunch/leisure (green) and home activities. In the evening, the end time of work activities follows a less steep distribution compared to the morning. After 19:00 h, the dominated activity (besides home) is leisure, which typically ends at between 21:00 h and 22:00 h. Only few persons in MTMC are participating at leisure activities after 24:00 h, which is not captured by the model since we constrain the time period to be 24 hours.

In general, travelling (dark grey) consumes a small percentage of the day of a full-time worker. But steep transitions from one activity type to another mean that a lot of persons are travelling during that time period, which causes the traffic peaks. Compared to the profile in the MTMC, the model tends to underestimate the travel episodes. This can be explained by the simplified assumption of a constant travel penalty $\beta_{travel} = -1$ (see Section 3.2.3) across all modes of transport. Due to this assumption, the individuals are just minimising travel time in the optimisation model and do not consider import mode choice aspects such as costs or personal preferences.

Another aim of the proposed framework is to capture the interactions in the temporal dimension. The temporal dimension mainly includes the choices of activity start times and durations. Travel episodes play a minor role in average as shown in Figure 2. Figure

(a) Simulated activity profiles in the optimisation model

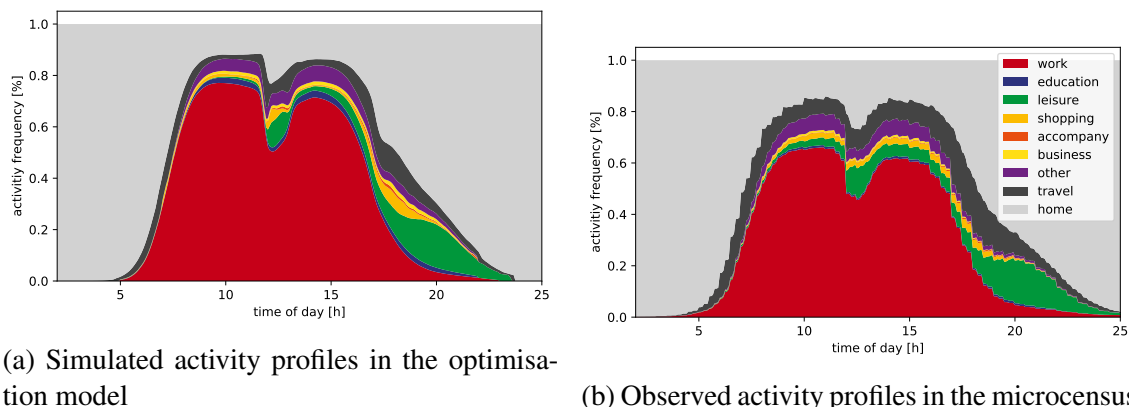(b) Observed activity profiles in the microcensus

Figure 2: Simulated activity profiles of the simulation compared to the Swiss microcensus.

3 demonstrates the correlation of start times and durations for work and leisure activities. The darker the area, the higher the number of observations that are made within this time frame. The figure compares the realised schedules of the optimisation model (left) to the observed schedules in the MTMC (right). Again, there is an excellent fit between the simulation and the empirical observation. Work activities (Figures 3a and 3b) show two different types of behaviour: (i) the work activities which are starting early in the morning and take place for around 9 hours; and (ii) the ones which are divided into two separate activities starting in the morning for 4 hours and than starting again in the afternoon for another 4 hours. The model slightly overestimates the number of persons who are working for the full 9 hours.

The leisure activities (Figures 3c and 3d) are also captured well by the model. The distribution shows a peak at lunch time for 30 to 60 minutes. In the evening, both short and longer leisure activities are taking place, and both types of leisure activities are well represented in the model. During the day, the model underestimates the participation at leisure activities of short duration (i.e. under 30 minutes).

5   Conclusion and future work

In this work, we demonstrate the capabilities of the presented framework to solve the activity-based scheduling problem and to reproduce empirical observations from the Swiss MTMC (BfS and ARE, 2017). Methodologically, this is accomplished by applying two modelling streams: (i) an estimation routine to quantify individual flexibility parameters based on maximum likelihood estimation; and (ii) a simulation approach that uses the flexibility parameters to resolve temporal conflicts (i.e. activity timings and durations) in order to maximise the schedule utility for each individual as part of a synthetic population. The simulation of the temporal choice dimensions is implemented as a MILP and builds on the work of Pougala et al. (2021a). For the non-temporal choice dimensions (i.e. activity participation, number and type of tours, and considered locations), we simulate a sequence of traditional discrete choice models as introduced by Scherr et al. (2020b).

Modelling time as a continuous variable with estimated flexibility parameters overcomes limitations of earlier approaches in the literature which either represent time as a dis-

(a) Simulated correlations for work activities in the optimisation model

(b) Observed correlations for work activities in the microcensus

(c) Simulated correlations for leisure activities in the optimisation model

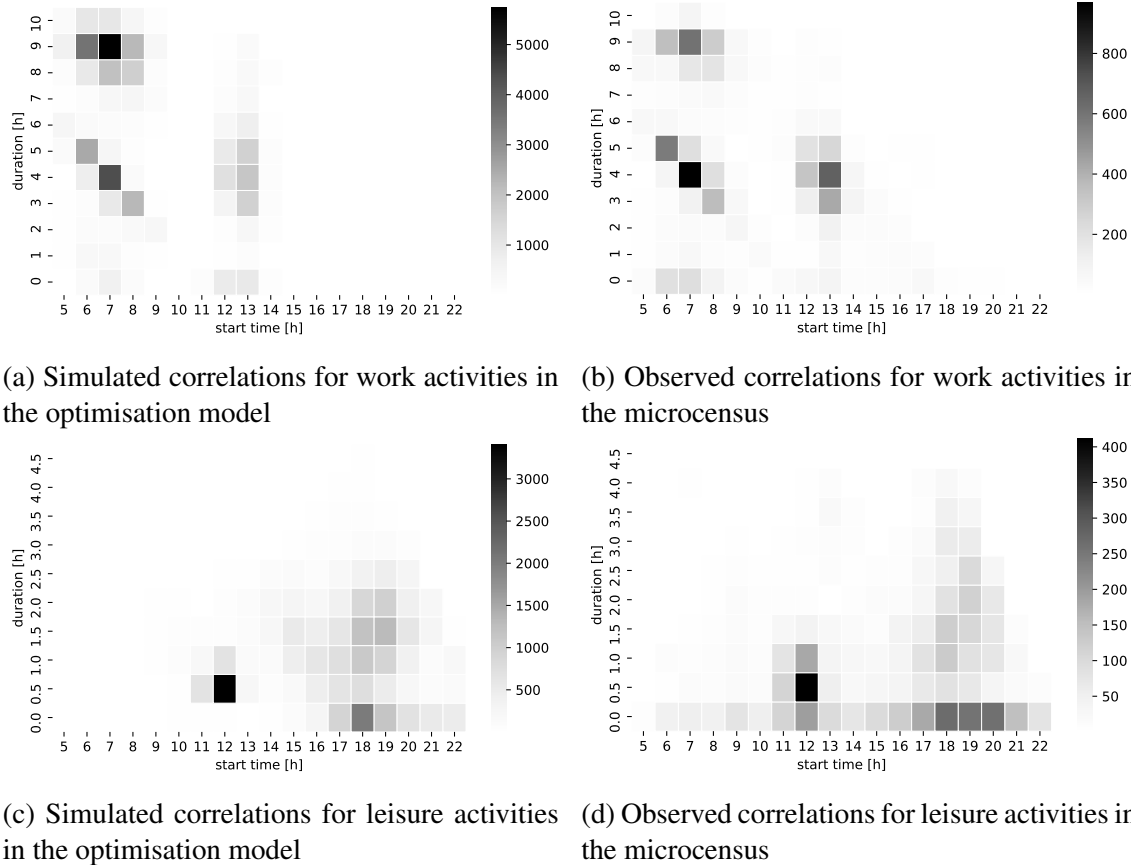(d) Observed correlations for leisure activities in the microcensus

Figure 3: Simulated correlation between start times and durations for selected activity types compared to the Swiss microcensus.

crete variable (Bowman and Ben-Akiva, 2001, Castiglione et al., 2015) or as a continuous variable in combination with heuristic flexibility parameters (Javanmardi et al., 2016, Esztergár-Kiss et al., 2020, Pougala et al., 2021a). By treating all temporal scheduling decisions simultaneously, our framework captures interactions between activity timings and durations in contrast to earlier work, e.g. Scherr et al. (2020b) who simulate activity timings as a sequential modelling step using fixed activity durations. Also, it goes beyond the theoretical framework of Pougala et al. (2021a) by proving the applicability of the framework in a real-world city-scale case study.

Our proposed method provides a holistic representation of the scheduling process and uses continuous temporal variables within a utility-maximisation framework. It allows for behavioural realism especially in the temporal dimension, which is often modelled based on simplified assumptions in the literature (Bhat et al., 2004, Auld et al., 2009, Saadi et al., 2016). The method is demonstrated to estimate flexibility parameters that produce close-to-reality results in a real-world application on a city level.

We identify the following paths for future work:

- *Utility specification*: All flexibility parameters in this work are defined to be linear penalties. Balać et al. (2018) suggest a non-linear utility function to score the activities in a schedule. It would be interesting to investigate if non-linear (or piece-wise linear) penalties are able to improve the model statistics.

23

- *Choice set generation*: We introduce the heuristic method of combining an existing activity-based model with a random alternative generator. This relies on already having a fully operational activity-based model. Since this is often not the case, a choice set needs to be generated from scratch containing alternatives that a decision-maker is likely to have considered for a reported schedule. A possible direction to design a choice set from scratch is to apply a Metropolis Hastings algorithm (Pougala et al., 2021b, Yamamoto et al., 2001).

- *Parameter estimation*: Our approach uses fixed desired activity timings and durations, which are found by manual and visual investigation of the empirically reported schedules. In the future, a possible enhancement of the estimation routine is implementing a mixed logit model (Ben-Akiva et al., 2001) to estimate distributed values for the desired timings and durations.

- *Mode choice utilities and travel times*: The focus of this work lies on activity timings and not on calibrating a mode choice model. A mode choice model would require an extension of the estimation routine to find mode choice parameters to express behavioural preferences and to quantify the influence of level of service indicators (e.g. parking costs, travel time or service frequency). Also, constant travel times are assumed over a 24h-day for all modes of transportation. Particularly the car travel times are depending on the time of the day as travel times differ a lot between peak and off-peak times. For this reason, time-of-day dependent travel times would be an enhancement of this framework when focusing on mode choice.

- *Coupling with traffic assignment*: The proposed framework cannot predict network loads, e.g. the number of rail passengers boarding a train at a certain time. For this purpose, the optimised schedules must be fed into a traffic assignment, e.g. using the software MATSim. The results of the traffic assignment can help transport planners to assess policy studies based on network loads.

- *Real-world policy study*: Based in this work, time-of-day relevant studies can be conducted. One example is the ongoing global COVID-19 pandemic, which is assumed to cause a higher prevalence in working from home. The increased flexibility in work scheduling decisions likely influence the ways in which people organise their day. It would be interesting to investigate how sensitive the proposed framework responds to questions such as (i) do people extend the time spent at home? (ii) do they travel longer distances to the work place? (iii) do they extend the duration of their leisure activities?

Acknowledgements

CRediT authorship contribution statement

Conflict of interest

The authors declare that they have no conflict of interests.

References

Achterberg, T., Berthold, T., Koch, T., and Wolter, K. (2008). Constraint Integer Programming: A New Approach to Integrate CP and MIP. In Perron, L. and Trick, M. A., editors, *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, pages 6–20, Berlin, Heidelberg. Springer Berlin Heidelberg.

Arentze, T. A. and Timmermans, H. J. (2004). A learning-based transportation oriented simulation system. *Transportation Research Part B: Methodological*, 38(7):613–633.

Auld, J., Mohammadian, A. K., and Doherty, S. T. (2009). Modeling activity conflict resolution strategies using scheduling process data. *Transportation Research Part A: Policy and Practice*, 43(4):386–400.

Axhausen, K. W. and Gärling, T. (1992). Activity-based approaches to travel analysis: conceptual frameworks, models, and research problems. *Transport reviews*, 12(4):323–341.

Balać, M., Janzen, M., and Axhausen, K. W. (2018). Alternative approach to scoring in MATSim and how it affects activity rescheduling. In *2018 TRB Annual Meeting Online*. Transportation Research Board.

Ballis, H. and Dimitriou, L. (2020). Revealing personal activities schedules from synthesizing multi-period origin-destination matrices. *Transportation research part B: methodological*, 139:224–258.

Bellemans, T., Kochan, B., Janssens, D., Wets, G., Arentze, T., and Timmermans, H. (2010). Implementation framework and development trajectory of FEATHERS activity-based simulation platform. *Transportation Research Record*, 2175(1):111–119.

Ben-Akiva, M., Bolduc, D., and Walker, J. (2001). Specification, identification and estimation of the logit kernel (or continuous mixed logit) model. *Working paper*.

BfS and ARE (2017). *Population's transport behaviour 2015. Key results of the mobility and transport microcensus*. Neuchâtel, Switzerland.

Bhat, C. R., Guo, J. Y., Srinivasan, S., and Sivakumar, A. (2004). Comprehensive econometric microsimulator for daily activity-travel patterns. *Transportation Research Record*, 1894(1):57–66.

Bierlaire, M. (2020). A short introduction to pandasbiogeme. Technical Report TRANSP-OR 200605, Lausanne, Switzerland.

Bodenmann, B., Bürki, P., Phillipp, C., Bernhard, N., and Müller, K. (2019). *Synthetische Population 2017 - Modellierung mit dem Flächennutzungsmodell FaLC*. Report for Bundesamt für Raumentwicklung ARE and Swiss Federal Railways (SBB), Bern, Switzerland.

Bowman, J. L. (2009). Historical development of activity based model theory and practice. *Traffic Engineering & Control*, 50(2):59–62.

Bowman, J. L. and Ben-Akiva, M. E. (2001). Activity-based disaggregate travel demand model system with activity schedules. *Transportation research part a: policy and practice*, 35(1):1–28.

Boyce, D. E. and Williams, H. C. (2015). *Forecasting urban travel: Past, present and future*. Edward Elgar Publishing, Cheltenham, UK.

Briem, L., Mallig, N., and Vortisch, P. (2019). Creating an integrated agent-based travel demand model by combining mobiTopp and MATSim. *Procedia Computer Science*, 151:776–781.

Castiglione, J., Bradley, M., and Gliebe, J. (2015). *Activity-based travel demand models: A primer*. Number SHRP 2 Report S2-C46-RR-1.

Davidson, B., Vovsha, P., and Freedman, J. (2011). New advancements in activity-based models. In *Australasian Transport Research Forum*, pages 28–30. Citeseer.

Drchal, J., Čertický, M., and Jakob, M. (2019). Data-driven activity scheduler for agent-based mobility models. *Transportation Research Part C: Emerging Technologies*, 98:370–390.

Esztergár-Kiss, D., Rózsa, Z., and Tettamanti, T. (2020). An activity chain optimization method with comparison of test cases for different transportation modes. *Transportmetrica A: transport science*, 16(2):293–315.

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

Hillel, T., Pougala, J., Manser, P., Luethi, R., Scherr, W., and Bierlaire, M. (2020). Modelling mobility tool availability at a household and individual level: A case study of switzerland. In *hEART conference. Lyon, France*.

Hörl, S. and Balac, M. (2021). Synthetic population and travel demand for Paris and Île-de-France based on open and publicly available data. *Transportation Research Part C: Emerging Technologies*, 130:103291.

Hörl, S., Balać, M., and Axhausen, K. W. (2019). Pairing discrete mode choice models and agent-based transport simulation with MATSim. In *2019 TRB Annual Meeting Online*. Transportation Research Board.

Horni, A., Nagel, K., and Axhausen, K. W. (2016). *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press, London.

Javanmardi, M., Langerudi, M. F., Shabanpour, R., and Mohammadian, A. (2016). An optimization approach to resolve activity scheduling conflicts in ADAPTS activity-based model. *Transportation*, 43(6):1023–1039.

Joubert, J. W. and De Waal, A. (2020). Activity-based travel demand generation using Bayesian networks. *Transportation Research Part C: Emerging Technologies*, 120:102804.

Liu, F., Janssens, D., Cui, J., Wets, G., and Cools, M. (2015). Characterizing activity sequences using profile Hidden Markov Models. *Expert Systems with Applications*, 42(13):5705–5722.

Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8(3):229–254.

Miller, E. J. and Roorda, M. J. (2003). Prototype model of household activity-travel scheduling. *Transportation Research Record*, 1831(1):114–121.

Moeckel, R., Kuehnel, N., Llorca, C., Moreno, A. T., and Rayaprolu, H. (2020). Agent-based simulation to improve policy sensitivity of trip-based models. *Journal of Advanced Transportation*, 2020.

Pougala, J., Hillel, T., and Bierlaire, M. (2021a). Capturing trade-offs between daily scheduling choices. Technical Report TRANSP-OR 210101, Lausanne, Switzerland.

Pougala, J., Hillel, T., and Bierlaire, M. (2021b). Choice set generation for activity-based models. In *Swiss Transport Research Conference 2021, Monte Veritá, Ascona, Switzerland*.

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

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

Recker, W. W. (2001). A bridge between travel demand modeling and activity-based travel analysis. *Transportation Research Part B: Methodological*, 35(5):481–506.

Rizopoulos, D. and Esztergár-Kiss, D. (2020). A method for the optimization of daily activity chains including electric vehicles. *Energies*, 13(4):906.

Roorda, M. J., Miller, E. J., and Habib, K. M. (2008). Validation of TASHA: A 24-h activity scheduling microsimulation model. *Transportation Research Part A: Policy and Practice*, 42(2):360–375.

Saadi, I., Mustafa, A., Teller, J., and Cools, M. (2016). Forecasting travel behavior using Markov Chains-based approaches. *Transportation Research Part C: Emerging Technologies*, 69:402–417.

Scherr, W., Manser, P., and Bützberger, P. (2020a). SIMBA MOBi: Microscopic Mobility Simulation for Corporate Planning. *Transportation Research Procedia*, 49:30–43.

Scherr, W., Manser, P., Joshi, C., Frischknecht, N., and Métrailler, D. (2020b). Towards agent-based travel demand simulation across all mobility choices - the role of balancing preferences and constraints. *European Journal of Transport and Infrastructure Research*, 20(4):152–172.

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

Vovsha, P., Bradley, M., and Bowman, J. L. (2005). Activity-based travel forecasting models in the United States: progress since 1995 and prospects for the future.

Yamamoto, T., Kitamura, R., and Kishizawa, K. (2001). Sampling alternatives from colossal choice set: Application of markov chain monte carlo algorithm. *Transportation Research Record*, 1752(1):53–61.

Ziemke, D., Nagel, K., and Bhat, C. (2015). Integrating CEMDAP and MATSim to increase the transferability of transport demand models. *Transportation research record*, 2493(1):117–125.