Resolving time conflicts in activity-based scheduling: A case study of Lausanne

Patrick Manser[a,*], Tom Haering[b], Tim Hillel[b], Janody Pougala[b], Rico Krueger[b], Michel Bierlaire[b]

[a]*Swiss Federal Railways (SBB), Hilfikerstrasse 1, 3000 Bern 65, Switzerland*

[b]*École Polytechnique Fédérale de Lausanne (EPFL), Transport and Mobility Laboratory (TRANSP-OR), Station 18, 1015 Lausanne, Switzerland*

Abstract

In this paper, we present a novel activity-based scheduling model that combines a continuous optimisation framework for 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 considered 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 derived utility. Our proposed framework has three primary advantages over existing activity scheduling approaches: (i) the time-conflicts between different temporal scheduling decisions are considered and resolved jointly; (ii) individual behavioural preferences are incorporated in the scheduling problem using a utility-maximisation approach; and (iii) the framework is computationally scalable and can be used to estimate and simulate a city-scale case study in reasonable time. We introduce an estimation routine for the framework that allows model parameters to be calibrated using real-world historic data, as well as an efficient mixed-integer linear solver to optimally resolve temporal conflicts in simulated schedules. The estimation routine is applied and calibrated to a set of observed schedules in the Swiss mobility and transport microcensus. We then use the optimisation program with the estimated parameters to simulate activity schedules for a synthetic population for the city of Lausanne, Switzerland. We validate the model results against reported schedules in the microcensus data. The results demonstrate the capabilities of our approach to simulate realistic, flexible schedules for a real-world case-study.

Keywords: activity-based schedules, discrete choice, maximum likelihood estimation, mathematical optimisation, mixed-integer linear program, city-scale application

1. Introduction

Planning the need for future transport infrastructure and efficient service concepts while considering possible changes in population (e.g. demographic shifts), policy or technology relies on quantitative forecasting of travel demand. For this purpose, simulation models are applied which closely represent mobility behaviour. Until today, the majority of these 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 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 instead treat mobility as arising from people's desire to participate in activities at different locations. These models have the potential to provide a higher resolution understanding of travel

*Corresponding author
Email address: patrick.manser@sbb.ch (Patrick Manser)

behaviour than traditional macroscopic models as they simulate each traveller as an autonomous decision-making unit and as they 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 established 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 are limited in modelling major system disruptions since they do not fully consider feedback between choice dimensions and are only partially based on behavioural preferences.

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. We adapt the framework introduced by Pougala et al. (2021) that simulates desired choice dimensions simultaneously using a mixed-integer linear program and that expresses behavioural 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 a reduced version of the framework, in which the non-temporal choice dimensions (i.e. activity participation, number of tours, and considered destinations) are simulated using existing discrete choice models, and are used as input for the optimisation program. This allows for a computationally efficient application to a synthetic population for a case-study city, with a detailed validation against empirically available observations.

The scope of this work is particularly relevant given the large-scale behavioural disruptions arising from the ongoing global pandemic, e.g. higher prevalence in working from home. With increased flexibility in work scheduling decisions, and no need to travel to the work place changes the ways in which people organise their day. This presents several questions, which cannot be easily answered using traditional modelling techniques, for example: (i) Do people extend the time spent at home? (ii) Do people extend the duration of their leisure activities? (iii) Do people travel longer distances to get to leisure activities? The model implemented in this paper can provide valuable new insights into these questions based on conflict-resolution preferences in the time dimension of daily activity scheduling.

The rest of the paper is structured as follows: firstly, 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 mixed-integer linear program. This framework is then applied to 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

The activity-based approach to travel demand modelling was first 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 by 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 highlighted by 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 choice.

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 MATSim (Horni et al., 2016). Most activity-based models are developed to generate schedules which are then fed into the dynamic network assignment of MATSim. 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 a framework 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 activity-based schedules are integrated into MATSim (Briem et al., 2019). A very similar approach is presented in Scherr et al. (2020b). They show a very 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.

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 mathematical programming 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) suggest 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. (2021). The main focus of their work is to combine multiple choice dimensions (activity participation, location, start time, duration and mode) into a single optimisation problem and to capture 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 the day. It represents all choices in the time dimension as continuous variables. The utility formulation is very flexible and preferences can be specified for each individual activity. The framework is implemented as a mixed-integer linear program that can be easily extended by additional constraints.

In contrast to Pougala et al. (2021), this work focuses on resolving scheduling conflicts in the time dimension (choice of activity timings and durations while considering travel times) and on computational efficiency to scale the model to a real-world large-scale simulation. The temporal dimension of the scheduling problem remains relatively underexplored within 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).

The aim of this work is to identify a well calibrated utility specification for temporal scheduling decisions, including the choices of activity start times and durations, while considering trip travel times between them. The parameters of the model are estimated based on empirically reported observations. Also, the work targets to apply the optimisation framework to a synthetic population and reproduce observations as reported in the survey. Therefore, we modify the optimisation framework to reduce the choice dimensions and to constraint some activity sequences in order to improve computation efficiency. The non-temporal choice dimensions in the scheduling process (activity participation, number and type of tours, and considered destinations) are generated using the well calibrated SIMBA MOBi model, as described by Scherr et al. (2020b), and fed into the optimisation problem as static input.

The contributions of this work are:

- *Parameter estimation*: An estimation routine is demonstrated that considers the temporal decision within the scheduling problem as a whole. It estimates schedule-based parameters and not – like most studies in literature – trip- or tour-based parameters (Bowman and Ben-Akiva, 2001; Bhat et al., 2004; Hilgert et al., 2017; Scherr et al., 2020b). The 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., 2021). This framework is hence responsive to disruptions in the temporal dimension.

- *Optimisation program*: In contrast to many studies (Bowman and Ben-Akiva, 2001; Bhat et al., 2004; Hilgert et al., 2017; Scherr et al., 2020b), the model demonstrated in this work allows for the optimisation of activities including changing the sequence of activities across multiple out-of-home tours. The implemented approach decouples the non-temporal decisions of activity participation from time-conflict resolution while considering a given schedule structure (tour types). This is achieved through reducing the decision variables and adding constraints in the framework proposed by Pougala et al. (2021). This simplification achieves two primary advantages: (i) a reduction in computational cost, and (ii) simplification of the calibration process for practitioners.

- *City-scale application*: This work applies the scheduling model to the full-time workers of a synthetic population of Lausanne, Switzerland. Substantial work has been put into calibration and fine-tuning of the parameters. The results are validated against empirically reported schedules in the Swiss mobility and transport microcensus (Federal Statistical Office, 2017). This extends the work of Pougala et al. (2021), who demonstrate the application of the framework for small samples of individuals of the testing set without calibration of parameters.

3. Methodology

This work applies an adapted version of the activity-scheduling framework proposed by Pougala et al. (2021) and simulates the non-temporal elements of the scheduling problem (i.e. activity participation, number and type of tours, and considered destinations) using sequential discrete choice models outside of the optimisation. They are fed as static input into the optimisation framework which solves all temporal scheduling decisions (start times and durations of each activity) while considering travel times simultaneously. The modified framework consists of two streams, as illustrated in Figure 1:

1. The first stream (top row in Figure 1) aims to estimate individual preferences based on reported schedules in the mobility and transport microcensus (Federal Statistical Office, 2017). Individual preferences are 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. The second stream (bottom row) optimises a schedule for each individual which is part of a given synthetic population. The methodology to generate a synthetic population goes beyond the scope of this paper. Each individual choose a set of activities they would like to perform during a 24h-day. The optimisation then resolves time conflicts based on the estimated flexibility parameters. More insight into the methodology is given in Section 3.3.

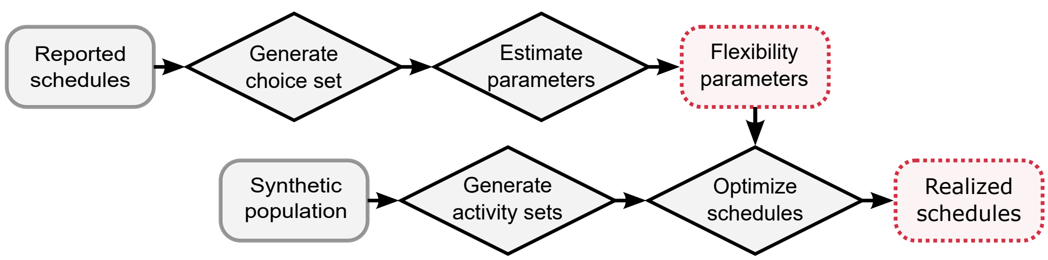


Figure 1: The two streams of the scheduling framework: calibration (top row) and simulation (bottom row).

3.1. General definitions

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

Schedule: Set of activities $\mathcal{A}$ with a given sequence without overlapping. Starts with a *dawn* activity at midnight and ends with a *dusk* activity at the defined time period. We consider two types of schedules:

Realised schedule: In the choice set of the parameter estimation, it is the reported schedule as found in a survey. In the optimisation context, it is the output of the mathematical program.

Feasible schedule: Any possible deviation from a realised schedule can be feasible as long as it fits into the time period constraint.

Activities: Consist of a specific type, start time, duration, and location. Locations and sequence of activities result in travel episodes with a mode and travel time. Grouped into three clusters:

Home activities: Set $\mathcal{H}$ with $\mathcal{H} \cap \mathcal{A}$ includes activity types *home*, *dusk* and *dawn*. 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 $h \in \mathcal{H}$ has a fixed location. Home activities do not take place consecutively, since this would not induce a trip.

Primary activities: Set $\mathcal{P}$ with $\mathcal{P} \cap \mathcal{A}$ includes activity types *work* and *education*. 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*: $\sum_{a\in\mathcal{P}} \tau_a = \tau_{\mathcal{P}}$. Each $p \in \mathcal{P}$ has a fixed location. Primary activities do not take place consecutively, since this would not induce a trip.

Secondary activities: Set $\mathcal{S}$ with $\mathcal{S} \cap \mathcal{A}$ includes activity types *leisure*, *shopping*, *accompany*, *business* and *other*. Have flexible locations. Secondary activities are not constrained to fill a given daily duration. This means that we consider the preferences of the individual towards the durations of each individual secondary activity rather than their sum. Are defined to be within a primary tour if there is no home activity between a primary activity and secondary activity.

Tour: Set $\mathcal{T}$ is the sequence of activities between two home activities. Size of set $\mathcal{T}$ equals the size of the home activity $\mathcal{H}$ set minus one ($|\mathcal{T}| = |\mathcal{H}| - 1$). $\mathcal{T}$ includes tour types *work*, *education* and *secondary*. Work and education tours include minimum one and maximum two primary activities of this type. Secondary tours include secondary activities only.

Sub-tour: Sequence of secondary activities between two primary activities without any home activity. The first primary activity is defined to be a sub-tour activity as well, while the second primary activity in not part of the sub-tour.

3.2. *Estimation of individual preferences*

3.2.1. *Choice set generation*

The way we define the choice sets 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 considering travel times between them. We can assume that the decision-maker only possesses a partial knowledge of the available opportunities. The set of all available combinations is defined as the *feasible schedule* set. In contrast to the universal set described by Shocker et al., this sample is combinatorial and potential infinite. In addition, the sample of alternatives that the individual actually considers at one point in time may not be fully known by the modeller, or readily accessible from traditional data sources.

The utility-based framework described in Pougala et al. (2021) 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. In order to estimate the parameters using classical methods such as maximum likelihood estimation, the choice set must be a strategically defined sample of the feasible set. It should contain the chosen alternative to be consistent with random utility theory. To get meaningful parameter estimates, the set should contain schedules with high utilities, as well as low-probability alternatives to avoid large biases in the estimators.

To do so, we use the heuristic approach of combining the following two methods:

1. *Random alternatives:* We start with generating random schedule alternatives by randomly modifying the activity start times and durations in the realised schedule with a simple scheduling algorithm and a given set of activities as observed in the realised schedule. The outcomes are likely to be low-probability alternatives.

2. *Likely alternatives:* We generate activity schedules with variations of activity start times, durations and travel times that are biased towards high probability schedules by applying the sequential activity-based demand model MOBi.plans as proposed in Scherr et al. (2020b) to the activities sets in the reported schedules.

3.2.2. *Utility specification*

As the adapted version of the scheduling framework in this paper puts the focus on temporal scheduling decisions while considering travel times, it makes use of a simplified utility specification from that proposed by Pougala et al. (2021).

Specifically, we include only the utility terms relating to schedule deviations from preferences for each activity a and the travel times between a and the consecutive activity in the utility specification for schedule alternative i:

$$U_i \ = \ \sum_{a\in\mathcal{A}_i} U_{\text{timing}}(x_a) \ + \ \sum_{a\in\mathcal{S}_i} U_{\text{duration}}(\tau_a) \ + \ \sum_{O\in\{\mathcal{P}_i,\mathcal{H}_i\}} U_{\text{duration}}(\textstyle\sum_{a\in O}\tau_a) \ + \ \displaystyle\sum_{a\in\mathcal{A}_i\setminus\{\text{dusk}\}} U_{\text{tt},a}(tt_a) \tag{1}$$

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 a priori negative flexibility parameters β^{early}_a and β^{late}_a which penalise the difference $|x^*_a - x_a|$.

$$U_{\text{timing}}(x_a) \ = \ \beta^{\text{early}}_a \max(0;\ x^*_a - x_a) \ + \ \beta^{\text{late}}_a \max(0;\ x_a - x^*_a) \tag{2}$$

- $U_{\text{duration}}(\tau_a)$ indicates the impact for 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 has been made for home activities. For secondary activities $a \in \mathcal{S}$, we instead use the individual desired durations (i.e., $\tau^*_a \neq \tau^*_b \forall a, b \in \mathcal{S}, a \neq b$). β^{short}_s and β^{long}_s indicate the loss in utility for deviating from a desired duration.

$$U_{\text{duration}}(\tau_a) \ = \ \beta^{\text{short}}_a \max(0;\ \tau^*_a - \tau_a) \ + \ \beta^{\text{long}}_a \max(0;\ \tau_a - \tau^*_a) \tag{3}$$

- $U_{\text{tt},a}(tt_a)$ is a dis-utility 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).

$$U_{\text{tt},a}(tt_a) \ = \ \beta_{\text{travel}}\ tt_a \tag{4}$$

3.2.3. *Model estimation*

Given the proposed utility specification, we need to estimate six parameters for each activity a based on its type: (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^{early}_a, (iv) β^{late}_a, (v) β^{short}_a, and (vi) β^{long}_a.

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 that activity type in historic data. We target to find a type for each activity a in order to ensure that the empirical distributions for each activity type are uni-modal.
2. Once the desired start times and durations are defined, the penalty terms (β^{early}_a, β^{late}_a, β^{short}_a, and β^{long}_a) can then be estimated from historic schedules using maximum likelihood estimation, according to the utility functions in Equations (2) and (3) in order to calculate the total schedule utility for each alternative i as defined Equation (1).

3.3. Schedule optimization

3.3.1. Activity set generation

As a static input for the schedule optimisation, a set of participated activities $\mathcal{A}$ is generated for each person. This comes with 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. Also, some general information about activity sequences is already available (e.g., a secondary activity can be fixed to take place in a primary tour). To generate the set of activities $\mathcal{A}$ for each individual, we simulate a sequence of multinomial logit models using the parameters from Scherr et al. (2020b) which are calibrated for the whole country of Switzerland.

In the first model step, each individual makes the long-term decision about possessing a mobility tool such as a driving license (Hillel et al., 2020) as well as about primary locations such as a work or school place. Based on these locations, a certain number and type of primary out-of-home tours is chosen. The third logit model simulates the number secondary tours. The total number of tours $|\mathcal{T}|$ defines the number of home activities $|\mathcal{H}| = |\mathcal{T}| + 1$ in the activity set $\mathcal{A}$. Then, the number and type of secondary activities is chosen for each tour type and the choice about participating at sub-tour activities is simulated for each primary tour. All choices depend on various socio-demographic attributes as well as level of service indicators such as accessibility. After this step, the number and types of primary activities $\mathcal{P}$ and secondary activities $\mathcal{S}$ are given. The last choice model generates a set of considered destinations for secondary activities. This choice heavily depends on the tour type. E.g., if a secondary activity takes place in a primary tour, the destination probabilities are highest between the home place and the primary location. The destination choice probabilities are the reason why we fix the tour the type in the scope of this work.

Table 1 gives an example for an activity set $\mathcal{A}$ which can be fed into the optimisation program. Empty cells mean that those variables are not given in advance and solved in the optimisation program. It contains four home activities, which equals three out-of-home tours. One tour is primary, and there is one sub-tour activity. Per definition, this lunch activity must always take place between the two work activities. The two work activities can still swap their position and it is still to be solved which one is part of the sub-tour. Also, it is not decided if the accompany activity takes place before or after the work activity at this point. Three activities take place within secondary tours and cannot be moved to the work tour. 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 flexible destinations, which will be decided later in the optimisation program.

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

Table 1: Example for a set of participated activities $\mathcal{A}$

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

3.3.2. Optimisation program

This section summarises the simplified version of the mixed-integer linear program (MILP) compared to the framework proposed by Pougala et al. (2021) and explains the extensions that are made to fulfil the requirements needed for this case study. 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). The objective for each individual is to maximise it's total schedule utility as defined in Equation (1):

$$\Omega \; = \; \max \; U_i(x_a, \tau_a, tt_a)$$

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

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

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:

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

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:

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

where $\mathcal{L}_a$ is the set of all available 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:

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

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 program, we linearise this multiplication with the help of an auxiliary variable.

The simplified optimisation problem is now subject to the reduced number of constraints compared to Pougala et al. (2021):

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

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

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

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

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

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

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

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

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

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

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) give certainty 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) 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 frame which.

3.3.3. Tour-based indicators and constraints

This work extends the optimisation program proposed by Pougala et al. (2021) by several additional indicator variables and constraints that aim to capture general schedule structure (e.g., the bundling an activity sequence into a predefined out-of-home tour). The first introduced tour-related variable is the tour-indicator variable:

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

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:

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

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 function as 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 want to 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:

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

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:

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

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:

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

together with the following constraints:

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

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:

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

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 Equation (22) to introduce a minimum a amount of sub-tour activities:

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

meaning 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 Equation (25). Lastly, since the two primary activities are not allowed to take place directly after each other (Equation (13)), at least one secondary sub-tour activity must happen between them.

4. Case study

This section shows the application of the proposed framework to 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 provides an empirical analysis for the clustering of the activity types which we use for both parameter estimation and schedule optimisation. In Section 4.3 and 4.4, the estimation routine consisting of the choice set generation and the estimation of the flexibility parameters is demonstrated. Finally, Section 4.5 gives an overview over the implementation of the optimisation program and Section 4.6 shows the results of applying the flexibility parameters to a synthetic population using the proposed mathematical program.

4.1. Overview

The proposed framework is tested for the city of Lausanne, Switzerland. With more than 140'000 inhabitants (as of 2017), it is one of the biggest cities in Switzerland. As shown in Figure 1, the framework needs two static inputs:

1. *Reported schedules*: We use the reported schedules from the mobility and transport microcensus (Federal Statistical Office, 2017). The computer-assisted telephone survey (CATI) 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.
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. Additionally, the long-term decisions of work and school locations as well as owning mobility tools such as a public transport subscription are simulated using existing discrete choice models (Scherr et al., 2020b; Hillel et al., 2020).

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 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 of public transport in Switzerland is very even 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.

Since behaviour heavily depends on specific person attributes – a pupil has a completely 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 in education). This group is the most active when it comes to mobility behaviour. In the synthetic population, almost 50'000 full-time workers are present in the city of Lausanne. The final realised set in the microcensus after the cleaning procedure contains 10'110 observed schedules for full-time workers in the whole country of Switzerland.

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

4.2. Empirical cluster analysis of activity types

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 estimation. 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 sub-activities. We use the available schedule information (e.g. number and type of tours) to search for sub-activity types that have uni-modal distributions in the dimensions of both activity start times and durations in the reported schedules. The aim is to find one clear distribution that uniquely defines both x_a^* and τ_a^*. 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.

Table 2 shows an empirical analysis of the activity types in the Swiss microcensus. We divide activities into sub-activities by applying a manual cluster analysis based on visually observing the distributions for activity timings and durations. The target 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 define a mean and a standard deviation for both x_a^* and τ_a^*. As shown in Figure 2, this assumption is valid for most start time distributions. For the distribution of activity durations, further work should also allow for other types of distributions (e.g. a Weibull distribution) in the context of this framework. Finding x_a^* and τ_a^* for each sub-activity type is done by iteratively assuming a value based on visual investigation and manually adjusting it based on the final results of the schedule optimisation. Hence, we repeated the cycle of parameter estimation and schedule optimization multiple times.

A straightforward example is the activity type work (see Table 2). It is divided into the two clusters of morning (first occurrence in the schedule) and afternoon (following occurrences) work activities. On a typical work day in Switzerland, 80 % of all full-time workers go to their work place at least once. In average, they start with their work activity at around 7:30 h. The observed distribution has a standard deviation of 0.9. Around 31 % have more than one work activity, meaning that they participate at least one activity between the two work activities. (which might returning home or going to restaurant for lunch). The average desired timing for returning to the work place is at 13:15 h and follows a stronger peak (standard deviation of 0.3). The total duration budget for full-time workers is distributed around 9.5 h according to the survey. In Figure 2, the observed distributions for both activity timings and durations as well as both work clusters are depicted.

Another notable example is the home activity. A typical person desires to be back home between 17:00 and 18:00 h. Together with the fact that people have a limited work duration budget, this is assumed to be the main driver for the mobility in the evening peak. Also, there is a reported home duration budget of around 13 h, meaning that even the very active group of the full-time workers in average desires to spend more around half of the day at home.

The only exception from the assumption of uni-modal peaks is the activity accompany. It has two strong peaks which cannot be explained based on the schedule structure. Both morning and evening peak are evenly likely and hence, we just make a random draw from this set to assign a desired start time to an accompany activity.

4.3. Choice set

The estimation of flexibility parameters relies on a competitive choice set as stated in Section 3.2.1. Competitive means that the alternatives should be biased towards high-probability schedules while still containing a number of unlikely alternatives. In this work, a choice set is generated consisting of a sample of 100 *random alternatives* and 100 *likely alternatives* besides the realised schedule. Therefore, we use the realised schedules as reported in the Swiss microcensus of all full-time workers and fix the general schedule structure as explained in Section 4.1 (e.g. the number of primary activities per tour). A choice set size of 200 alternatives has been found to be good number to find robust

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

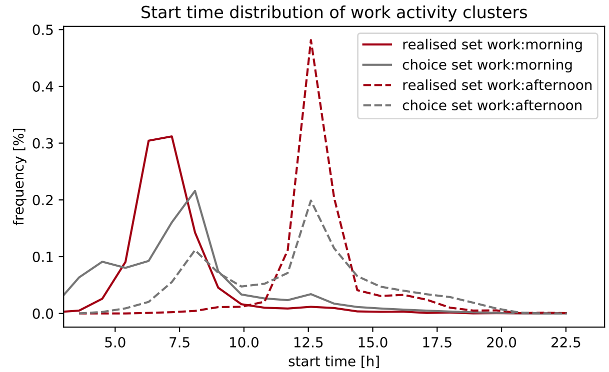
Activity	**Sub-activity**	**Description**	**Avg. occurrence**	**Des. timing x^* [h]**		**Des. duration τ^* [h]**	
	indicator		per schedule	mean	std	mean	std
work	p_1	morning	0.80	7.4	0.9		
	$p_2, ..., p_n$	afternoon	0.31	13.2	0.3		
	$\sum_{a \in \mathcal{P}} \tau_a$	duration budget				9.5	0.5
leisure	$\psi = 1$	sub-tour	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
	$\lvert\mathcal{P}\rvert = 0$	no primary activity	0.16	13.0	4.0	1.5	0.7
shopping	$\psi = 1$	sub-tour	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
accompany		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	$\lvert\mathcal{T}\rvert > 1$	multi tours	0.41	15.4	1.0		
	$\lvert\mathcal{T}_{work}\rvert = 1$	one work tour	0.12	18.0	0.5		
	$\lvert\mathcal{T}_{secondary}\rvert = 1$	one secondary tour	0.47	16.4	0.7		
home	$h_1 \wedge \lvert\mathcal{T}_{work}\rvert \geq 2$	between work	0.14	12.0	0.1		
	$\lvert\mathcal{T}_{work}\rvert > 0$	after work	0.25	17.2	0.2		
	$\lvert\mathcal{T}_{work}\rvert = 0$	no work	0.10	17.3	0.4		
	$\sum_{a \in \mathcal{H}} \tau_a$	duration budget				12.9	2.0

parameters within a reasonable computation time. A sensitivity analysis to further investigate the impact of the number of alternatives of each type is part of future work.

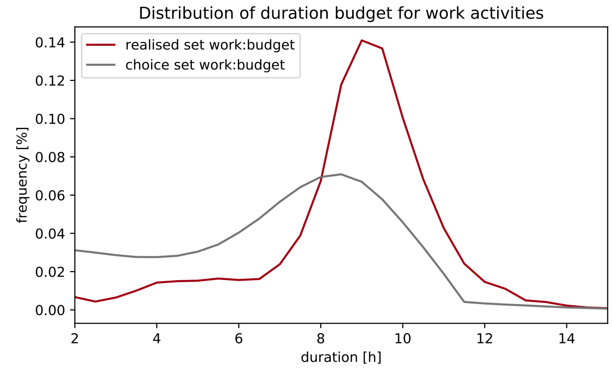
Figure 2 shows the distributions for activity timings and duration for selected activity types. It compares the distributions of the realised schedules (red) to the distributions of schedules in the full choice set (gray). Different clusters within specific activity types have different line styles.

The distribution of work start times (Figure 2a) clearly shows the two peaks in the realised set as already discussed in Section 4.2. It also shows how the desired start times for both work activity clusters are derived. The morning peak has a mean desired start time x_a^* of 7.4, while the peak for the afternoon activity is distributed around a mean x_a^* of 13.2 (see Table 2). The standard deviation in the morning is higher than in the afternoon peak. The grey distribution shows how the start times are distributed in the choice set. The strong peaks are still visible, but reduced towards unlikely start times because of the random alternatives. The random alternatives seem to have a tendency towards very early start times, which could be improved with a more elaborated algorithm to generate unlikely alternatives. Figure 2b illustrates the distributions for the total duration budget for participating at work activities. While the realised set has a strong peak around 9.5 h, the full choice set is biased towards smaller work durations.

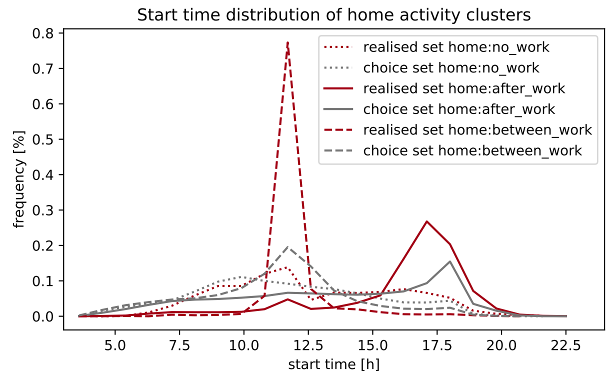
Similar observation are made for both home and leisure activities. The distributions in the full choice set have smaller, but still visible peaks compared to the realised set. Therefore, we make the assumption that requirements of a competitive choice set are fulfilled as they are biased towards likely schedules with certain number of low-probability



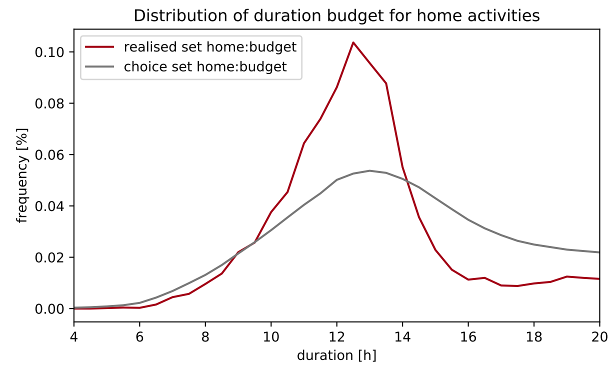
(a) work start time distribution



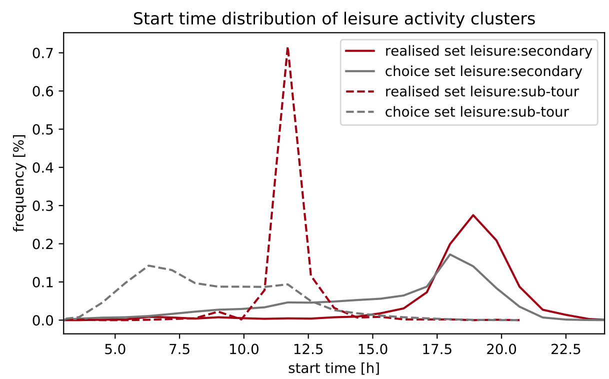
(b) work duration budget



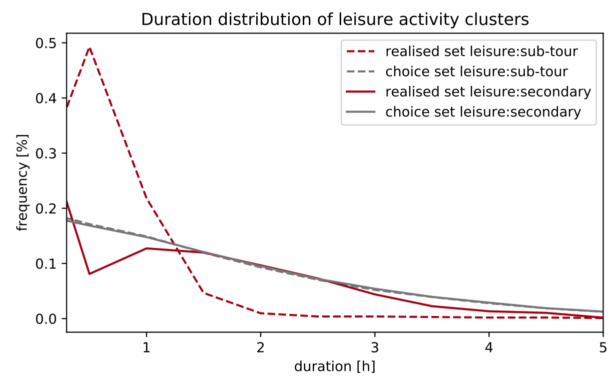
(c) home start time distribution



(d) home duration budget



(e) leisure start time distribution



(f) leisure duration distribution

Figure 2: Distributions for realised schedules compared to distributions of schedules in the choice set for selected activity clusters.

schedules as well.

4.4. Flexibility parameters

Having defined a choice set (see the previous Section 4.3), the activity types and their corresponding desired timings x_a^* and durations τ_a^* (see Table 2), we can now apply Equation (1) to each alternative in the choce 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, the less flexible the person is to deviate 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: 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:morning	-0.615	-0.436		
work:afternoon	-0.406	0		
work:duration budget			-0.022	0
leisure:sub-tour	-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:sub-tour	-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
accompany	-0.175	0	0	-2.120
business	-0.135	-0.267	-0.361	-0.863
education	-0.246	-0.376	-0.376	-0.939
other	-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:between work	-2.040	-0.929		
home:after work	-0.073	-0.596		
home:no work	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

Table 3 shows the resulting flexibility parameters. The complexity of the nature of this problem raises high computational times. On a big machine with 40 cores, it took more than 13 hours to estimate the parameters. In terms of being early or late, the lunch activities (either at home or in a restaurant) are the most rigid. This seems to be a very fixed point in time in the schedule of a full-time worker. Also, starting work in the morning is a given timing compared to other activity types. Interestingly, the estimation indicates that people are very flexible in extending their lunch break and starting work in the afternoon. But since full-timers have a minimum work duration budget and want to be back home in the evening after work, they have to go back to work after the lunch break. The flexibility parameter for being late at home after work confirms the fact that people want to be back home on time. However, it is notable that the dusk activity is not penalised for being late. Secondary activities are generally flexible, with some being more flexible than others. Especially leisure activities are surprisingly flexible, while shopping activity have more rigid timings.

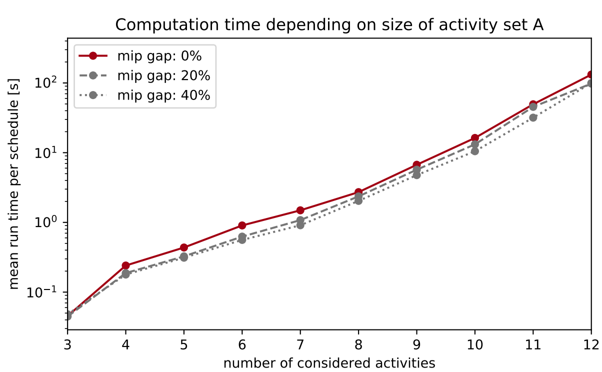
In terms of deviating from the desired duration of an activity, some activity types are very rigid. The leisure activity in a sub-tour (between the work activities, i.e. lunch) has a very specific desired duration. Not being long enough at work is penalised by a very small dis-utility. Accompany (e.g. dropping of a child at the kindergarten) is a very short activity, and people are not willing to extend it. Very interesting is the total budget to spend time at home. Full-time workers are fine with being at home for a very short amount of time. However, they don't want to spend too much time at home, meaning that they actively want to participate at other activities outside of the home for a certain time of the day.

4.5. Implementation of the optimisation program

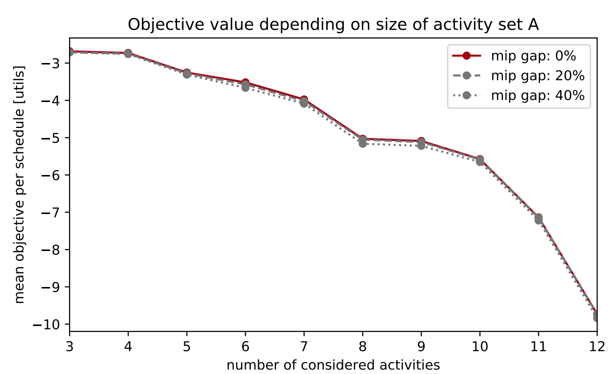
For this case study, we implemented the mixed-integer linear program as described in Sections 3.3.2 and 3.3.3 in the framework OR-Tools[1] for Python. The general library makes it possible to easily test the performance of the optimisation program by applying multiple solvers. To forestall the license burden for practitioners, our goal was to use a non-commercial solver. For our optimisation framework, the SCIP-solver (Achterberg et al., 2008) turned out the be the most efficient in terms of computational time.

Figure 3a shows the performance depending in the size of the set of activities $\mathcal{A}$ and the optimality gap defined in the solver. Since the y-axis (average run time in seconds) is logarithmically scaled, the time to optimise one schedule increases exponentially with number of activities in $\mathcal{A}$. For $|\mathcal{A}| = 5$, the average time to find an optimal solution (mip gap: 0 %) per schedule is 0.43 seconds. In the case of $|\mathcal{A}| = 10$, the average solving time is 16.27 seconds. An optimality gap of 20 % reduces the average computation time by around 25 % and a mip gap of 40 % by around 30 %, respectively. We implemented a possibility to apply parallel computing (using the library Ray[2]), which reduces the computation time efficiently by the amount of cores.

The influence of the optimality gap on the objective value (total schedule utility) is depicted in Figure 3b. Whilst the objective value clearly decreases with an increasing size of $\mathcal{A}$, the influence of the mip gap is very small. However, the impact on the results and the validation of the schedules has to be investigated in future work.



(a) Mean computation time per schedule



(b) Mean objective value per schedule

Figure 3: Mean computation time and objective value per schedule depending on the size of the activity set for different optimality gaps.

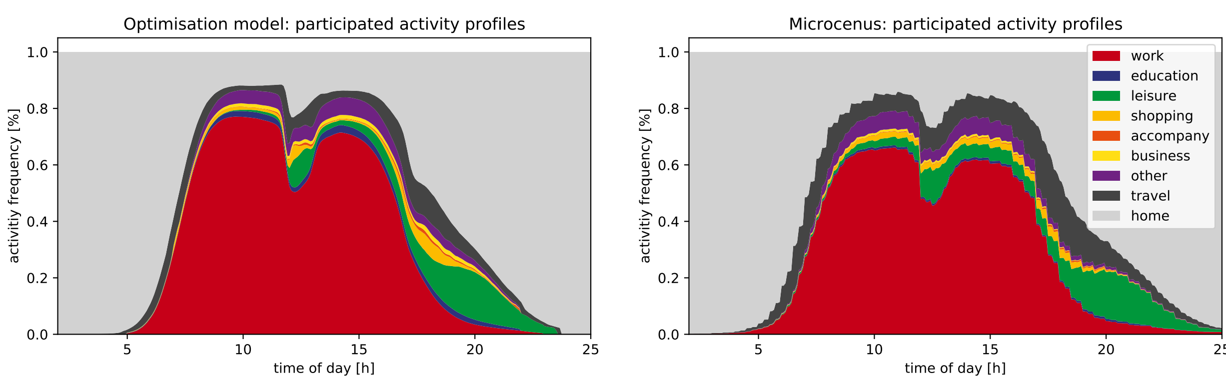
[1] `https://developers.google.com/optimization`, last accessed: 29.08.2021
[2] `https://ray.io/`, last accessed: 29.08.2021

4.6. Optimised schedules

Having all the inputs including flexibility parameters (Table 3) and an implementation of the optimisation program (previous Section 4.5), the framework is applied to each full-time worker as part of the synthetic population for the city of Lausanne, Switzerland. Firstly, we generate an activity set $\mathcal{A}$ for each individual, as described in Section 3.3.1. For this purpose, we use a sequence of existing discrete choice models (Scherr et al., 2020b) to simulate the decisions of activity participating, tour types and considered destinations. For each activity $a \in \mathcal{A}$, we make a random draw from a normal distribution to assign a desired timing x_a^* and duration τ_a^* for each activity based on its type and cluster based on the parameters as given in Table 2. The proposed framework then targets to resolve time conflicts which arise from overlapping desired timings and durations while considering the travel times for the trips between two consecutive activities. The result is a realised schedule for the individual of the synthetic population.

A minority of the full-time workers make the decision of not participating at any activity outside of the home during the day. After removing them from the optimisation program, 46’970 schedules have to be simulated. We are defining the optimality gap to be 0 % to reach a guaranteed optimum. Each individual chooses 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 *car as a driver* is also possible. The mode has to remain the same throughout a 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 4 depicts the final results of the time-conflict resolution for each full-time worker in Lausanne. It compares the results of the model (left Figure 4a) to the observed activity profiles in the Swiss mobility and transport microcensus (right Figure 4b). It shows the frequency of individuals who are participating at certain activity type at a given time of the day. Overall, the model fits the curve of the microcensus very 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 increase in participating at 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. In general, full-time workers have very few time to participate at secondary activities during the week and besides working, it is essential to have some time at home. Also, travelling consumes a (surprisingly) small percentage of the day since there are only few persons who commute long distances.



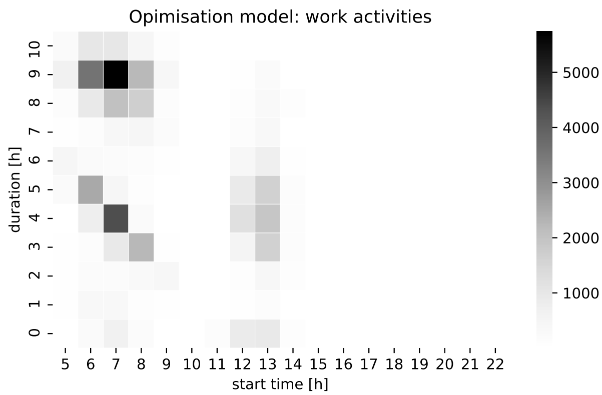
(a) Simulated activity profiles in the optimisation model

(b) Observed activity profiles in the microcensus

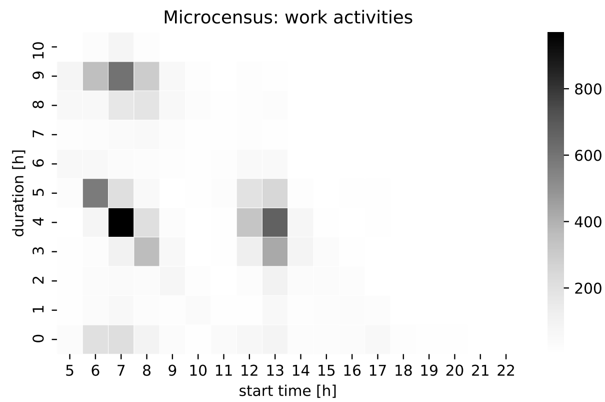
Figure 4: Activity profiles of the simulation using the proposed optimisation framework compared to the Swiss microcensus.

The proposed framework targets to capture the interactions in the temporal dimension. The dimension mainly includes the choices of activity start times and durations. In Figure 5, we show the correlation of start times and durations for the work and the leisure activities. The darker the area, the more observations are made within this time frame. Also, we compare the realised schedules of the optimisation model to the observed schedules in the microcensus. Again, there is a very nice fit between the simulation and the empirical observation. Work activities are separated into two different types: (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.

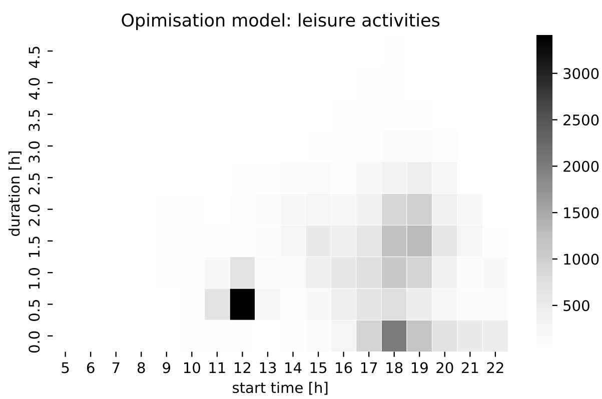
The leisure activities (Figure 5c and 5d) are also captured well by the model. The distribution shows a peak at lunch time for around 30 to 60 minutes. In the evening, both short and longer leisure activities are taking place, and both types of leisure activities are well represented in the model. During the day, the model underestimates the participation at very short leisure activities.



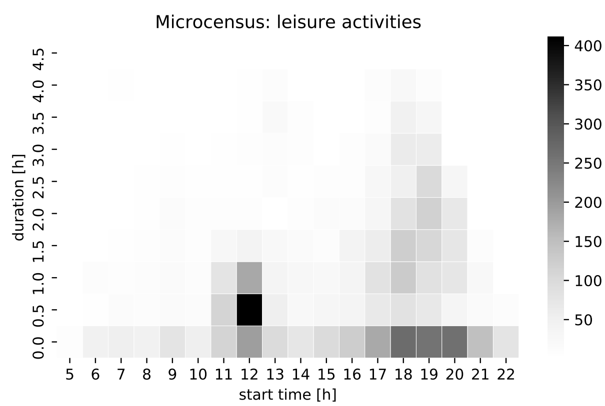
(a) Simulated correlations for work activities in the optimisation model



(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 5: Correlation between start times and durations for selected activity types in the proposed optimisation framework compared to the Swiss microcensus.

5. Conclusion and future work

In contrast to other choice dimensions in the process of activity-based scheduling such as activity participation or destination choice, the temporal dimension is largely unexplored. This work fills the named gap by proposing a framework that represents time as a continuous variable and that expresses temporal decisions with behavioural parameters. Our activity-based scheduling model combines an optimisation framework for 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 considered destinations). For each individual, it resolves 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 derived utility.

This presented approach has three main advantages over existing activity scheduling approaches:

1. The concept of utility-maximising individuals in combination with a continuous representation of time makes the framework fully responsive to disruptions in the temporal dimension. It is hence able to predict the impact of e.g. having more flexibility to start with the work activity.

2. Considering the interactions between all temporal decisions simultaneously provides a holistic overview over the temporal dimension. As a result, the proposed model is able to explain the correlation between activity timings and durations.

3. Coupling the optimisation framework with existing discrete choice models allows for computational efficiency and for city-scale simulation in reasonable time. Also, it simplifies the process of calibrating parameters based on real-world observations for practitioners.

We conduct a case study for the city of Lausanne and present a working solution for each component of the framework as shown in Figure 1. The application of the presented simulation framework to a synthetic population reliably reproduces real-world observations from the Swiss mobility and transport microcensus (Federal Statistical Office, 2017). For the calibration of the flexibility parameters, we demonstrate an estimate routine that is applied to the group of full-time workers as observed in the microcensus. Using these parameters, we then resolve time conflicts for each individual as part of a synthetic using a mathematical program that maximises the utility of the whole schedule.

We identify the following paths for future work:

- *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. In the future, advanced techniques based on a Metropolis Hastings algorithm might provide a way to find a competitive choice set for a reported schedule from scratch.

- *Parameter estimation*: The estimation of the parameters is based on fixed desired activity timings and durations. These are found by manual and visual investigation of the empirically reported schedules. It would be interesting to apply distributed values in the future. The distribution type includes the normal distribution but also other types of distributions like Weibull, especially for the activity durations. Also, an investigation of the robustness of the parameters depending on the size of the choice set is part of future work.

- *Generation of activity sets*: Activity participation is not included in the optimisation program to reduce the size of the activity set and computational complexity. A relaxation of this constraint while still having a reasonable computation time would give an more holistic view of the activity-scheduling process.

- *Optimisation program*: The run time is a bottleneck for large-scale applications. The impact of less optimal solutions on the validation results by increasing the optimality gap should be explored. Also, different ways of setting up the program, e.g. as a Constraint Program, might help to reduce computation time.

- *Travel times and mode choice utilities*: Time-depending travel times would be an enhancement of this framework in the future. Also, an estimation routine to find mode choice parameters depending on various level of service indicators would be a substantial contribution.

- *Coupling with network assignment*: Transport planners typically assess policy studies based on network loads. For this purpose, the optimised schedules must be fed into a network assignment, e.g. using the software MATSim.

Acknowledgements

This work is part of the collaborative project *Optimization of individual mobility plans to simulate future travel in Switzerland* between the Swiss Federal Railways (SBB) and EPFL, supported by the Swiss Innovation Agency (Innosuisse) (Project number: 2155006680).

A special thanks goes to the management of SBB – Patrick Buetzberger and David Henny – who support the investment of many working hours into this project. Also, the authors wish to thank the advisory board of the project, namely Michel Bierlaire, Wolfgang Scherr, Patrick Buetzberger, Grégoire Vionnet, Rolf Moeckel and Allister Loder.

References

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

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

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

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

Ballis, H., 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., Timmermans, H., 2010. Implementation framework and development trajectory of FEATHERS activity-based simulation platform. Transportation Research Record 2175, 111–119.

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

Bierlaire, M., 2020. A short introduction to pandasbiogeme. A short introduction to PandasBiogeme .

Bodenmann, B., Bürki, P., Phillipp, C., Bernhard, N., 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.

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

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

Briem, L., Mallig, N., 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., Gliebe, J., 2015. Activity-based travel demand models: A primer. SHRP 2 Report S2-C46-RR-1.

Davidson, B., Vovsha, P., Freedman, J., 2011. New advancements in activity-based models, in: Australasian Transport Research Forum, Citeseer. pp. 28–30.

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

Federal Statistical Office, 2017. Population's transport behaviour 2015. Key results of the mobility and transport microcensus. Neuchâtel, Switzerland.

Hilgert, T., Heilig, M., Kagerbauer, M., Vortisch, P., 2017. Modeling week activity schedules for travel demand models. Transportation Research Record 2666, 69–77.

Hillel, T., Pougala, J., Manser, P., Luethi, R., Scherr, W., 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., 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., 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. pp. 19–02409.

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

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

Joubert, J.W., 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., Cools, M., 2015. Characterizing activity sequences using profile Hidden Markov Models. Expert Systems with Applications 42, 5705–5722.

Miller, E.J., Roorda, M.J., 2003. Prototype model of household activity-travel scheduling. Transportation Research Record 1831, 114–121.

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

Pougala, J., Hillel, T., Bierlaire, M., 2021. Capturing trade-offs between daily scheduling choices .

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

Recker, W., Duan, J., Wang, H., 2008. Development of an estimation procedure for an activity-based travel demand model. Computer-Aided Civil and Infrastructure Engineering 23, 483–501.

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

Rizopoulos, D., Esztergár-Kiss, D., 2020. A method for the optimization of daily activity chains including electric vehicles. Energies 13, 906.

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

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

Scherr, W., Manser, P., 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., 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, 152–172.

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

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

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