Personalized multi-activity scheduling of flexible activities

Sergio Arturo Ordóñez Medina

Future Cities Laboratory **July 2015**

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Future Cities Laboratory

Personalized multi-activity scheduling of flexible activities

Sergio Arturo Ordóñez
Medina
Future Cities Laboratory
(FCL)
Singapore ETH Centre (SEC)
138602 Singapore
phone: +65-9298 7627
fax: +41-44-633 10 57
ordonez@ivt.baug.ethz.ch

July 2015

Abstract

Understanding how the individuals from a certain city or region plan their activities is very useful for many applications, such as travel demand modelling, land use planning and redevelopment, or market research and analysis. Accurate multi-activity scheduling is a non-trivial problem for its number of dimensions, and the amount of information involved in human decisions. The concept of fixed and flexible activities allows to study this problem at two levels. It can be assumed that mandatory activities like rest, work or study, are already prearranged, and non-mandatory activities like shopping, eating or running errands are planned on-the-fly. This paper proposes an algorithm that emulates personalized multi-activity scheduling of flexible activities within well defined time windows. The algorithm uses a mental map composed by an *Activity agenda* and a *Set of known places* obtained from real data. It calculates the maximum utility activity-trip chain according to temporal and spatial conditions of the decision. It solves the number of the activities to be performed, the sequence, the locations, the start times and the activity durations. Transportation modes of the trips between consecutive activities are also predicted. This method (i) doesn't prioritize or fix any scheduling dimension, (ii) uses common available data (i.e. travel surveys and land use datasets) as its input, (iii) generates personalized solutions, and (iv) can be configured to schedule flexible activities of a full population in tractable times. In order to show the efficiency of this approach flexible activity patterns from Singapore were extracted from a travel survey carried out in 2012, and the method was tested with a fraction of the survey observations which were not used for the estimation of the models. Five types of flexible activities were included. Similar shapes of general travel time and activity duration distributions between real and predicted activity-trip validate the potential of the proposed algorithm. Finally, the method was used for within day re-planning of a synthetic population using MATSim, a well known agent-based transport simulation platform. Results show that the algorithm is computationally feasible for large-scale scenarios.

Keywords

Activity scheduling, Utility maximization, Location choice, Secondary activities, Time geography

1 Introduction

Understanding how the individuals from a certain city or region plan their activities is very useful for many applications, such as travel demand modelling, land use planning and redevelopment, or market research and analysis. These models become a powerful tool when they can be used to perform thousands or millions of predictions leading to macroscopic insights from microscopic changes. Accurate multi-activity scheduling is a non-trivial problem for its number of dimensions, and the amount of information involved in human decisions. These decisions depend on intrinsic variables related to the behaviour of the person who is planning, and on extrinsic variables related to the state of the rest of the world where the decision is taken. As mentioned by (Feil, 2010) the fundamental problem of activity scheduling is its combinatorial complexity due to its number of dimensions (activity durations, locations, number of activities, activity types, activity sequence, etc.). The modeller needs to develop solutions which allow solving the problem in tractable time.

Complex methods and algorithms have been proposed to model these multi-activity scheduling decisions with spatio-temporal restrictions (Arentze and Timmermans (2000), Doherty et al. (2002), Miller and Roorda (2003), Arentze and Timmermans (2004), Feil (2010)). Other methods are based on addressing only the next activity (Kuhnimhof and Gringmuth (2009), Arentze and Timmermans (2009)) or performing continuous planning without taking into account activity locations (Märki et al. (2014)). Many of these models are based on strong assumptions, such as a fixed number of the activities to be scheduled, or fixed activity durations, producing restricted results. One of the most common assumptions imposed in some heuristic or probabilistic models, such as Miller and Roorda (2003), Arentze et al. (2010), Kuhnimhof and Gringmuth (2009), is the prioritization of some scheduling dimensions during the decision process. That means, the scheduling process is carried out in a predefined and fixed order e.g. activity type -> location -> duration -> activity type -> etc. With this restriction, decisions in which for example a location is a priority cannot be modelled.

The concept of fixed and flexible activities has been used in several activity-based modelling studies (Arentze and Timmermans (2000), Miller (2005), Chen and Kwan (2012), Doherty et al. (2002)). These concepts allow studying activity scheduling at two levels. The first level is focused on the activity skeleton, which is composed of mandatory or fixed activities. The most important or primary human activities, like Home, Work and Study activities, belong to this level. Primary activities have been extensively studied by transportation modellers and urban planners (Hansen (1959), Small (1982), Vovsha and Gupta (2013), Ordóñez Medina and Erath (2013)). They try to find spatial patterns within geographical regions like cities, and temporal patterns during one day or longer periods of time, for example weeks. The second level is focused on non-mandatory or flexible activities. Most of the secondary human activities belong to this level. The challenge at this level is the combinatorial variability of these activities, which makes it more difficult to find a small number of patterns that can explain flexible activity decision making. For example, when analysing a travel survey containing daily activity chains of 1% of the population of a city, there are more than 900 combinations of activities. By contrast, there are just 40 combinations of primary activities after removing secondary activities. If primary activities in this dataset are just categorized as Home and Work (i.e. Studying is a type of work), only 10 combinations can be found. Thus, it seems a good strategy to study separately fixed and flexible activities to face this problem. This study is focused on flexible activities scheduling, assuming that fixed activities are already scheduled.

This work proposes a multi-activity scheduling method for flexible activities which (i) doesn't prioritize or fix any scheduling dimension, (ii) uses common available data (i.e. travel surveys and land use datasets) as its input, (iii) generates personalized solutions, and (iv) can be configured to schedule flexible activities of a full population in tractable times (although its complexity grows with the length of the time window).

Each prediction is personalized, i.e. it uses socio-demographic characteristics and matched behavioural parameters of the decision maker. In the first step of the prediction, this information is used to estimate an *Activity agenda* of flexible activities and a set of possible activity locations for them (choice set). An *Activity agenda* is a collection of possible activities to be performed with a range of possible durations and frequencies as proposed by (Axhausen, 2006). These two elements model a simple mental map of the decision maker. This first step won't be fully described in this paper, but details can by found in Ordóñez Medina (2015). In the second step a graph-based utility maximization algorithm takes the *Activity agenda*, and the *Set of known locations*, the locations where the decision maker plans to start and end his/her activities, and the available time budget, to find an optimal flexible activity-trip chain (the prediction). This utility maximization algorithm is based on (i) the time geography concepts introduced by (Hägerstrand, 1970) and developed by (Wilson, 2008), and (ii) the idea of using spatio-temporal graphs for activity-trip chain optimization proposed by (Arentze and Timmermans, 2004). This algorithm also uses the functions described by (Charypar and Nagel, 2005) to measure activity utilities and trip disutilities. Information about car-availability and dynamic travel times can be included in this algorithm, as they would affect the choices of location and transportation mode. In simple words, if the method is told who the decision maker is, where he/she is, how much free time he/she has and where he/she has to be after that time, the model predicts a possible and detailed flexible activity-trip chain which on average corresponds to the patterns of the city or region, and maximizes his/her utility.

Next section describes the multi-activity scheduling algorithm in detail, and results of basic tests with controlled inputs are presented. Then, to show the applicability and efficacy of this approach with real data, flexible activity patterns for Singapore were extracted from the household interview travel survey carried out in 2012 (HITS 2012). This survey contains reported motorized trips of 1% of the population during one day. A dataset with information of more than one hundred thousand activity locations, and estimated dynamic travel times from a large-scale agent-based transport model implemented for Singapore (Erath et al. (2012)) were also used for more realistic calculations. Five flexible activity type (eating, shopping, social activities, running errands, recreation) were included in the model. Activity duration and travel time distributions were also extracted from the travel survey. Thus, in the third section results of testing the method with a fraction of the survey observations which were not used for the estimation of the models, are presented and analyzed. In the section four, to assess the computation feasibility and efficiency of the approach, the method was used for within day re-planning of a synthetic population using MATSim, a well known agent-based transport simulation platform for large-scale scenarios. Finally some conclusions and future work are discussed in the last section.

2 Spatio-temporal network algorithm to calculate optimal activity-trip chains

In this paper a multi-activity scheduling algorithm is proposed. It finds an approximately optimal activity-trip chain for a given time window, according to its activity perform utilities and its traveling dis-utilities. If a person has more than one free time period during one day, this process must be run more than once. Next, the proposed algorithm is explained in detail and tests with controlled inputs are presented.

2.1 Input information

The inputs include the origin o and destination d, start time st and latest end time et, an *Activity agenda* and a *Set of known places*.

An *Activity agenda* G is a set of n activities intended to be performed by a person. However, these activities are not set in time (start time and duration) or space (location). The information related to each activity a_i is a typical duration d and a typical frequency f. These two attributes can be modeled as random variables, estimating probability distributions from observations. Introduced by (Axhausen, 2006), an *Activity agenda* provides a level of abstraction to model flexible activity decisions. It restricts the combinatorial problem from a universal set of possible activities to a controlled set of activities included in the agenda. Given an activity skeleton with fixed activities planned during a defined time period and time windows without planned activities, the agenda can be used to plan flexible activities during these free times. Models by (Nurul Habib and Miller, 2009) and (Nijland et al., 2012) also aim to generate personalized activity agendas with an econometric and a Bayesian approach respectively. In TASHA, an activity-based model proposed by (Miller and Roorda, 2003), they developed similar procedures for the proposed activity generation step.

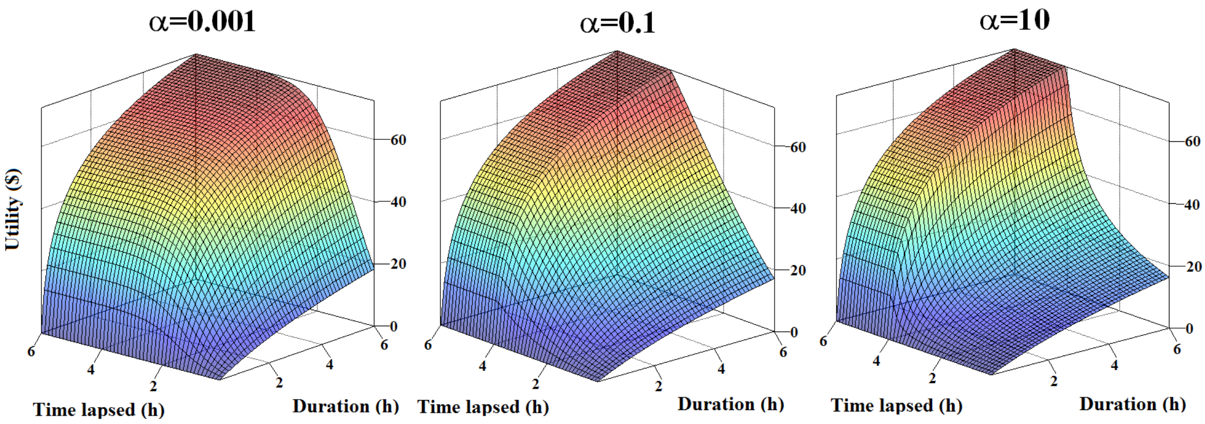
As this work also aims to resolve the locations where these flexible activities are performed, another concept must be introduced, that is the *Set of known places*. When a person plans flexible activities, it can be assumed that he/she only takes into consideration a limited number of activity locations. Known places is not the best name for this set, as a decision maker can know many places which he doesn't take into consideration or even is reluctant to travel to. Regarding the set of alternatives in a general choice, (Narayana and Markin, 1975) introduce 5 subsets: *Unawareness set*, *Awareness set*, *Inept set*, *Inert set* and *Evoked set*. *Unawareness set* and *Awareness set* are exclusive and, as the names imply, if the decision maker is aware of the alternative, it belongs to the *Awareness set*, or otherwise to the *Unawareness set*. *Inept set*, *Inert set* and *Evoked set* are subsets of the *Awareness set*. If the decision maker has a negative image of an alternative it belongs to the *inept set*, if the image is neutral to the *Inert set* and if it's positive to the *Evoked set*. It is therefore necessary to point out that in this paper the term *known places* only refers to the locations where the decision maker is willing to travel to, i.e. locations that belong to the *evoked set* in the decision maker mental map. Thus, a small set of m known places P is the second piece of input information given to this multi-activity scheduling method.

2.2 Spatio-temporal network definition

The main algorithm consists of recursively constructing a spatio-temporal network or directed graph (V, E) inside the space-time prism defined by the initial and final states (see Hägerstrand (1970)). Each node is defined by a geographic location and a time stamp. Nodes are only created at known places and at specific times according to a defined time bin δt, i.e. $\forall(p, t) \in V, p \in P \wedge t = st + k * \delta t$ with $k \in \mathbb{Z}$. This controls the size of the network. Trips and activities are represented by links of the network, i.e. $V = A \cup T$ where A is the set of activity-links and T is the set of trip-links. The nodes of an activity-link must have the same geographical location ($\forall l \in A | fromNode(l) = (pF, tF) \wedge toNode(l) = (pT, tT), pF = pT \wedge tF < tT$) and the nodes of a trip-link must have different geographical locations ($\forall l \in T | fromNode(l) = (pF, tF) \wedge toNode(l) = (pT, tT), pF <> pT \wedge tF < tT$). Each activity-link has also an activity type or purpose associated $type(l)$, and trip-links have a transportation mode associated $mode(l)$.

Utilities of the activity links are calculated using an extension of the activity utility function presented in Charypar and Nagel (2005). This version, presented in Equation 1 and illustrated in Figure 1, incurs a penalty if the same activity is performed more than once during a period of time T. Hence, the utility of an activity-link within an activity-trip-path depends on previous activity-links of the path from the origin. This restricts the information sharing between paths with common nodes and the Dijkstra algorithm cannot be applied to simplify the network construction. Travel disutility functions from (Charypar and Nagel, 2005) were also employed

Figure 1: Activity performing utility as a function of the its duration and the time lapsed since the last time the same activity activity was performed.



for the travel link costs, and using travel times between known places.

$$
\begin{aligned}
&U(t_d, t_l) = \beta T_x \left(\frac{t_d + g(t_l)}{T_0 + g(t_l)} \right) \\
&g(t_l) = \begin{cases} \alpha e^{\gamma(T - t_l)} - \alpha, & \text{with } t_l < T \\ 0, & \text{with } t_l \geq T \end{cases} \\
&\gamma = \frac{1}{T} ln \left(\frac{T + \alpha}{\alpha} \right) \\
&\text{Thus: } g(t_l)t_l = \begin{cases} 0, & \text{when } t_l = T \\ T, & \text{when } t_l = 0 \end{cases}
\end{aligned}
\tag{1}
$$

Where t_d is the duration of the activity, T_x is the typical duration, T_0 is the minimal duration, β is the marginal utility of performing the activity, t_l is the time lapsed since the last time that activity was performed, $g(t_l)$ is its penalty function, α is a parameter to determine how strong is the applied penalty (see Figure 1), and γ is just a function of α.

2.3 Network construction algorithm

The process starts at the initial node (o, st), and the objective is to find the path to the final node (d, et) with the maximum utility. From a node all possible activity links and trip links can be created. When a new node is created during the recursive algorithm, the path from the initial node is saved. When any previously created node is reached again through a new path from the origin, just the path with the maximum utility is maintained. The simplified pseudocode of this recursive algorithm is presented in Algorithm 1 and Algorithm 2.

Algorithm 1 Maximum utility path

```
1: constants:
2: time_bin ← Fixed time bin
3: procedure getMaximumUtilityPath
4: parameters:
5:     startTime ← Start time
6:     startPlace ← Origin
7:     endTime ← End time
8:     endPlace ← Destination
9:     agenda ← Activity agenda
10:    places ← Set of known places
11:    modes ← Transportation modes
12: begin:
13:    global path ← null.
14:    currentPath ← empty.
15:    firstNode ← createNode(startPlace, round(startTime, time_bin)).
16:    addNodePath(currentPath, firstNode).
17:    addLinkNode(currentPath,round(endTime,time_bin),endPlace,agenda,places,modes).
18:    return path.
```

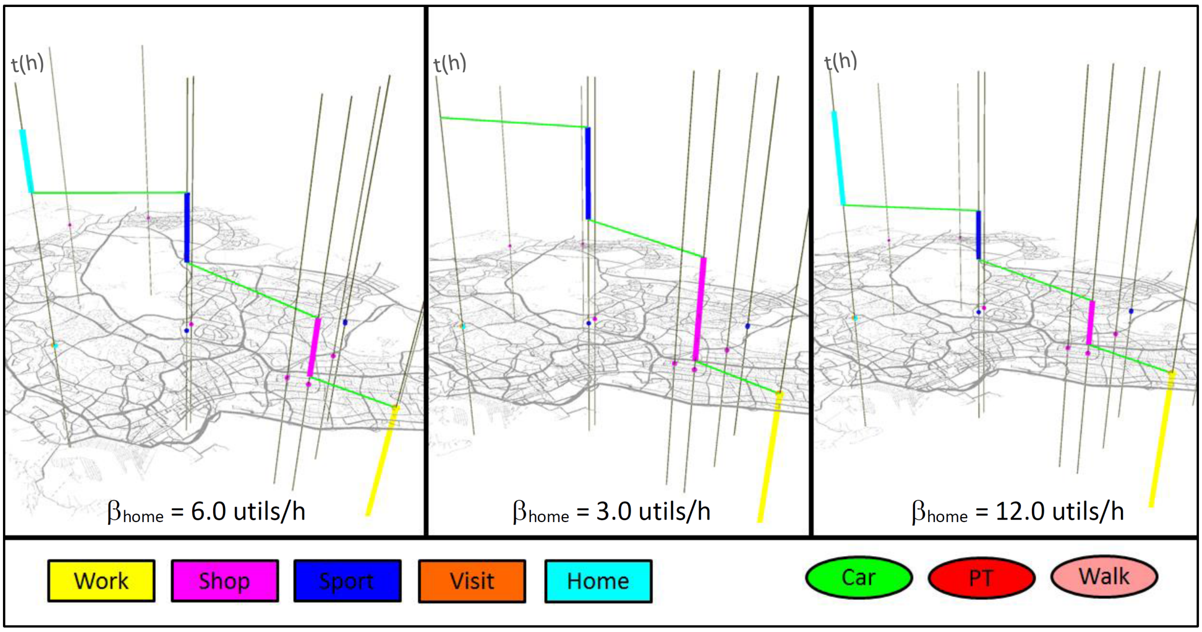
Algorithm 1 prepares the input data, the output variable ($path$) and calls the recursive Algorithm 2 at line 17. Algorithm 2 checks firstly if the current node is the final node (same time and geographic position) and checks if this complete path ($currentPath$) has a better utility than the current maximum utility path ($path$). If so, the maximum current path is updated at line 13. If the current node is not the final node, the network can grow in two ways: (i) staying at the same location creating activity links and (ii) moving to another location creating trip links.The loop starting at line 17 creates all possible activity links and continue the process recursively at line 24. The loop starting at line 27 creates all possible trip links and continue the process recursively at line 34. The *Activity agenda* controls the type of the activities and their duration at line 21 while the *Set of known places* controls the destinations at line 27. Another key point in the algorithm is the time discretization using the constant $time\_bin$. Continuous times are rounded according to this value at the line 15 of Algorithm 1 and at lines 16, 30 of Algorithm 2.

The implemented recursive algorithm is more restrictive than the presented pseudocode. It cuts out a branch when the utility of the current path plus an expected utility according to the available time and current location doesn't exceed the current maximum utility. Transportation modes are

Algorithm 2 Add next link and node

```
procedure addLinkNode
parameters:
    currentPath ← Path from the origin to the current node
    endTime ← End time
    endPlace ← Final destination
    agenda ← Activity agenda
    places ← Set of known places
    modes ← Transportation modes
begin:
    currentNode ← lastNode(currentPath).
    if currentNode.place == endPlace AND currentNode.time == endTime then
        if path == null OR utility(currentPath) > utility(path) then
            path = currentPath.
    else
        lastTravelTime ← bestTravelTime(currentNode.place, endPlace, modes).
        time ← round(endTime − lastTravelTime, time_bin).
        while time > currentNode.time do
            node ← createNode(currentNode.place, time).
            duration ← time − currentNode.time.
            for activity ∈ currentNode.place.activities do
                if possibleActivityDuration(agenda, activity, duration) then
                    createActivityLink(currentNode, node, activity.type).
                    addNodePath(currentPath, node).
                    addLinkNode(currentPath,endTime,endPlace,agenda,places,modes).
                    removeNodePath(currentPath, node).
            time ← time − time_bin.
        for place ∈ places do
            for mode ∈ modes do
                travelTime ← travelTime(currentNode.place, place, mode).
                time ← round(currentNode.time + travelTime, time_bin).
                node ← createNode(place, time).
                createTripLink(currentNode, node, mode).
                addNodePath(currentPath, node).
                addLinkNode(currentPath,endTime,endPlace,agenda,places,modes).
                removeNodePath(currentPath, node).
```

Figure 2: Sensitivity of the highest utility activity-travel path to the marginal utility of performing the home activity (β_{home}). Time window from 17:00 to 23:00.

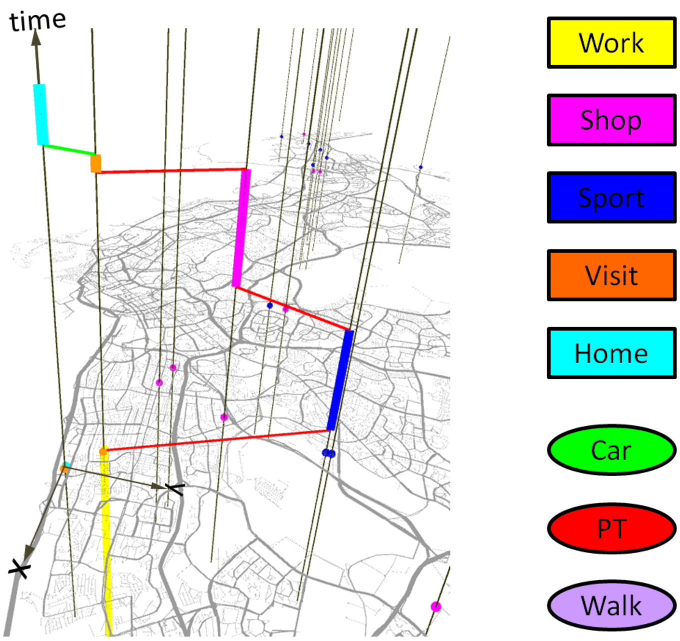


also better modeled, saving the location of the car (if it was available in the beginning). When the mode of a trip link is not car, the last car location is saved. Thus, the car mode is not available for the next locations till the saved car location is reached again. The implemented algorithm also ensures that no path has two consecutive trip links. Maximum number of activities, or minimum activity duration can also be defined.

Thus, when the construction process finishes, the path with the maximum utility to the final node is found automatically. This path models an optimal and fully characterized activity chain of flexible activities, including the number of the activities, the order, the start times, the duration of the activities and the places where these are performed. Optimal transportation modes of the trips between the locations are also part of the solution. Figure 2 illustrates an example of shortest paths of activities and trips found by this algorithm. Known places of two types of flexible activities (shopping and practicing sports) are represented by colored dots on the map. After varying the marginal utility of performing the home activity, utility locations, number of activities and sequence remain the same, but activity durations vary according to how valuable is to go back home.

Figure 3 shows another characteristic of the proposed algorithm mentioned before. In this test the utility of an activity-trip path is penalized if the car mode is available but the vehicle doesn't end in the final destination. The algorithm tracks the location of the vehicle. Besides, working facilities are intentionally located near home, and the car trips are modeled 250% more expensive

Figure 3: Activity-trip chain scheduled by the proposed algorithm, with a high price for car mode and working facilities located near home. Time window from 17:00 to 23:00.



than public transport trips. For these reasons the algorithm returns an activity-trip path in which the decision maker prefers to travel by public transport to perform activities near his/her work location, return to his/her work location for the car and perform a short trip from there to home.

3 Application of the algorithm using real data from Singapore

To apply the proposed multi-activity scheduling algorithm in real scenarios it's necessary to estimate *Sets of known places* and *Activity agendas* (*Mental maps*) from real data. Figure 4 summarizes the processes developed (bright boxes), and data used (dark boxes) for this purpose in the Singapore scenario. The multi-activity scheduler is represented by the light box on the bottom-left corner. This method uses commonly available datasets represented by the two dark boxes on the top: (i) the spatial information (transportation network or travel times, and activity facilities information) on the left and (ii) a travel or activity survey on the right. Specialized surveys could also be carried out for better estimations.

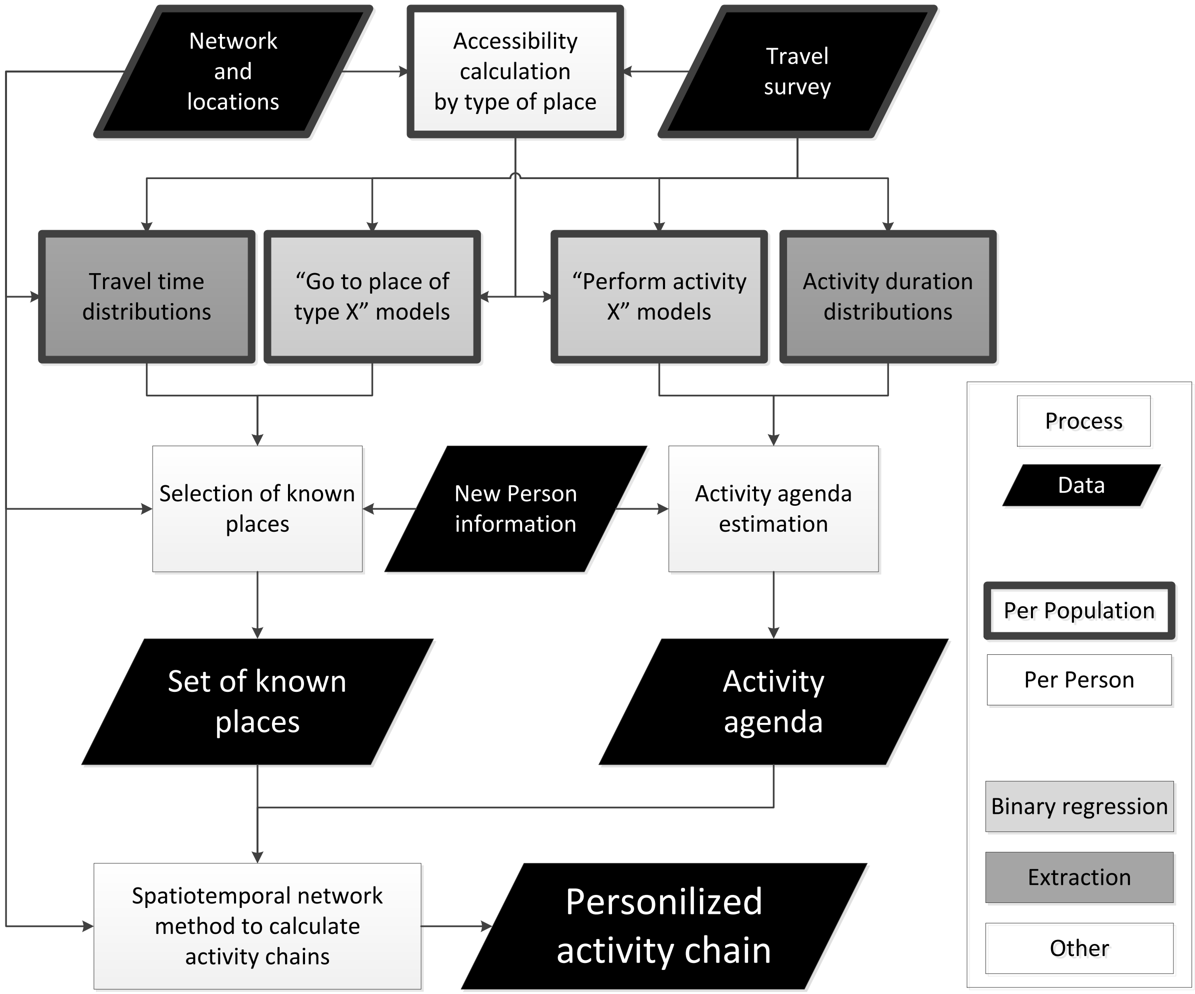
The boxes with a thick border on the upper half of Figure 4 represent data and processes applied to a population dataset. Then, to extract temporal and spatial activity patterns of the population, these processes only need to be executed once. In contrast, the boxes at the bottom half represent the data and processes which must be applied for every activity scheduling, using the results from the population processes. The *"Go to place of type X" models* are binary logistic models to predict how likely it is for a person to visit a certain type of place based on his/her socio-demographic characteristics. The *"Perform activity X" models* are similar binary logistic models, but to predict how likely it is for a certain decision maker to perform a certain activity. These two sets of models, along with the extraction of activity duration and travel time distributions, generate the input parameters for the *Selection of known places* and the *Activity agenda estimation*. Next, these models design to extract *Sets of known places* and *Activity agendas* (*Mental maps*) are summarized; for more details please check Ordóñez Medina (2015).

3.1 Mental map extraction models

The binary logistic models are estimated with 90% of the respondents from a travel survey (training set), a database of activity facilities, and estimated multi-modal travel times between these facilities. The remaining 10% of the observations are used for validation in the next section. The following personal attributes from the Singaporean *(*Household interview travel survey) carried out in 2012 (HITS 2012) were included:

- Age
- Gender
- Car availability
- Ethnicity: The reported ethnicities are Chinese, Indian, Malay and others.
- Accessibilities: For each type of place, a measurement of accessibility was calculated

Figure 4: Processes for multi-activity scheduling using common datasets.



from each primary activity locations of every person (i.e. residence, workplace, school, university, etc). The general accessibility of a person to a place type is the maximum accessibility among all his/her primary locations. Thus if for example, the workplace of a person is very accessible to shopping places, but the home residence is not, the shop-accessibility measurement for that person is still high. Each measurement was calculated with the following formula: $Acc(p_i, t) = \sum_{p_j \in P(t)} (A(p_j)^{\beta}) e^{\alpha T(p_i, p_j)}$. In this equation t is a type of place, $P(t)$ is the set of facilities of type t, $T(p_i, p_j)$ is the travel time between the facilities p_i and p_j, and $A(p_j)$ is the attraction level of the jth facility. Finally, α and β are parameters calibrated to optimize the linear relation of the accessibility measurement with the duration of activities performed at each place type.

- Size of household: Number of people in the household
- Role in the household: Four roles were defined. A *main* role is assigned to the person with highest income in the household. A *partner* role is assigned to the person with most similar age to the person with the main role. Members younger and older than the main

Place type	Purpose	Age	Gender	...	Outcome	Description
Park	rec	40	Female	...	0	Rec is related to community centre
Community centre	work	35	Male	...	N/A	Work is not a flexible activity
Shopping place	shop	45	Male	...	N/A	Shop is not related with community centre
Community centre	social	35	Male	...	1	Social is related with community centre

Table 1: Some examples of the classification of trips for the *"Go to community centre" model*.

and partner persons are assigned to the *younger* and *older* roles respectively.

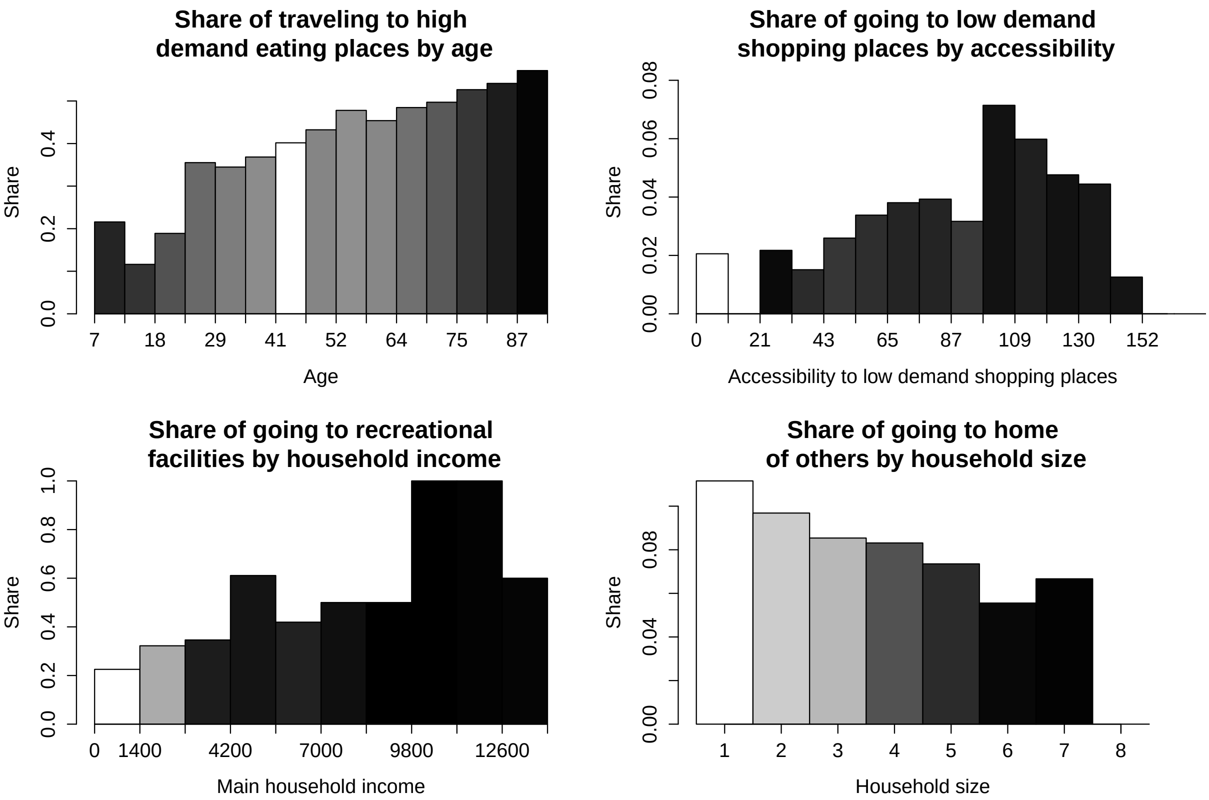
- Income: Net income of the person
- Main income: The income of the person with highest income within the household
- Home time: The time the person stays at home during the day
- Work time: The time the person spends working or studying during the day

The place types included in the method are recreational facilities, parks, community centres, homes of others, high-demand shopping places, low-demand shopping places, high-demand eating places, and low-demand eating places. The categorization of a place as high demanded or low demanded is determined by its size and/or the number of trips reaching that place according to the travel survey. One place (location or building) can be classified into multiple types. Flexible activities included are eating, shopping, social meetings, recreational activities, and running errands. A binary logistic model was developed for each type of place, as well as for each flexible activity. With 8 place types and 5 activities, a total of 13 models were estimated. Each trip to a flexible activity in the travel survey can represent an observation (a set of input variables and an outcome) for these models. For each reported trip the activity purpose and the place type of his/her destination are known. The table 1 shows some examples of the classification of trips for the *"Go to community centre" model*. Binary choices for the *"Perform activity X" models* are much simpler, if a person travels to perform a certain flexible activity during the reported period (e.g. one day), that person is counted as a positive observation for the corresponding activity model and vice versa.

3.1.1 "Go to place of type X" models

Figure 5 shows some of the dependencies of socio-demographic characteristics on travelling to or not to certain place types. Using these figures, 8 linear utility functions were formulated. They represent the influence of some of the mentioned personal attributes on travelling or not to each

Figure 5: Some dependencies on the willingness to travel or not to travel to places of different types, based on socio-demographic attributes. The height of each bar represents the fraction of the number of observations which the person goes to the corresponding place type, over the total number of observations in the corresponding category. The color of each bar represents the absolute number of observations in the corresponding category, wither color means more observations.



place type. *logistic regressions* were used to estimate the coefficients of these functions. For the sake of brevity table 2 summarizes the significance of the 11 socio-demographic attributes in the 8 *"Go to place of type X" models*. For the coefficient estimation, dummy variables were created for categorical variables while linear scaling was applied to continuous variables to limit their ranges to the interval [0, 1].

Activity	Age	Female	Car availability	Ethnicity	Accessibility	Household Size	Household role	Employed	Income	Time at home	Time at work	R^2
Rec	s				s			s			s	0.296
Park	s	s			s			s	s	s	s	0.536
CC	s				s					s		0.682
HoO					n/a	s		s	s	s		0.647
ShH	s		s	ocmi	s	s	ympo	s	s	s	s	0.207
ShL	s				s				s			0.816
EatH	s		s	imco	s	s		s	s	s	s	0.273
EatL			s		s					s		0.665

Table 2: Summary of the *"Go to place of type X" choice models* absolute impact. In this table, green represents a positive dependency while red refers to a negative one. Blue color means that the relation between the parameter and the outcome is not always increasing or decreasing. Deeper shades represent higher impact or higher absolute value of the utility parameter. The letter "s" means that the variable is significant according to the t-test. Yellow cells are significant categorical variables. The text in these cells represents the order from the most to the least significant category. For ethnicity "c" means Chinese, "i" refers to Indian, "m" is for Malay and "o" means other. For household role "m" is for the main role, "p" means the partner role, "o" stands for the older role and "y" is the younger role.

3.1.2 "Perform activity X" models

In a similar way to the *"Go to place of type X" models*, linear utility functions were modelled for each of the 5 "Perform activity" models to represent dependencies. Figure 6 presents some of the dependencies found in the travel survey. Table 3 summarises the priorities of the 11 attributes in the performance or non-performance of these 5 flexible activities. For the coefficient estimation, dummy variables were created for categorical variables and linear scaling was applied for continuous variables to fix their ranges to the interval [0, 1].

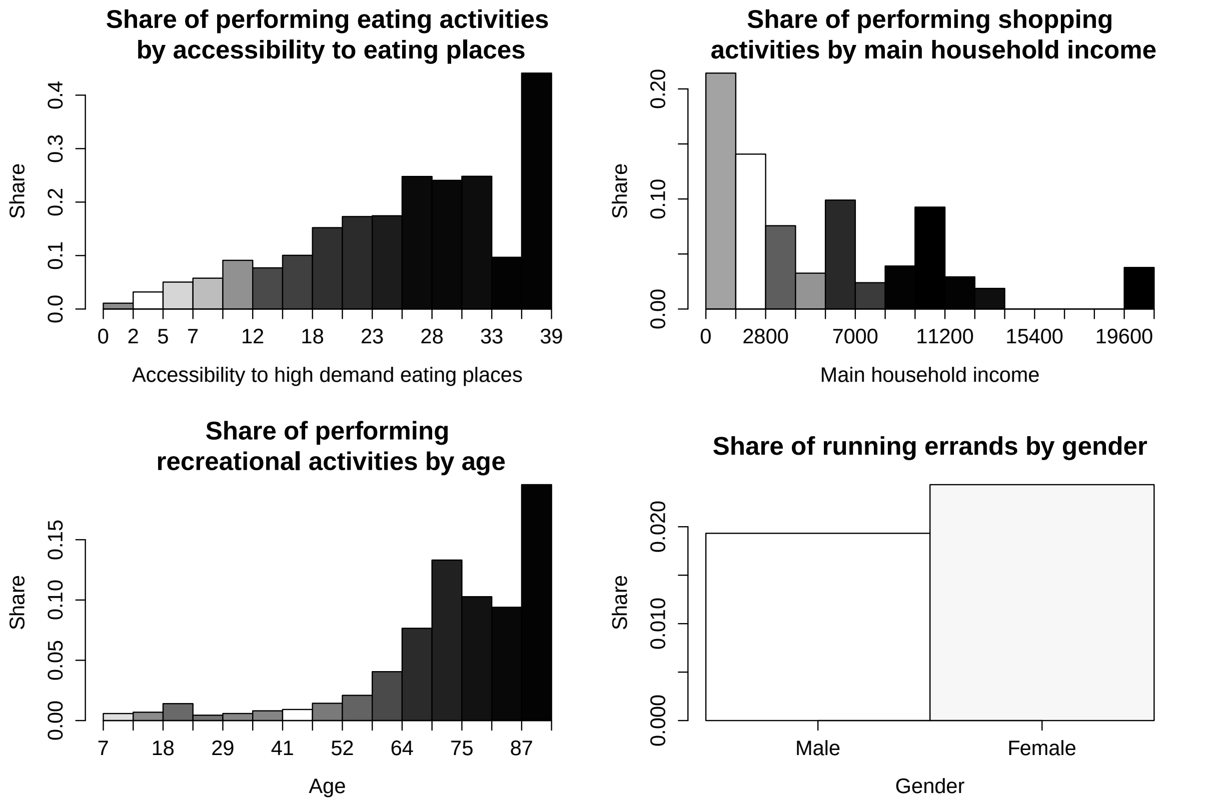


Figure 6: Some dependencies on the performance or non-performance of flexible activities based on socio-demographic attributes. The height of each bar represents the fraction of the number of observations which the person performs the corresponding activity, over the total number of observations in the corresponding category. The color of each bar represents the absolute number of observations in the corresponding category, wither color means more observations.

3.2 Validation

As mentioned before, 10% of the travel survey respondents were taken aside from the beginning of the process for validation purposes (test set). To validate the efficiency of the proposed methodology the multi-activity scheduling algorithm was executed for each respondent in the test set (566 persons). These predictions took 31 seconds using a one-thread java program running on a windows machine with a common processor (intel i7) and running one thread; the full population of Singapore would take 8.6 hours with this rate. Furthermore, for each person in the test set, other maximum utility activity schedules were calculated, but using different selection methods for the known places and randomly constructed agendas. Figure 7 presents the result of comparing these methods with the real schedules. For the chart at the top, flexible activity durations were extracted from the reported trips, from random input predictions, and from the developed systematic predictions. Systematic predictions were more accurate, with very short activity durations for the random draws. Although the same activity utility function

Activity	Age	Female	Car availability	Ethnicity	Accessibility	Household Size	Household role	Employed	Income	Time at home	Time at work	R^2
Eat	s		s		s	s			s			0.709
Shop	s	s	s				ompy	s	s	s	s	0.753
Soc	s		s			s	opmy		s			0.957
Rec	s		s	cimo				s	s	s	s	0.926
Err	s	s	s	icmo	s			s			s	0.897

Table 3: Summary of the *"Perform activity X" choice models* absolute impact. In this table, green represents a positive dependency while red refers to a negative one. Blue color means that the relation between the parameter and the outcome is not always increasing or decreasing. Deeper shades represent higher impact or higher absolute value of the utility parameter. The letter "s" means that the variable is significant according to the t-test. Yellow cells are significant categorical variables. The text in these cells represents the order from the most to the least significant category. For ethnicity "c" means Chinese, "i" refers to Indian, "m" is for Malay and "o" means other. For household role "m" is for the main role, "p" means the partner role, "o" stands for the older role and "y" is the younger role.

was used for both estimations, with the same duration distribution for each flexible activity, the number of activities predicted with random inputs resulted higher. These results were expected because the agendas were not restricted for the random input predictions and the utility raises when many short activities are scheduled (due to the performing activity utility function). On the other hand, restricted agendas and the extension of the performing activity utility function, which penalizes continuous activities of the same type, generate shorter activity chains with longer activity durations. The second bar chart shows that the number of predicted activities were quite similar, except for running errands with 40% less observed activities.

Figure 8 presents a crucial issue in transportation studies, travel times. This figure shows comparisons of travel time distributions of two flexible activities, on the top eating activities and below shopping activities. These comparisons are directly related to the selection of known places. The first method selects the known places randomly (i), the second selects places according to travel time distributions (ii), the third randomly selects using the results from the binary regressions described before (iii), and the fourth selects the 20 best locations according to type and travel time distributions (iv)(deterministic). Full random selection (i) can not reproduce the observed distributions at all (as expected), while optimal deterministic selections (iv) fails reproducing the long tail of the observed distribution. In contrast using travel time distributions (ii) and (iii) from the training data (5100 people) reproduces better the observed distributions of the 566 testing records.

Figure 7: Comparison of maximum utility predictions made with socio-demographics and geographical information against maximum utility predictions using random inputs.

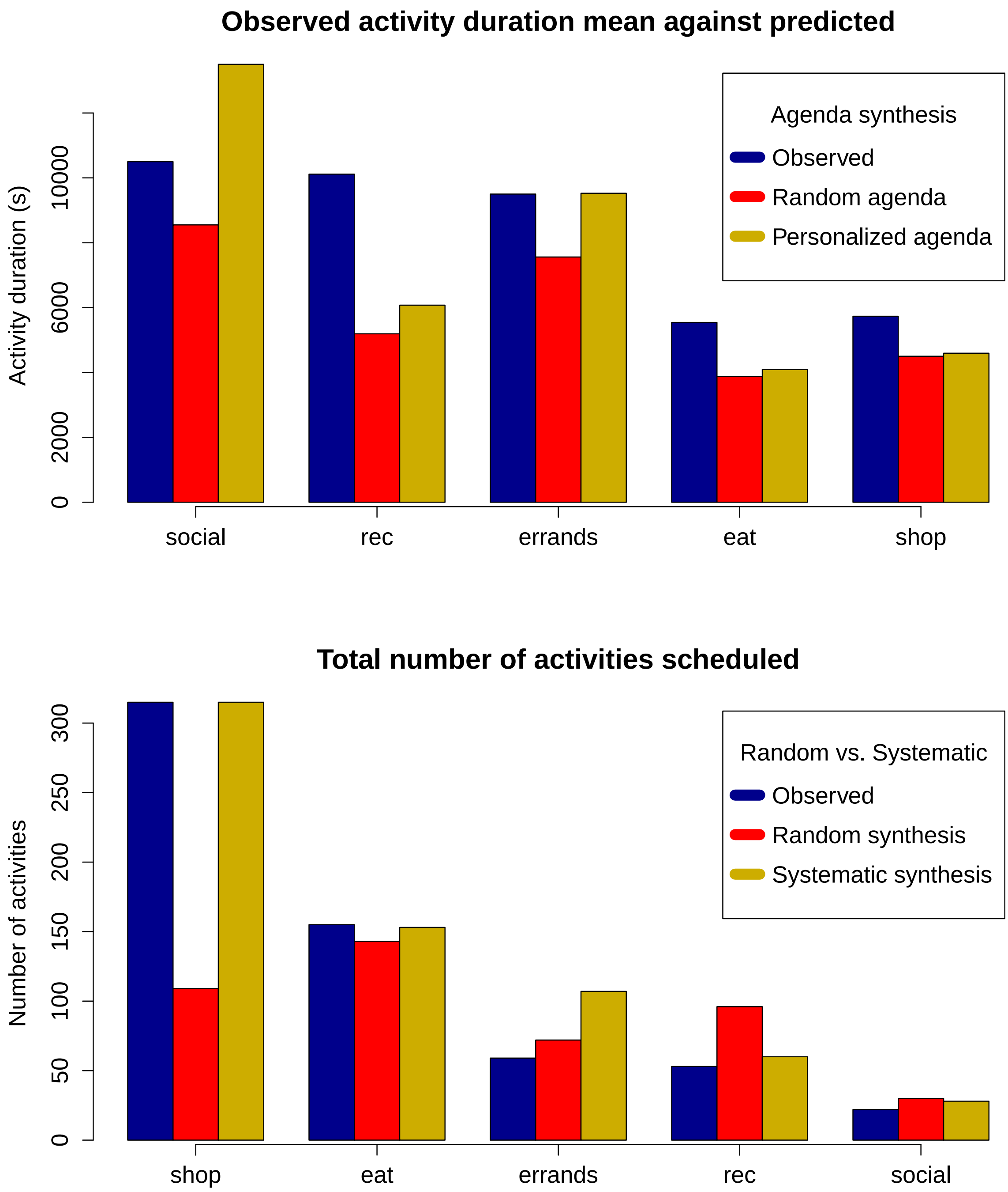


Figure 8: Travel time kernel densities of eating and shopping activities. Prediction capabilities after using four different known place selection methods can be compared.

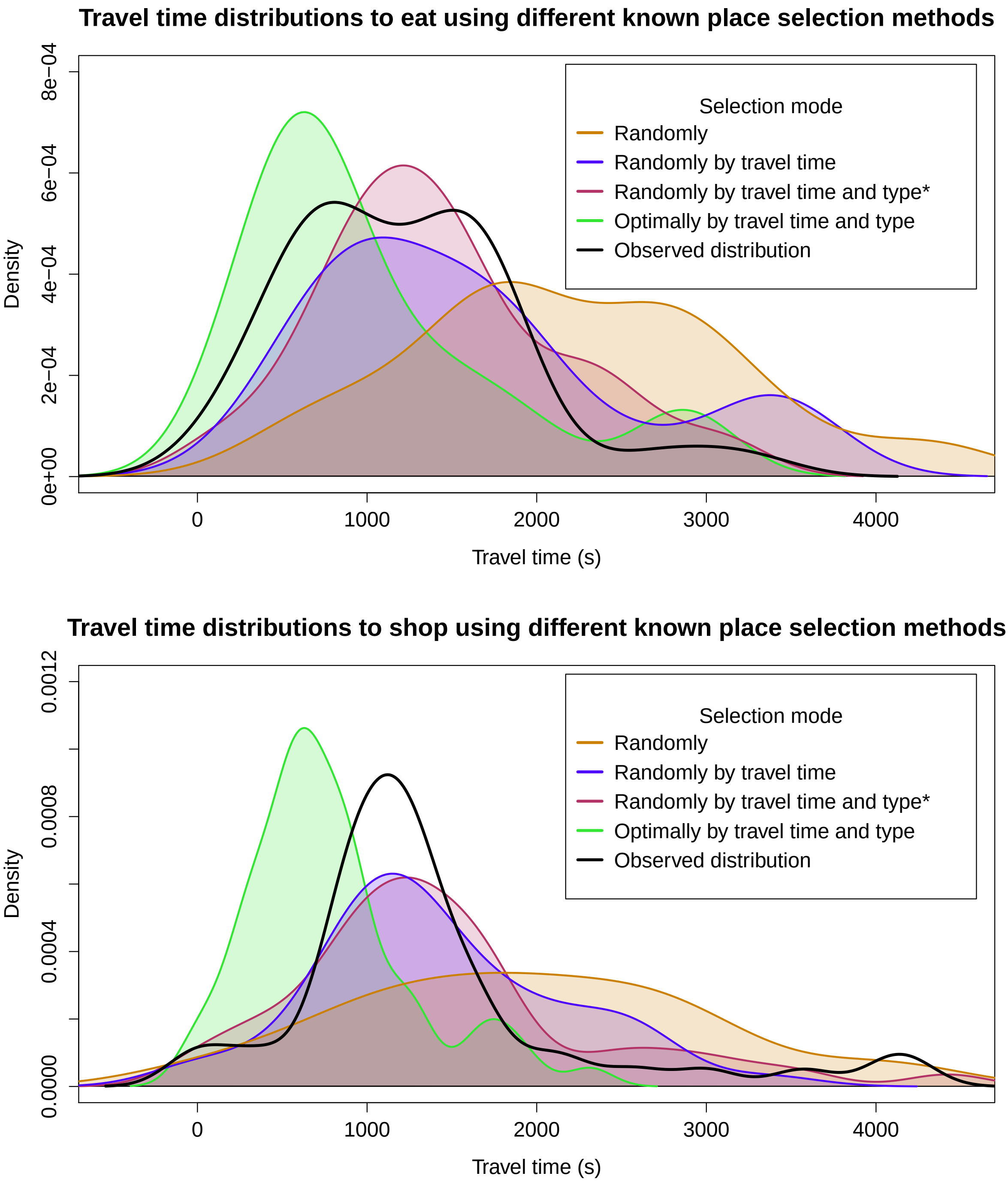
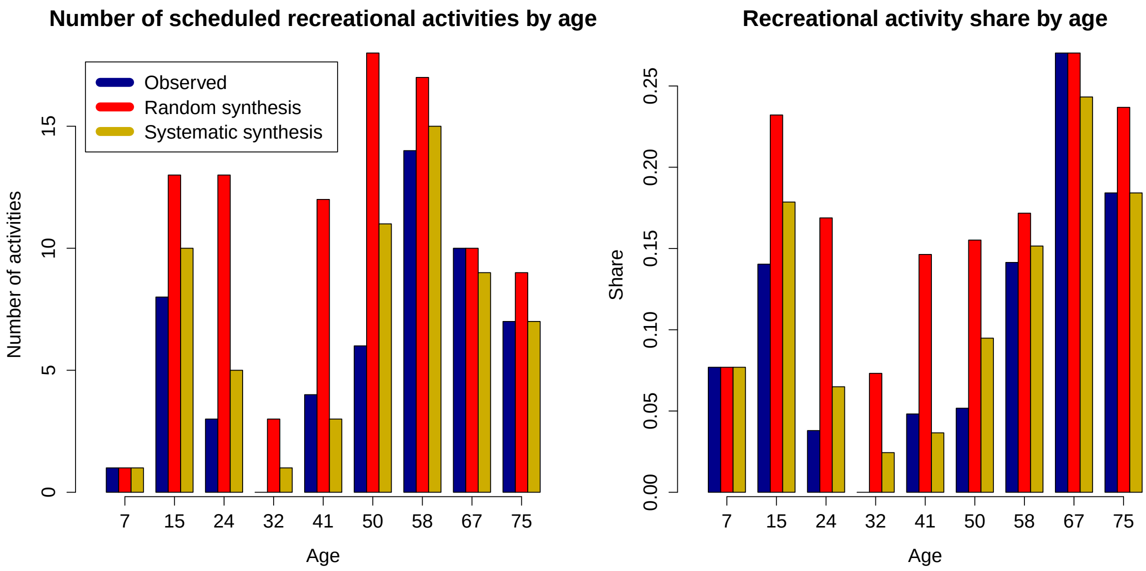


Figure 9: Prediction comparison of the total number of recreational activities and activity performance share by age. For each interval the share is obtained dividing the number of people with a recreational activity scheduled over the total number of people in that interval



Finally to assess the correlation of socio-demographic characteristics on the predicted schedules, Figure 9 compares the random input prediction with the systematic prediction at 10 different age intervals. The systematic method resulted very accurate using the high correlation of age in performing recreational activities and going to community centres. On the other hand, the random input prediction is just related with the number of respondents at each age interval. This can be appreciated comparing the total number of activities on the left with the activity performance share on the right.

4 Agent-based simulation

In this section the proposed scheduler is incorporated to MATSim for within day re-planning of secondary activities. MATSim is an agent-based transport simulation platform for large-scale scenarios. Figure 10 summarizes the MATSim process. MATSim is basically an evolutionary algorithm for the activity plans of the agents. This means the daily plans of the agents evolve when many simulations of the same day are executed and scored. In the MATSim mobility simulation (Mobsim) thousands or millions of agents representing a population interact in a highly detailed transport supply. Then, re-planning modules mutate some of the plans in different ways. The plans change in Time, Routes, Modes, Activities, etc. After many iterations the general utility of the population converges obtaining a user-equilibrium.

To test the multi-activity scheduler the MATSim process was executed with a 1% sample of the Singaporean agents (20,000). This extended version of MATSim allows agents to start the process with incomplete activity plans (a plan with fixed activities or a skeleton as called by (Doherty et al., 2002)). Each iteration, 10% of them were selected randomly to modify their flexible activities i.e. to start the simulation with incomplete plans. With this, scheduling events are highly heterogeneous due to the different conditions of the agents (agendas, known places, experiences travel times). In other words, the decisions are personalized, as every agent experiences the day in a different way (even if they live and work at the same locations). Thus, when an agent finds free time during its plan the proposed algorithm schedules activities on-the-fly. This process in MATSim is commonly called within-day re-planning.

Results of these initial experiments show firstly that the proposed scheduler is computationally feasible using parallel processes in charge of several scheduling tasks. Each iteration took in average 20 mins using a computer with 24 cores. Furthermore, the utility of agents increases because their activity schedules complexity increases. Figure 11 shows the evolution of the utilities after running 50 MATSim iterations with fixed agendas, fixed primary activities, but sharing information of secondary activity locations and experienced travel times among themselves. Future work is needed to include the mental map models of the previous section in MATSim and validate those results.

Figure 10: MATSim process.

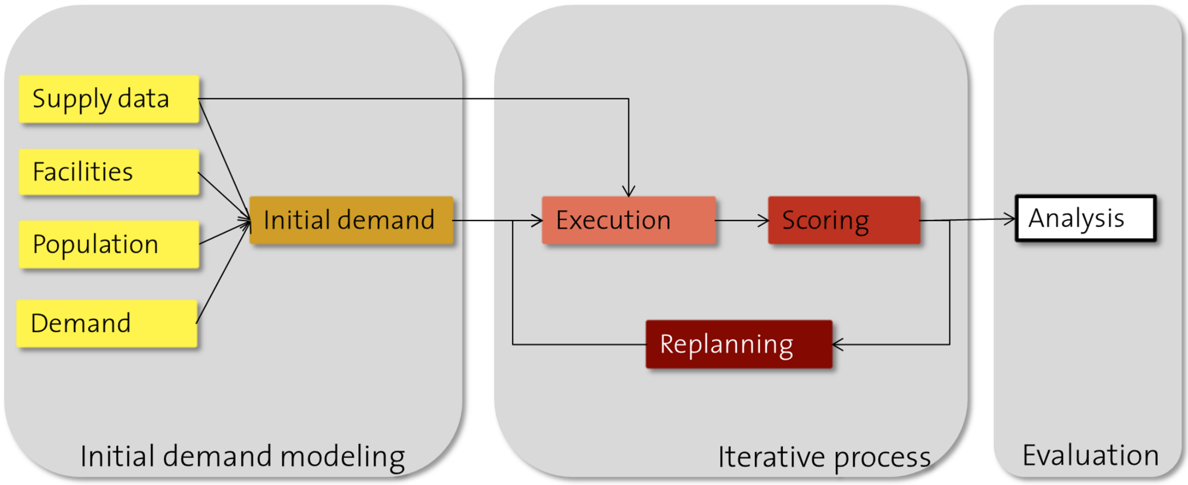
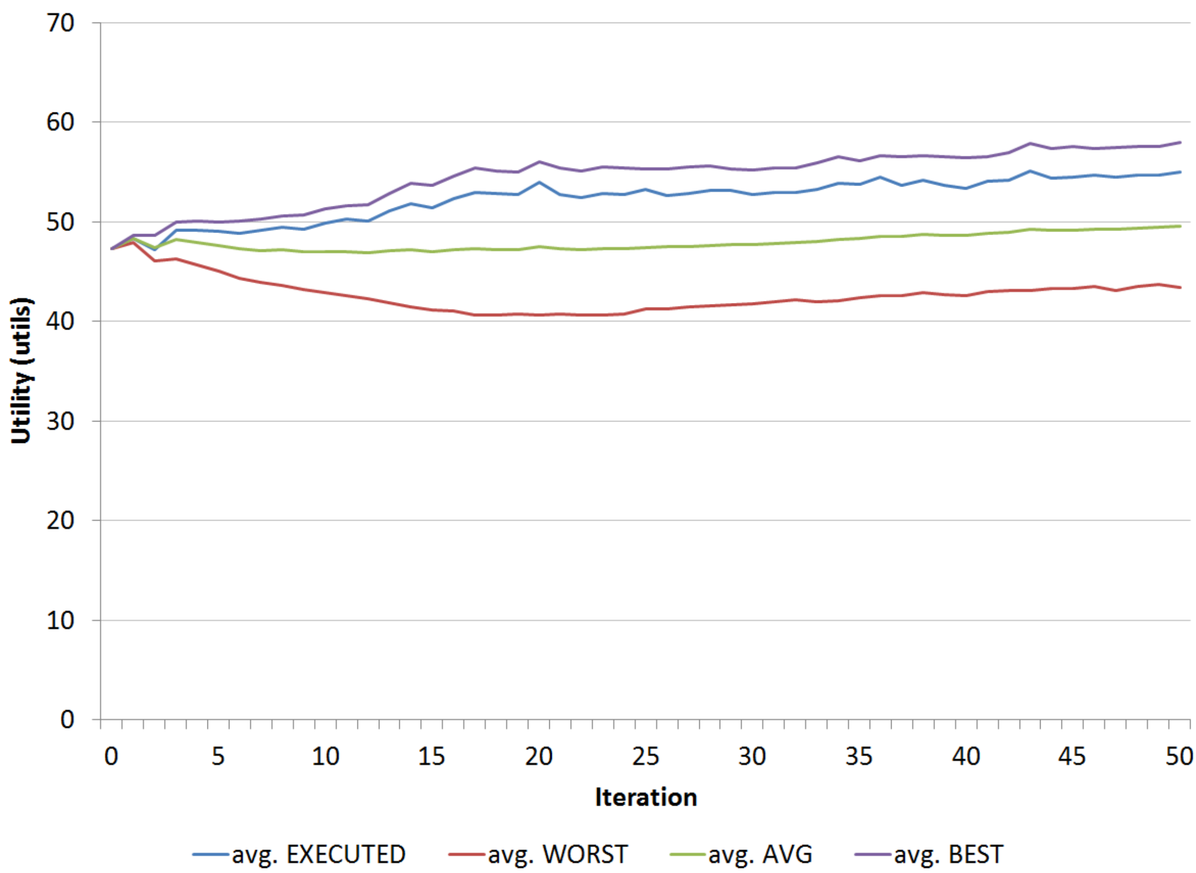


Figure 11: Evolution of the daily utilities of 20,000 agents within the MATSim evolutionary algorithm, due to the increase of their activity schedules complexity.



5 Conclusions and outlook

In this paper, personalized flexible activity scheduling was studied. A recursive multi-activity scheduling algorithm that doesn't prioritize any scheduling dimension was design and implemented. The utility-maximization algorithm returns the sequence of activities without imposing any size and the trips between them. For each activity the method calculates start time, duration, location and type; and for each trip the algorithm returns the travel time and the mode. Tests with controlled inputs demonstrated the potential of the model. It is possible to calibrate the number and durations of specific activities varying the corresponding utility parameters. Spatial, temporal and tour restrictions can also be imposed.

One of the main goals was to use commonly available data to validate the approach. Flexible activity patterns of the city of Singapore where extracted, estimating 13 *Binary logistic regression* models from the *Household interview travel survey* carried out in 2012. Results successfully show that using socio-demographic and geographical characteristics as an input of a maximum-utility activity scheduler, improves prediction capabilities. The *Activity agenda* concept successfully allowed to restrict the activity type and accurate activity durations were also achieved. The *Set of known places* concept was employed to model destination choice. It is based on the assumption that the type of places where people travel is related with their socio-demographic characteristics. Accurate travel time distributions by activity type were predicted.

Initial tests using an activity-based multi-agent simulation platform (MATSim) were carried out. The multi-activity scheduler were included in MATSim for within-day re-planning of flexible activities. Results show that the process is computationally feasible (20 minutes for iteration of a 1% sample using 24 cores for parallel within-day replanning). Furthermore, MATSim agents increase their utility just increasing the complexity of their activity chains.

The final goal of this work, which is not presented in this paper, is to use the proposed multi-activity scheduler and the *Mental map* concepts together to model flexible activity demand in MATSim. With this, longer time horizons can be simulated as fixed activities present less variability. Furthermore, decisions such as change the order of activities to avoid traffic congestion or a location choice based on future opportunities will be able to be modeled. On the computation issue, a duration of 8 hours predicting flexible activities of a full population is a tractable time. Furthermore, the method is fully parallelizable as the proposed activity scheduling for one person doesn't depend on other people. If shorter computation times are needed the activity scheduling algorithm can be relaxed, assuming that the perform activity utility doesn't depend on previous activities. In this case dynamic programming can be applied to find the optimal activity-trip chain, and the complexity would drop from $O(2^n)$ to $O(nlog(n))$.

6 Acknowledgements

The author would like to thank the following Singaporean Authorities for providing access to data and valuable review: Land Transport Authority, Urban Redevelopment Authority, Singapore Land Authority (SLA Digitised Land Information). Financial support comes from the National Research Foundation (NRF) of Singapore and ETH Zurich research fund.

7 References

Arentze, T. A., T. A. Pelizaro and H. J. P. Timmermans (2010) An agent-based micro-simulation framework for modelling of dynamic activity-travel rescheduling decisions, *International Journal of Geographical Information Science*, **24** (8) 1149–1170.

Arentze, T. A. and H. J. P. Timmermans (2000) *ALBATROSS: A Learning-Based Transportation Oriented Simulation*, EIRASS, Eindhoven.

Arentze, T. A. and H. J. P. Timmermans (2004) Multistate supernetwork approach to modelling multi-activity, multimodal trip chains, *International Journal of Geographical Information Science*, **18** (7) 631–651.

Arentze, T. A. and H. J. P. Timmermans (2009) A need-based model of multi-day, multi-person activity generation, *Transportation Research Part B: Methodological*, **43** (2) 251–265.

Axhausen, K. W. (2006) Moving through nets: An introduction, in K. W. Axhausen (ed.) *Moving Through Nets: The Physical and Social Dimensions of Travel*, 1–7, Elsevier, Oxford.

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

Chen, X. and M.-P. Kwan (2012) Choice set formation with multiple flexible activities under space-time constraints, *International Journal of Geographical Information Science*, **26** (5) 941–961.

Doherty, S. T., E. J. Miller, K. W. Axhausen and T. Gärling (2002) A conceptual model of the weekly household activity-travel scheduling process, in E. Stern, I. Salomon and P. H. L. Bovy (eds.) *Travel Behaviour: Patterns, Implications and Modelling*, 233–264, Edward Elgar, Cheltenham.

Erath, A., P. J. Fourie, M. A. B. van Eggermond, S. A. Ordóñez Medina, A. Chakirov and K. W. Axhausen (2012) Large-scale agent-based transport travel demand model for Singapore, paper presented at the *13th International Conference on Travel Behaviour Research (IATBR)*, Toronto, July 2012.

Feil, M. (2010) Choosing the daily schedule: Expanding activity-based travel demand modelling, Ph.D. Thesis, ETH Zurich, Zurich.

Hägerstrand, T. (1970) What about people in regional science?, *Papers of the Regional Science Association*, **24** (1) 7–21.

Hansen, W. (1959) How Accessibility Shapes Land Use, *Journal of the American Institute of Planners*, **25** (2) 73–76.

Kuhnimhof, T. and C. Gringmuth (2009) Multiday multiagent model of travel behavior with activity scheduling, *Transportation Research Record*, **2134**, 178–185.

Märki, F., D. Charypar and K. W. Axhausen (2014) Agent-based model for continuous activity planning with an open planning horizon, *Transportation*, **41** (4) 905–922.

Miller, E. J. and M. J. Roorda (2003) A prototype model of 24-hour household activity scheduling for the Toronto area, *Transportation Research Record*, **1831**, 114–121.

Miller, H. J. (2005) A measurement theory for time geography, *Geographical Analysis*, **37** (1) 17–45.

Narayana, C. L. and R. J. Markin (1975) Consumer behavior and product performance: An alternative conceptualization, *Journal of Marketing*, **39** (4) 1–6.

Nijland, L., T. A. Arentze and H. J. P. Timmermans (2012) Incorporating planned activities and events in a dynamic multi-day activity agenda generator, *Transportation*, **39**, 791–806.

Nurul Habib, K. M. and E. J. Miller (2009) Modelling activity generation: a utility-based model for activity-agenda formation, *Transportmetrica*, **5** (1) 3–23.

Ordóñez Medina, S. A. (2015) Predicting personalized flexible activity patterns, paper presented at the *14th International Conference on Travel Behaviour Research (IATBR)*, Windsor, July 2015.

Ordóñez Medina, S. A. and A. Erath (2013) Estimating dynamic workplace capacities by means of public transport smart card data and household travel survey in Singapore, *Transportation Research Record*, **2344**, 20–30.

Small, K. A. (1982) The scheduling of consumer activities: Work trips, *American Economic Review*, **72** (3) 467–479.

Vovsha, P. and S. Gupta (2013) A model for work activity schedules with synchronization for multiple-worker households, *Transportation*, **40** (4) 827–845.

Wilson, C. (2008) Activity patterns in space and time: Calculating representative Hagerstrand trajectories, *Transportation*, **35** (4) 485–499.