# Representing location choice within activity-based models

**Nicolas Salvadé, Tim Hillel, Janody Pougala, Tom Haering, Michel Bierlaire**

**STRC conference paper 2022** **May 11, 2022**

**STRC** | **22nd Swiss Transport Research Conference**
Monte Verità / Ascona, May 18-20, 2022

# Representing location choice within activity-based models

Nicolas Salvadé, Tim Hillel, Janody Pougala,
Tom Haering, Michel Bierlaire
Transport and Mobility Laboratory (TRANSP-OR)
EPFL
{nicolas.salvade,tim.hillel,janody.pougala,tom.haering,michel.bierlaire}@epfl.ch

May 11, 2022

## Abstract

Since the 70s, there has been a growing interest in activity-based modelling. This approach models the need to travel as a result of performing daily activities (Bowman and Ben-Akiva, 2001). Nevertheless, the activities need to be scheduled which involves a lot of variables and results in a huge number of unique alternatives (Pougala *et al.*, 2021). Among these variables, the number of possible locations is usually bigger than other variables, motivating the use of a choice set for locations. However, this choice set of locations is usually not known by the modeller (Pagliara and Timmermans, 2009), so there is a need to recreate it. In addition, it would be useful for two purposes: simulation of daily schedules, and estimation of the parameters of an activity-based model based on an underlying choice model. For the first one, alternatives in the choice set must be competitive, to generate realistic schedules, as for the latter one, it should also contain unlikely alternatives to estimate unbiased parameters. In this paper, a methodology to generate a choice set of destinations suitable for both purposes is presented. The choice set is generated with a choice model and can be transformed afterwards to include unlikely alternatives. The methodology is validated using the 2015 Swiss Mobility and Transport Microsensus (fédéral de la statistique and fédéral du développement Territorial, 2017) dataset.

# 1 Introduction

Since the 70s, there has been a growing interest in activity-based modelling. This approach models the need to travel as a result of performing daily activities (Bowman and Ben-Akiva, 2001). Nevertheless, prior to performing activities, they need to be scheduled. The scheduling process for individuals is an important part of activity-based models and involves a lot of variables. The type of activity, the activity location, the mode of transport or the activity timing are all examples of variables that an individual considers in order to schedule their day. Interactions between these components are highly complex. For example, having access to a car can make a location available that would otherwise be difficult to reach by public transport; or needing to travel during a congestion peak could increase the attractiveness of soft modes and closer destinations. In this paper, an alternative is defined as a combination of the choice of activity, destination, mode of transportation and timing. Due to the number of variables, the scheduling process is highly combinatorial and results in a huge number of unique alternatives (Pougala *et al.*, 2021).

There are two main streams of activity-based models: rule-based models and utility-based models (Pougala *et al.*, 2020). In the latter one, the scheduling process is obtained by assuming that individuals take part in activities that maximise their utilities. There is therefore a choice model with parameters to be estimated beforehand. As mentioned before, the number of alternatives can be extremely high, which causes problems for the estimation of the choice model. More specifically, in the context of destinations, the universal choice set, which is composed of all the possible locations for an activity, is usually much larger than the ones for other variables like activity type, transportation mode or discretised timings. Thus, it would be unrealistic and too computationally expensive to estimate parameters of the activity-based choice model with this set applied to every individuals. Furthermore, the choice set of individuals, which is made of destinations that an individual is actually considering choosing, is usually not known by the modeller. For all of these reasons, a choice set is needed when estimating parameters of an activity-based choice model. In addition, unlikely alternatives should be included in the consideration set so that the activity-based choice model can capture why the chosen alternative is superior to other alternatives. Moreover, this allows to realistically simulate the scheduling process and to avoid biased parameters (Pagliara and Timmermans, 2009). One must then reconstruct a choice set that includes competitive and more random alternatives.

In this paper, we propose a methodology to rebuild a choice set that is suitable for estimating parameters of a choice model. Ideally, to reconstruct how individuals build their choice set as accurately as possible, the relations between the different variables of the activity-based model should be considered. While it would be very challenging to estimate everything at once, we propose a new approach to jointly generate a destination and mode choice set from a trip diary dataset. In doing so, we tackle the problems associated with estimating a choice model in the context of a utility-based ABM. This also provides a behavioural interpretation of how an individual generates his choice set and gives more flexibility than a sequential reconstruction of mode and destination choice sets. The choice set is obtained by estimating a choice model, which provides a probability of choosing a destination with a specific mode of transportation. The given probability distribution is then used to form a choice set through importance sampling of alternatives. This choice set is composed of destinations linked to a mode of transportation.

The rest of the report is structured as follows: section 2 highlights what has been undertaken in the literature until now. In section 3, the methodology used to resolve the problem is presented. The results and discussion of a case study are described in section 4. Finally, a conclusion and the future work are provided in section 5.

## 2 Literature review

According to Pagliara and Timmermans (2009), there are three important ways to generate a spatial choice set.

1. a **naive** way, where all the alternatives are simply enumerated.
2. a **deterministic** way, where the choice set is derived according to some rules. For example, the choice set could be constrained in distance and time.
3. a **stochastic** way, where the choice set is obtained probabilistically. For example, the spatial choice set can be derived with utility-maximisation principles, where individuals maximise the utility that they receive when performing activities at specific locations.

The naive way is generally used in practice only if the number of alternatives is low. Indeed, if the number of alternatives increases, the computational time also increases, and can quickly become out of control; which motivates the use of the two other ways in order to generate a choice set.

For the deterministic part, the choice set is usually generated according to some rule-based algorithms that constrain the destination location to be within a distance or a travel time. The concept of *time-space prism*, introduced by Hagerstrand (1970), where the chosen destination is forced to be within a perimeter of edges defined by the time and distance from the origin, is vastly used to constrain the choice set of destinations. The *Activity space*, which is the area where activities are performed, is also a concept to define the choice set in a deterministic way. Schönfelder and Axhausen (2003) measured destinations locations with a geocoded dataset of trips, and showed that they would fall within an activity space of three different forms: a two-dimensional confidence ellipse, kernel densities, and an area covered by a buffer around the shortest paths network. Another way to constrain the choice set destinations is the *rubber banding* method (Scherr *et al.*, 2019). Using home and primary activity (work or education) location as anchor points, they assume that a third destination would be situated inside the perimeter of these two locations.

On the other hand, one can also use a stochastic way to generate the choice set. This way is mostly associated with the utility that one can derive from going to a destination. The choice set is based on the assumption that individuals consider locations that maximise their utility. Swait and Ben-Akiva (1987) used the concept of *constraints* (mode of transportation, income, accessibility, knowledge, social, psychological, ...), but in a stochastic way. The constraints generating the choice set would be generated randomly. They argued that some constraints might be well suited for a deterministic approach, such as distance, but that appears to not be that straightforward for social, informational or psychological constraints. This idea is extended by representing unobserved constraints or elimination criteria by indicators that would be derived from the data (Ben-Akiva and Boccara, 1995). For example, driving alone would be available if and only if the random constraint is above a threshold defined by the modeller. In this example, the threshold depends on the number of cars in the household associated with parameters to be estimated. Swait (2001) argued that the choice-set is taste-driven. He used this assumption to derive the *GenL* model, which is a cross-nested logit model. Nests are the different possible choice sets, and alternatives belong to each nest where they appear. Supplementing the previously mentioned work, Scrogin *et al.* (2010) used stochastic efficient frontier analysis to derive a choice set. Alternatives that would provide a sufficient quality with respect to a certain level of travel time would be kept in the choice set. Overall, we have to keep in mind that these models have been developed to predict a chosen destination afterwards, in a specific context (mode choice, shopping location, etc.). Not much work has been done in the context of activity-based model, and especially not to address the estimation of a discrete choice model's parameters.

When using a choice set to estimate parameters, the choice set must contain competitive alternatives, so that the model captures trade-offs made by choosing an alternative, but also worse alternatives, so that the model captures why the chosen alternative is more interesting. When doing importance sampling, i.e. sampling alternatives that are more likely to be chosen, a correction term in the maximum likelihood estimation must be added (Bierlaire and Krueger, 2020). This correction term is obtained from the probability of alternatives to be sampled before the estimation of the model. It considers that the individual can only choose amongst a reduce number of alternatives. This allows to avoid biased parameters. The sampling protocol must be defined and the probability to get the choice set must be calculated, in order to derive the correction term. Frejinger *et al.* (2009) is an example of importance sampling to derive the choice set, where a route choice set is built accordingly to a probability of choosing links that will compose the route. This probability depends on how far the link diverges from the shortest path. This sampling protocol allows to compute the correction term due to this alternative sampling. A method based on the same idea has been used by Pougala *et al.* (2021) to generate a choice set of activity schedules. Starting with a given schedule, and using a Metropolis-Hastings algorithm, several transformations were applied to the schedule, such as time discretisation, activity assignment and swapping, changes of activity duration mode or location and a combination of all these changes. The resulting schedules are accepted according to a probability of acceptance, and the probability to apply these changes are defined by the modeller, which permits to compute the correction term.

In this paper, we will use a stochastic model to derive a joint choice set of destinations and mode, by getting a first choice set from a discrete choice model. This allows to perform importance sampling over alternatives, and spot the one that could be more attractive to a specific individual. Then we introduce perturbations in the choice set to include alternatives with lower utility, with the motivations raised in the previous paragraph. The aim of the project is to use the flexibility of a joint estimation of the destination and the mode of transportation choice set, while still keeping a balanced choice set, in order to avoid biased parameters in the activity-based model. Using a choice model should also give a behaviorally more realistic explanation on the choice of alternatives.

# 3 Methodology

This section presents the methodology used for the model. First, in subsection 3.1, the behavioural assumptions used are introduced. Then, the problem is formalised in 3.2. Afterwards, the model framework is presented in 3.3, as well as how to obtain the choice set from the model in 3.4. The probability to obtain the choice set from the sampling protocol is also stated down in this section.

## 3.1 Behavioural assumptions

In order to recreate the choice set of an individual, several behavioural assumptions are made:

1. **Mode and location choices are inherently linked**: we assume that mode and location choices are linked, with the underlying interpretation that the transportation mode influences the destination choice and vice versa. For example, one could choose a car to access a destination zone that is easily reached by car but not by public transport. On the other hand, one could choose a close destination because they do not have an easy access to car or public transport services. Additionally, this gives a better flexibility than modelling them sequentially and captures potential interactions between these variables.
2. **The utility derived by the choice of a location, with a transportation mode, depends on the travel time and some measures of attractiveness of the zone**: we assume that an individual chooses its mode of transportation and destination according to the travel time, the type of zones (urban, rural or intermediate), the population and job density of the zone. The underlying interpretation of this assumption is that an individual prefers to minimise the time lost in transportation. Furthermore, the type of zones and a higher or lower job or population density would increase or decrease a destination utility and justify a longer or shorter travel time.
3. **Activity choice is given and impacts the utility function**: the type of

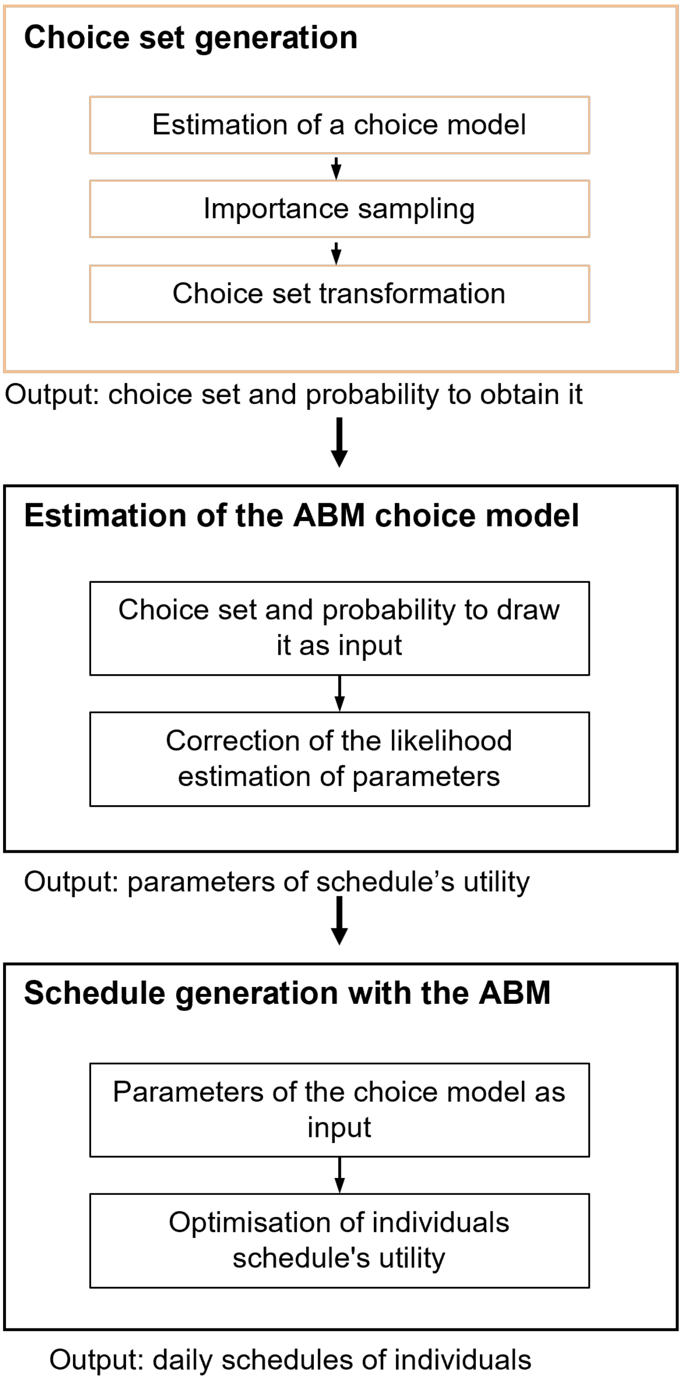
activity is assumed to be given in order to choose the destination and the mode of transportation. It is also an explanatory variable in the utility function. This implies that the choice of transportation mode and destination depends on the type of activity. For example, with an education activity, one would prefer to travel by public transport but for a shopping activity, a soft mode might be preferred.

4. **Alternatives with the same transportation mode are correlated**: we assume that alternatives with the same mode of transportation are correlated. This is interpreted as a similar comfort or perception of travel time in the same mode.
5. **Alternatives are spatially correlated**: we define different groups of destinations where we assume that there is a spatial correlation amongst zones that are in the same group of destinations. For example, zones that are close to a water plant share the same advantages such as a swimming possibility or beautiful views.

## 3.2 Problem formalisation

We adopt a definition of the universal choice set similar to the one in Shocker *et al.* (1991), which is the set of all alternatives. Shocker *et al.* (1991) introduces three others sets; the *awareness* set, which consists of the alternatives that are known by the individual, the *consideration* set, composed of alternatives that are considered by the individual within a specific context (point of time, given activity, and so on), and the *choice* set, which contains alternatives from the consideration set that are still relevant directly before making the choice. A choice set is therefore a specific type of consideration set. In this paper, we generate a choice set of destinations linked to a mode of transportation for an individual $n$ from a given activity $a_n$. We define $Z$ the set of all possible destination zones $z$ and $M$ the set of all transportation modes available $m$. There are therefore $Z \cdot M$ alternatives in the universal set. We also introduce $C_n$, the choice set of an individual $n$ with regard to an activity $a_n$. Since this choice set is not fully observed by the modeller, we introduce a choice set for a given type of activity $\hat{C}_n$ generated for all individuals. The aim of the methodology is to recreate $\hat{C}_n$, such that it contains alternatives that are competing, hence that could potentially be in $C_n$, and that are less likely to be chosen, so that $\hat{C}_n$ can be used to estimate the parameters of the ABM choice model. To include unlikely alternatives, the choice set needs to be transformed with a sampling protocol such that it is always possible to calculate the probability to obtain it. Finally, we define $P(\hat{C}_n)$, the probability to obtain the transformed choice set of an individual.

Figure 1: Modelling pipeline of the full activity based model




Outlined in orange, the scope of this paper

Figure 1 summarises the full modeling pipeline. The first part is to generate a choice set, with a probability to obtain it. Then, this choice set is used to estimate the parameters of the activity-based choice model. Finally, the ABM optimises individuals daily schedules. The scope of this paper is to focus on the first part of the whole process, outlined in orange.

### 3.3 Modelling the framework

In order to derive the choice set of an individual $n$, the utility related to a destination zone $z$ and a mode of transportation $m$ is defined as follows:

$$
\begin{aligned}
V_{zmn} = ASC_m + \beta_{TIME_m} \cdot TT_{zm} + \beta_{TYPE} \cdot T_z + \omega_{work} \cdot \beta_{JobD_{WORK}} \cdot JobD_z + \\
\omega_{education} \cdot \beta_{PopD_{EDU}} \cdot PopD_z + \omega_{leisure} \cdot \beta_{JobD-PopD_{LEI}} \cdot (JobD_z + PobD_z) + \\
\omega_{shopping} \cdot \beta_{JobD-PopD_{SHO}} \cdot (JobD_z + PobD_z) + \omega_{other} \cdot \beta_{PopD_{OTH}} \cdot PopD_z \quad (1)
\end{aligned}
$$

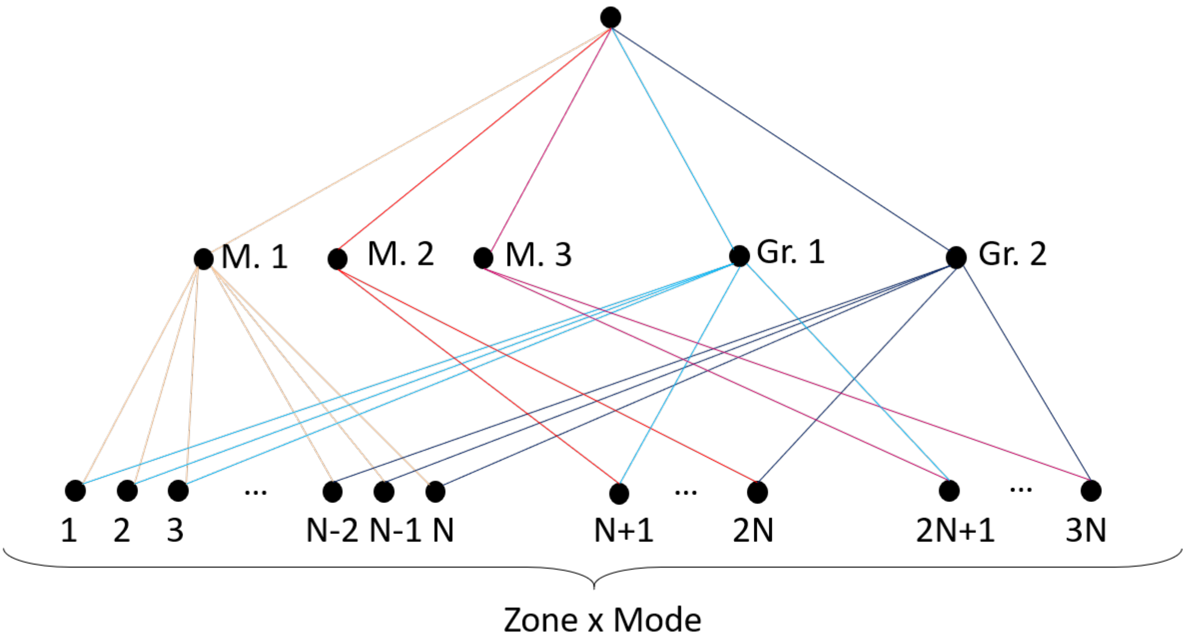
where in this equation:

- $ASC_m$: a constant for the transportation mode
- $\beta$: parameters to be estimated from the dataset
- $\omega$: dummy variable, equal to 1 if the corresponding activity is chosen
- $TT_{zm}$: travel time from an origin to the zone $z$ with mode $m$
- $T_z$: type of zone $z$
- $PopD_z$: population density of zone $z$
- $JobD_z$: employment density of zone $z$

The explanatory variables have been chosen according to the behavioural assumptions described in subsection 3.1. The relationship between activities and population density or employment density has been arbitrarily decided. Note that a few combinations have been tried before arriving to this final utility. It is clear that, with more data, such as the number of schools and universities per zone, the number of shops per zone or the number of leisure places per zone, this utility function could be improved.

A cross-nested logit model (CNL) is then used, with this utility specification, with the aim to develop a probability distribution to choose a zone. The use of a cross-nested logit model follows the behavioural assumptions. First, it allows to give a behavioural

Figure 2: Structure used for the cross-nested logit model



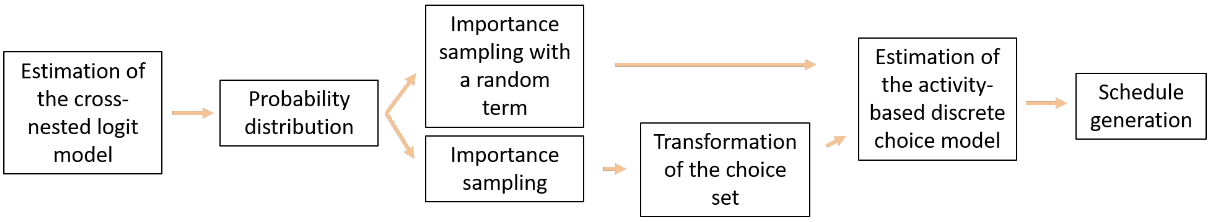

Each alternative belongs to one mode (M) and one group of transportation (Gr)

interpretation of the choice set. Then, the cross-nested structure allows to deal with both mode correlation and spatial correlation assumptions. This structure can be found in Figure 2. In this figure, all the alternatives are enumerated, and each alternative belongs to two nests: one corresponding to its mode (M), and one corresponding to its group of destinations (Gr). Note that, in this example, there are 3 modes and N zones, which is why there are 3N alternatives in total.

## 3.4 Generating the choice set

This section is separated in two parts: in 3.4.1, it is explained how importance sampling is performed with a choice model, and in 3.4.2, how the choice set is transformed with the aim to be suitable for parameters estimation. A summary of the whole process is drawn in Figure 3.

Figure 3: Summary of the methodology




### 3.4.1 Importance sampling

Once the choice model parameters are estimated, importance sampling can be performed over all alternatives by using the probability distribution obtained from the choice model. From Ben-Akiva *et al.* (1985), if the error term follows an extreme value distribution *i.e.*, with $\epsilon \sim EV(\mu, 0)$, the probability to choose alternative $i$ for an individual $n$ with the logit model is:

$$P_n(i|C_{ns}) = \frac{e^{\mu V_{in}}}{\sum_{j \in C_{ns}} e^{\mu V_{jn}}} \tag{2}$$

And for the cross-nested logit model, with $K$ nests, we have:

$$P_n(i|C_{ns}) = \sum_{k=1}^{K} \frac{\alpha_{ik}^{\frac{\mu_k}{\mu}} e^{\mu_k V_{in}}}{\sum_{j \in C_{ns}} \alpha_{jk}^{\frac{\mu_k}{\mu}} e^{\mu_k V_{jn}}} \frac{\left(\sum_{j \in C_{ns}} \alpha_{jk}^{\frac{\mu_k}{\mu}} e^{\mu_k V_{jn}}\right)^{\frac{\mu}{\mu_k}}}{\sum_{p=1}^{K} \left(\sum_{j \in C_{ns}} \alpha_{ik}^{\frac{\mu_p}{\mu}} e^{\mu_p V_{jn}}\right)^{\frac{\mu}{\mu_p}}} \tag{3}$$

Where:

- $\alpha_{jk}$: degree of membership of the alternative $j$ to the nest $k$, must be between 0 and 1.
- $\mu_k$: scaling parameter of the nest $k$. This parameter comes from the correlation amongst alternatives in the same nest and must be greater than one.
- $\mu$: parameter of the extreme value distribution.
- $V_{jn}$: deterministic utility of alternative $j$ for an individual $n$.
- $C_{ns}$: choice set where alternatives can be sampled from. It represents which alternatives are available to an individual.

Note that in all these formulas, $\mu$ is normalised to 1, and $\alpha = \frac{1}{2}$. If we assume that all the alternatives are available to anyone, the probability to draw the choice set of size $N$ with important sampling is:

$$P_{IS}(\hat{C}_n) = \frac{\Pi_{j=1}^{N} P_n(j)}{\Pi_{j=1}^{N-1}(1 - \sum_{l=1}^{j} P_n(l))} \tag{4}$$

In the above formula, the upper part represents the probability to draw alternatives in the choice set, according to the probability distribution of the corresponding choice model. The lower part comes from the fact that alternatives cannot be replaced, meaning that the probability distribution needs to be re-normalised after each draw.

#### 3.4.2 Perturbation of the choice set

From section 2, we have seen that the choice set needs to be transformed to include unlikely alternatives. Practically, this is done in two different ways. First, by adding a random term in the computation of the probability distribution, which leads to a different importance sampling. Second, by transforming the choice set with 3 operations, occurring with a probability defined by the modeller. The 3 transformations considered are to add a random alternative, to swap the transportation mode and/or to change the zone of an alternative to a neighbouring one.

The probability to get the choice set with a random term is the same as in equation 3 and 4, but with different values of the utility. For the second technique, we have:

- **Adding random alternatives**: if there is a number of zones $Z$ and modes $M$, and the size of the choice set is $N$, the probability to add an alternative $j$ which is not in the choice set $\hat{C}_n$ is $P_{add}(j) = \frac{1}{Z \cdot M - N}$. Note that if we want to add an alternative from a specific nest of size $Z_k$ and $M_k$, the probability is $P_{add}(j) = \frac{1}{Z_k \cdot M_k - N_k}$, where $N_k$ is the number of alternatives belonging to the nest already in the choice set.
- **Swapping the transportation mode**: if there are $M$ transportation modes, and the choice set $\hat{C}_n$ is of size $N$, then the probability to swap the mode of transportation of an alternative is $P_{swap}(j) = \frac{1}{N}\frac{1}{M-1}$, where the first part is the probability to select this specific alternative in the choice set, and the second the probability of choosing another mode of transportation. Note that this transformation is only valid if and

only if the resulting alternative is not already in the choice set.

- **Changing the zone to a neighbouring one**: if the alternative $j$ has $\text{Nei}_j$ neighbouring zones, the probability to change a zone is $P_{change}(j) : \frac{1}{N}\frac{1}{\text{Nei}_j}$. Note that this transformation is only valid if and only if the resulting alternative is not already in the choice set.

These transformations happen with a probability defined by the modeller. If these probabilities are respectively $p_{add}$, $p_{swap}$, $p_{change}$, we can define the probability of a transformation $t$ for alternative $j$:

$$P_t(j) = \begin{cases} p_{add}P_{add}(j) & \text{if } t = add \\ p_{swap}P_{swap}(j) & \text{if } t = swap \\ p_{change}P_{change}(j) & \text{if } t = change \end{cases}$$

Finally, we can compute the probability to obtain the transformed choice set $\hat{C}_n$ after $T$ transformations, which is:

$$P(\hat{C}_n) = P_{IS}(\hat{C}_n) \cdot \Pi_{j=1}^{T} P_t(j) \tag{5}$$

Note that the chosen alternative must be part of this choice set, but it does not modify the probability to draw the choice set. Indeed, if the chosen alternative is in the choice set, nothing more needs to be done. If it is not in the choice set, we add the chosen alternative with a probability equal to 1, which gives the same probability to draw the choice set.

# 4 Case study

In this section, a case study on Lausanne, Switzerland is conducted. First, simplifying assumptions are presented in 4.1, and details on the dataset used to estimate the cross-nested logit model are shown in 4.2. The results of the parameters estimation, as well as a specification testing of the cross-nested logit model against the logit model are conducted in 4.3 and 4.4. Then, a choice set is obtained in 4.5 and a sensitivity analysis on the choice set perturbation is presented in 4.6. Finally, a schedule is generated in 4.7 and assumptions are discussed in 4.8.

## 4.1 List of assumptions

Assumptions used to simplify the model implementation are listed and explained. Their impact on the model, and some ideas to relax them will be discussed in section 4.8.

1. **Zones with the same mode of transportation share the same alternative specific constant**: this assumption is made to avoid the number of parameters to explode. Indeed, there is only 3 modes of transportation, while the number of zones can be substantially larger.
2. **All modes and destination zones are available to everyone**: we assume that every individual has access to a car, to public transportation and can walk or bike to any destinations.
3. **Destinations are zones**: defining location as an area (zone) instead of a point location is not a loss of information.
4. **Soft modes include biking and walking**: in order to avoid the number of transportation to be too large, we regroup walking and biking in the same mode of transportation, where we assume that their utility does not differ significantly.
5. **Soft modes travel time is derived from the car travel time**: the soft modes travel time is not available from our data. Therefore, it is assumed that soft modes travel time is colinear to car travel time.
6. **Public transport mode includes rail, bus, tramway and postal bus**: we assume that all these modes have a similar impact on the utility, so we regroup them in one transportation mode, allowing to reduce the number of alternatives and increasing the number of trips from the dataset for this mode of transportation.
7. **Trips made by car, motorbike, or being a car passenger, or a motorbike passenger are all seen as the car transportation mode**: again, we assume that trips made by these different modes are not significantly different, allowing for more trips in the database and less complexity in the model (because the number of transportation modes is lower).
8. **The error term is assumed to be random following an extreme value distribution**: this assumption allows to estimate the difference between individuals that cannot be captured by the model.
9. **Population density and jobs density are used instead of absolute values**: since zones are of different sizes, the density is assumed to represent the attractivity of the zone in a better way.

## 4.2 Dataset

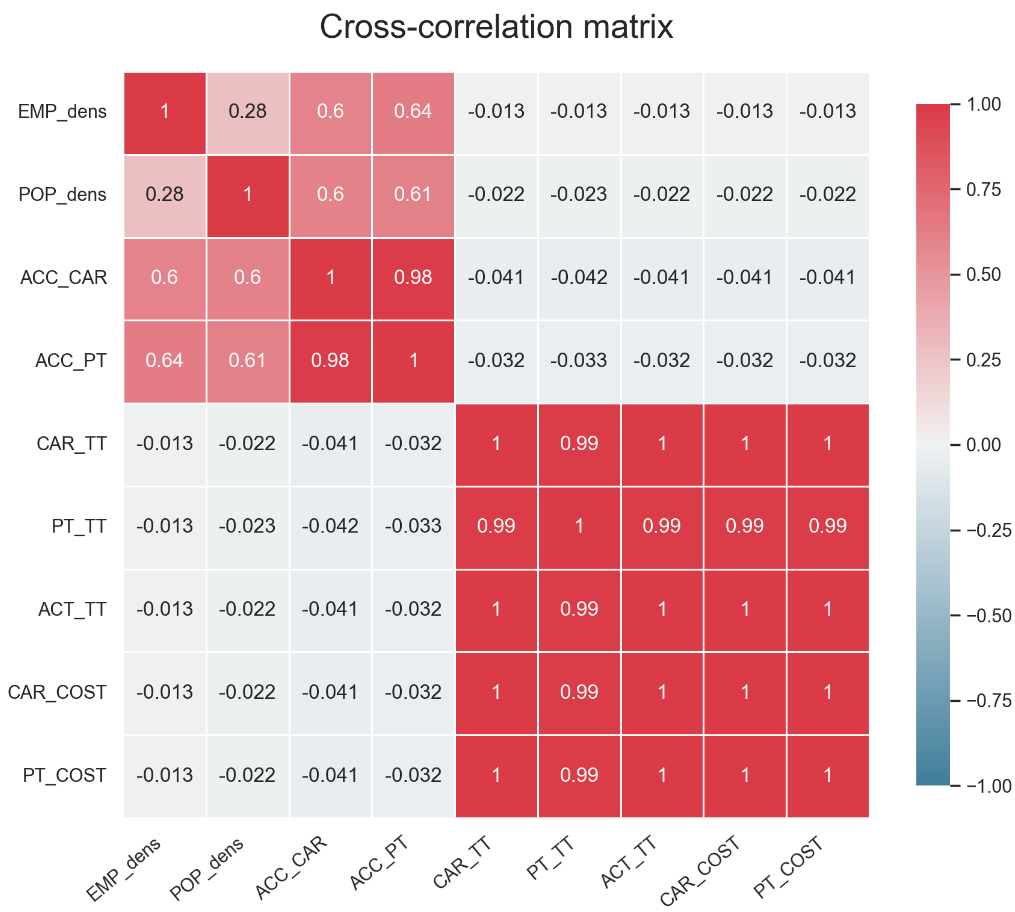
Data used for this project comes from three main sources.

1. The 2015 Swiss Mobility and Transport Microsensus (OFS and ARE, 2017) dataset, which is a trip diary dataset containing more than 100000 trips from about 50000 households in Switzerland. This dataset is mainly used to estimate parameters from the cross-nested logit model. For the case study, only trips with Lausanne as a destination and one of the mode of transportation and activity mentioned in section 3 are kept. After this pre-processing, there are 3536 trips left for the estimation of parameters.
2. Destination zones are the ones defined by the Swiss government (VM-UVEK, 2019). Switzerland is divided in approximately 8000 zones, with 88 in Lausanne, which will be used in this case study. They are shown in Appendix C. This dataset also provides the type (rural, intermediate, urban) of each zone, which is used in the utility function.
3. Zone-to-zone travel time and distance, as well as some attractivity measures of zones, such as the population or job density, and the accessibility of each zone are provided by the SBB. Zone-to-zone distance was used to convert it as an estimation of the cost of each trip. However, after analysis of the cross-correlation of these features, the cost, as well as the zone accessibility, are dropped due to their cross-correlation with travel time (see Figure 4). Note that the travel time (which is given in minutes) is scaled down by a factor of 10, in order to keep values around 1 for the parameters estimation. The distance between zones is calculated from MATSim data. In MATSim, distances are computed between two points. Therefore, to compute the distance between zones, 5 random points per zones are taken. To calculate the zone-to-zone distance, the mean of all point-to-point distances is taken.

## 4.3 Parameters estimation of the cross-nested logit model

Parameters are estimated with trips from the 2015 Swiss Mobility and Transport Microsensus (OFS and ARE, 2017 2017), that have a destination in Lausanne. They can be seen in Table 1. Destinations have been separated into three groups that are shown in Appendix C. Note that the alternative specific constant of soft modes ($ASC_{act}$), $MU_{act}$, the scaling parameter associated to the soft modes nest, and $\beta_{TYPE}$ have been normalised to 1 since

Figure 4: Cross-correlation matrix of different zones features




they were not significant in previous estimations. Also, the estimation of parameters has been done using the biogeme package on python (Bierlaire, 2020). We observe that all the parameters have a significant p-value with a 95% level of confidence. However $MU_{car}$, the scaling parameter of the car nest, is not significant according to its robust p-value. We observe that the value associated with this parameter is quite high (7.06) and with a big standard error (2.89). An explanation on why this parameter is not significant, is the number of trips to estimate parameters of the model. There are less than 1316 car trips, with 276 alternatives, and car trips can be more diversified than public transport trips (where they usually follow public transport network). The other nest parameters are all in a similar range (between 1.63 and 2.05), implying some correlation between alternatives of the same nest. We observe that both car and public transport alternative specific constant are negative, which means that with everything being zero, soft modes are the preferred ones. The impact of the job density is positive for the working activity, meaning that a zone with a higher job density is more likely to be chosen for a working activity. The same observations are made for a shopping activity, where both job and population density increase the chance of a destination for a shopping activity. On the other hand, the population density decreases the utility of a zone with an education, leisure or other (such as escort children or people with a disability) activity. This implies that these activities are usually performed in zones that have less inhabitants. Finally, the impact of travel time is always negative on the utility, which can be interpreted as further destinations being less attractive.

Table 1: Parameters of the cross-nested logit model

| **Name** | **Value** | **Std err** | **t-test** | **p-value** | **Rob. Std err** | **Rob. t-test** | **Rob. p-value** |
|---|---|---|---|---|---|---|---|
| ASC_c | -1.99 | 0.127 | -15.7 | 0 | 0.118 | -16.8 | 0 |
| ASC_pt | -1.73 | 0.166 | -10.5 | 0 | 0.19 | -9.12 | 0 |
| B_JobD_PopD_SHOP | 0.00712 | 0.00203 | 3.5 | 0.00046 | 0.00334 | 2.13 | 0.0328 |
| B_JobD_WORK | 0.00335 | 0.001 | 3.34 | 0.000842 | 0.00136 | 2.48 | 0.0133 |
| B_PopD_EDUC | -0.0251 | 0.00591 | -4.24 | 2.19e-05 | 0.00576 | -4.36 | 1.29e-05 |
| B_PopD_LEIS | -0.0189 | 0.00289 | -6.55 | 5.66e-11 | 0.00291 | -6.51 | 7.32e-11 |
| B_PopD_OTHE | -0.0178 | 0.00424 | -4.2 | 2.63e-05 | 0.00475 | -3.75 | 0.000174 |
| B_TIME_act | -0.537 | 0.0107 | -50.1 | 0 | 0.0124 | -43.3 | 0 |
| B_TIME_c | -0.457 | 0.0595 | -7.69 | 1.49e-14 | 0.0818 | -5.59 | 2.21e-08 |
| B_TIME_pt | -0.327 | 0.0429 | -7.61 | 2.84e-14 | 0.0589 | -5.55 | 2.91e-08 |
| MU_CAR | 7.06 | 2.89 | 2.44 | 0.0146 | 4.99 | 1.41 | 0.157 |
| MU_CENTER | 1.66 | 0.132 | 12.6 | 0 | 0.118 | 14.1 | 0 |
| MU_EAST | 2.05 | 0.223 | 9.18 | 0 | 0.186 | 11 | 0 |
| MU_PT | 1.63 | 0.296 | 5.51 | 3.5e-08 | 0.433 | 3.76 | 0.00017 |
| MU_WEST | 1.98 | 0.209 | 9.48 | 0 | 0.177 | 11.2 | 0 |

## 4.4 Specification testing of the CNL model

A simple joint logit model has been estimated (the result of the estimation are in Appendix B), and we will test the cross-nested logit model according to a log likelihood ratio test, with a 95% confidence interval. The log likelihood of the logit model $\mathcal{L}_{logit}$ is -16755.54 and the one from the cross nested logit model $\mathcal{L}_{cn-logit}$ is -16732.25. To obtain the null hypothesis $H_0$, we first need to set the alphas to 1 or 0, which transforms the cross-nested logit in a nested logit model. Then, the nest parameters must be equal to 1, which means that $MU_{CAR} = MU_{PT} = MU_{EAST} = MU_{CENTER} = MU_{WEST} = 1$, which transforms the nested logit model to a simple logit model. Since no alphas were estimated, we have 5 less parameters in the null hypothesis, which leads to a degree of freedom of 5 and a $\chi^2$ value of 11.07 for a confidence interval of 95%. The likelihood ratio test is:

$$LR = -2(\mathcal{L}_0 - \mathcal{L}_1) = -2(\mathcal{L}_{logit} - \mathcal{L}_{cn-logit}) = -2(-16755.54 + 16732.25) = 46.58 \quad (6)$$

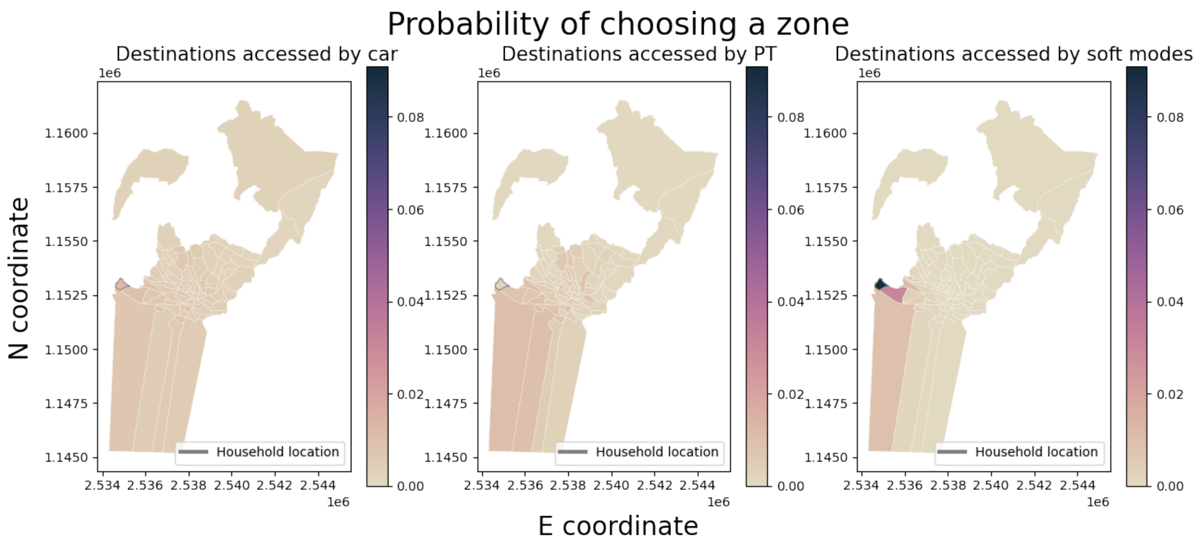
$$LR = 46.58 > \chi^2(95\%, 5) = 11.07 \quad (7)$$

Therefore, we can reject the null hypothesis at a 95% interval of confidence.

## 4.5 Choice set considerations

In this section, we present an example of a choice set, and the probability distribution associated to the cross-nested logit model for a household located in the western part of Lausanne and a working activity. The probability distribution (Figure 5) indicates that there is a higher chance to choose a zone next to the household location with soft modes. However, the distance has a strong impact (a figure of the distance distribution can be seen in Appendix A), and zones that are further away from the household location are not likely to be chosen. With driving, we observe that the probability to choose a zone is not significantly different. This unveils that a car gives a good accessibility to most of the zone in Lausanne, and is quite independent of the travel time. Finally, with public transport, zones that have a higher chance to be chosen are less homogeneous. This comes from the public transportation network, giving a better accessibility to zones that are served with subways and bus lines than others.

Figure 5: Probability distribution of choosing a zone




From the probability distribution shown in Figure 5, we draw a choice set of 15 alternatives. It is shown in Figure 6. We find a similar distribution as to what was previously observed in the probability distribution. First, zones that are accessed by soft modes, are zones close to the household. Then, zones accessed by driving seem to be evenly distributed. Finally, zones accessed by public transport appear to be on bus lines, and closer to the city center.

Using the choice set in Figure 6, we show an example of how we can include unlikely alternatives. Here, we apply 10 transformations, with the same probability to be chosen for an addition of alternative, a swap of transportation mode, and a change of zone to a neighbouring one. This results in 5 changes of mode of transportation, 3 swap of zones, and 2 random addition of alternatives. The new choice set can be observed in Figure 7. While the choice set is still similar to the one observed in Figure 6, some new and less likely alternatives are introduced.

Finally, using the same activity and household location, a perturbed choice set is drawn with a random term in each utility function. The random term is normally distributed,

Figure 6: An example of a choice set after importance sampling

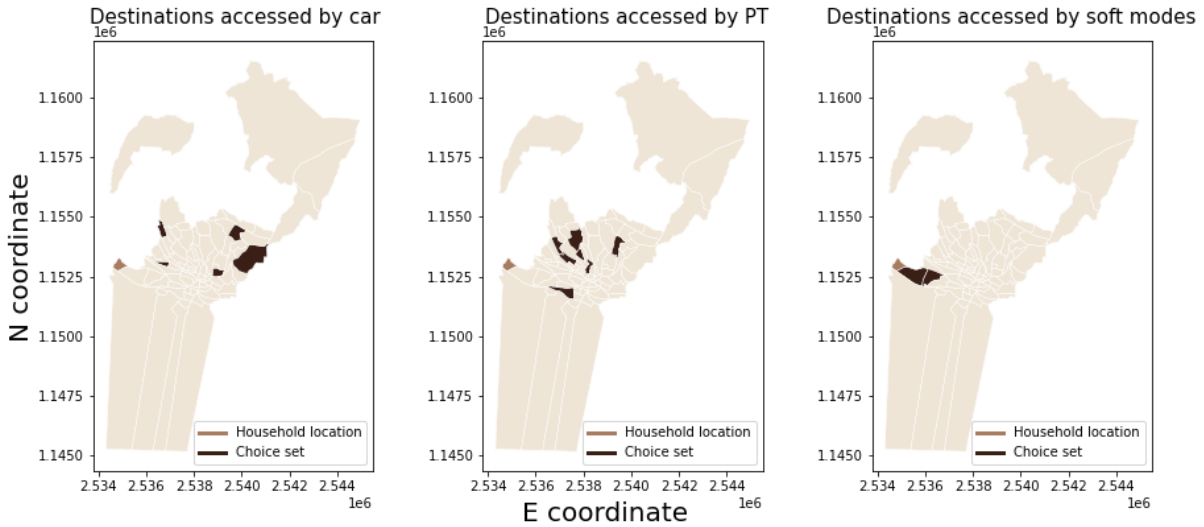
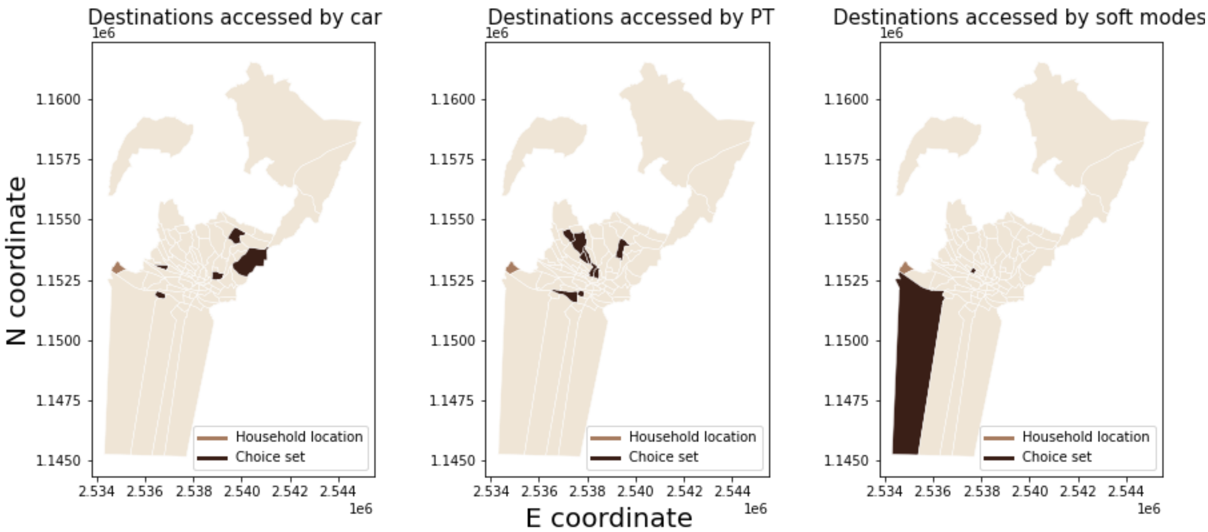


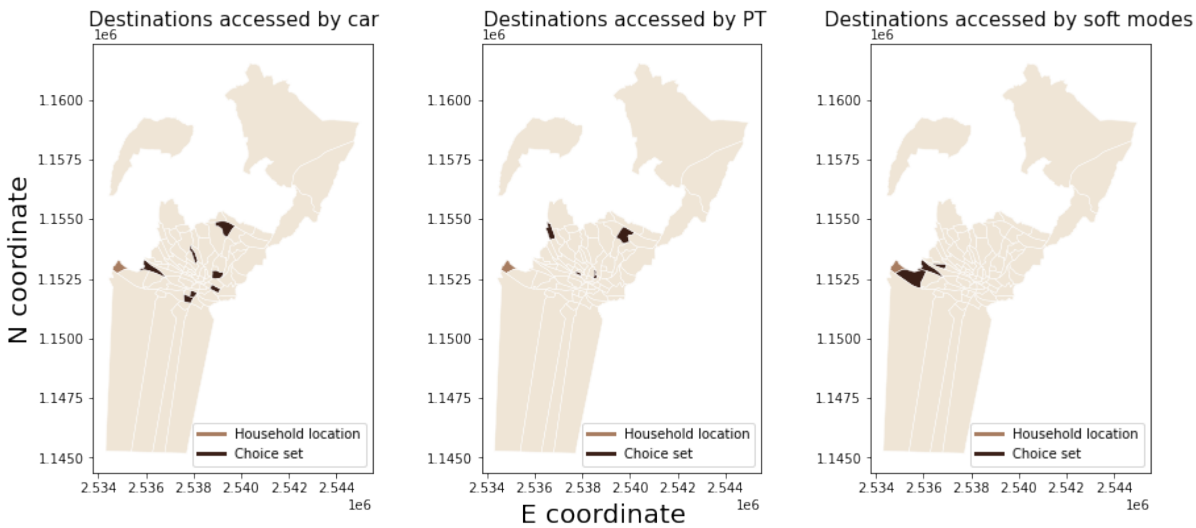

Figure 7: An example of a choice set after being transformed




centered around 0 with a scale of 1. The obtained choice set can be seen in Figure 8. Alternatives chosen are mostly different than the ones chosen in Figure 6 and 7, which implies that they would be less likely alternatives. However, it is difficult to assess on how much it differs. A deeper analysis will be conducted in 4.6.

Figure 9 compares the mode shares obtained from the MTMC dataset, the probability distribution and the choice sets generated. Note that this comparison is only done for the working activity. Mode shares in the MTMC are not the same depending if the trip

Figure 8: An example of a choice set with a random term in each utility function

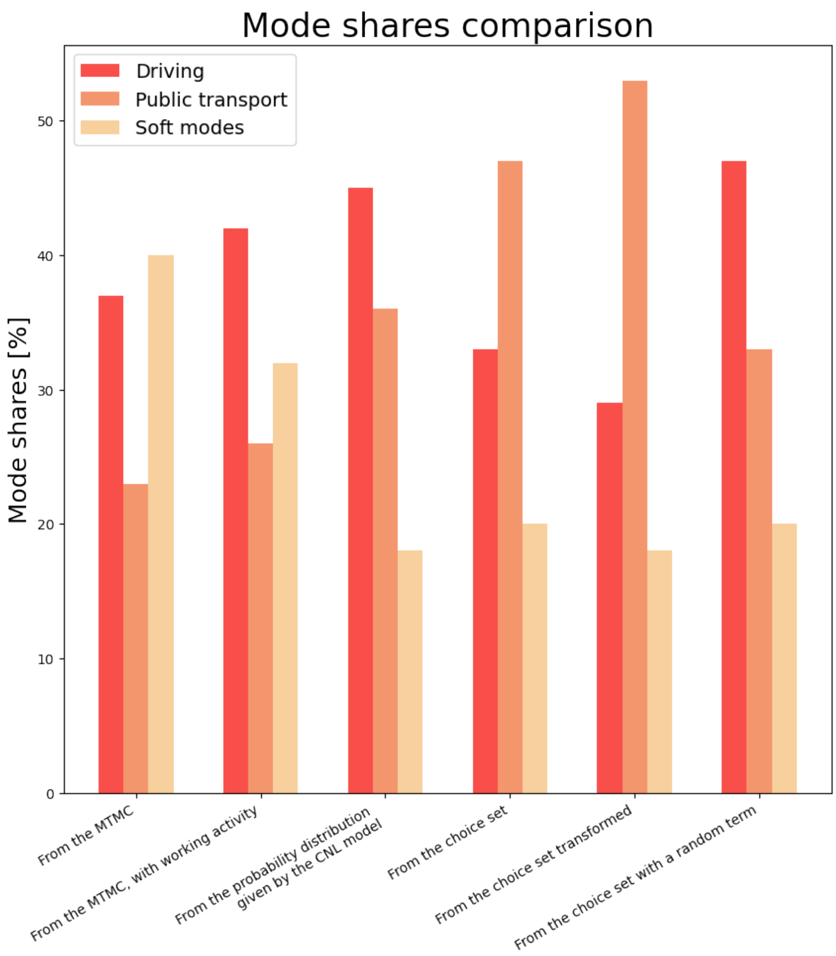



purpose is a working activity or not. We observe that they are not following exactly the one from the Microsensus, but this can be explained by the household location, being far away from the city center and, therefore, reducing the attractiveness of soft modes (since the number of job locations is higher in the city center). The proportion of alternatives in the choice set is not following the probability distribution due to the stochastic phenomena of importance sampling. If we were to draw a higher number of choice sets, and take the mean of their mode shares, the result would be much closer to the one from the probability distribution. We note that the transformations of the choice set take them even further away from the probability distribution, but that the random term gives much more similar proportions.

## 4.6 Sensitivity analysis and validation of the choice set

Since the choice set of individuals is not known, it is very difficult to assess the quality of a choice set. Nevertheless, the observed choice of individuals is in the dataset. In order to perform a sensitivity analysis, we make the assumption that the aggregate value of individuals choice sets should follow the same distribution than the distribution of chosen alternatives in the dataset. Ideally, we would need to record the choice set of individuals beforehand, and use this distribution for the comparison. Practically, we draw a thousand choice sets of 10 alternatives, for random household locations in Lausanne,

Figure 9: A comparison of mode shares




and calculate the final distribution of alternatives from the choice sets. This is done for all 5 activities, and for 3 different types of draws: importance sampling with no perturbation, importance sampling with some transformations, importance sampling with a random term. For the second one, three different numbers of transformations are applied, and for the last one, it is a normally distributed random term with three different values of $\sigma$. Note that for the transformation part, a change of zone, a swap of transportation mode and an addition of a random alternative have all the same probability to happen. To compare values in an aggregate way, the mean absolute error (MAE) is taken between each calculated distribution, and the one from the MTMC. Results are on Table 2.

Overall, the mean absolute error is the lowest with no perturbation, and increases almost always with the perturbation increasing. It means that if the perturbation increases, the choice set is more likely to include alternatives that are less chosen in the MTMC, hence

Table 2: Sensitivity analysis on the choice set perturbations

| **Type of draw** | **Parameters value** | **Work** | | **Education** | | **Shopping** | | **Leisure** | | **Other** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | MAE | Comput. time [s] | MAE | Comput. time [s] | MAE | Comput. time [s] | MAE | Comput. time [s] | MAE | Comput. time [s] |
| **No perturbation** | - | **0.263** | **40.991** | **0.426** | **43.343** | **0.365** | **43.183** | **0.247** | **43.770** | **0.308** | **43.519** |
| **With transformations** | 10 transfo. | 0.341 | 46.403 | 0.488 | 45.101 | 0.419 | 44.986 | 0.339 | 45.514 | 0.403 | 44.741 |
| | 15 transfo. | 0.350 | 45.576 | 0.487 | 46.042 | 0.436 | 45.770 | 0.353 | 45.552 | 0.404 | 45.555 |
| | 20 transfo. | 0.373 | 46.167 | 0.498 | 46.572 | 0.446 | 46.099 | 0.376 | 46.370 | 0.427 | 46.588 |
| **With a random term** | Sigma = 1 | 0.317 | 44.006 | 0.465 | 44.016 | 0.385 | 44.492 | 0.296 | **43.756** | 0.350 | **43.850** |
| | Sigma = 2 | 0.357 | 44.514 | 0.471 | **43.769** | 0.407 | 44.437 | 0.322 | 44.494 | 0.388 | 44.033 |
| | Sigma = 5 | 0.404 | 44.650 | 0.499 | **43.623** | 0.451 | **43.895** | 0.375 | 44.400 | 0.431 | 44.191 |

Table 3: Sensitivity analysis on transformations

| **Type of transformation** | **Work** | | **Education** | | **Shopping** | | **Leisure** | | **Other** | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MAE | Comp. time [s] | MAE | Comp. time [s] | MAE | Comp. time [s] | MAE | Comp. time [s] | MAE | Comp. time [s] |
| No transformation | 0.263 | 40.991 | 0.426 | 43.343 | 0.365 | 43.183 | 0.247 | 43.770 | **0.308** | 43.519 |
| All transformations | 0.350 | 45.576 | 0.487 | 46.042 | 0.436 | 45.770 | 0.353 | 45.552 | 0.404 | 45.555 |
| Only addition of an alternative | 0.466 | 44.021 | 0.542 | 50.533 | 0.506 | 47.780 | 0.464 | 46.609 | 0.509 | 46.967 |
| Only change of transportation mode | **0.251** | 49.536 | **0.413** | 46.701 | **0.356** | 54.303 | **0.245** | 46.721 | 0.320 | 46.873 |
| Only swap of zone to a neighbour | 0.254 | 46.745 | 0.425 | 47.190 | 0.372 | 47.210 | 0.254 | 46.443 | 0.320 | 47.348 |

less likely alternatives. Except for a working activity, introducing a normally distributed random term with a $\sigma = 5$ seems to perturb the choice set as much as applying 20 transformations. But the perturbation with a random term is faster, by 2-3 seconds, than applying 20 transformations. Note also that changing $\sigma$ does not significantly change the computational time, where applying more transformations increase computational time. Therefore, perturbing the choice set with a random term seems more suitable than applying simple transformations.

A deeper sensitivity analysis is conducted on the type of transformation, where each transformation is compared to both no transformation and all transformations at the same time. Results are shown on Table 3. We observe that 15 additions of random alternatives give a choice set with worse MAE for all activities. Changing the transportation mode or swapping a zone to a neighbouring one leads to a similar MAE as if there was no transformation. We conclude that only adding random alternatives successfully perturbs the choice set, whereas other transformations are not strong enough. For computational time, there are no significant tendencies that can be identified, except that it is almost always worse than applying the three transformations at once.

## 4.7 Schedule generation

To give an example of the full modelling pipeline (see Figure 1), we will generate schedules using the CNL model estimated previously. Note that the step 2, which is estimating the activity-based choice model parameters will not be done, due to computational limitations. The ABM used to generate the schedule is the one defined in Pougala *et al.* (2021), which optimise schedules, with respect to their utility, defined as:

$$U_S = U + \sum_{a=0}^{A-1}(U_a^1 + U_a^2 + U_a^3 + \sum_{b=0}^{A-1}(U_{a,b}^4 + U_{a,b}^5)) \tag{8}$$

where in this equation, there are $A$ activities, and:

- U: a generic utility.
- $U_a^1$: utility associated from the activity participation.
- $U_a^2$: utility associated with the starting time of the activity.
- $U_a^3$: utility associated with the activity duration.
- $U_{a,b}^4$: utility associated with the trip from the location of $a$ to the location of $b$, but without considering travel time.
- $U_{a,b}^5$: utility associated with the travel time from the location of $a$ to the location of $b$.

For the generic utility, and utility 1 to 4, the parameters are left untouched. Here, we make the implicit assumption that these parameters would not be significantly different if we re-estimate them with a choice set. This assumption is made mainly because of computational purposes, since estimating an activity-based choice model involves a lot of variables, and is therefore computationally very costly. On the other hand, $U_{a,b}^5$ is replaced by the utility described in section 3.3, with the parameters estimated from the CNL model (see section 4.3):

$$U_{a,b}^5 = V_{a,b}^5 = V_{zmn} \tag{9}$$

Note that the utility is scaled down by a factor of 10, in order to keep a similar weighting than in the initial utility of schedules. Finally, to avoid long computational times, the ABM has been run with a choice set of 10 alternatives, for each activity. These choice sets have been generated with importance sampling, as in section 3.4.1. They are stated down in Appendix D.2. Note that the choice set is different for each activity (since it is an exogenous variable in the utility function), and that it does not need to be perturbed,

Figure 10: Final schedule for Alice

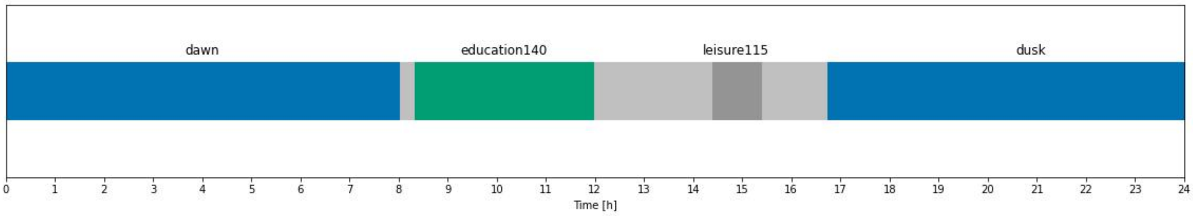
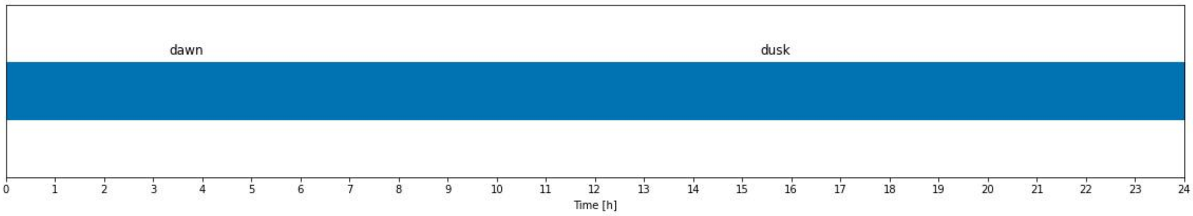



Figure 11: Final schedule for Brian




because it is not used for parameters estimation. Also, for simplicity, the chosen alternative is not added in the choice set, since most of activity destinations were not in Lausanne. In the ABM, schedules have been optimised for 5 iterations.

Schedules have been generated for two different individuals, who will be called Alice and Brian. Their initial schedules from the MTMC, as well as their preferences, are shown in appendix D.1. For more informations about how the preferences were chosen, we refer to Pougala *et al.* (2021). Their schedules after optimisation are drawn in Figure 10 and 11. For Alice, we observe that she is going by public transport to her education and leisure activities. For the education, a destination in the city center of Lausanne (zone 52, see Appendix C) is chosen. The choice seems suitable, since there are various types of schools in this zone. The leisure destination is in the northern area of Lausanne (zone 27, see Appendix C), in the forest of Sauvabelin. On the other hand, Brian preferred to stay at home. We can imagine that he did his education activity online, and canceled its shopping activity.

If we compare the computational time between Alice and Brian in Table 4, we observe a huge difference. This might be due to one more mandatory activity to do for Alice, which increases the number of potential alternatives. Indeed, there are 20 alternatives for the morning and afternoon education activities, and 10 alternatives for the leisure activity. In

Table 4: Summary of schedules generation

| | **Number of alternatives to choose from** | **Number of activities initially** | **Number of activities in the final schedule** | **Computational time [hours]** |
|---|---|---|---|---|
| Alice | 34 | 7 | 4 | 7.089 |
| Brian | 23 | 5 | 2 | 0.176 |

addition, there are the 4 initial home activities. This larger number of alternatives could mean that the computational limits of the computer (8Go of RAM memory, 4-core at 1.8GHz processor) have been reached. Otherwise, we observe that the number of activity to perform is lower than the one schedules originally, meaning that the utility to perform an activity can be more costly than doing nothing.

## 4.8 Discussion of assumptions

In this subsection, we estimate the impact of the assumptions (see section 4.1) on the model. In other words, how strongly these assumptions can cause the model to be wrong. We also provide an estimation on the complexity surrounding their resolution, meaning how easily we could relax these assumptions, with some suggestions on how to achieve it. Typically, a high complexity to resolve assumptions requires to collect various data within the population, or, alternatively, to have a big computational power. A low complexity means to adapt the methodology which is achievable but requires some implementation time. Table 5 summarises this.

Table 5: Estimation of the impact on the model of assumptions and complexity to resolve them

| **Assumption (see Section 4.1)** | **Impact on the model** | **Why does it impact the model** | **Complexity to resolve** | **Idea to resolve** |
|---|---|---|---|---|
| 1 | Medium | Unique informations about zones cannot be captured | Medium | Unique ASC for all alternatives, but it requires a high computational power |
| 2 | Medium | While this assumption is very wrong, it might only concern a minority of trips, which mitigates its impact | Medium | Addition of some rules, such as car licence or owning a car for car availability, or to limit travel time to zones over a specific threshold |
| 3 | Low | From the literature, it is possible to rebuild a point location afterwards | Low | Rebuild the point destination after computing zones in the choice set |
| 4 | Medium | While this assumption can be valid for short range trips (less than 1km), it becomes harder to justify it for mid range trips (1-5km) | High | Collect informations about biking and walking travel time as well as biking availability |
| 5 | High | This could be very wrong, especially in a urban context | High | Collect data about the actual travel time with soft modes |
| 6 | Low | Travelling time and comfort of these different transportation modes are not significantly different | High | Travelling with public transport often includes several modes of transportation. Detailed data about trips would be needed |
| 7 | Medium | Travel time is not substantially different, but the comfort can change | High | Separating these modes would require more data and would increase the number of alternatives |
| 8 | Low | Allows to compute probability of choosing alternatives | Medium | Increase model complexity |
| 9 | Low | Smaller zones attractivity is shown in a better way | Low | The results with density have shown to be significant |

Assumption 14 has been dealt with by performing a specification testing that shown the significance of using a CNL model. Assumptions 1, 2, 3 and 8 could be dealt with assuming that one has big computational power available, and more time. For assumptions 4, 5, 6 and 7, it would require to collect more data, which is a burdensome task.

# 5 Conclusion and further work

In this paper, we have shown a way to generate a choice set of destinations, jointly with the transportation mode. By estimating a cross-nested logit model first, and then perform importance sampling, we could obtain a choice set for an individual, given the origin of the trip and the activity performed. We have also provided the probability to compute the choice set. This probability is crucial if the choice set is then used to estimate a choice model. It allows to correct the maximum likelihood estimation to avoid biased parameters. A case study over the city of Lausanne, Switzerland has been conducted. This showed how to draw an example of a choice set and how to transform it. The sensitivity analysis revealed that perturbing the choice set with a normally distributed random term leads to better results than with simple transformations. Parameters of the cross-nested logit model have been reused to modify the ABM defined by Pougala *et al.* (2020), and generate schedules with the methodology. While deeper investigations would be needed, we could still observe the trade off that an individual do when scheduling activities, especially the cost from travelling.

While we have shown examples of the methodology in the case study, future work could be conducted in order to validate the cross-nested logit model on a larger scale, and the schedules that are generated. This might be done by finding the same distribution of activities, as in the Microsensus dataset. Deeper investigation could be conducted to compare the transportation mode and destination shares obtained with this methodology and the one observed in the data, over Switzerland. This would require to adapt the methodology with some rules, so that the computational time is reasonable.

Additionally, further work could be done to tackle some of the assumptions mentioned in section 4.1 and 4.8. Special attention could be put on the variables of the utility function, by investigating the addition of other variables. For example, we could add a proper trip cost, trip comfort, or having a travel time that depends on a congestion level. With unlimited resources, we should also collect data about the number of schools or universities per zone, as well as the number of shops, in order to combine them to an education or shopping activity. Additionally, with better computational resources, activities could be jointly estimated in the cross-nested logit model, and more variables in the model could be added, such as alternative specific constant for each alternative.

Finally, in the case study, the activity-based choice model has not been re-estimated. The bias coming from the choice set are considered with a random parameter in the model. Ideally, one should re-estimate the model with the generated choice set as input, and with a correction to the likelihood estimation. Since the choice set probability can be computed, the parameters of the activity-based choice model could be re estimated if the computational resources are sufficient.

# 6 References

Ben-Akiva, M. and B. Boccara (1995) Discrete choice models with latent choice sets, *International Journal of Research in Marketing*, **12** (1) 9–24, ISSN 0167-8116.

Ben-Akiva, M. E., S. R. Lerman, S. R. Lerman *et al.* (1985) *Discrete choice analysis: theory and application to travel demand*, vol. 9, MIT press.

Bierlaire, M. (2020) A short introduction to PandasBiogeme, 22.

Bierlaire, M. and R. Krueger (2020) Sampling and discrete choice, 31.

Bowman, J. and M. Ben-Akiva (2001) Activity-based disaggregate travel demand model system with activity schedules, *Transportation Research Part A: Policy and Practice*, **35** (1) 1–28, ISSN 09658564.

Frejinger, E., M. Bierlaire and M. Ben-Akiva (2009) Sampling of alternatives for route choice modeling, *Transportation Research Part B: Methodological*, **43** (10) 984–994, ISSN 01912615.

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

Hagerstrand, T. (1970) What about people in regional science, *Ninth european congress of the regional science association*, 7–21.

Pagliara, F. and H. Timmermans (2009) Choice set generation in spatial contexts: a review, *Transportation Letters*, **1** (3) 181–196, ISSN 1942-7867, 1942-7875.

Pougala, J., T. Hillel and M. Bierlaire (2020) Capturing trade-offs between daily scheduling choices, 27.

Pougala, J., T. Hillel and M. Bierlaire (2021) Choice set generation for activity-based models, 28.

Scherr, W., C. Joshi, P. Manser, N. Frischknecht and D. Métrailler (2019) An Activity-based Travel Demand Model of Switzerland Based on Choices and Constraints, 13.

Schönfelder, S. and K. W. Axhausen (2003) Activity spaces: measures of social exclusion?, *Transport Policy*, **10** (4) 273–286, ISSN 0967070X.

Scrogin, D., R. Hofler, K. Boyle and J. Walter Milon (2010) An efficiency approach to choice set formation: theory and application to recreational destination choice, *Applied Economics*, **42** (3) 333–350, ISSN 0003-6846, 1466-4283.

Shocker, A. D., M. Ben-Akiva, B. Boccara and P. Nedungadi (1991) Consideration set influences on consumer decision-making and choice: Issues, models, and suggestions, 17.

Swait, J. (2001) Choice set generation within the generalized extreme value family of discrete choice models, *Transportation Research Part B: Methodological*, **35** (7) 643–666, ISSN 01912615.

Swait, J. and M. Ben-Akiva (1987) Incorporating random constraints in discrete models of choice set generation, *Transportation Research Part B: Methodological*, **21** (2) 91–102, ISSN 01912615.

VM-UVEK (2019) Verkehrsmodellierung im uvek; daten zustand 2017, `https://doi.org/10.5281/zenodo.3846547`.

# A Distance distribution

Figure 12 shows the zone-to-zone distance distribution from the household location used in section 4.5.

# B Logit model parameters

Table 6 shows the value of parameters estimated in the Logit model.

Table 6: Logit model parameters

| **Name** | **Value** | **Std err** | **t-test** | **p-value** | **Rob. Std err** | **Rob. t-test** | **Rob. p-value** |
|---|---|---|---|---|---|---|---|
| ASC_c | -3.88 | 0.0721 | -53.8 | 0 | 0.0785 | -49.4 | 0 |
| ASC_pt | -2.66 | 0.129 | -20.7 | 0 | 0.135 | -19.8 | 0 |
| B_JobD_PopD_SHOP | 0.0174 | 0.00189 | 9.21 | 0 | 0.00176 | 9.92 | 0 |
| B_JobD_WORK | 0.00759 | 0.00161 | 4.72 | 2.37e-06 | 0.00173 | 4.39 | 1.11e-05 |
| B_PopD_EDUC | -0.0371 | 0.00871 | -4.26 | 2.05e-05 | 0.0081 | -4.58 | 4.59e-06 |
| B_PopD_LEIS | -0.0281 | 0.00403 | -6.96 | 3.42e-12 | 0.00424 | -6.62 | 3.68e-11 |
| B_PopD_OTHE | -0.0313 | 0.00641 | -4.88 | 1.04e-06 | 0.00625 | -5.02 | 5.29e-07 |
| B_TIME_act | -0.592 | 0.0119 | -49.9 | 0 | 0.0142 | -41.6 | 0 |
| B_TIME_c | -0.733 | 0.0419 | -17.5 | 0 | 0.0451 | -16.3 | 0 |
| B_TIME_pt | -0.525 | 0.0276 | -19 | 0 | 0.028 | -18.7 | 0 |

# C Group of destinations and their index

Figure 13 shows the index number of zones, as well as how they are grouped for the cross-nested logit model.

Figure 12: Zone-to-zone distance distribution

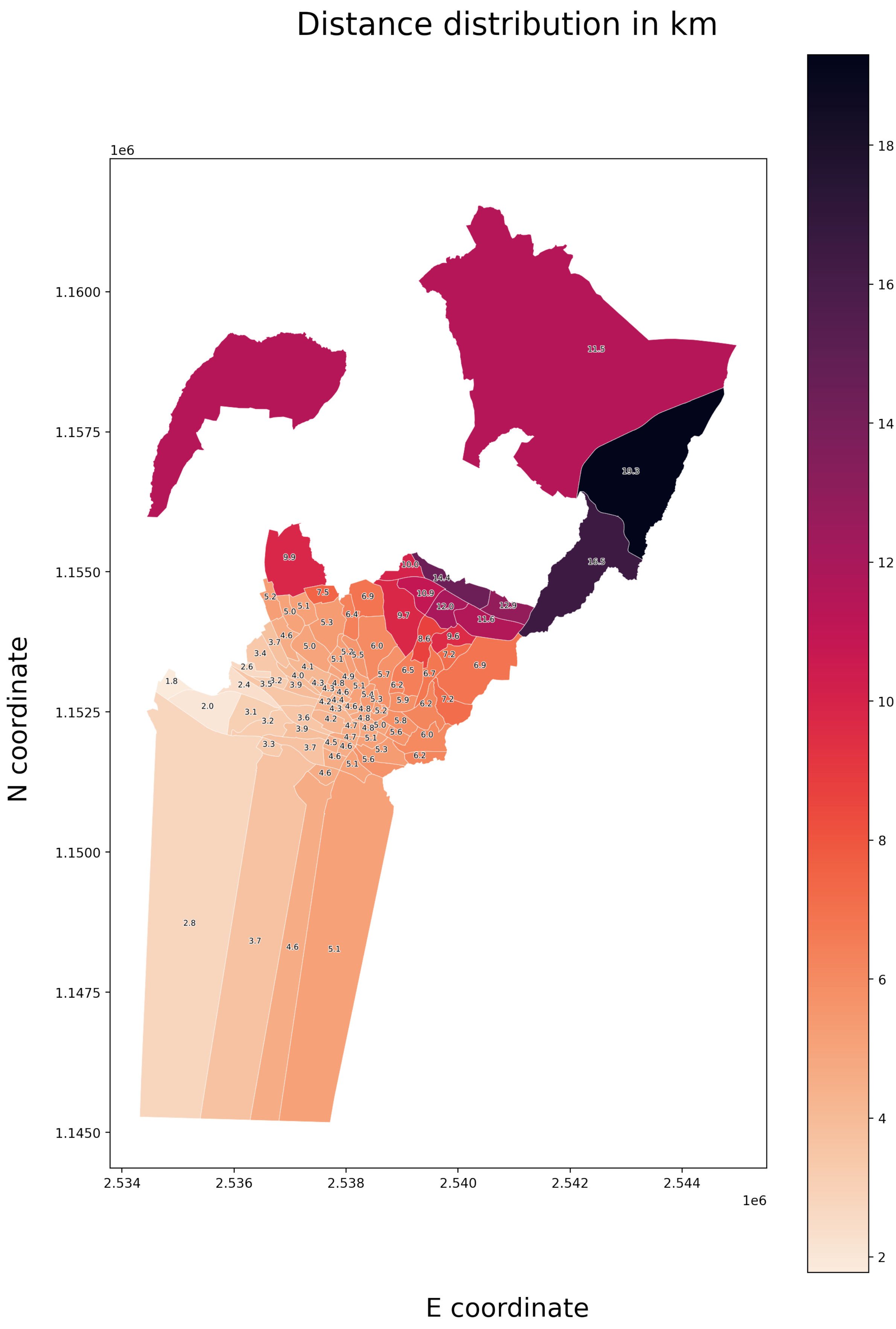
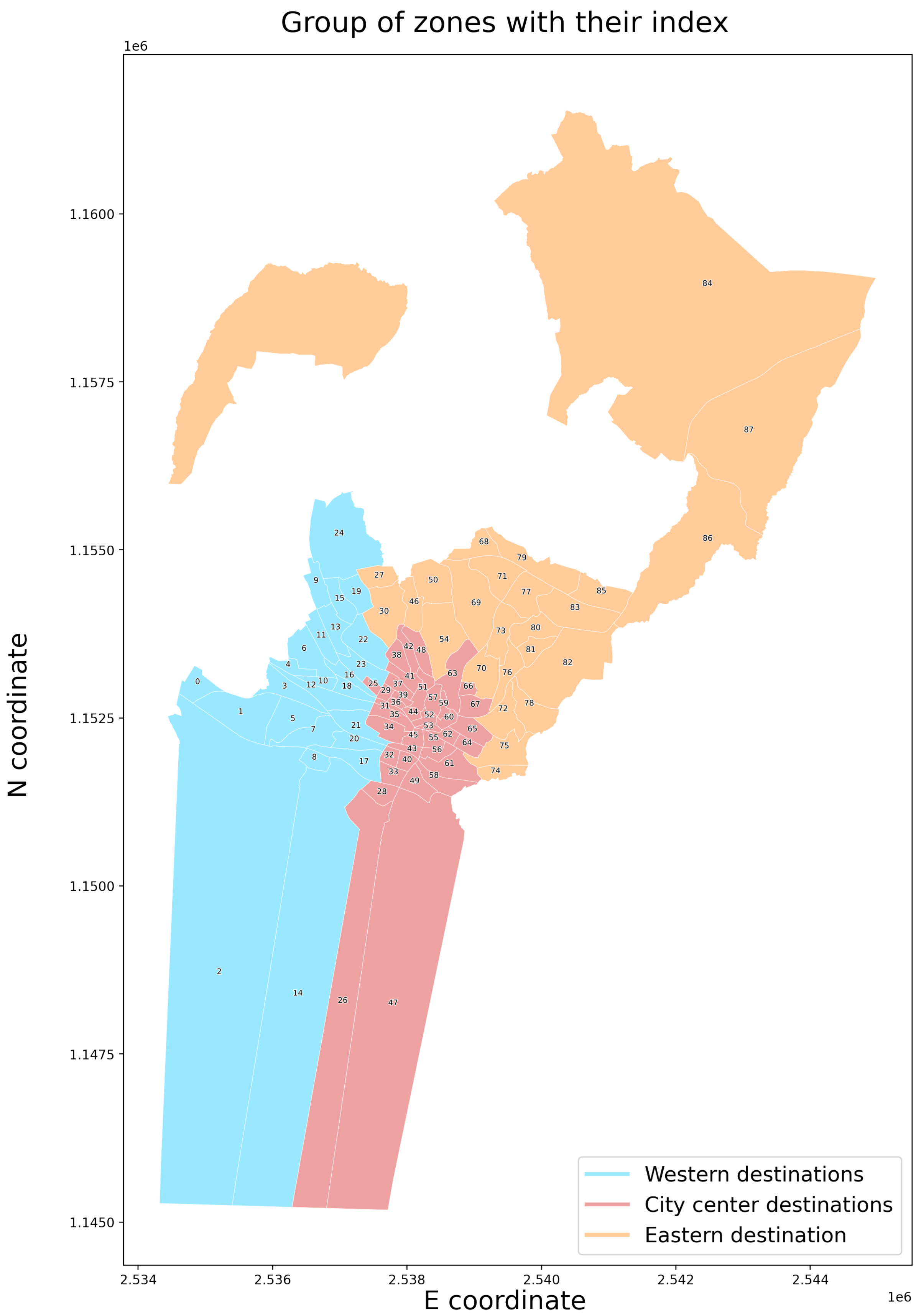


Figure 13: Destinations by group and with their index



# D Original schedules and choice set for Alice and Brian

## D.1 Original schedules

Table 7: Initial schedules and preferences of Alice

| **Activity number** | **Activity label** | **Label** | **Start time** | **End time** | **Duration** | **Feasible start** | **Feasible end** | **Location** | **Categories** | **Flex early** | **Flex late** | **Flex short** | **Flex long** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | home | dawn | 0 | 8.25 | 8.25 | 0 | 24 | (46.6033, 6.67643) | discret | F | M | F | F |
| **2** | education | education1 | 8.33 | 12 | 3.63 | 7 | 23 | (46.6014, 6.68191) | mandat | R | R | R | M |
| **3** | home | home1 | 12.2 | 13.2 | 1.08 | 0 | 24 | (46.6033, 6.67643) | discret | F | M | F | F |
| **4** | education | education2 | 13.5 | 16.2 | 2.75 | 7 | 23 | (46.6014, 6.68191) | mandat | R | R | R | M |
| **5** | home | home2 | 16.5 | 17 | 0.5 | 0 | 24 | (46.6033, 6.67643) | discret | F | M | F | F |
| **6** | leisure | leisure | 17.1 | 18 | 0.917 | 0 | 24 | (46.6014, 6.68191) | discret | F | M | F | F |
| **7** | home | dusk | 18.1 | 24 | 5.92 | 0 | 24 | (46.6033, 6.67643) | discret | F | M | F | F |

Table 8: Initial schedule and preferences for Brian

| **Activity number** | **Activity label** | **Label** | **Start time** | **End time** | **Duration** | **Feasible start** | **Feasible end** | **Location** | **Categories** | **Flex early** | **Flex late** | **Flex short** | **Flex long** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | home | dawn | 0 | 7.33 | 7.33 | 0 | 24 | (46.5509, 6.63299) | discret | F | M | F | F |
| **2** | home | home1 | 7.42 | 7.42 | 0 | 0 | 24 | (46.558, 6.63459) | discret | F | M | F | F |
| **3** | education | education | 7.47 | 12.1 | 4.62 | 7 | 23 | (46.558, 6.63459) | mandat | R | R | R | M |
| **4** | home | home2 | 12.3 | 16.2 | 3.92 | 0 | 24 | (46.5509, 6.63299) | discret | F | M | F | F |
| **5** | shopping | shopping | 16.5 | 18.5 | 2.03 | 7 | 20 | (46.5226, 6.62739) | discret | F | M | F | F |
| **6** | home | dusk | 19 | 24 | 5 | 0 | 24 | (46.5509, 6.63299) | discret | F | M | F | F |

## D.2 Choice sets

Table 9: Choice sets per each activity for Alice and Brian

| | **Activity** | | **Choice set** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| **Alice** | Education | Zone ID | 75 | 87 | 82 | 59 | 84 | 60 | 51 | 67 | 54 | 34 |
| | | Transportation mode | Car | Car | Car | Car | Car | Car | PT | PT | Car | PT |
| | Leisure | Zone ID | 71 | 77 | 59 | 11 | 13 | 3 | 40 | 26 | 46 | 72 |
| | | Transportation mode | Car | Car | PT | Car | Car | Car | PT | PT | PT | PT |
| **Brian** | Education | Zone ID | 22 | 71 | 79 | 22 | 37 | 24 | 68 | 81 | 64 | 44 |
| | | Transportation mode | SM | Car | PT | Car | PT | Car | SM | Car | Car | Car |
| | Shopping | Zone ID | 64 | 65 | 87 | 4 | 30 | 28 | 62 | 50 | 54 | 60 |
| | | Transportation mode | PT | Car | Car | Car | PT | PT | SM | PT | Car | Car |