PhD-FSTM-2022-030
The Faculty of Sciences, Technology and Medicine

DISSERTATION

Defence held on 01/04/22 in Esch-sur-Alzette

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN SCIENCES DE L'INGÉNIEUR

by

Ariane SCHEFFER
Born on 2nd April 1992 in Luxembourg, (Luxembourg)

DYNAMIC ACTIVITY-BASED DEMAND ESTIMATION CONSIDERING DESTINATION AND MODE CHOICE

Dissertation defence committee

Dr Francesco Viti, dissertation supervisor
Professor, Université du Luxembourg

Dr Joachim Arts, Chairman
Professor, Université du Luxembourg

Dr Marialisa Nigro
Professor, University of Roma Tre

Dr Guido Cantelmo
Assistant Professor, Technical University of Denmark

Dr Richard Connors
Research Scientist, Université du Luxembourg

Table of Content

Table of Figures

Table of Tables

Notations

Table I. 1 Table of notations

Notation	Description
t	Discrete time of day $t \in \{t_0, t_0 + \Delta t, t_0 + 2\Delta t, \dots, t_0 + (N-1)\Delta t\}$
Δt	Time-step of discretisation
$[a_t^i] = a^i$	Activity of individual i at time t
$[o_t^i] = o^i$	Origin zone of individual i at time t
$[d_t^i] = d^i$	Destination zone of individual i at time t
$[A_j^i] = A^i$	Sequence of (chained) activities of individual i of length n^i
$n^i = \lvert A^i \rvert$	Number of distinct activities in activity chain of individual i
$t^{i,j}$	Time points of the j-th activity in chain of i
$t_s^{i,j}, t_e^{i,j}$	Start/End time of j-th activity in chain of i
$t_s^{i,yz}, t_e^{i,yz}$	Start/End time of trip for individual i travelling from zone y to zone z
tt_a^{yz}	The travel time from zone y to zone z to engage in activity a
c_t	Cost of travelling
$Q^y(t)$	The number of people engaged in non-travel activities in zone y at time t
D_a^z	The number of people per day engaging in activity a in zone z
$T^{yz}(t)$	Total number of trips departing zone y at time t to travel to zone z
$T^{y\rightarrow}(t)$	Total outgoing trips from zone y at time t
$u_{a,z}(t, t_s)$	Marginal utility for performing an activity a in zone z at time t, having started at time t_s
u_t	Marginal utility of travel is constant ($u_t \leq 0$)
$U_a(t_s, t_e, z, m)$	The utility accrued from engaging in activity a from time t_s to t_e in zone z, travelling with mode m
U^i	The total utility accrued for the entire activity chain of individual i

$P_a^{yz}(t_s, t_e, m)$	Probability of departing from zone y to go to zone z with mode m in order to engage in activity a with start time t_s and end time t_e.
$P_a^{yz}(t)$	Probability of departing from zone y at time t, to go to zone z for activity a.
$P_a(t_s, t_e)$	Probability of engaging in activity a from time t_s, to time t_e
σ	Scale parameter of the Multinomial Logit model
W_a	Probability distribution of activity a starting time

Parameters of the Activity-Specific Utility Functions

U_a^{max}	Magnitude-parameter of the marginal utility for activity a
α_a	Axis location-parameter of the marginal utility We will also sometimes omit for sake of readability and without loss of generality the subscript a
β_a	Steepness-parameter of the marginal utility. For the same reason above, subscript a can be at times omitted.
τ_a	Starting time impact-parameter on the marginal utility. For the same reason above, subscript a can be at times omitted.
γ_a	Skewness-parameter of the marginal utility. For the same reason above, subscript a can be at times omitted.
$\upsilon_{a,z}$	Relative attraction factor of zone z for activity a
$\mathbf{\Theta} = (U^{max}, \alpha, \beta, \tau, \gamma)$	Set of parameters to be estimated for each marginal utility function of interest
θ	Any single element of $\mathbf{\Theta}$
θ'	Proposed parameter for the Markov Chain related to element θ
$P(\mathbf{\Theta})$	Probability distribution function of the parameter independently from any evidence
$P(D)$	Probability distribution function of D independently from any parameter value
$\pi(\mathbf{\Theta})$	Prior density of the parameter $\mathbf{\Theta}$
Δ_θ	Step size of parameter θ

i	Iteration number
r	Residual
ω	Weight of the scoring function
ρ_n	Scale factor applied at the n^{th} component of the likelihood
$\mathcal{L}$	Likelihood function
S	Total score value
φ	Acceptance ratio

Chapter 1

Highlights of the chapter

1. Issues and challenges of mobility and its modelling in today's world motivate the development of new approaches able to assess new transport and mobility services
2. Advantages of a macroscopic activity-based demand models are discussed, in particular in terms of mathematical tractability and estimation accuracy
3. Application opportunities of the proposed model are introduced

I. INTRODUCTION

I.1. Context and motivation

Despite the growing ecological responsibility, the development of new technologies, new modes of transport and new mobility solutions, traffic caused by people's journeys is still a source of considerable problems in dense urban areas. Not only are the environmental consequences significant, but also the time lost, and the dissatisfaction of users have a negative impact on society despite very high investments cost for transport in Western Countries. To address issues related to mobility, we believe it is essential to better understand how human movements work and what their spatial and temporal interactions are, to better estimate and predict the demand for transport and mobility services.

A way to better understand how these movements work is to understand their cause, their type and above all to know what the alternatives are and what would make these alternatives more attractive in terms of, for instance, mode choice. In the current context, better understanding is also made difficult by the growing complexity of the mobility offer caused by the increase of multimodal trips, shared mobility, and the complex patterns emerging from individual activity-trip chains. In Luxembourg, the levels of congestion and car usage are particularly high. To counter that, the country set a series of strategies aimed at achieving objectives to address these issues and align themselves with European sustainability objectives. The mid-term mobility targets are ambitious and require adequate strategies to be achieved as they cannot be only accomplished through current infrastructural response. These objectives concern modal share for commuting with a particular attention given to the reductions of car drivers share with respect to car passengers. This cannot be done without increasing car occupancy drastically, e.g. via shared mobility services. They also concern the decarbonisation and reduction of CO_2 generated by transport which is a direct consequence of the reduction of car usage and the electrification of the national fleet.

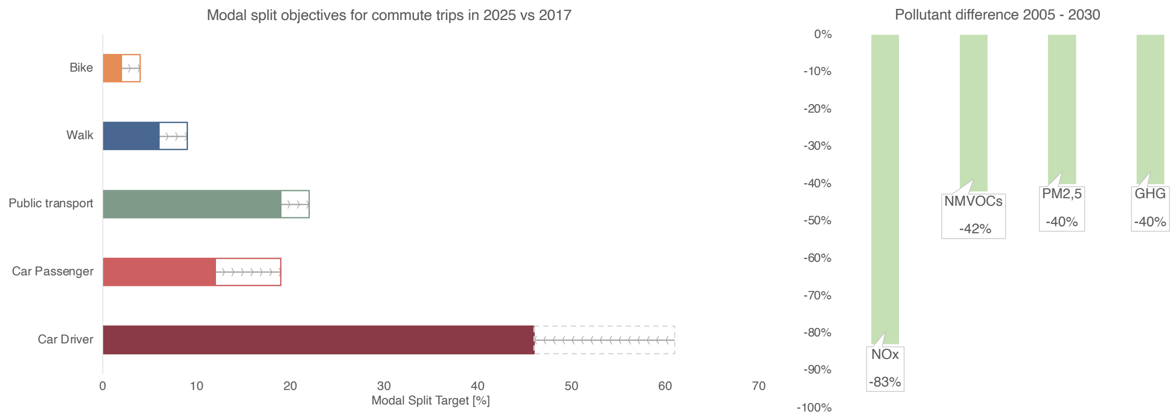


Figure I. 1 Objectives 2025-2030 in Luxembourg (source MoDu 2.0) (a) Modal Split (b) Emissions

The usage of electric vehicles and services like carpooling bring challenges in terms of demand modelling. Demand modelling and forecasting are usually used to help making right decisions in terms of transport planning and policy making by understanding how many trips people will make and what will be their characteristics. We believe that an improvement of the current situation in terms of mobility can be achieved with the help of the three following steps:

1. Firstly, estimating and modelling current and future mobility patterns in terms of flows and modal choice,
2. then, assessing the impact of planning and management solutions, and finally
3. recommending how to redesign future networks.

In this thesis we focus mostly on the first step while keeping in mind the necessary application and thus implementation constraints. That is, we focus on the applicability of the proposed model in the perspective of large-scale transport network planning. In the current context, especially with resource sharing and other policies encouraging the use of electric vehicles, it is important to know not only the characteristics of the trip itself but also those of the idling periods and of the following trips. It is important to do so since, for instance, it may have an impact on the charging needs.

While for years traffic was modelled without much concern for the reason of a movement but focusing more on individual trips and their effects on the network, the trend today is quite different. People move to perform specific activities and many modellers seek to understand what aspects of their activity chain conduct them to making decisions related to their related movements chain. In this work, we want to identify the ideal point where we can in some way benefit from two well-established modelling approaches: trip-based models and activity-based models. Both bearing several advantages and disadvantages, we propose a model that allows us to reproduce the desired aspects of these two branches. To do so, it is reasonable to consider the following assumptions:

- Travelling is a derived demand generated from the need of performing or participating to an activity
- Travellers are rational decision-makers who try to maximize their utility
- This need varies from one activity type to the other (from basic needs to self-fulfilment needs)
- The travel itself brings a disutility to the user
- This disutility is perceived differently depending on the activity executed at the destination
- Attraction of a destination is higher for certain areas which can be worth longer trips

These assumptions are common in demand estimation for transport planning in order to model users' behaviour (Ortuzar and Willumsen 2011). We aim at developing a model which can consider these hypotheses inside the demand estimation framework and be applied for short-term to long-term planning and resolution strategies.

I.2. Objectives and scope

To ensure the possible use of the proposed model in applications where the available resources do not allow for the representation of each individual, we propose in this work to focus on a macroscopic representation of the demand. This important issue, in the context of data protection is also motivated by the computational complexity of modelling single individuals and their interactions. In order to develop a robust model that can be calibrated and integrated with traffic assignment models, we also choose to aggregate demand at the scale of analysed traffic zones (TAZ). At the temporal level, we target a modelling of a 24-hour scenario in order to be able to capture the flows emerging from full daily travel chains inside the study area. The dynamics addressed in this thesis are thus spatial on the one hand and temporal on the other. Furthermore, we give a particular importance to mode choice modelling as one of the major solutions to reducing emissions and congestion is to shift from single car use to other modes of transport.

The main research question we want to answer in this thesis is therefore:

"Is it possible to capture spatial and temporal distribution of activity- and mode-specific flows over a day from aggregate data?"

The complexity in answering this question lies in the fact that the modal choice is very complex and involves many criteria that are characterised by different levels of importance and have different degrees of complexity to model, especially at a macroscopic scale. This is why we seek to reproduce emergent behaviours, i.e. patterns that appear when all individuals are seen as a whole and their singular behaviour result in recognizable trends. Also, even the simplest individual choices are correlated with each other and therefore we are looking for a "mesoscopic" approach to meet the feasibility criteria described above. Even though this term is usually used in traffic modelling, we apply it to demand modelling including behavioural individual characteristics to be used at the macroscopic application level. The trade-off between the rigorous level of detail obtained by activity-based models and that required for the validation of conventional traffic assignment models is ideally incorporated in the proposed estimation model. In order to answer this main question, we propose four intermediate questions that allow us first to explain, then to model and quantify and finally to predict these dynamics.

RQ1: Can we quantify how much commuting mode choice has an impact on other activities or trips?

This first question concerns the understanding of the existing dynamics and especially the correlations underlying modal flows. In particular, we seek to understand to what extent the modes of transport used are governed by characteristics specific to a chain of trips and activities, beyond the characteristics of the trip itself. In reference to the described dichotomy between modelling approaches, one of the main aspects of activity-based models is that they are also, often tour-based models. This means that all the trips belonging to a same tour, i.e. the chain of trips done starting and ending at a given location are jointly modelled. The car usage reduction objectives focus on the commuting trip. Indeed, the highest levels of congestion are often observed during the morning peak hour and these trips are often the most constrained ones in terms of destination, starting time and frequency. It is rational that modal choices are correlated to each other and that decisions are often made for an entire agenda and not just a single trip. It is therefore important to know and quantify the impact of commuting mode choices on the entire trip chain. We are therefore looking to find a way of linking, at the population level, the characteristics of a journey with the places visited so as to be able to use clear and precise indicators in predicting them. A sensible approach is to use existing models of destination choices and definition of user activity space in order to adapt them to homogeneous groups within the studied population. To answer this question, we use a multi-day travel survey collected in Belgium and analyse in particular the daily movement of commuters belonging to this database. This question is addressed in the first part of the thesis.

RQ2: Can the trade-off governing individual activity scheduling explain emerging travel flows?

The second question is motivated by empirical observations described in the first part of this work but introduces the issue of modelling. The first step is to exploit the empirical analysis in order to point out the desirable features of a mesoscopic modelling framework as well as the possible modelling assumptions needed for feasibility issues. In addition, to answer this question we need to look at how existing models can take interdependencies into account and what their limitations are. The connection between individual tours and dynamic flows at the scale of zones has consequently a series of requirements in terms of modelling which can be done at different grades. A first way to connect activity scheduling to travel flows is simply to estimate activity-specific daily profiles of departure rates. A way to answer the trade-off aspect in scheduling more in detail is to propose a model that allows to work on two successive trips together and to estimate the different activity-specific demand in a certain environment. In the second part of this thesis, two important elements referring to those two aspects are introduced: trip- and utility-primitives. They allow the consideration of activities in the observed traffic dynamics and, more specifically, the activity-specific choice process which can be modelled at the population level in order to connect individual behaviours and travel flows.

RQ3: Can we model mode dynamics considering explicitly their correlation with activity-travel chaining?

This question is strongly linked with the previous one but integrates other facets of trips performed by individual decision-makers. The activity-travel chaining binds successive trips and put their characteristics in relation. If the temporal succession of trips can simply be estimated through an estimation considering a full day, the spatial connections and correlations require more hypothesis. As with traditional demand modelling models, we focus on the generation and distribution of demand before we can incorporate the modal choice aspect into the proposed system. Each different step requires different degrees of input and modelling standards. The distribution and mode choice modelling are described in the third part of the thesis.

RQ4: How can an activity-based aggregated model estimate and predict the effect of disruptive policies and new services?

The last question, more than a methodological aspect, corresponds to a possible application to case studies of the demand generated according to the previous point. First of all, it requires creating a realistic base scenario containing a modal choice linked to the tour, i.e. where successive choices are constrained to each other. Given this scenario, we can observe to what extent it can be used to create value in a series of applications with respect to demand models currently in use. These applications can either concern supply characteristics with variations in their level of service for different modes, or concern an analysis of the model itself to estimate, for example, the management of parking or charging stations.

I.3. Thesis contributions

The main contribution of this dissertation is the description of the proposed demand model. In traditional demand modelling, mode choice is typically evaluated as the last aspect of the individual decision process before the route choice. Also, shifting to modes other than private car, even if it is one of the main targets of European governments, is far from being the sole solution for achieving greener mobility. Therefore, this thesis goes beyond mode choice modelling and integrates more levels of the decision-making process. This leads to a better approximation of mode choice on one side but also to intermediary results which can be used in some of the applications opportunities. Another contribution of this thesis, presented as first in this manuscript, is the empirical analysis of a multi-day travel survey which provides the basis and justification for the model proposed hereafter. We describe in detail the dynamics that exist in terms of workers' trips during business days. With a particular focus on mode choice, we analyse the impact of trip characteristics (e.g., time of day, activity at destination) but also of the trip chain (e.g., places visited, number of activities, modes used at home) on individual choices. The lessons learnt from this analysis allow to highlight the fundamental aspects to be integrated in the proposed macroscopic activity-based approach.

To describe the contribution made in this thesis, we then propose an analogy with the traditional "four-step model" of demand modelling (Figure I. 2).

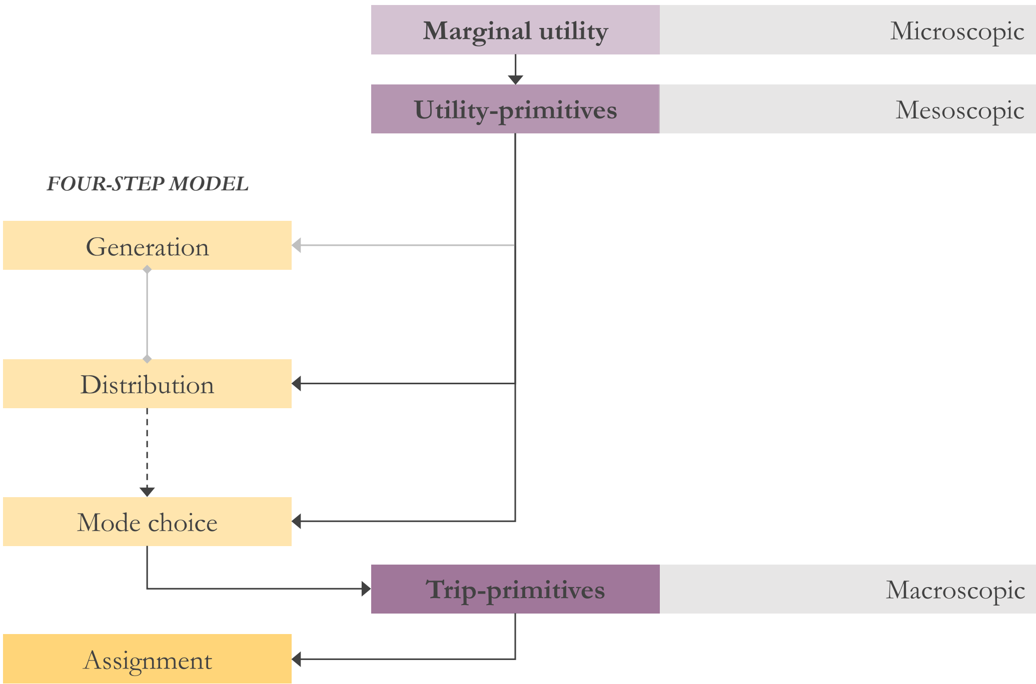


Figure I. 2 Methodological Framework

On the left we can see the four elements of this modelling technique, commonly used by practitioners. On the right side are depicted three fundamental concepts used in the developed model, each of them corresponding to a given modelling scale. Starting from a well-established concept, commonly used at the individual level (marginal utility), we propose a definition of "utility-primitives" in order to generate "trip-primitives", i.e., the single components of the daily demand profiles, each corresponding to either a given activity, mode of transport, OD pair or a combination of those. Given its aggregate nature, it can be used at a macroscopic level, as input to network loading models. The concept of utility-primitives can be used at different levels of the four-step model, depending on the modelling assumptions, available input, and desirable output.

Generation: for a given study area or zone, we propose an estimate of the generated trips specific to a given activity type and by time of the day. For this purpose, the concept of "activity trip-primitives" is defined and adopted to model daily demand profiles. For each given period of the day, we generate the number of trips to reach a destination where to perform an activity.

Distribution: the choice of destination varies from one type of activity to another and here, we estimate dynamic, activity-specific origin-destination matrices. The technique used depends mainly on the activity itself, while taking into account the journey costs. In this case, the generation and distribution of the demand is jointly modelled, and utility-primitives are used within the departure time and location choice model in order to generate activity-trip primitives as well as activity-specific OD specific demand.

Modal split: The last aspect modelled in this thesis is mode choice. For this component, we focus more on the components related to the trip itself (e.g., duration, access to the different modes, distance…) than in the two previous steps. In this case it is hardly conceivable to model the alternative between different choice options only through variations in terms of accumulated positive utility at the trip destination. In opposition to generation and distribution which are jointly modelled, mode choice is subsequent to those two steps and the calibration process is done in two distinct stages.

Traffic Assignment: Finally, the traffic assignment, last part of the four-step approach is out of the scope of the proposed model, as our approach is mainly concerned with pre-trip choices. In this thesis, we include however an example of an integration of our generated trip-primitives and a traffic flow simulator in the case study section of the dissertation.

The advantages of the proposed multi-scale approach are threefold:

- First, we estimate activity-specific components of aggregated generated trips that are both consistent with individual mobility patterns (thanks to the utility functions used as a mainstay for the estimation process), and with the observed aggregated demand flows (thanks to the input data used as a reference for calibration).
- Second, the primal outputs of this method are dynamic activity-specific demand flows. They can certainly be deployed at the origin-destination level and used as input for applications such as dynamic within-day Origin-Destination (OD) estimation, for creating synthetic trip data for transport simulation models, or even for reconstructing activity schedules (Ballis and Dimitriou 2020b).
- Finally, thanks to the adoption of underlying functions often used for representing the marginal utility gained by individuals, we can estimate relevant behavioural parameters and acquire insight into complex decisions such as when to schedule activities and the related departure and arrival times, or which mode(s) to use in a daily trip chain.

Application opportunities are therefore numerous. By estimating activity demand and typical duration, we can develop dynamic resource modelling and management strategies.

I.4. Thesis outline

This manuscript is divided in four parts and consists of nine chapters (Figure I. 3). Firstly, after this introduction we review the state of the art, with a particular focus on existing demand models and the integration between microscopic representation of the demand and usage of emerging behaviours for planning purposes among others. The third chapter answers to the first research question (RQ). This allows us to sketch the conceptual framework of the modelling approach and specific indispensable functions to be integrated. In terms of modelling, we have thereafter two branches, on one side the temporal aspect (relating to the second RQ) and on the other side the spatial aspect (third RQ). While the temporal aspect is often dealt with after the distribution of trips inside the study area, we decide here to follow the parallel presented above and introduce as first the concept of trip-primitives. Indeed, the generation step is intrinsically time-dependent inside the proposed approach. Chapter 4 describes this concept and introduces a simplified case study with the estimation of a mixture model. The estimation process itself is also introduced in this chapter on the Markov Chain Monte Carlo calibration of the model's parameters. The application of this calibration method is then described in the fifth chapter in the context of a utility maximization approach. This approach is used with the help of utility-primitives for estimating the distribution of trips in chapter 6 and the modal split in chapter 7. The last part of this thesis concerns the application opportunities an application on the network of Luxembourg (chapter 8) followed by conclusive remarks including possible improvements of each step of our macroscopic activity-based model.

Part I
Introduction

Chapter 1
Context and motivation

Chapter 2
Review of the state of the art

Chapter 3
Empirical analysis

Part II
Temporal Aspect

Chapter 4
Trip primitives

Chapter 5
Utility primitives

Part III
Spatial Aspect

Chapter 6
Destination Choice

Chapter 7
Mode Choice

Part IV
Application opportunities

Chapter 8
A case study of Luxembourg

Chapter 9
Conclusions

Figure I. 3 Outline of the thesis

Chapter 2

Highlights of the chapter

1. Advanced models often require very detailed information, which may not be readily available or may require expensive surveys
2. Utility maximization models can describe well individual decision processes, and relate them to explanatory and observable variables
3. Limitations of integrated models due to balance between complexity and completeness

In this chapter, we briefly describe the state of the art in the different areas related to the choice facets modelled in this thesis. That is, generation, distribution, and modal choice. The last aspect of the 4-step model, namely route assignment, is outside the scope of this work. Instead, we also discuss integrated models between traffic assignment and advanced demand models and what are the different types of proposed solutions. That is what are the best models to combine sophisticated methods of individual planning with road network loading or multimodal assignments. This chapter aims to highlight the research gap that motivates this PhD thesis and outlines the modelling strategy.

II. STATE OF THE ART

II.1. Introduction

As described in the introduction of this thesis, we argue that dynamic information on trip purpose needs to be included in macroscopic, aggregated models. It is essential to evaluate spatial and temporal travel demand dynamics and consequently to better predict the long-term impacts of planning, policy, and traffic management measures. In fact, research in demand modelling has evolved substantially during recent decades with growing interest in Activity-Based Models (ABM) as opposed to the more elementary Trip-Based Models (TBM) (Bowman and Ben-Akiva 2001b; Timmermans, Arentze, and Joh 2002). ABM usually reflect the scheduling of individual agendas and the derived trips in time and space. Because they are modelled in a disaggregate manner, the activity-travel prediction has a strong behavioural component and a high level of detail. In this chapter, we emphasize the differences between those two modelling approaches and the gap existing between them.

Figure II. 1 shows on the left the level of detail in terms of schedule modelling in the existing literature. On one side activity-based models which can be used as tour-based models are the most advanced approaches to describe individual behaviours but usually rely on a simplified supply model. On the other side, dynamic traffic assignment models often use a less detailed representation of the demand as input but are very useful tools for modelling daily variations of the network state and predict congestion. Because the goal of this thesis is to propose an activity-based approach of demand modelling able to seamlessly accommodate a Dynamic Traffic Assignment (DTA), the chart also includes the existing integrated models. Few works focus on the actual integration between DTA models and activity-based demand models (Boyce 1986; Lin et al. 2008).

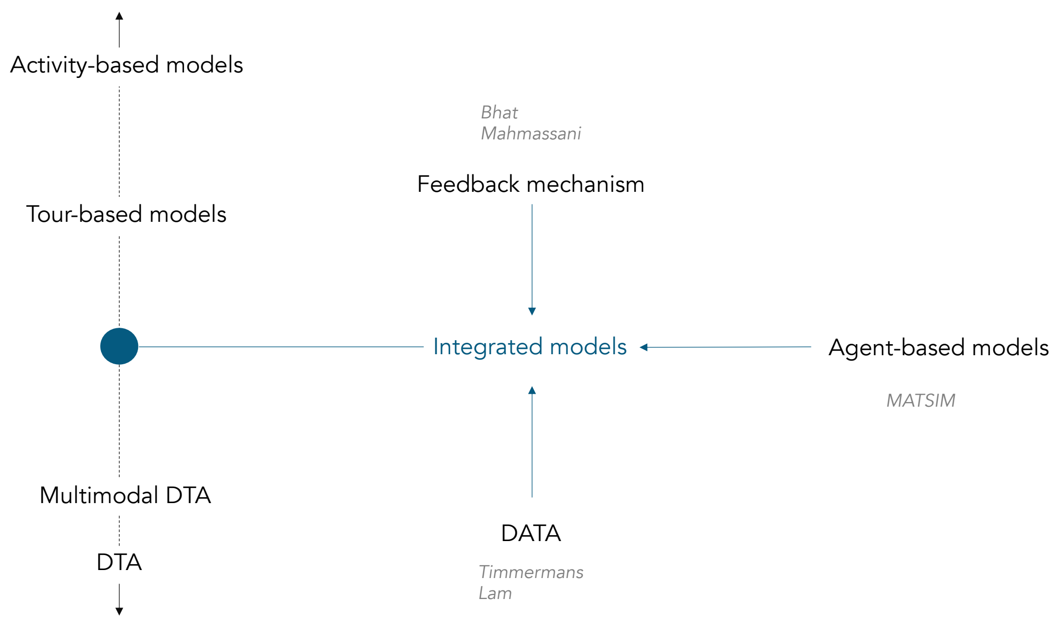


Figure II. 1 Chart of relevant literature branches

However, other approaches such as agent-based models can embed these two aspects in one singular simulation framework. By limiting our focus on the problem of linking input data and demand models, research can be divided in two main branches: including information on the trip purpose within standard trip-based models (*macroscopic* representation of the demand) and generating a fully disaggregate population of agents to evaluate tour of activities (*microscopic* representation of the demand).

In this section of the thesis, we will describe these different existing models and the assumptions on which they are built in order to highlight their advantages, to be retained in the following proposed model as well as their disadvantages to be overcome. Following the conventional demand modelling framework presented in the introduction, we present a short description of alternative generation, attraction, and mode choice modelling approaches. Finally, the assignment step is handled within a short review of integrated models.

II.2. Generation

The first step in the four-step modelling approach is the generation model. In the case of activity-based models, it consists in the creation of a population for which long-term, short-term plans and by consequence trips will be generated. Concretely, procedures are mostly based on the creation of a virtual set of agents and households, provided with specific attributes and subject to activity scheduling (T. A. Arentze and Timmermans 2004a; Bowman and Ben-Akiva 2001a). This founding step of disaggregate demand modelling is well known in literature as "Synthetic Population Generation". The starting point of synthetic population is usually composed of both aggregate socio-economic characteristics and disaggregate information of a sample of the population. Merging aggregate data from different sources means incorporating strong assumptions on their distributions (Farooq et al. 2013). The two main options to generate this data are synthetic reconstruction (and in particular Iterative Proportional Fitting (IPF)) and reweighting methods like Combinatorial Optimization (CO) (Mueller and Axhausen 2011). Farooq et al. (2013) introduced a third category: Markov process-based methods. In the IPF algorithm, a contingency table is evaluated iteratively, based on the correlation of attributes in the sample; a population is created by replicating the sample accordingly, it has been used since a long time and is still recently (T. Arentze, Timmermans, and Hofman 2007; Duguay, Jung, and McFadden 1976; Ye et al. 2009). In the CO, a weight is linked to the sample to select a combination of households from the dataset (Voas and Williamson 2000).

The fact remains that, if these complex models can yield good results, major drawbacks appear when evaluating their usage. On one hand, the aggregate data required needs to be very consistent and extremely accurate. On the other hand, data sources need to be of good quality, recent and representative of the entire population. To get viable output of such models, a large amount of information is needed and the performance classically increases with the quantity, quality and precision of inputs (Barthelemy and Toint 2013; Farooq et al. 2013). Furthermore, in many countries, privacy restriction are so tight to make almost impossible to implement ABM without running conventional (expensive) travel surveys (Barthelemy and Toint 2013). In order to overcome this issue, sample-free synthetic reconstruction methods have recently appeared (Barthelemy and Toint 2013; Gargiulo et al. 2010). They overcome the restriction of micro-samples or travel-survey necessity but are still based on very specific probability distributions. Besides, experiments concluded their lower performance in comparison to sample-based approaches (Lenormand and Deffuant 2012).

In principle, Synthetic Population Generation model discussed so far may be applied for introducing a trip purpose specification within time dependent OD matrices. However, since they require a large amount of data, they are usually implemented only to calibrate Activity-Based micro simulators, while time-dependent OD matrices are usually estimated using simpler approaches. Usually approximated for different trip purposes, the number of attracted and generated trips by TAZ depends on characteristics of this zone and of its residents.

Table II. 2

	ABM	TBM
+	Accuracy and level of detail	Low data requirement
-	Complexity of the models	Risk of consistency lack

II.3. Distribution

The balance between applicability and level of detail applies naturally as well to trip distribution or destination choice model. We can oppose on one side individual destinations within scheduling choice modelling (Bowman and Ben-Akiva 2001b) or agent-based simulators (Horni, Nagel, and Axhausen 2011) and on the other side the distribution part of the four-step model and the so-called "Gravity model" (Voorhees 2013).

Even though destination choice model (in the form for example of logit models) and distribution models can be seen as similar approaches using comparable trip characteristics when applied to large homogeneous groups, a discrete choice approach can model the impact in terms of utility. This improvement given certain attributes and explanatory variables is nonetheless at the cost sometimes of higher computational times and information requirements (Jonnalagadda et al. 2001). Even if this travel decision is not the main focus of disaggregated modelling approach, it can result in better estimation than gravity model (C. Bhat, Govindarajan, and Pulugurta 1998). This allows indeed to incorporate for example trip chaining in the destination model (Bernardin, Koppelman, and Boyce 2009) accounting for actual dependency between trips at the individual level (Kitamura 1984a). Time-geography theories can also be included for choice set simplification in the case of agent-based simulations (Horni et al. 2009). The usage of such theories derived from (Hägerstrand 1970) and especially space-time prism is particularly relevant to constrain the modelling of activities. The travel time budget, key element of these concepts, can be linked this way to the time allocation problem used in this work to understand the possible trade-off in time and space.

Nevertheless, the most common approach to generate origin-destination matrices remains in practice the usage of impedance function of distance between zones but also cost and time of the trips and constrained iterative fitting. This often has limitations in terms of accuracy and refinement because the level of aggregation is usually very high. These issues are raised particularly when dynamic OD matrices need to be generated.

Table II. 3

	ABM	TBM
+	Consistency to individual decisions and variables	Easy to implement
-	Choice set formation	Low demand elasticity Aggregation problem

II.4. Mode choice

While in the past travel demand was mostly captured by single modes users, the development of new systems such as Mobility-as-a-Service, as well as policies aiming to reduce car use and ownership, enhance the multimodal behaviour of travellers. By definition, activity-based models ensure consistency of successive trips which are triggered for undertaking activities (Ben-Akiva and Bowman 1998).

Mode choice is a complex decision that involves many determinants at various levels. Obviously, the level of service and trips characteristics impact the choice set and relative attractiveness of options, but the decision involves socio-demographic characteristics and personal preferences or habits as well (Tyrinopoulos and Antoniou 2013). (De Witte et al. 2013) undertook a comprehensive review and highlighted that, among a large number of interacting parameters, departure time and even more so, trip chaining are too often ignored, e.g. trip chaining is considered meaningful for 80% of the cases but included in only 20% of the papers. While many variables can be included at a macroscopic level, some characteristics are linked to individuals or households and their long-term decisions. For this reason, household travel surveys are commonly used for studying single modal choices. (Pucher and Renne 2003) show regional variations but also significant variations in modal split by trip purpose, nonmotorized modes being less represented for going to work. In the case of Melbourne, (Currie and Delbosc 2011) observe that complexity of tours is lower for car users and underline the need of adding a spatial perspective. On the contrary, (Hensher and Reyes 2000) rely on time budget and value of time and model mode choice using different logit models in order to demonstrate to what extent trip chaining is a barrier to public transport use. In the same vein (Krygsman, Arentze, and Timmermans 2007) show that public transport hinders the inclusion of secondary activities in work tours, notably because of mode (un)availability at the workplace. Furthermore, they use a co-evolutionary approach to conclude that the intermediary activity decision is made most of the times before the mode decision. Similarly, (Ye, Pendyala, and Gottardi 2007) use econometric modelling to explore the directionality of the relationship between mode choice and complexity of trip chaining patterns using micro census travel survey and show that trip chain complexity precedes mode choice.

The complexity of trip chaining and its impact on mode choice (as well as destination choice) can also been linked to the concept of activity space (AS)(Tsoleridis, Choudhury, and Hess 2022). The notion of time-space prisms, introduced by (Hägerstrand 1970) illustrates for example the necessity of carrying out "[roles] within a given duration, at given times and places" and constrains them with "geometrical shapes in terms of location in space, areal extension, and duration in time" whose parameters depend on the available modes of transport. This concept, refined by (Lenntorp 1976) has been widely used notably in the activity-based approach to travel demand modelling (Bowman and Ben-Akiva 2001b). The AS geographical aspect has been described in many ways and applied for various purposes. (Patterson and Farber 2015) reference 66 applications of AS and potential path areas in diverse fields and highlight 4 main methods to estimate AS: ellipses and circles; network-based approaches; kernel density approaches; minimum convex-hull polygons. Some authors compared different models based on a series of criteria, for example (Schönfelder and Axhausen 2003) regret the rigid assumptions of the confidence ellipse that makes the AS's size too high. However, the versatility of the AS concept prevents reaching a consensus on a single set of criteria appropriate for all use cases. In order to represent better AS, new geometries are still recently proposed (Li and Tong 2016) and notably for integration in MATSim (Rai et al. 2007). To reflect better AS, (Perchoux et al. 2014) combines different methods to create a set of indicators in order to qualify individual space-time patterns. This study also concludes that active modes are used in small AS centred around home and that larger AS correspond to the use of motorized modes. While most of explorations on AS focus on individuals, (Harding, Patterson, and Miranda-Moreno 2013) assess the relationship between mode choice and AS at the city level. They selected convex hull for describing AS of sampled individuals and use 'area to compactness ratio' for comparing transit, car, and active modes. While they mostly have expected outcomes, transit users are associated to compactness ratios closest to 1.

Making specific mode choices for an earlier trip influences following decisions, for example because of the consequent availability of transport alternatives at the origin or destination of the trips. Few models are able to handle all feasible mode combinations for tour-based choices (Vovsha et al. 2017). Even though, to handle the aspects and correlation in mode choice, tour-based models seem to be the most relevant. When a tour-based mode choice modelling approach is adopted explicitly in activity-based models, the mode choice usually follows an activity scheduling model. The output (i.e. agenda of individuals) is then fixed and used as starting point to estimate mode selections. For example, (Miller, Roorda, and Carrasco 2005) consider the schedule as an external input to their choice model. Also, the mode choice may sometimes be limited to a single choice for the whole day. In many other cases tours are specified, based on the commute mode-choice only and without the possibility to change mode. This reduces the approach to unimodal tours without combinations available (T. A. Arentze and Timmermans 2004b; Chandra R. Bhat 1997; Bowman and Ben-Akiva 2001b). More advanced models include a "main mode" which characterizes the tour and may then be used inside trip-based mode choice models, that are constrained by that choice (Bradley, Bowman, and Griesenbeck 2010).

Table II. 4

	ABM	TBM
+	Activity space and trip chaining has a strong impact on mode choice	Mode choice is affected by many non-individual factors and data is easily collected
-	Often simplified into single mode usage	No explicit correlation between trips within a tour

II.5. Models' integration with assignment models

Trip-based origin-destination demand flows are the typical input for advanced DTA models, which have become the most commonly used tools for planning, optimizing and managing transportation networks (Peeta and Ziliaskopoulos 2001). DTA frameworks including mode choice under equilibrium conditions and including activity-related choices are an efficient way to represent full-day schedules of a whole population, considering feasibility constraints. The usage of variational inequality dynamic user equilibrium model has been applied as first by (Lam and Yin 2001) to propose a concept associating activity-based demand on one side and time-dependent traffic assignment on the other side. The applied elastic demand is used in such way that, at equilibrium, perceived utilities of activities are maximal with a minimal travel time on selected paths. This *supernetwork* representation allows to model different choices as path choice through the constructed network, offering a basis for activity-travel assignment. For example it allows to model multimodal travel choice problems in an integrated framework (Carlier et al. 2003). Although there is no critical distinction between a conventional traffic network and supernetworks, the latter are capable of illustrating the transition and interactions between distinct modes and so model multimodal journeys. Different facets of travel choice are turned into path choice through the different layers of constructed network, offering a basis for activity-travel assignment. Activity-travel patterns are input of particular networks that range from PT-only networks with mode choice (Fu and Lam 2018), to multi-state supernetworks (Liao, Arentze, and Timmermans 2013). However, all these models are constructed at the individual level, together with the scheduling and activity chain planning. Time-based utility profiles can in this way be used for estimating scheduling as well as in a combined activity and route choice problem through a Dynamic Activity Travel Assignment (DATA) (Liu et al. 2016). The dynamic activity-travel assignments models are arguably the most advanced methods for capturing multiple choice dimensions such as activity sequence and duration that can be estimated through the DTA itself. However, because these networks are time expanded and at the individual level, size and complexity can grow extremely rapidly.

Another common approach to integrate activity-specific aspects into the route choice modelling is to include a departure time aspect and combined activity-based and a traffic assignment models through feedback mechanisms (Lin et al. 2008). At each time interval the time-dependent activity-travel demand is introduced as input into a distinct traffic flow model. A stochastic dynamic user equilibrium model which uses the travel plan at a user level is used as input demand to evaluate departure time, activity and route choice, for a limited number of activities and time periods (Abdelghany and Mahmassani 2003). More recently, (Halat et al. 2016) proposed an integration of the full activity-trip chain demand inside the assignment process. A dynamic traffic simulator (*Dynasmart*) is used in combination to the *CT-RAMP* activity-based demand model to use origin-destination demand consistent with full-day schedules of the population, iteratively evaluated through a bi-criterion dynamic user equilibrium. Agent-based traffic simulators are one of the most advanced approach for modelling complex choices and their interactions (Patwary, Huang, and Lo 2021). As an example, the population-based trip micro simulator *MITO* used as input to a DTA makes the demand sensitive and accurate (Moeckel, Huntsinger, and Donnelly 2017). In can be used as well to generated demand and activity chains for mesoscopic traffic simulators such as MATSim (Llorca et al. 2019).

In order to estimate the aggregated demand emerging from these individual decisions, often complex frameworks are used. (Hesam Hafezi et al. 2021) include among others agenda formation, tour simulator or trip scheduler modules with various inherent techniques.

Observed traffic data, on the contrary, does not suffer this limitation as it is a consequence of travellers having diverse characteristics and needing to move for subjective reasons. While most of the above-mentioned methods rely on agent-based models, it is possible to incorporate information about daily activity scheduling and duration in order to derive dynamic OD flows, which are consistent across time periods (Cantelmo et al. 2018). In this way the relation between utility and dynamic user equilibrium is exploited without generating an SP. Works focusing on the estimation of activity schedules based on aggregated OD matrices are very few, but recent works use those data to convert purpose-dependant demand into realistic activity schedules (Ballis and Dimitriou 2020a; 2020b).

II.6. Trip- vs. activity-based models

Demand modelling in most cases is used in order to be associated with a supply model and estimate route choices to eventually assess congestion and service usage issues.

Last decades have witnessed an immense effort in bringing travel demand modelling to a new level of comprehension. While, since the early 80s, person travel has been modelled with a trip-oriented rather than activity-oriented specification, this approach has been universally criticized for being unrealistic. Because of these challenges linked to application, traditional TBM continue to be widely adopted for forecasting travel demand (Ortuzar and Willumsen 2011). Relying on the traditional four-step demand model for transport planning (McNally 2007), TBM are sometimes inadequate for planning purposes and for incorporating departure time choice in demand estimation (Lindveld 2003) as they usually provide too coarse a representation of travel demand. For example, the introduction of mobility services like car sharing involves strongly interdependent decisions such that modelling the entire daily mobility patterns with their spatial and temporal variations is recommended. The main problem is that while researchers agree that travels are derived from the demand for activities and services (Axhausen and Gärling 1992), conventional macroscopic TBM do not explicitly account for trip-purpose, activity scheduling, nor duration constraints (Cantelmo et al. 2018). Despite these limitations, TBM provide essential application and interpretability opportunities, and are convenient to calibrate using observed traffic flows or other aggregate data, compared with agent-based models.

Thus, advanced ABM allowing schedule-based and tour-based approaches, are nowadays the most active research branch in demand modelling with new tools still being developed. These models often contain a set of detailed models (from synthetic population generation and long-term decision making to daily activity patterns and singular trip choices) which made their application limited due to the high cost of collecting input data (Bowman and Ben-Akiva 2001a). ABM are by nature able to integrate this behavioural component originally lacking in TBM. Tour-based approaches model individual choices and disaggregated mobility patterns, and have been developed to capture detailed individual responses, for instance, within the framework of agent-based simulation (Charypar and Nagel 2005). These detailed models can generate plans at the individual or household level, allowing explicit representation of agents' interactions (Khan and Habib 2021), which entails large numbers of parameters. This level of detail is hard to apply in large-scale analyses where calibration and consistency with macroscopic emergent behaviour remains an issue (Moeckel, Huntsinger, and Donnelly 2017). The application of these models also requires resources and skills which, while attainable in major cities like Zurich or Toronto, are often not available in small and medium sizes cities (H. Becker et al. 2019; Gao, Balmer, and Miller 2010). The advent of new technologies allowed researchers to gather immense volume of information, making possible to implement ABM on large urban and regional scenarios. Despite this technological advantage, to calibrate a reliable demand model is still a difficult challenge (Toole et al. 2015). One of the main reasons is that different assumptions within the demand model will lead to different errors. For instance, a disaggregate model can be extremely powerful for long-term prediction of the demand or for evaluating phenomena such as activity relocation. However, when the congestion during the rush hour is the main concern of the modeller, this level of detail is not required and may even lead to biased analysis. It is thus difficult to estimate activity-specific trips at the aggregate level where the available information does not include such features (e.g., scheduling or trip chaining information).

Because the application of the four-step model is nearly universal, various methods are still developed to tend improving the efficiency of this model (Lim and Kim 2016). However, all these models provide a limited insight in terms of temporal distribution of the demand and its activity specification. In this sense, the most established approach to reconstruct a realistic temporal profile for the travel demand is the dynamic OD estimation problem (Antoniou et al. 2016; Cascetta, Inaudi, and Marquis 1993). While many works focus on improving the consistency between OD matrix and traffic performances (Antoniou et al. 2015), there are only a few works dealing with the critical point of explicitly capturing the behavioural component of the demand (Flötteröd 2009a). Alternative demand generation methods emerged, which integrate inside the traditional Four-Step-Model a microscopic alternative (Vrtic et al. 2007). These models still rely on highly specific probability functions, which are very hard to calibrate without a large volume of data.

Due to anonymization, trips data are limited in providing relevant information such as the purpose of the trip, and the motivation for performing that trip with a specific mode of transport at a certain time of the day. This limitation is not likely to be overcome by new technologies (big data) due to privacy restrictions.

In conclusion, we can say that because of the complexity of data collection and computational implementation, an alternative to synthetic population-based ABM is needed and that this alternative needs a higher level of consistency than the traditional TBM. We propose to define an activity-based trip model which has the following properties with respect to the described literature:

Table II. 5 Summary of gap model's features from the literature review

Significant features	**Inapplicable features**
Aggregate demand model	Synthetic population generation
Adjustable data requirement	Individual information requirement
Flexibility to be applied to different modelling scales	Necessity to model the most disaggregate aspect for every application case
Modelling of successive trips interactions	Full tour-based model
Integrated model	Succession of individual modules
Computational simplicity	Risk of combinatorial explosion
Activity-based model	Trip-based model

Chapter 3

Highlights of the chapter

1. Empirical analysis of a rich dataset consisting of 5848 home-based tours is performed to gain insight into emerging regularities in the trip patterns
2. Introduction of an aggregated probabilistic activity-space approach to quantify the probability of performing a certain activity given the locations of home and work
3. Highlight the importance of tour-based mode choice modelling to explain individual choices of sequences of modes

Individual chains of activities are governed by spatial and temporal constraints reflected in the successive mode choices as well as the locations of activities. To understand the measure of this phenomenon, we start this doctoral thesis with an analysis of a dataset collected in Belgium in 2008. The main goal is to underline and quantify this impact in the emerging behaviour at the scale of the population of an average Benelux city.

A first phase of this project consists of cleaning the processed data. This includes not only encoding in a format chosen to facilitate data analysis, but also checking the plausibility of given responses and finally grouping certain variables into consistent and uniform units.

A cluster analysis made it possible to select the types of activities to be analysed together, according to classic criteria such as the frequency and duration of the activities but also according to their start and end time profiles. Likewise, the modes of transport used have been grouped manually according to the proximity of the type and estimated level of service. A check on the distribution of distances and travel times confirmed the validity of the selection. The trip analysis was carried out on the basis of 404 workers who described several consecutive days of their routine.

This study allowed us to highlight factors that affect car use. Beyond an obvious link between the use of the car and topological data related to the trip made, a connection between the use of the car and the time of day as well as the type of activity was observed. In addition, a strong impact of vehicle ownership was seen as part of the chain of modes of transport employed throughout the day. We estimated that using a bicycle or car as the first mode of transportation of the day caused users to continue their tour with the same vehicle.

In a context other than that linked to the analysis described above, the impact of the complexity of the activity chain on the use of the car was established. Despite the size of the dataset used, a Structural Equation Modelling (SEM) analysis allowed us to measure the weight of the number of both work-related, non-work-related and late-hour activities on the rate of use of the car and usage amounts.

This chapter makes it possible at the same time to motivate the modelling of the trip choices with a tour-based approach, even at a macroscopic scale, and to justify some assumptions within the modelling framework. Beyond that, a concept proposal is proposed in order to apply the notion of "activity space" to a group of people sharing certain characteristics. Indeed, we have extended the calculation of ellipses (SDE) in order to have a probabilistic approach and to estimate certain factors characterizing them according to the place of work and residence. The centre, shape factor and orientation can be characterized with the home-work journey (location of the two points and mode chosen).

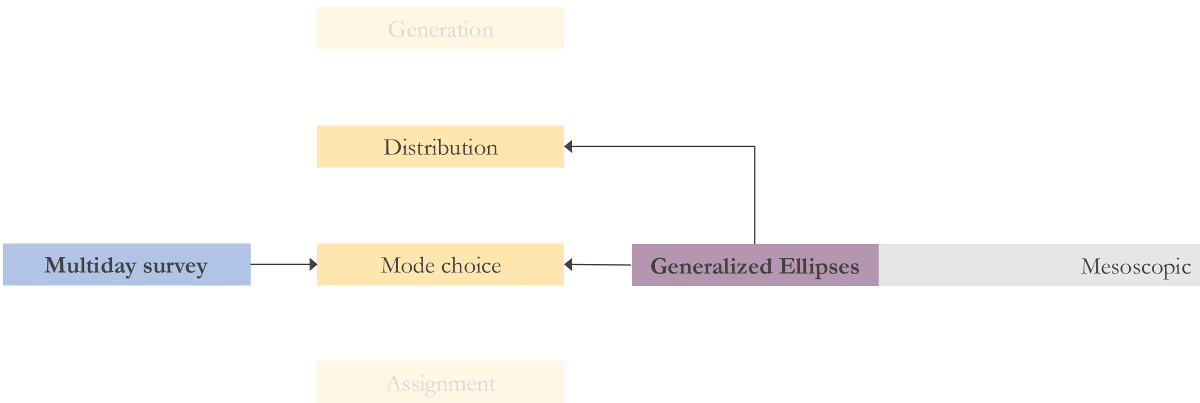


Figure III. 1 Thesis framework chapter 3

On the thesis framework, we include on the left in blue the input data used in this chapter and the methodological aspect proposed, with respect to the modelling step. Because the analysis is done on mode choice in this chapter, only this cell is highlighted, even though we can apply the concept of generalized ellipses in the distribution part in the case of secondary activities location choice.

The work presented in this chapter has been described in the following paper:

<u>"Trip chaining impact on within-day mode choice dynamics: Evidences from a multi-day travel survey"</u>

Transportation Research Procedia

23rd EURO Working Group on Transportation Meeting, EWGT 2020, 16-18 September 2020

III. EMPIRICAL ANALYSIS

III.1. Introduction

Mode choice is influenced by a large variety of factors, as for example users' socio-economic attributes or level of service for the different alternatives. In order to understand better what leads to temporal and spatial variations of modal split, we propose in this chapter an analysis of a multi-day travel survey, with a series of descriptive statistics as well as inferential analysis on the correlation between mode choice and tour-specific attributes at both spatial and temporal levels. This chapter discusses the importance of considering tour-based mode choice not only because it brings consistency between successive mode choices but also allows the inclusion of relevant tours' characteristics such as activity types, distances, time of the day, and previous mode choices. A total of 5848 home-based tours done in 2008 are studied in the area of Ghent, Belgium. Identified patterns show the importance of modelling dynamic mode choice with trip chaining and time of the day. The modal share of car drivers differs of more than 40% between hours of the day and about 30% between different activities. Furthermore, the definition of activity spaces by principal mode choice and home-work locations introduces the calibration of probabilistic aggregate Gaussian fit to visited points.

In this section, we also propose a new definition of AS and propose to link it with mode choice through common factors for groups of users. Through analysis of multiday survey data, we quantify how the sequence of modes is impacted by the sequence of activities generating the travel need. The "Behaviour and Mobility within the week" (BMW) database used in this work is a multiday travel survey collected in Ghent in 2008 (Castaigne 2009). The opportunity to use the BMW database is twofold, not only the size of the database is large with representative modal split but also the duration of the survey allows to define AS for each individual in a robust way. Few similar databases exist and for example the "Mobidrive" study (Axhausen et al. 2002) constitutes an excellent reference in such analysis, as six weeks-diaries for 317 people were recorded and it was used for analysing both activity scheduling (Cinzia Cirillo and Axhausen 2010) and mode choice of complex tours (C. Cirillo and Axhausen 2002). However, the Mobidrive database contains a share of 86,1% unimodal tours, while the BMW database reaches 33% multimodal tours. This is an opportunity to study more in detail the correlations and interconnection between successive mode choices, time of the day and activity characteristics.

III.2. Methodology

In order to test the following hypotheses, we analyse a travel diary collected for 707 individuals in the city of Ghent (Belgium) in 2008 (the city was about 237.250 inhabitants at that time). A description of the database and analysis of the variability of daily activity-travel pattern is available in (Raux, Ma, and Cornelis 2016). The goal here is to observe emerging behaviours from these respondents in order to detect quantitative characteristics to be applied in future aggregated dynamic mode choice models. The tested hypotheses are the following:

1. Modal split changes over the day and is statistically correlated with activity choice dynamics
2. The mode chosen for a trip strongly depends on the mode chosen at an earlier time of the day
3. Owned resources such as bike and car increase this dependency
4. AS varies with the most frequently mode used and can be described by knowing home and work locations

III.2.1. <u>Data</u>

We focus on the Home-Work (HW) trip chain of workers from a spatial and temporal dimension. 404 respondents were selected for analysis who described at least one trip going to work, resulting in an average of 4 working days per person and 2.8 trips within the HW tour. 3543 unique points addresses of trip destinations were translated to GPS coordinates. The following assumptions are considered:

- Each trip is described by its origin, destination, starting, ending and travel time, modes, activity at origin and destination and their duration.
- Distances are the direct aerial distances.
- Home/work locations were identified for each worker as the most visited locations they referred to as home/work.
- The data includes 7 different modes of transport and their combination: car (or motorbike) driver, car passenger, train, bus (and tram), bike, walk and other
- 12 activity types could be recorded in this multiday survey.

The focus not being on activity choice, a simplification has been made by clustering the activity types based on the following criteria:

- Occupancy,

- starting time profile,
- weekly frequency (distribution from one to seven days in which the activity is done),
- distribution of the duration.

These four criteria have been normalized to one and grouped using k-means clustering into five groups of activities (home, work, daily tasks, personal business and eating). Figure III. 2 shows the original list of activity types and their observed weekly frequency.

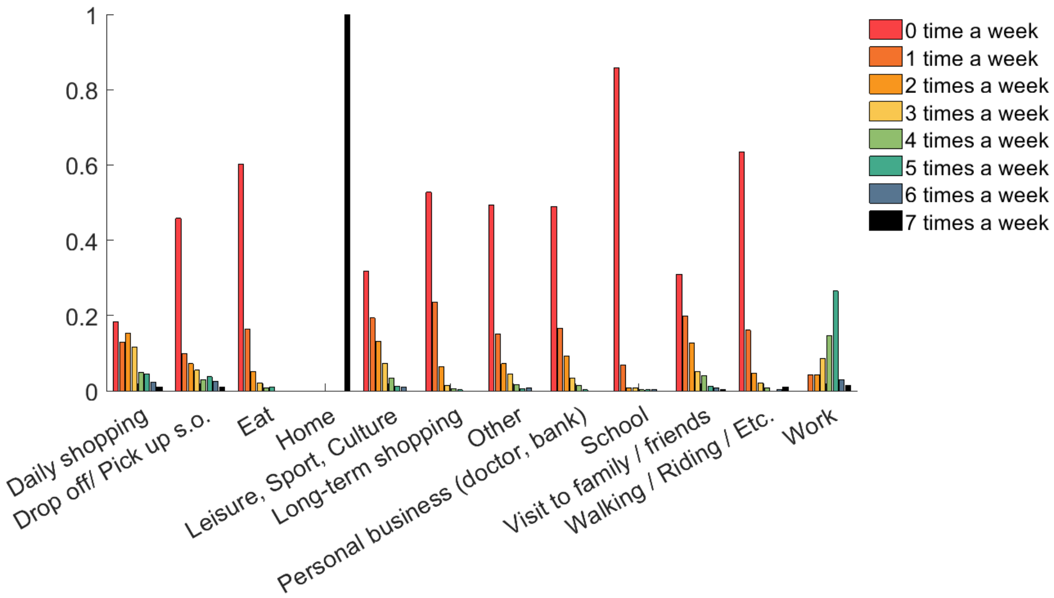
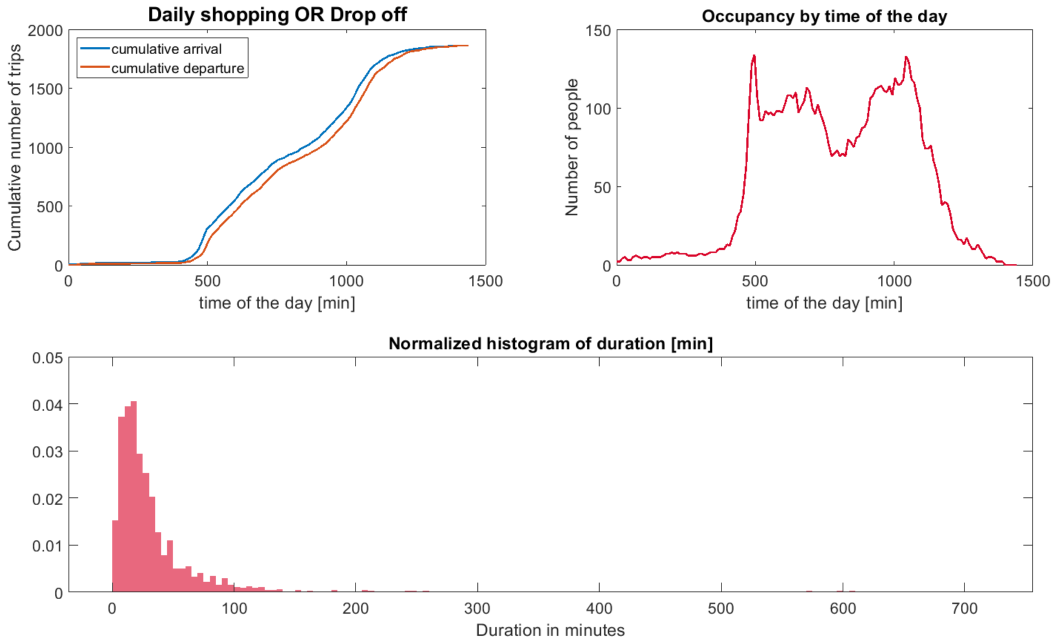
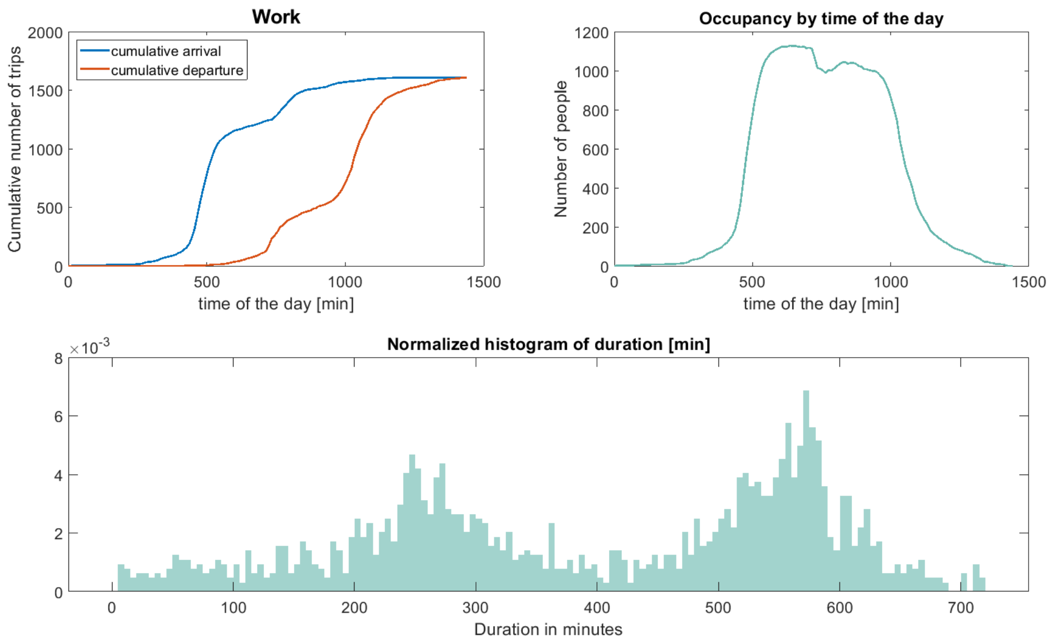
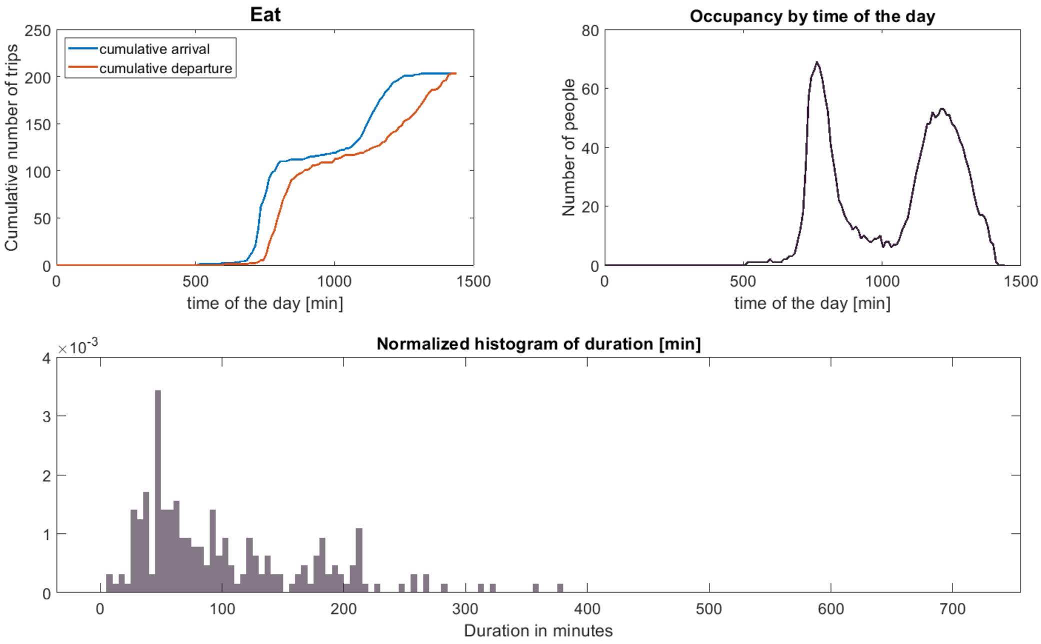


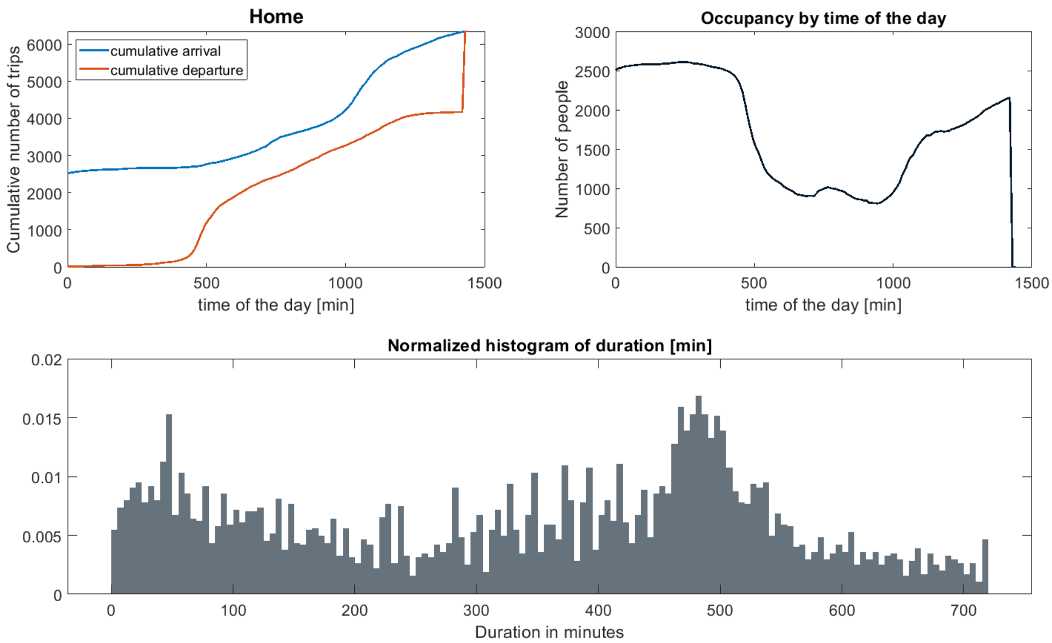
Figure III. 2 Weekly frequency of 12 original activity types

The occupancy is the difference between cumulative people starting and people ending the activity. This indicator represents, at each time of day, the potential number of people who may start a trip at later time, in order to start a subsequent activity. The profiles of different activities suggest that the negative correlations and offset between these curves translate activity chaining. Figure III. 3 shows the three clustering criteria for the five studied activity types.


Daily shopping OR Drop off
cumulative arrival
cumulative departure
Cumulative number of trips
time of the day [min]
Occupancy by time of the day
Number of people
time of the day [min]
Normalized histogram of duration [min]
Duration in minutes


Work
cumulative arrival
cumulative departure
Cumulative number of trips
time of the day [min]
Occupancy by time of the day
Number of people
time of the day [min]
Normalized histogram of duration [min]
$\times 10^{-3}$
Duration in minutes


Eat
cumulative arrival
cumulative departure
Cumulative number of trips
time of the day [min]
Occupancy by time of the day
Number of people
time of the day [min]
Normalized histogram of duration [min]
×10-3
Duration in minutes


Home
cumulative arrival
cumulative departure
Cumulative number of trips
time of the day [min]
Occupancy by time of the day
Number of people
time of the day [min]
Normalized histogram of duration [min]
Duration in minutes

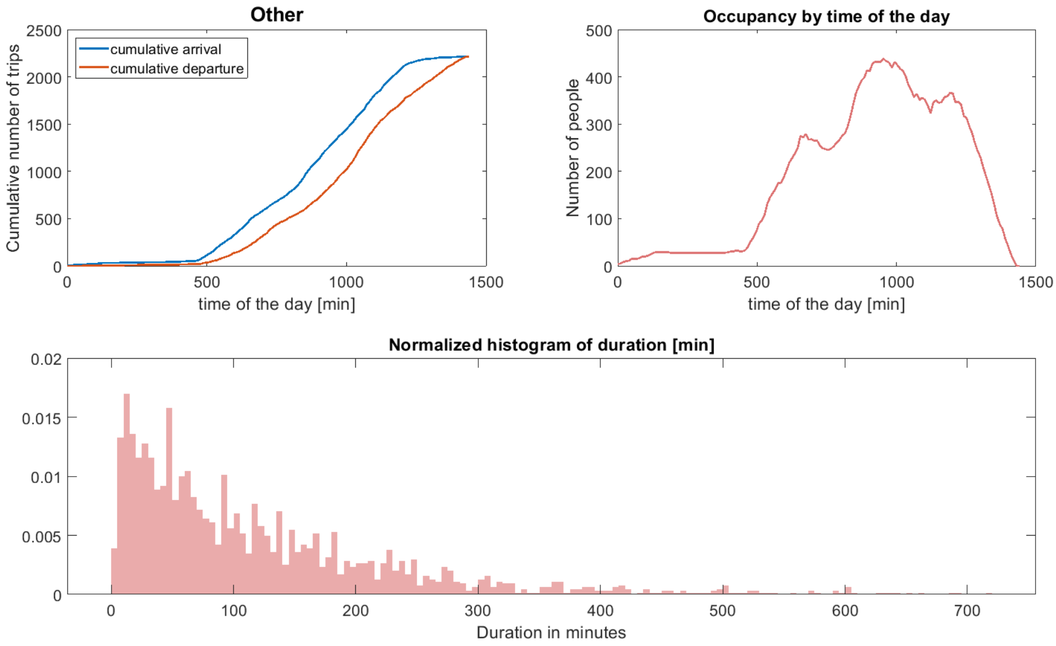


Figure III. 3 Clustered activities characteristics

K-means clustering was also applied to the 404 workers, to group them according to their mode use and their home and work locations. Because of the limited number of respondents belonging to each sub-group, geographical and behavioural attributes resulted in two independent groupings, applied for distinct parameters comparison. Users are firstly separated into groups based on the proportion of their trips done by each of the seven modes, resulting in the following five groups: car users, soft mode users, public transport users, train users and multimodal users (Figure III. 4 a). Secondly, users were first classified by municipality. For those living and working in Ghent, home and work GPS coordinates and distance between the two points has been used. Using the Calinski-Harabasz criterion, suggested 5 groups. An example of resulting clusters is shown on Figure III. 4b.

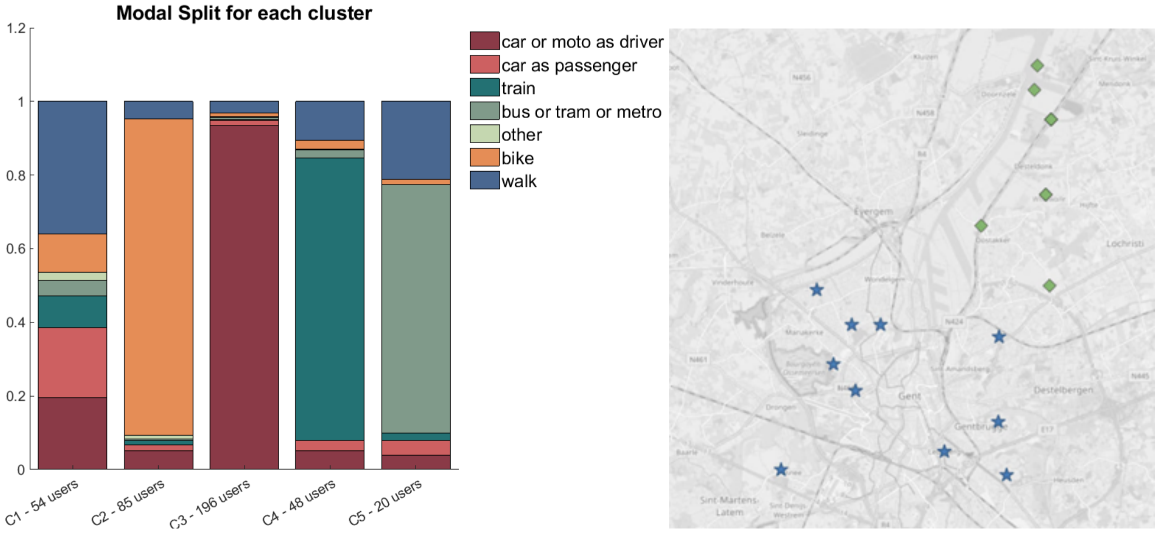


Figure III. 4 (a) modal split of 5 mode-based clusters (b) example of a Ghent HW locations-based cluster

III.2.2. Temporal analysis

Intuitively, different modes have different typical usage patterns, in particular based on the land-use and supply characteristics. In order to observe how they combine and how usage differs by time of day, we calculated the usage profiles for those modes and compared these patterns. The results are matched to individual observations through the transition probabilities. The transition matrix was calculated for all successive trips, for only commuting trips and finally for only intermediate trips, in order to estimate to what extent sets of modes complement one another and which are the most binding for round trips. To include all sizes of tours in the transition calculation, a notional mode corresponding to ending the trip chain was added. Additionally, we show that mode choice varies by trip purpose and secondly that the full activity sequence of workers impacts choices on the set of modes used. The way activities have been clustered distinguishes naturally systematic and non-systematic activities as can be seen in the weekly frequency (Figure III. 2). We believe that this difference impacts the travel time budget, due notably to the location that is more or less free to choose and leads to additional travel time for non-systematic activities higher than for systematic activities. This can explain, based on geographical dispersion, the relationship between the previously highlighted factors. In order to quantify the spatial aspect of mode choice dynamics, the AS of those workers has been analysed.

III.2.3. Spatial analysis

To show the link between temporal and spatial dynamics, the dispersion of each user's AS has been estimated using the method of 95% confidence ellipse. Such ellipses have few parameters to estimate and are commonly used for estimating AS. The 95% confidence ellipse ensures that number of visited points covered by the surface is high enough and so that it describes better the AS. The centre of the ellipse is the centre of mass of all visited points within the HW tour for the complete study period, weighted by the number of visits. The ellipse's axes are the square root of the eigenvalues of the covariance matrix, and the ellipse is then scaled to get the desired confidence interval, based on the number of degrees of freedom and the Chi-Square distribution.

The resulting ellipses are compared based on:

- orientation angle
- minor axis
- major axis
- aspect ratio
- centre.

The hypothesis is that the ellipse is strongly defined by the home and work location, considered as anchor points for the user. This can be observed through the position of the centre of the ellipse and its orientation.

Because the multiday travel survey contains on average four working days for each worker, we first applied an outlier detection method, such that non-recurrent behaviours do not unduly impact the estimation of AS. This detection is based on the Orthogonalized Gnanadesikan-Kettenring robust estimator (Maronna and Zamar 2002) using the Mahalanobis distance:

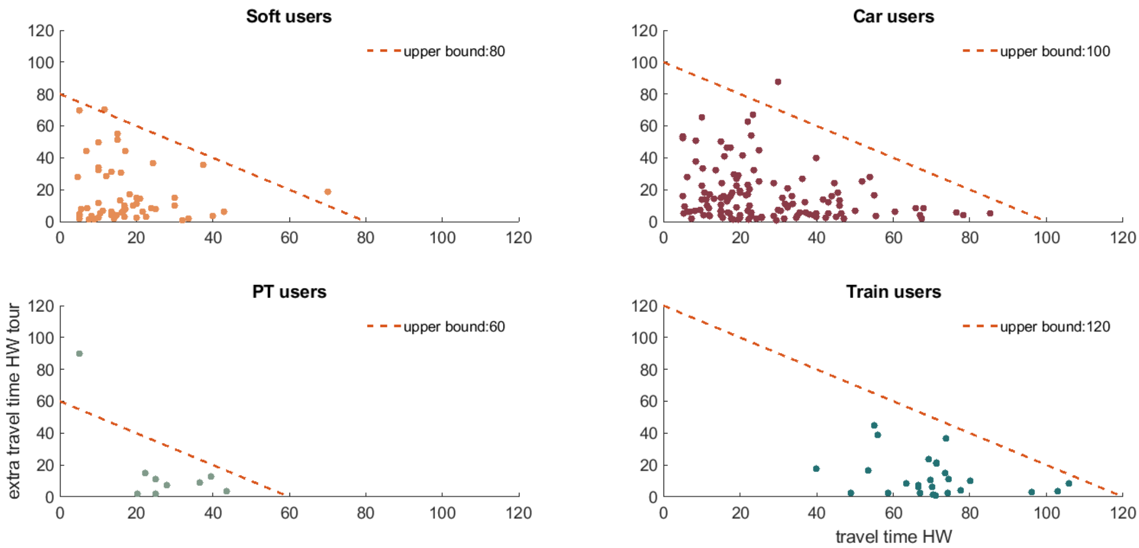
$$d_{(\mu,\Sigma)}(x_i)^2 = (x_i - \mu)'\Sigma^{-1}(x_i - \mu)$$

with respect to location μ and covariance Σ. Detected outliers are not always deleted. Firstly, home and work locations are never discarded. Secondly, points closer than a given threshold to home/work are kept. If the number of visits to an outlier point is two or more, it is kept. Finally, for each user we try to retain enough points to compute the AS. Finally, 181 points have been removed: 26% of the personal business locations, 16% of daily tasks locations, 12,5% of eat locations and 8% of work locations.

Properties of individual ellipses can reveal emerging trends for groups of individuals. However, the application to aggregate location choice models is limited with such strongly constrained geometrical units. In order to apply the observations to groups of users and to use soft constraints, a gaussian fitting has been chosen to make ellipses probabilistic. The multivariate gaussian describes plausible visited areas given home and work regions. It is again parametrized by the centre of mass of the visited points and the model is estimated by the maximum likelihood, using the expectation-maximization algorithm.

III.3. Results

In this section, a selection of the most significant results is presented in order to support the validation of hypotheses noted above and highlight decision needed for modelling these behaviours. The section is separated into temporal and spatial analysis. Firstly, indicative results about the correlation between these two aspects are described in order support the link between them (Figure III. 5).



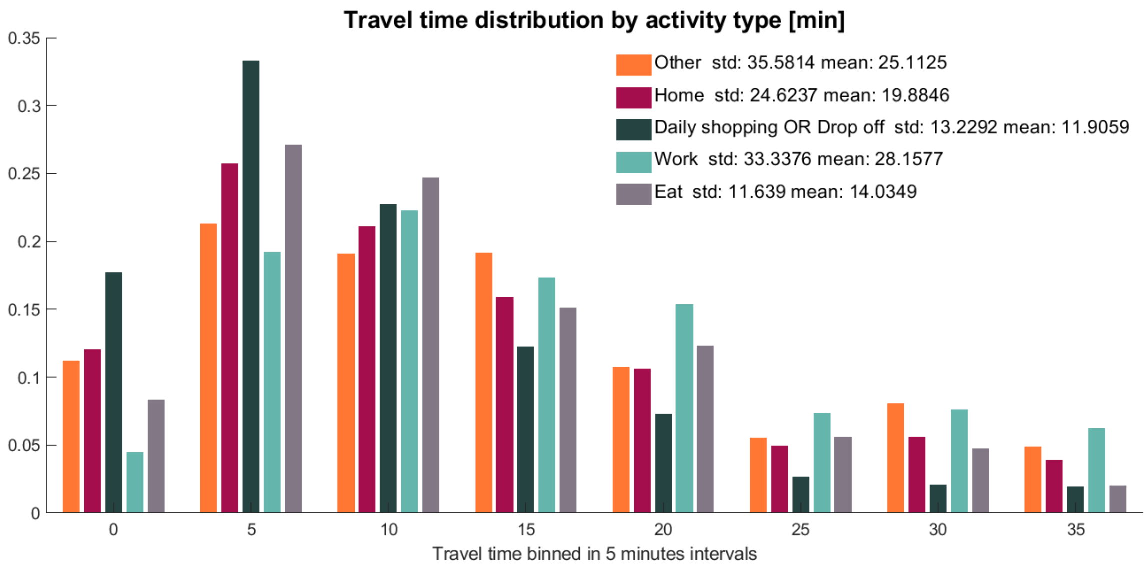


Figure III. 5 (a) HW travel time vs. additional travel time for four mode-based clusters (b) Travel time distribution by activity type

First of all, the travel time inside the HW tour has been separated into two components: HW direct trip and extra travel time. For each of the five groups of users (according to their mode choice), the results indicate that there exists an upper bound, which differs with respect to the chosen mode. Public transport users reserve a large part to the HW travel time itself, in opposition to bike and car users who have an upper bound of respectively 80 and 100 minutes. This can be related to the question of the travel time budget and gives the idea of a threshold travel time which if not reached for HW trips can be allocated to secondary trips. This threshold seems to vary from one mode to another.

Results indicate also that the travel time is also following different distributions with respect to trip purpose. For eating or daily tasks, the travel time is on average less than fifteen minutes while the average travel time for work-related trips approaches half an hour.

III.3.1. Emerging behaviours at temporal levels

We believe that these differences can be seen with emerging behaviours at the temporal level. Indeed, Given the information described on Figure III. 3, different activity types are not performed at the same time of the day and for the same duration. For example, eating activity has a strong peak around 12AM which result in a distinct usage of modes at this time of the day.

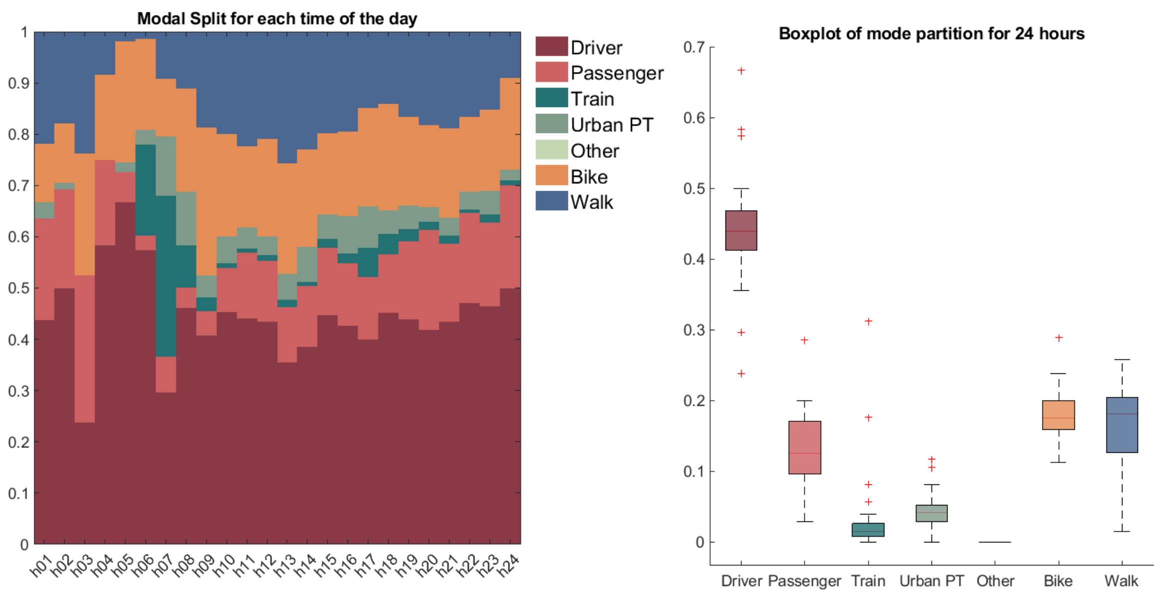


Figure III. 6 (a) Modal split by time of the day (b) Boxplot of distribution of modal share

Figure III. 6a shows the hourly values of the modal split for starting trips described in the observed database. We can see that modal share significantly varies along the day. On Figure III. 6b, these variations are shown for each of the studied modes of transport. As an example, car use as driver varies between 23.8% and 66.7%. This variance is lower for urban public transport users which could reflect a more multipurpose mode of transport.

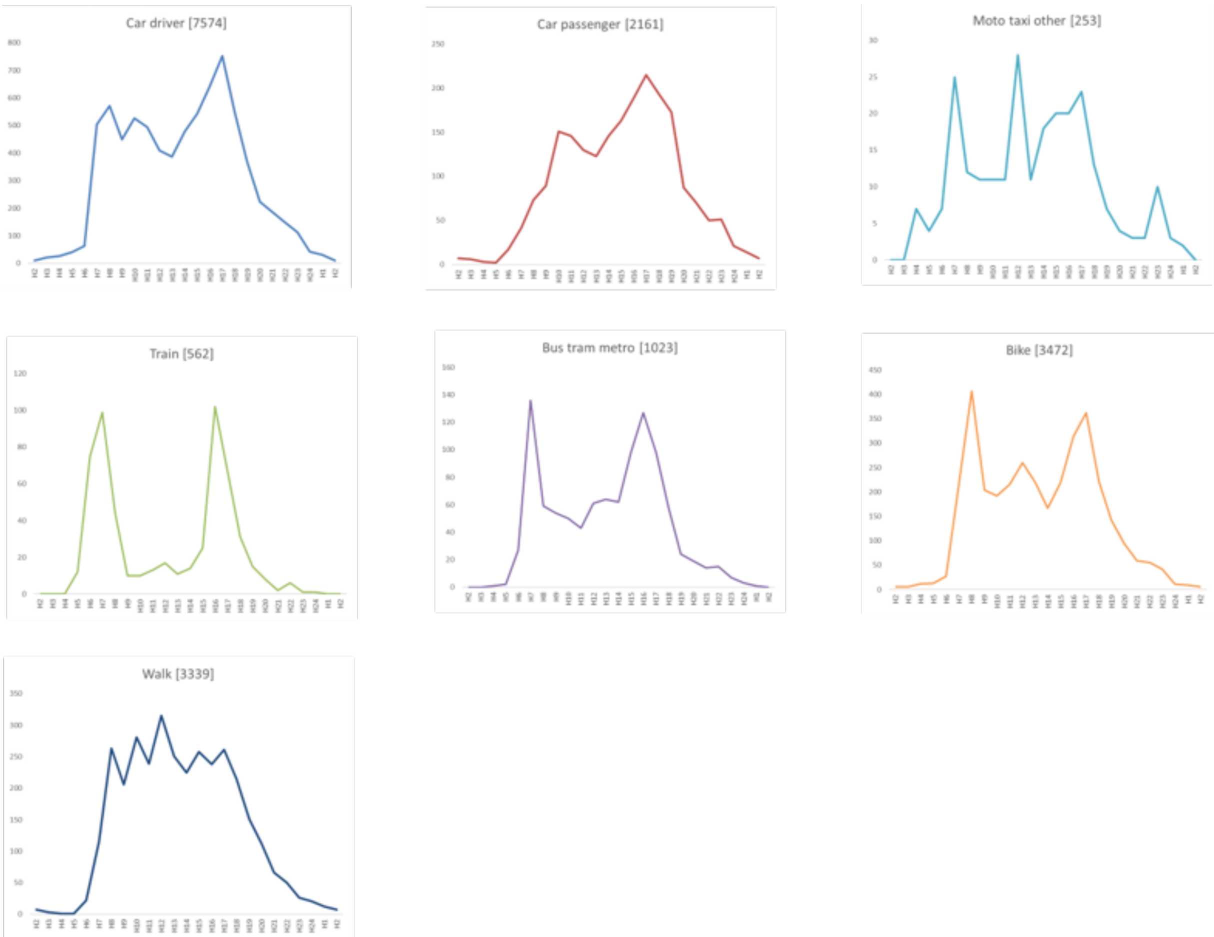


Figure III. 7 Usage profile of different modes

On Figure III. 7, the profile of each mode shows that they are used at different times of day which can on the contrary, be explained by different usage types of modes. For example, train seems to be used for commuting, while owned vehicles are more like the general demand profile. Walking seems to be used as complementing mode during off-peak hours.

The difference between theoretical modal split by activity is indeed particularly significant for the following combination: train to go to work and walk to eat (Figure III. 9a). This matrix is calculated based on the contingency table of the variables “mode choice” and “activity type” from the list of available trips. We first calculate the observed frequencies $F_{m,a}$ and then the expected (theoretical) frequencies according to the following formulation: $E_{m,a} = \frac{M_m A_a}{sample\ size}$. The matrix on Figure III. 9 shows the difference in percentage of the expected value: $\frac{F_{m,a} - E_{m,a}}{E_{m,a}}$.

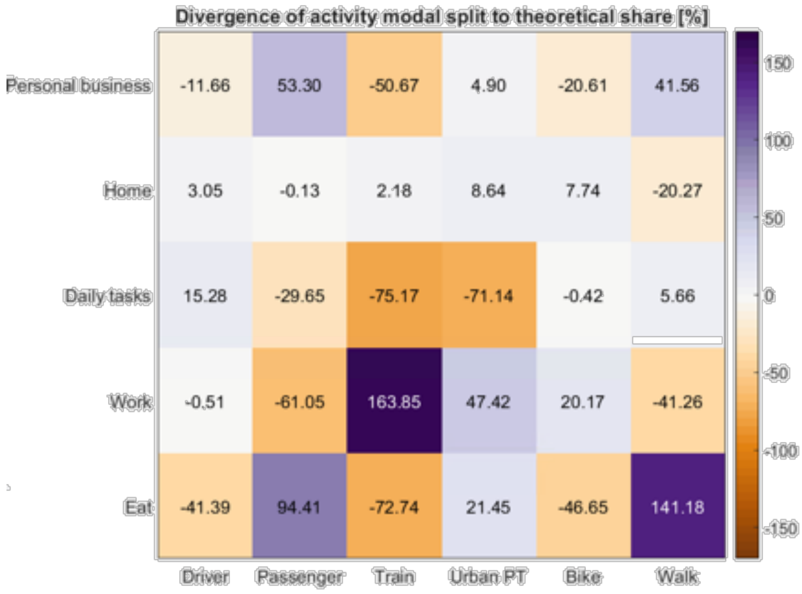
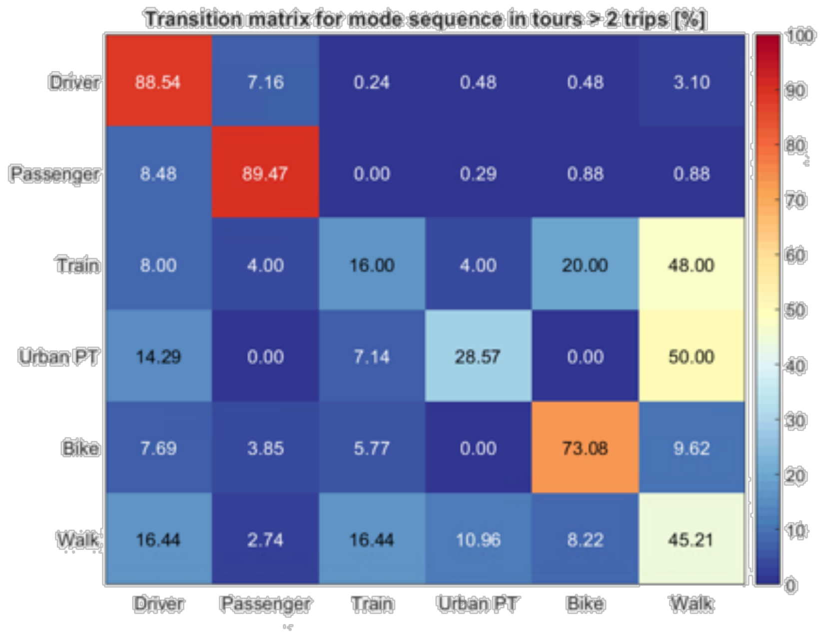


Figure III. 8 Divergence of activity modal split to theoretical share

In terms of mode choice, this can be observed through the sequence of trip modes within tours. Figure III. 9 shows the transition matrices for choosing a mode after having used any other mode. The highest probabilities are walking after using public transport and for the rest is to use the same mode for successive trips. When looking at the two commuting trips, the probability to choose the same mode is always higher than 50% and higher than 90% for owned resources (reaching 95% for car users).



(b)

Figure III. 9 Transition matrix for mode sequence

Figure III. 10 shows the example of car commuters in the form of a probability tree. From the dataset, we calculated, from the first trip of the day starting with the car, the share of travellers by decision. These decisions include :

- continuing their chain with the car p(car)
- continuing their chain with another mode than car p(other)
- ending their tour, which in this particular case results in returning home p(end)

When car is chosen to leave home the probability to choose the car later on the trip chain is much higher than other modes. Furthermore, when another mode is introduced in the chain, the probability of returning home is lower than to ultimately go back to car before ending the chain.

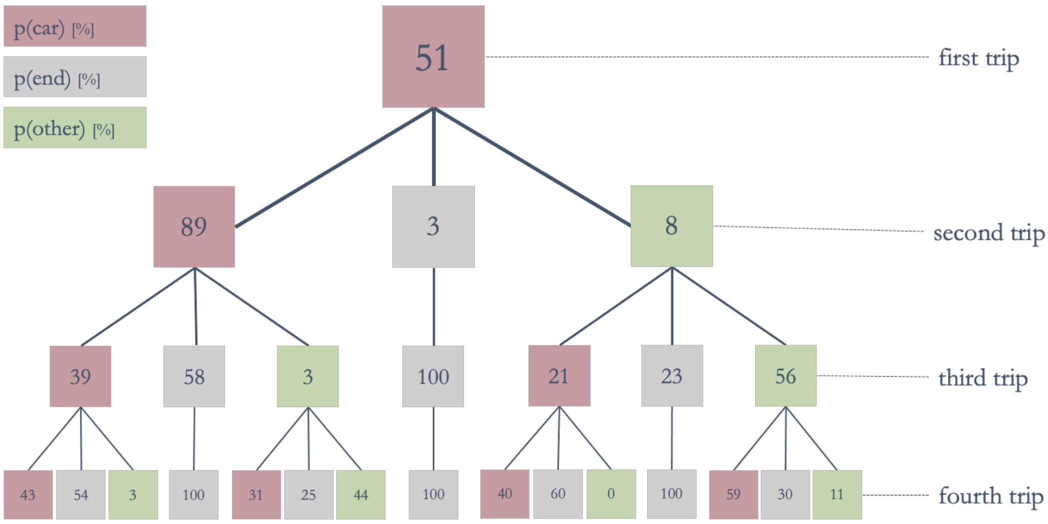


Figure III. 10 Car use probability tree

This shows the impact of owned resources on the propensity to have a multimodal trip chain. When a person leaves home with a bike or car, (s)he will return home with this mode, and mostly use this mode for intermediate trips. However, there is not a big difference in the total number of trips made during the day with respect to that decision.

III.3.2. <u>Emerging behaviours at the spatial level</u>

AS were calculated only when there were enough unique points, i.e. 283 workers were used for ellipses parameter estimation. Figure III. 11 shows the difference between the orientation of the calculated AS and the orientation of the HW direct segment. The points are displayed with increasing eccentricity, this means that when the ellipse is elongated, the difference tends to be very low while the highest errors appear when the ellipse has more of a round shape. The strong correlation of 0.8 between the two values supports the use of HW axis as an approximation of AS orientation.

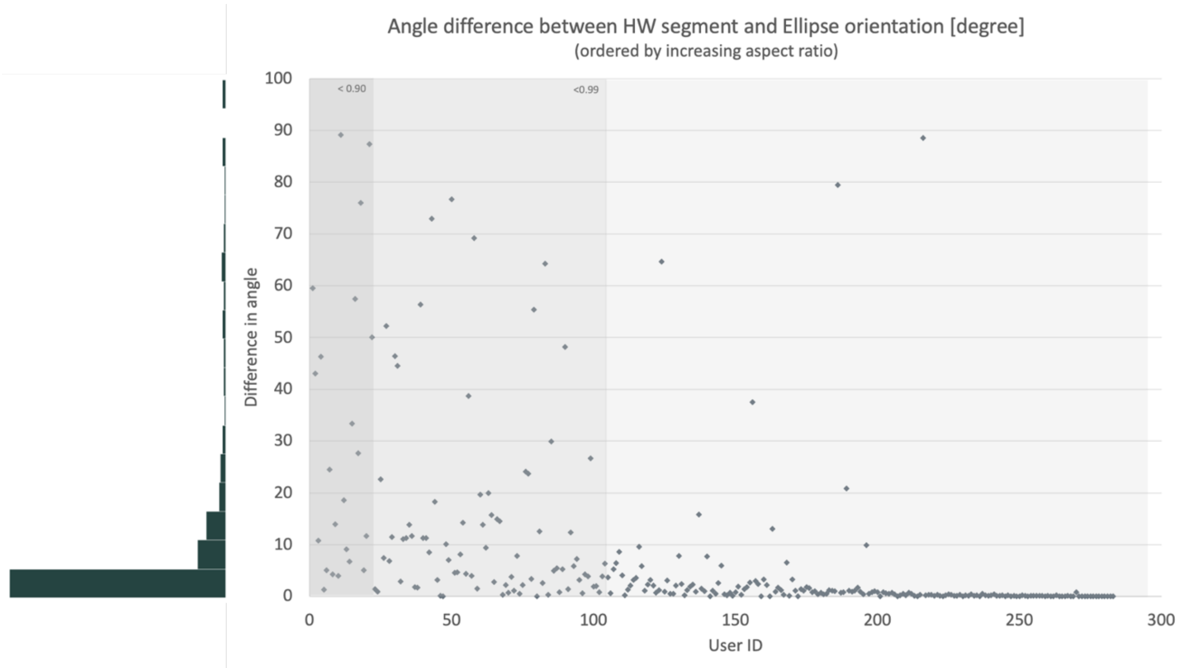


Figure III. 11 Orientation vs HW axis (a) histogram of the distribution of angle difference (b) individual differences

The correlation is even higher between the centre of the AS and the middle point of the HW segment Figure III. 12. This can be explained because of the ellipses which have a less elongated shape, the centre remains fixed while variations in term of angle do not have a strong impact. Figure III. 12 shows the distribution of the distance between these two points. The median is 1.1km, being higher for two types of users: the commuting distance is high (more than 100 km) or home and work are both located in peripheral areas.

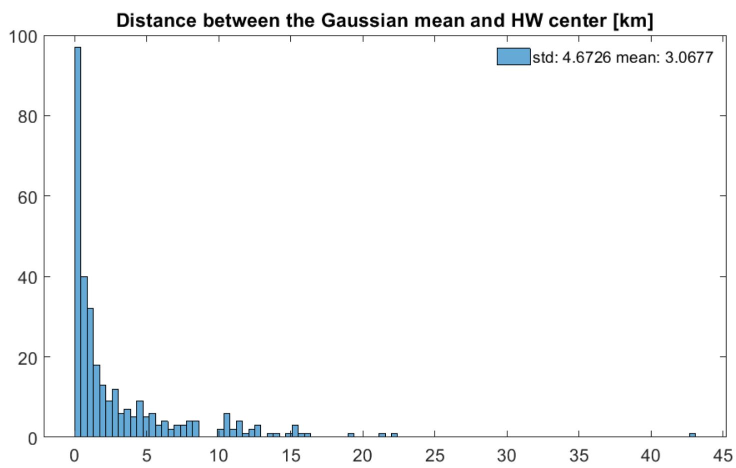


Figure III. 12 position of the centre vs HW centre

Figure III. 13 shows an indicator of ellipse's radii and aspect ratio for different user types. We clearly see that the aspect ratio is low for public transport (close to zero for all train users) and that the area is also the smallest for urban public transport users. The size of the Ellipse is also small for bike users, which have a more round AS, this can be explained because they are not subject to any limitation for the direction they can go, such as train/bus line or even roads. A high aspect ratio (eccentricity tending to zero and circular shape) reflects an AS less impacted by home and work locations with secondary activities potentially off the HW direct route (typically car and bike users). In opposition, if the AS has an eccentricity closer to 1, it means that the secondary activities are more likely to be located on the HW segment or close around those two focal points. This is the case for public transport users.

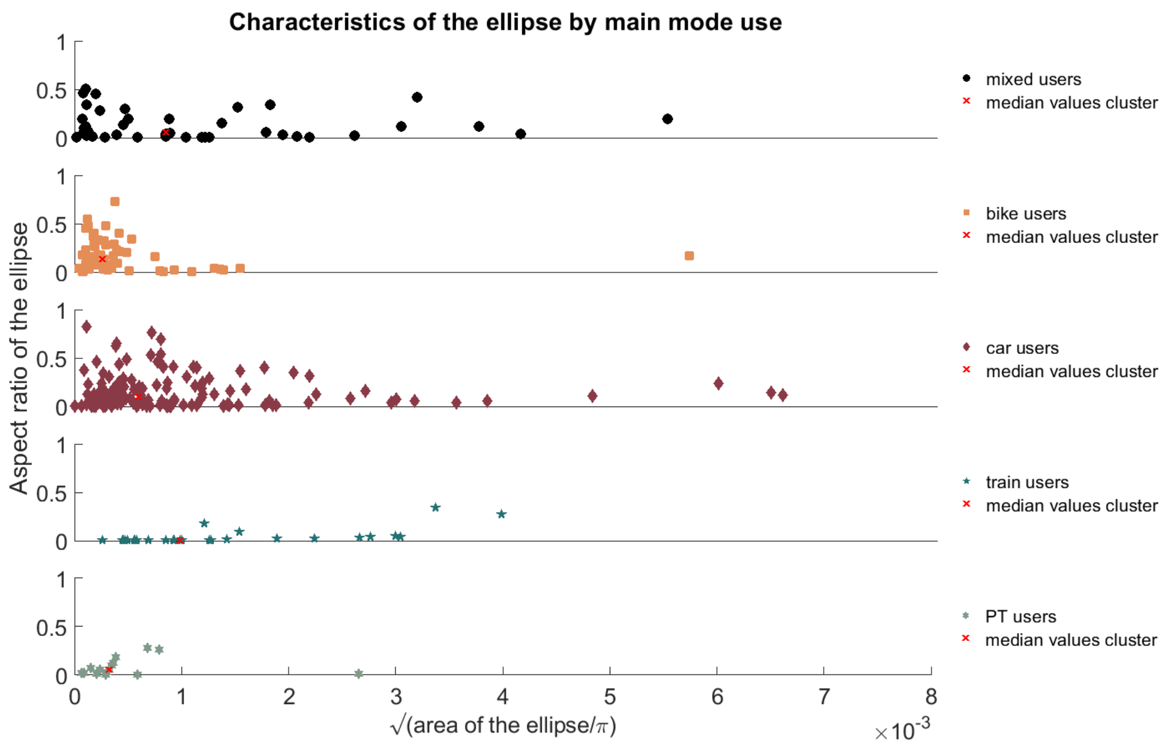


Figure III. 13 Characterization of ellipse shape by mode

The clustered users reveal that the typical ellipses for the mode-specific groups have characteristic shapes: thin elongated ellipses for public transport users; larger for train and smaller for bus users. Ellipses are more rounded for private transport users: larger for car and smaller for bike users (Figure III. 14).

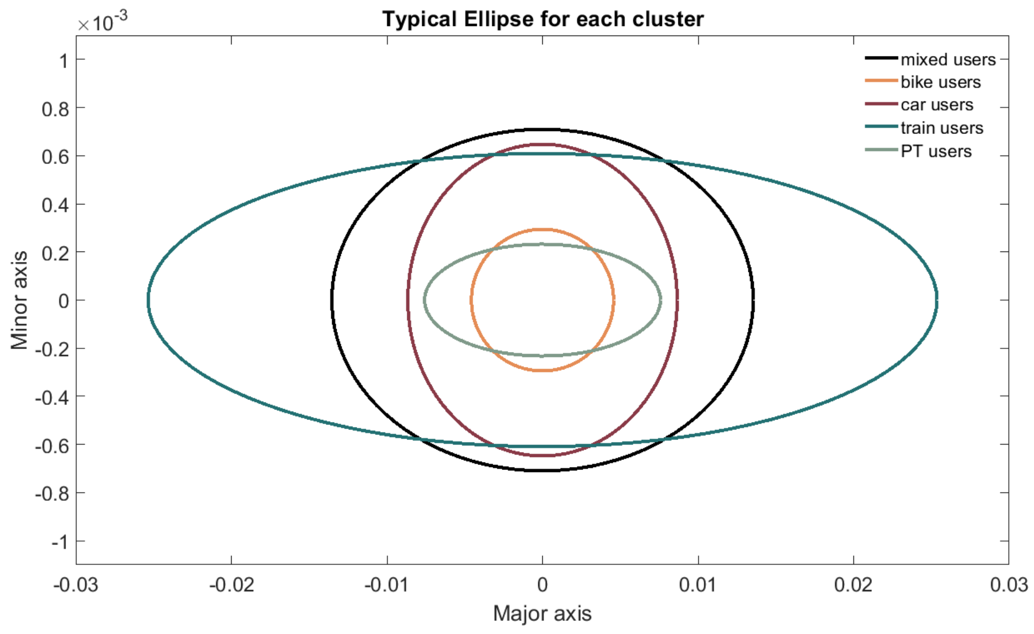


Figure III. 14 Typical ellipse for 5 mode-based clusters

30% of the travellers visited only two points in their HW tour, each day and resulted in the estimation of a segment; in order to relax the constraint of choosing an activity location strictly between the home and work location for these users and estimate AS even with low number of observations, we use the revealed regularities of similar users and proposed a probabilistic approach to the AS estimation. Figure III. 15 shows the output of aggregate probabilistic AS for the same group of users as seen on Figure III. 4b and for which only seven out of eleven users where subject to the single AS calculation.

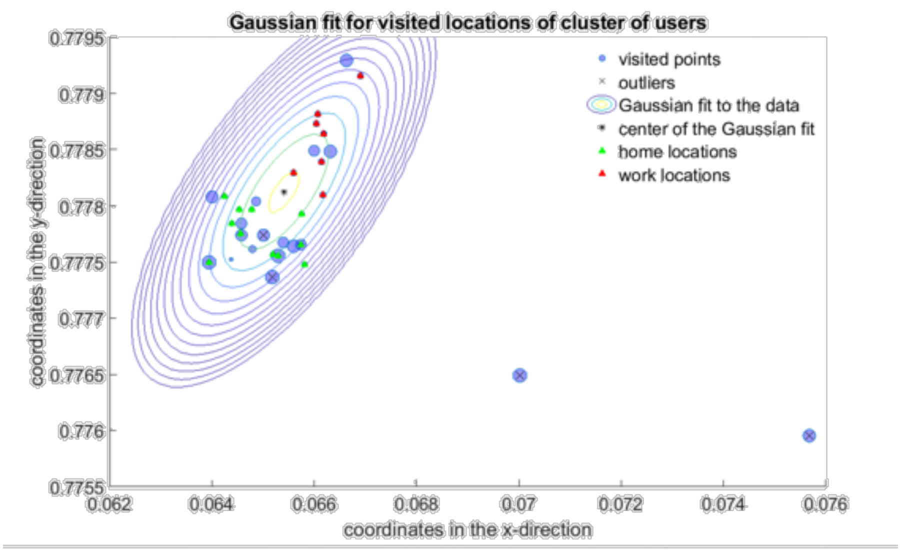


Figure III. 15 Gaussian fit for visited locations of a Ghent HW cluster of users

We can see that, characteristics of the fitted gaussian after outlier detection and removal follows the same rules as what has been observed for individual workers. The centre is situated between the home and work areas and the contour lines of the distribution are oriented along this HW axis. Additionally, a single ellipse does not represent properly the option of 'crossing the line' and perform and activity outside the AS limits. This feature is very desirable when applied to modelling demand. That is why the estimation of distributions provides a more realistic perspective to be used in future research among other in order to estimate secondary activities' locations.

The following key aspects in terms of input-output of the demand modelled described in this thesis can be underlined:

Table III. 6 Summary of conclusions for the modelling approach

Input possibilities	**Desirable output**
Clustered activities	Occupation of the zone by time of the day
Characterization of generalized AS	Modal split by time of the day
Correlation mode and distance travelled	Usage profiles by mode
Private vehicle constraint	Activity durations by zone and time

III.4. Conclusion

This preliminary work presented descriptive statistics and empirical analysis on mode choice in relationship with time of day and chains of activities as well as spatial distribution of visited locations in the HW tour. Starting hypotheses have been supported by a sample of Ghent population described in the BMW database and these empirical observations can be used as the basis for more accurate travel demand modelling. The modal split varies throughout the day and successive mode choices are strongly correlated to each other. This stands in particular for owned vehicles (bicycle or car), as there is a constraint of carrying the resource around. The usage profile of modes and the transition matrix show complementarity of modes, in particular walking which can be considered as a mainstay of many multimodal trip chains. Activities which are most frequent are usually located within a shorter distance from home or work, while longer distances are travelled for infrequent activities for which the destination cannot be substituted, like business trips or visits to family or friends. Parameters defining AS can be estimated by knowing the home and work locations of an individual which gives a good approximation of its centre, orientation and one of the axes of the ellipse. Aspect ratio and area are governed instead by the commuting mode choice. For applying these observations to aggregated groups and using AS as a soft constraint for choice modelling, we fit, instead of a single ellipse, a bivariate normal distribution for groups of users. These findings can be used for estimating mode specific travel demand by time of day and regions.

Chapter 4

Highlights of the chapter

1. Observable variations of daily demand emerge from distinct temporal profiles, defined as trip-primitives
2. Description of a simplified activity-based approach to estimate dynamic trip generation
3. We show Markov Chain Monte Carlo method's ability to calibrate parameters with aggregated dynamic trip data

In this chapter, we introduce two fundamental concepts applied to this thesis. First of all, the principle of "trip-primitives" which allows us to model the trips generated during the day according to the activities at their destination. A second pillar of the proposed methodological framework is the choice of the estimator applied to these functions.

We introduce here a simplified approach, through a mixture model to introduce the estimation process. "Trip-primitives" are parametric functions that allow an ABM trip to be incorporated into a model based on an approach based on individual trips. Thus, we model a correlation between trips at different times of the day, even if this link is not formally modelled through explicit individual trip chains. In order to describe the primitive functions, we have selected a generalized extremum law for each of the five selected activities. The distribution of demand during the day is calculated by summing the estimated distributions for each activity, multiplied by a fixed factor, corresponding to the overall probability of performing one activity rather than another.

The chosen probability density formula in this chapter was selected primarily for its simplicity of form as it is only described by two parameters. We also looked for a shape characterized by a maximum point at a certain time of day with variable variance and asymmetry. The two parameters are estimated by a Markov Chain Monte-Carlo (*MCMC*) method.

This method is used in various fields of transport engineering and in particular in demand modelling, for example in the context of the generation of synthetic populations. Here, we use the *MCMC* method to calibrate the two parameters of each of the "trip-primitive" functions. To do this, we select a probability density function to define the plausible domain of each of the parameters and through an iterative learning algorithm, we update this density in order to obtain the "posterior". The update is made by comparing the profiles generated by the addition of primitives with the observed departure rates. The distributions obtained as a result of this calibration process make it possible in this case to describe, for example, the variance of the time at which a starting peak for a given activity will occur.

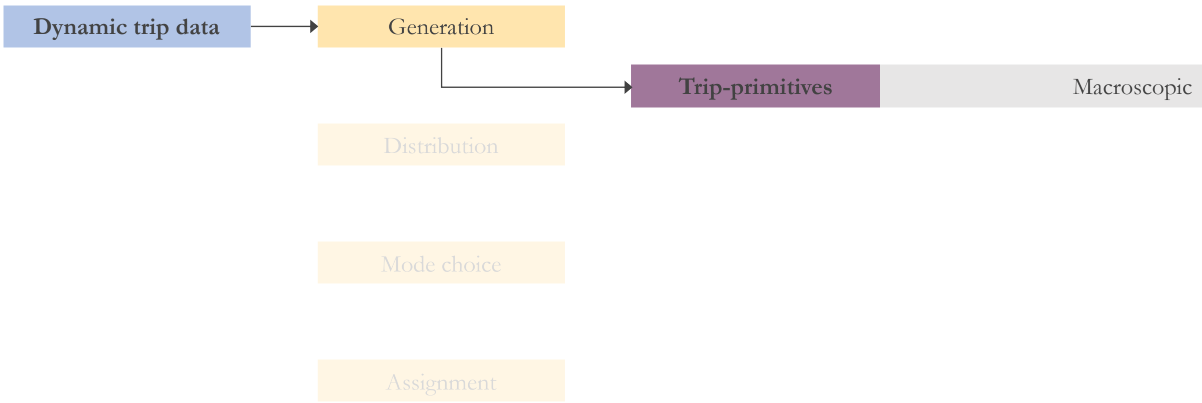


Figure IV. 1 Thesis framework chapter 4

The work presented in this chapter has been described in the following paper:

<u>“Generating Macroscopic, Purpose-Dependent Production Factors Through Monte Carlo Sampling Techniques”</u>

Transportation Research Procedia
20th EURO Working Group on Transportation Meeting, EWGT 2017, 4-7 September 2017

IV. TRIP-PRIMITIVES

IV.1. Introduction

While estimating origin-destination (OD) demand flows usually requires a large amount of data, nowadays a key issue in traffic engineering is to estimate the trip purpose while protecting user privacy. The aim of this chapter is to derive from macroscopic and aggregate information production distribution for each Traffic Analysis Zone (TAZ) of a study area, with a trip-purpose specification.

This chapter presents the possibility to use Monte Carlo simulation to characterize demand flows. We suggest a procedure for estimating activity-production factors with a Markov Chain Monte Carlo (*MCMC*) procedure. This approach is used to approximate a set of functions that describe the production of trips from one specific zone along the day. This method requires a low level of information and computes the likelihood with respect to the number of generated and attracted trips, reducing as much as possible the required information at the individual level.

Since the proposed approach tries to estimate the number of (activity-based) components within an OD matrix, it can remind of the well-known Gaussian mixture model. However, an *MCMC* model has the advantage of including additional information in the estimation and so reduce the number of unrealistic solutions with respect to the Gaussian mixture model. *MCMC* methods are more and more used in the travel behaviour modelling and in the creation of synthetic population (Beckman, Baggerly, and McKay 1996; Saadi et al. 2016; Farooq et al. 2013) since they allow to draw samples and to estimate discrete outcomes from known probabilities for the different variables used to qualify agents of the disaggregate models.

The rest of this chapter is structured as follows. The next section introduces the concept of trip primitives and the calibration process. Then, we test the reliability of the proposed approach and the case study. Lastly, conclusions are discussed to the possible development of the model.

IV.2. Methodology

IV.2.1. Trip primitives

When modelling traffic demand at an aggregated level, the area of study is usually subdivided into zones of origin and destination (TAZ). Continuous-in-time mobility patterns are discretised and collected into matrices, which represent the OD flows within a certain time period. In this matrix representation, the distribution of the mobility patterns is assumed stationary, i.e. intra-period dynamics are usually neglected.

In our approach the demand is assumed to be derived through a convolution of different activity patterns, whose dynamics emerge from individual activity-travel behaviour. The methodology applied in this chapter is based on the concept of **trip-primitives**. The trip primitives are functions describing the variations in terms of generated trip in function of the time of the day. They are defined for a given activity type and their sum is equal to the total observed demand at every time of the day. An example of trip-primitives is shown in Figure IV. 2 representing the mobility in Luxembourg as described in the MODU 2.0. On this figure we can see three distinct primitives: work, duty, and leisure, each revealing a different dynamic.

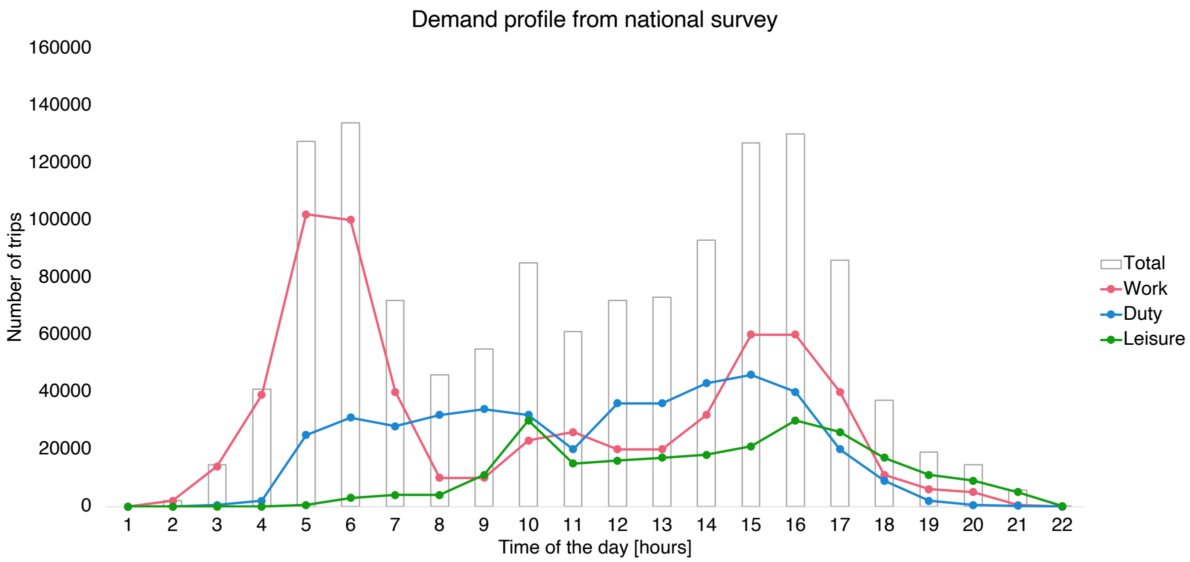


Figure IV. 2 Trip Primitives in MODU 2.0

In this study, we consider that every zone may have a different profile and consequently trips to and from those zones may be derived from different trip-primitives. Trip-primitives can be defined in different ways which may depend on the generation model used. In this application, we want to show the possibility of modelling trip-primitives in the simplest way, without any data or behavioural assumption needed at the individual nor zonal level. To do so, we assume that the trip-primitives can be described by a generic probability density function. A distribution W is selected according to recommended characteristics such as the form and number of parameters. Each primitive could theoretically be modelled by a different kind of function. For each zone, we consider that the complete demand T is the combination of the demand for n activity types, being each characterized by a probability of appearance D_a and an individual probability distribution W_a.

$$T = \sum_{a=1}^{n} D_a \; W_a \tag{4.1}$$

The distribution of the travel demand along the day is therefore modelled as the sum of the five distributions, multiplied by a factor corresponding to the global probability of performing one activity or the other. In the simple setup presented here, no variations of the activity type's spread are taken into consideration. However, the value of D_a are estimated as unknown variables.

IV.2.2. Calibration process

To separate the demand into activity-based flows, the distributions along the day are approximated through an MCMC. Figure IV. 3 describes the process used in this methodology. The filled blue boxes are the only necessary inputs, described in the following section. In yellow are the two main outputs. The first one "updated parameter distribution" is the raw product of the *MCMC*, used to define the trip-primitives. The last three elements are the actual core of the model. Once the generation model is selected, and the parameters defining it are chosen, the sampling process is used to calculate a proposed value of simulated trip and compare the outcome to a set of criteria in the Bayesian update. This mechanism is described thereafter.

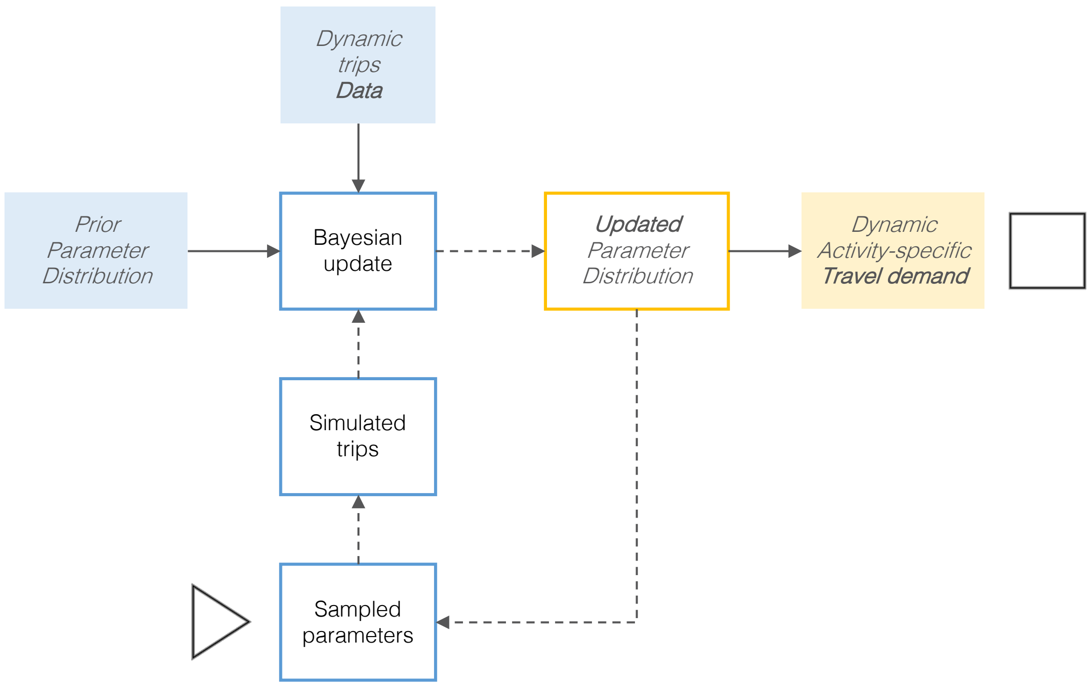


Figure IV. 3 Process of the MCMC calibration

IV.2.1. Input

The necessary input of the process outlined in Figure IV. 3 can be the observed time-dependent aggregated demand flows, which are used as datapoints to fit, for each time of the day corresponding to an observation. This data can be obtained from an available dynamic OD matrix or from observed movements (e.g., probe data from smartphones, travel diaries).

The prior information $\pi(\mathbf{\Theta})$ can be of a different kind. It is a density function defined for each of the estimated parameters and reflects the a-priori knowledge about its behaviour. In the least demanding case in terms of data, it can simply be the modeller's intuition, result from logic or be reduced to the "non-informative" case, where the observed data govern fully the estimation. Alternatively, a population sample, aggregate information or previously estimated posterior distributions can be used, if available.

Finally, the initial values can be selected according to the prior and are used as a starting point for each Markov chain and can significantly influence the efficiency of achieving convergence, given the non-linear nature of the estimation problem, and the likely under-determinacy due to a limited set of observations.

IV.2.2. Output

The primary output of the iterative process is the posterior distribution value of every parameter resulting from the Markov Chain process. Its main characteristic is that it is sampled from a probability distribution which should be representative of the posterior distribution. Its average value is used as a point estimate to decompose the observed demand by activity type and trip type (start or end) and generate the utility primitives which can be used to estimate the activity-specific trips, or the departure times to /from an activity. The posterior can be regarded as a distribution over the population.

IV.2.3. Operations

Candidate parameters are drawn from a distribution function of possible values. They are either accepted or rejected according to a rule which depends on two components: the likelihood and the plausibility of the set of parameters.

Because the target distribution $P(\mathbf{\Theta})$ is unknown, direct sampling is not possible, unlike the case of standard Monte Carlo sampling methods. Additionally, the procedure is more efficient thanks to the availability of a prior. The used statistical inference (equation(4.2)) is based on the Bayes' formula which describes how the prior belief $\pi(\mathbf{\Theta})$ is adapted thanks to observed data $\boldsymbol{T}$ and the calculated likelihood $\mathcal{L}$, into the posterior $P(\mathbf{\Theta})$, for all parameters of the utility functions. It also includes the probability distribution of $\boldsymbol{T}$. The term $P(\boldsymbol{T})$ is not known and fully independent from any parameter value $\mathbf{\Theta}$ so it does not impact the output and is not included in the estimation process. The posterior is thus a product of the likelihood and prior only.

$$P(\mathbf{\Theta}) = \frac{\mathcal{L} \cdot \pi(\mathbf{\Theta})}{P(\boldsymbol{T})} \tag{4.2}$$

The initial belief $\pi(\mathbf{\Theta})$ is updated in successive steps, which create a Markov Chain whose stationary distribution converges to the target distribution $P(\mathbf{\Theta})$.

To apply equation (4.2), we define the likelihood, priors, and a transition operator which allows the creation of a chain that converges to this condition. In this paper we make use of the most common sampling method developed for MCMC based on the Metropolis-Hastings algorithm (Metropolis et al. 1953).

IV.2.4. Sampling

For each of the parameters, a prior probability density function and starting point is defined. The following proposal function is used in each iteration to draw a new sample:

$$\theta'_i = \theta_i + Normal(0, \Delta_\theta) \tag{4.3}$$

This form justifies a key property of the Markov chain: given θ_i, the i^{th} value of the chain, a proposed value θ'_i is generated in function only of the previous element of the chain defining the sampling density function at iteration i. The choice of Δ_θ is adjusted in order to speed up the process and reduce autocorrelation. The variance influences the step size, i.e. the possible difference between the proposed value of a parameter and the current one. A step size which is too small takes too long to explore the feasible space while a high step size allows easy escape from local minima but struggles to converge. The objective being to have a mixed and converging chain, the value is selected such that it leads to an acceptance rate between 10% and 70% for each dimension.

The model is applied to the proposed set of parameters in order to evaluate them against observed data.

IV.2.5. Score and update

To reflect how the sampled parameter is related to the modeller's assumption, the overall score is calculated as follows, implementing the prior as a log-likelihood penalization. Because of the logarithmic form of the likelihood and prior, equation (4.2) results in an additive score (equation (4.4)). The first element of the score reflects how close the modelled daily profile is with respect to observed data, and the second element serves as an evaluation with respect to expected values of each parameter.

$$S_i = \frac{\mathcal{L}_i}{\omega} + \sum_{\boldsymbol{\Theta}} \log\left(\pi(\boldsymbol{\Theta}_i)\right) \tag{4.4}$$

In some cases, the likelihood's order of magnitude means that the impact of the prior is negligible. In order to balance the effect of data $\boldsymbol{T}$ and prior π, a scaling parameter ω is added and tuned (with a trial-and-error method) to obtain a reasonable acceptance rate. Its value depends on the number of observed estimated parameters and datapoints as well as their order of magnitude.

The score is then compared to the one evaluated in the previous iteration (equation (4.5)) through a logarithmic conversion of the acceptance ratio. On one hand, if the score improves, the proposed value is included in the Markov chain to build the posterior distribution and used as reference for the next candidate sampling. On the other hand, values leading to a decreasing score, even repeated, may or not be included, based on the rejection rule (equation (4.7)) to preserve the distribution density, and add flexibility.

$$\varphi = \exp\left(S_i - S_{i-1}\right) \tag{4.5}$$

$$\mathrm{p}(\theta, \theta'_i) = \min\{\varphi; 1\} \tag{4.6}$$

$$\theta_{i+1} = \begin{cases} \theta'_i \; with\; probability\; p(\theta_i, \theta'_i) \\ \theta_i \; with\; probability\; 1 - p(\theta_i, \theta'_i) \end{cases} \tag{4.7}$$

With this rejection rule, the Markov Chain approach can avoid remaining blocked into local optima and ends with a sample from the posterior target distribution. The chain will not stop at a "true solution" but sample around a plausible true solution. There is no pre-specified stopping criterion, the number of iterations or only the number of accepted steps is chosen beforehand. The necessary number of iterations rapidly increases with the size of the problem and the number of variables to estimate. In order to avoid obtaining a biased posterior due to the initial parameter selection, a burn-in period is also defined, and these first steps are removed from the chain because they are likely less representative of the posterior.

IV.3. Case Study

IV.3.1. Database

The data set used for validating the proposed model is the "Behaviour and Mobility within the Week" (BMW) dataset (Castaigne et al. 2009), which was collected by the KU Leuven and the University of Namur in 2008. 717 valid travel diaries were collected, which describe a one-week period for the city of Ghent, Belgium. This dataset is the same as the one used for the empirical analysis described in the previous chapter.

In order to apply the proposed methodology, the first necessary step is to create a dynamic OD matrix. After ensuring the consistency of the database, we artificially generated an OD matrix by aggregating all trips occurring during weekdays of the study period. In the BMW study, the area was composed of 17 zones. Even though these zones are difficult to be identically recreated, postal codes have been used to cluster the respondents in 17 artificial zones. Five supplementary geographical units were created for the city centre of Ghent, inside which trips represent 61% of the complete demand.

The activities considered in this case study are the following:

- AT1 - Activities usually located in residential areas i.e. "Home" and "Visit to family of friends";
- AT2 – Activity "Work";
- AT3 - Leisure activities, such as "Walking/riding", "Leisure/sport/culture" and "other activities";
- AT4 - Regular and unavoidable activities, such as "drop off/ pick up" and "eat";
- AT5 - Activities often located in town centres, such as "Shopping", "School" and "Personal business"

IV.3.2. MCMC settings

W is selected to follow an Extreme Value distribution. This form has been chosen for the following characteristics:

- Low number of parameters (2)
- One parameter describing the position of a peak on the x-axis (μ)
- One parameter describing the dispersion (σ)

- A non-symmetric behaviour around the peak.

The effect of the two parameters can be seen on Figure IV. 4.

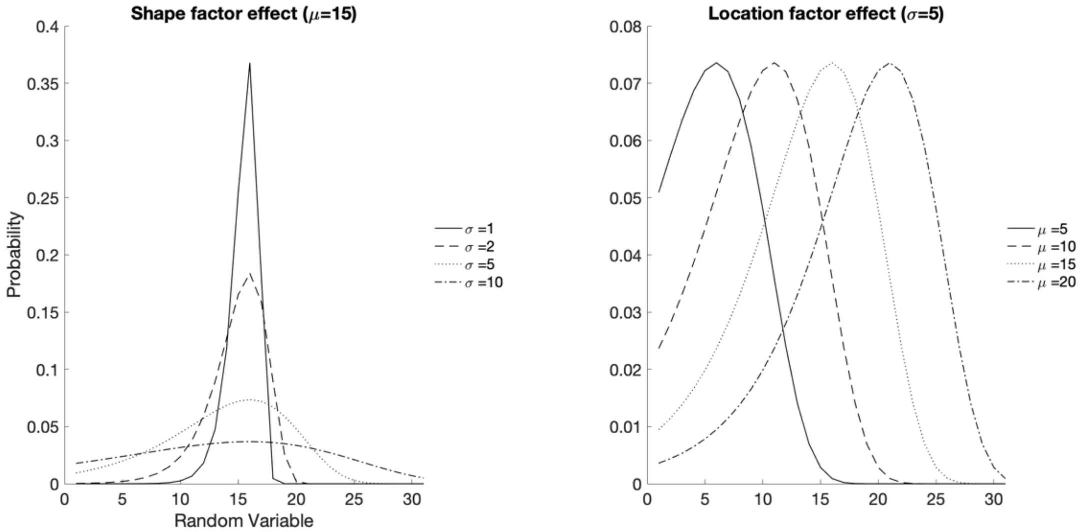


Figure IV. 4 Parameters of the Extreme Value law

In each iteration of the *MCMC*, parameters are renewed, and their combination is evaluated according to the a-priori information. The likelihood of the set of parameters is calculated with respect to evidence. The ideal configuration is a dynamic OD matrix, nevertheless the method can be adapted to handle GSM data or loop detector, for example.

The set of parameters is accepted or rejected as a whole, and each parameter serves as a starting point for the following proposed parameter. The algorithm can be performed for each zone in parallel, independently. For this reason, the methodology doesn't get more complex, and the rapidity is stable with and increasing number of areas.

The prior distribution of the sigma values is a uniform distribution, with values between zero and 10. The location parameter, on the contrary, has a more meaningful interpretation. Because it relates directly to the typical departure time of a specific activity type, this is where the knowledge and assumptions are the most easily introduced. In this experiment, the approximation is limited to the determination of activities being more likely started in the morning, evening, or afternoon.

IV.4. Results

To test the method and observe results variations, the MCMC has been used to determine the primitives of trips departing from the largest zone outside Ghent. 1120 trips were analysed, and departure times were used to determine two parameters of five functions. The total distribution between activities is fixed, based on the complete survey proportions. Analysis of the quality of the results is simply based on the comparison of profiles between the real demand and estimate generated demand for the traffic zone.

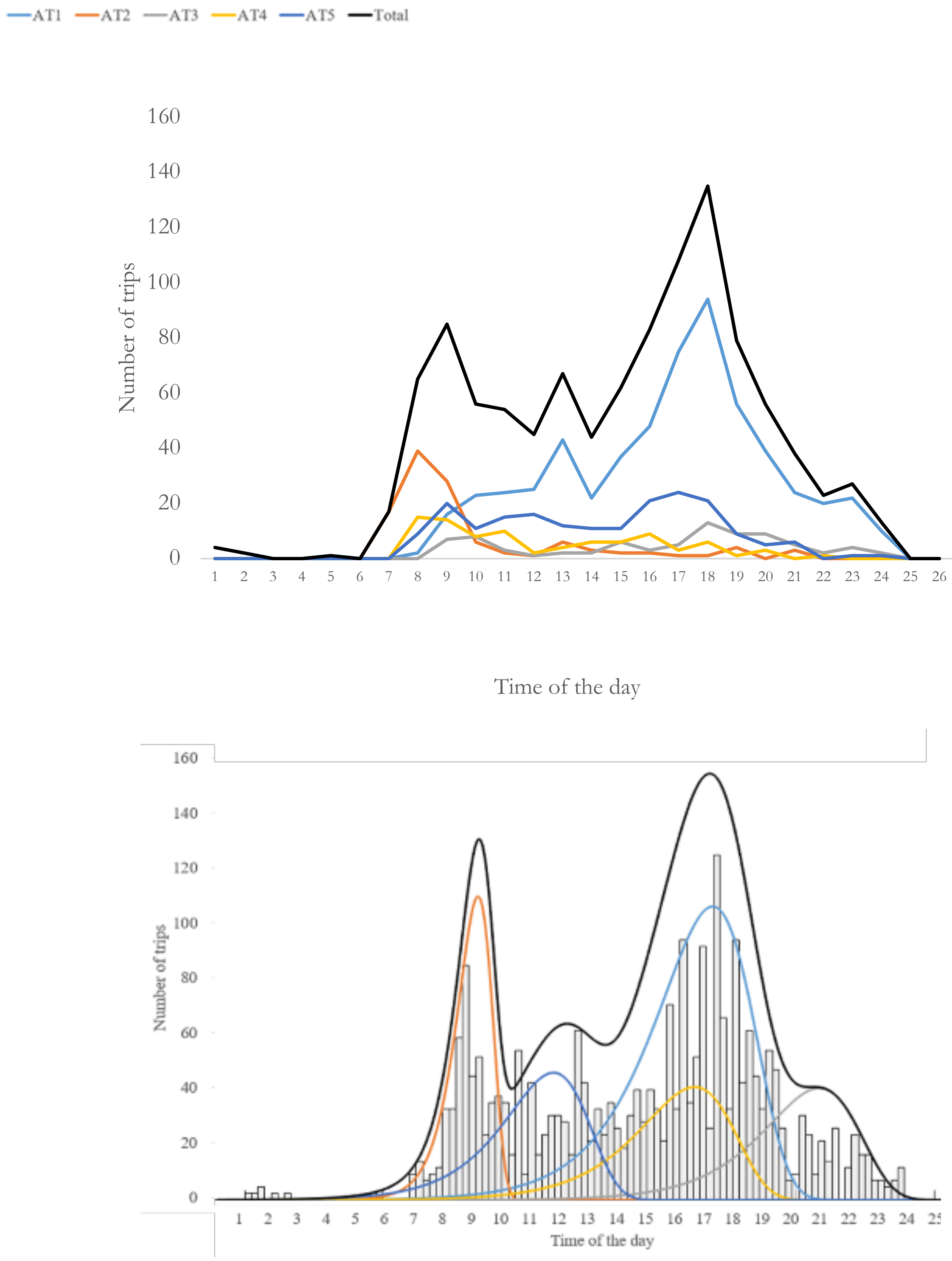


Figure IV. 5 Departure profile by activity for zone 1 (a) From the survey; (b) From the simulation.

Comparing these two figures, it appears that the proposed methodology identifies activity patterns of the zone in their general shapes. In particular, peaks are identified to the appropriate trip-purposes: the most straightforward are indeed "work" in the morning and "going home" in the afternoon. The model recreates the increased demand at lunch time and in the evening without being able to accurately determine the other activity components. Two reasons explain this observation; not only fewer observations exist for these activities but also the restriction about the form of the function is not as suited as for the peaks. Another test for the simulation quality is the comparison between the estimated location parameter of each evaluated distribution and the typical departure time of each activity calculated from the database. Here again, the two major activity types are correctly characterized, whereas the three secondary purposes compose peaks at regular interval during the day, without specific relationship to the real meaning of trip-purpose. It is important to note that the procedure treats without much difference these activity-types: same prior, almost same percentage of users, which can explain the inaccuracy to distinguish them. In sum, this experiment shows that this MCMC-based methodology is promising and allows determining activity-based priors, giving acceptable results for the most characteristic activity-types. Nevertheless, more precision in the selection of probability distribution form and activity clustering would be necessary for a better recognition of the complete demand. Moreover, an advanced generation model would allow to distinguish better the activity profiles.

IV.5. Conclusions

In this chapter, we presented a novel procedure to estimate purpose specific flows in time with respect to the zone of departure of a journey and without information at the user-level. A Monte Carlo technique is proposed for evaluating dynamic classification of flows, with respect to trip-purpose. For these simulations, flows are disaggregated and used as evidences for the calibration of an unknown distribution, with all the available information they can contain. Assumptions might be added to the models in order to improve their reliability. The validity of the models gets subject to particular care in case of a large number of parameters and requires a high amount of iterations of the simulation to get a stable result. It is still generally able to recognize typical characteristics, such as morning commute, but the link to specific activity can be concluded over a second phase. The components representing activities without preferred starting time and less occurrences are indeed harder to be identified in these settings. A drawback is that the quality of representation and evaluation of less common cases. In particular, small zones, off-peak hours as well as activities with fewer observations. Nevertheless, by choosing more constraints and selecting each parameter with a great precision a priori, the model itself can also become more reliable. Furthermore, the inherent shapes of the selected functions do not have enough flexibility to reproduce complex demand profiles and so other formulations may be needed to recreate better observed demand. The proper distributions are function of the possible knowledge and assumption that can be made on the behavioural components, each activity can possibly be represented by a different kind of demand model. However, this could mean handling more parameters for some cases and so, an increase in the complexity.

Chapter 5

Highlights of the chapter

1. Application of marginal utility formulation to model heterogenous groups of people is proposed to derive behavioural-based trip primitives
2. Utility-primitives are introduced and defined at an aggregated level
3. Departure time choice model calibration with MCMC is developed and assessed using multiday travel diary data.

In order to reinforce the behavioural aspect of the proposed method, we integrate a third pillar, i.e. “utility maximisation”, into the "intermediate" model framework presented in the previous chapter. This chapter is also based on the assumption that the total daily demand is composed of activity-specific profiles. These profiles emerge in our approach from aggregated utility functions by adopting a probabilistic Logit model that gives them their functional form, instead of using the generic law of Extreme Values as it was done in the previous chapter. Here we present the methodology that allows to calculate these functions and the parameters estimated through the MCMC. The level of complexity of this model extension increases severely with the introduction of this method since the functions are parameterised by at least 12 parameters instead of the previous two.

Indeed, for a given type of activity, we define three marginal utility functions. The main one corresponds to the activity in question itself, one to the whole chain of trips and activities before the trip to this activity and finally the chain after this activity. This simplification of a complex tour in which there is one type of activity, not only simplifies the problem but also allows for the inclusion of any type of chain and therefore any individual programme. This hypothesis is fundamental in order to apply advanced formulations from a behavioural point of view to the whole population, and to apply it to the whole area under study.

The marginal utility functions used in this thesis are taken from the literature and some of their characteristics are fixed and selected according to the type of activity. The other variables, as well as some aggregate demand values, are estimated by the MCMC process. In the same way as in the previous chapter, the only information needed for the calibration, apart from the a priori estimates of densities, are profiles of the generated demand.

This methodology is developed to estimate the total demand generated in a study area by time of day and type of activity.

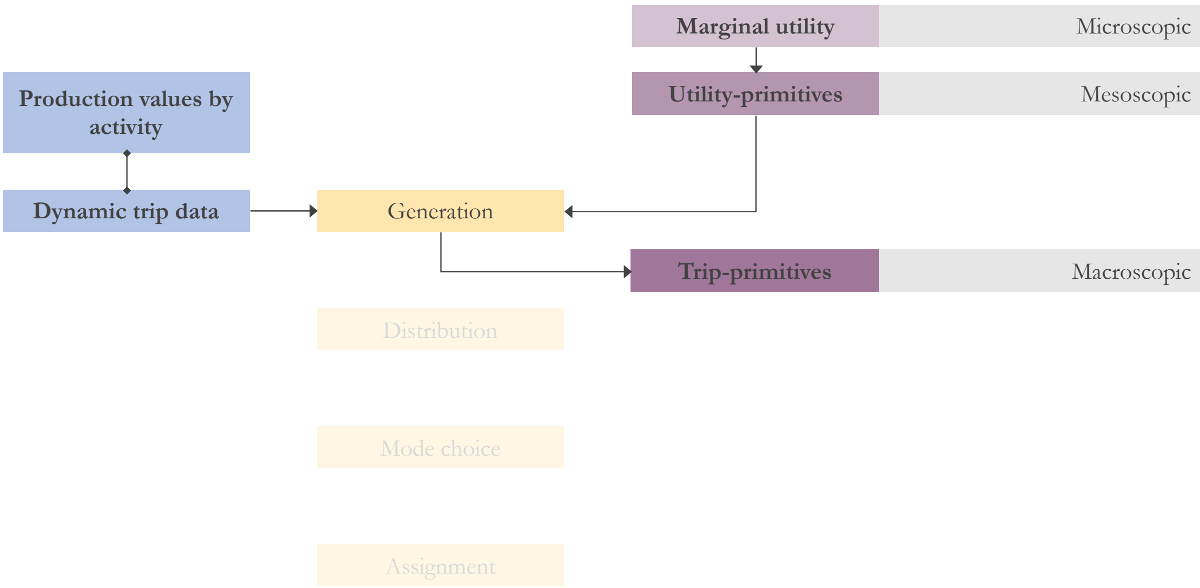


Figure V. 1 Thesis framework chapter 5

The work presented in this chapter has been described in the following paper:

“A Markov Chain Monte Carlo Approach for Estimating Daily Activity Patterns”

Poster at Scientific congress:
Transportation Research Board TRB Annual Meeting 2019

And

Symposium of the European Association for Research in Transportation 2021

And submitted to

Transportation Research Part A: Policy and Practice (October 2021)

V. UTILITY-PRIMITIVES

V.1. Introduction

To capture individual daily scheduling and travel interdependencies, advanced demand models typically adopt the concept of tours or trip chains which require decisions to be modelled at a microscopic level. The objective of this section is to enhance the representation of the macroscopic aggregated model presented before and infer the trip purposes of a population, by employing utility theory and applying the same Bayesian approach to generate trips. A utility function that includes activity-specific terms is used to model departure times for different activities, at an aggregate level, and its parameters are estimated using dynamic trip counts. This methodology is characterised by low data requirements and is shown to be flexible and easy to implement. In this work, we do not seek to reproduce individual behaviour, and we avoid time-consuming simulations. Instead, we aim to directly model mobility patterns emerging from activity-travel choices of a heterogeneous population and calibrate them against aggregated traffic data using an iterative Bayesian estimation scheme. The proposed methodology is characterised by low data requirements and is shown to be flexible and easy to implement. The well-established utility maximisation principle is used to model departure times specific to an activity at origin or destination, and each activity-specific utility is defined for a group of users, as proposed in past research (Adnan, 2010; Cantelmo and Viti, 2019). The only input data needed to be collected in our approach can thus be aggregated trips, which can be collected in a variety of ways. These trips are used to calibrate the parameters of activity-specific marginal utility functions, which in turn are applied in a departure choice model relying on utility maximization principles, used to generate activity-specific dynamic demand patterns. This chapter proposes a combination of established methods that enhances macroscopic demand modelling with interpretable results, consistent with utility maximisation. Furthermore, due to the flexibility of the process, it provides practical guidelines on how to use data in a sensible way. Hence, we refer to the research question on the modelling aspect and aim at answering to a specific proposition: "is it possible to model aggregate mobility patterns emerging from activity-travel choices of a heterogeneous population in a parsimonious way, avoiding the need to reproduce individual behaviour?". To answer this question, we adopt principles relying on individual behaviour, (i.e. utility theory) and generate utility-primitives using macroscopic data, including heterogeneity through a Bayesian estimation process. This methodology is tested with case studies to validate its relevance.

This chapter of the thesis is structured as follows. After a brief review of existing marginal utility functions, we describe the methodological approach. The second part includes synthetic case studies showing the opportunities of the method, followed by an application to a multiday survey collected in the province of Ghent, Belgium (2008) which shows potential for estimating purpose-specific demand patterns, consistent with observed individual behaviour.

V.1. Marginal utility functions

The core of the decision model used in this work relies on utility maximization principles. Many formulations have been proposed to be used in probabilistic models for travel behaviour. In the most simplified case, the scheduling problem can be linked to the preferred arrival time and include early and late arrival penalties due to congestion dynamics (Vickrey 1969; Small 1982). This often dubbed bottleneck model has been used to estimate departure time choices (Fosgerau and Engelson 2011; Tseng and Verhoef 2008). Activity time allocation models dealing with time as a finite resource (G. S. Becker 1965) are essential to focus on activities' utility gain as the main reason for travel (Supernak 1992) and reflect the utility gain and satiation effect (Charypar and Nagel 2005). The same connection and trade-off process can be extended to successive trips when including the activity duration (Zhang et al. 2005). The satisfaction linked to activity duration and, on the opposite, fatigue effect can however well be reproduced simply by a logarithmic decrease in time (Kitamura 1984b; C.R. Bhat and Misra 1999), which can be used for time allocation and for describing departure time and route choices (Yamamoto et al. 2000a). The trade-off between travel time and scheduling decisions can be modelled, instead of including early or late arrival penalties, by linking marginal utility and time of the day (Polak, Jones, and Vythoulkas 1993). In this approach, the departure time choice model includes a marginal utility for each time of the day, which expresses the utility gained from one-time unit of activity participation (Wang 1996). Nonetheless, flexible preferred hours can be modelled by relating the location of the satiety point to the starting time of the activity (Ettema and Timmermans 2003) or including delay and duration factors in activity-travel scheduling (Ettema et al. 2007). Depending on the adopted functional relations and assumptions, these models can be divided into either *duration-* or *clock-based*. A model being both duration and clock-based, can be achieved by combining a duration model and a discrete choice model to the scheduling of tours and individual daily travel patterns (Vovsha and Bradley 2004). More recently, an hybrid formulation that included a logarithmic decrease with respect to starting time, employing a time of the day-dependent marginal utility form was proposed (Cantelmo and Viti 2019). Another approach, which relates to satiation with activity frequency, is to combine baseline utility for an activity type and additional utility reflecting satiation (Nurul Habib and Miller 2009), allowing introducing utility-maximization for unplanned activities within a time budget constraint.

The utility-maximization process subject to such "space-time-needs" constraints can also capture heterogeneity and joint choices in activity travel decisions using inventory routing problems approach (Chow and Nurumbetova 2015). Even if utility models are often used for modelling disaggregated individual activity-scheduling (Arentze and Timmermans, 2004; Pendyala et al., 1998), they sometimes are used within an assignment or an aggregated route choice model (Cantelmo and Viti 2019; Yamamoto et al. 2000a) and at the individual level, activity scheduling can be modelled including the effect of road network congestion (Adnan, 2010). Yet, few works have focused on the calibration of agent-based models (Flötteröd 2009b; Patwary, Huang, and Lo 2021). (Yasmin, Morency, and Roorda 2017) focus on the aggregation level and different methods for validating the TASHA activity-based model (Toronto area) in Montreal Island. Without re-calibrating parameters and without considering all the facets of the choice model, they determined a good transferability of the results in different contexts, at different scales. However, the calibration remains a complex, underdetermined problem when applied to both traffic and behaviour-related features.

V.2. Methodology

To develop a model that is still capable of representing and explaining the heterogeneous trips in a study area we propose a framework to partition time-dependent generated trips into activity-specific components (Figure V. 2). Our model can take as input simply aggregated trips, collected for instance via traffic counts, but it can be used with any other type of data related to dynamically generated trips (mobile phone data, floating car data, license plate recognition data, etc.). Using a probabilistic approach based on *MCMC* sampling techniques, the parameters of pre-specified marginal utility functions are assigned in a stochastic way to match resulting demand to observed flows, via optimisation.

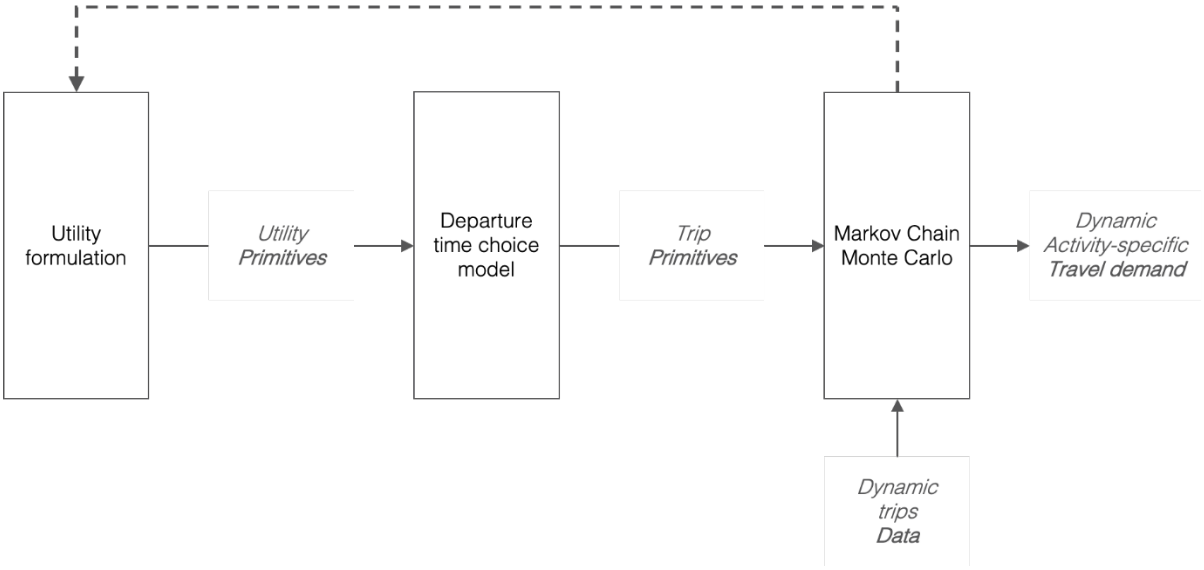


Figure V. 2 Methodological overview

To connect the proposed methodology to well-established behavioural frameworks and make it sensitive to characteristics of both demand and supply, utility maximization principles underpin this approach and are applied to the activity-specific departure time choice process. The challenge is to use an individual-based formulation applied in an aggregated way in order to embed heterogeneity characteristics. This will allow revealed and observable preferences such as departure time and chosen mode, to be formally linked with underlying, latent characteristics, such as (marginal) utility gains/losses. Assuming that people seek to maximise their total daily utility, including travel time costs and (positive) accumulated utility, they will try to optimise their travel choices, in this case the trip schedule. The optimality of any trip schedule certainly depends on the marginal utilities of the sequence of activities. At an aggregate level, this phenomenon can be captured by departure time probabilities, resulting in emerging trip rates.

The three main components of the proposed approach (Figure V. 2) are described in the following three sections, with special attention given to the estimation process which iteratively connects the utility formulation and departure choice model inside the *MCMC* sampling.

This Bayesian approach results in both a distribution and point estimate for the model parameters. This is a method commonly used in SP creation, and in demand estimation when traffic counts are available, and we demonstrate here that it can be used with aggregated trip rates to calibrate the complex underlying behavioural functions. In order to apply traditional concepts from utility theory and further generalize them in the context of this paper, the following section introduces definitions of fundamental concepts used in the proposed methodology.

V.2.1. Definitions

Marginal utility is an established economic concept, widely used in transportation science. It represents the satisfaction or benefit of a consumer from using a unit of a service. Often in travel demand modelling and in particular in this work, service corresponds to participation in an activity for a unit of time. The marginal utility will differ across individuals from a heterogeneous population, and this can be represented by having distributions of parameters for a common marginal utility functional form.

A **Utility-primitive** describes the marginal utility of performing a specific type of activity at a certain location, for a certain time period within an activity-travel chain. A utility-primitive corresponding to a given activity type is thus determined by a set of three marginal utility functions, the central one corresponding to the "activity" itself, while the other two represent all other activities performed before and after the activity to be estimated. These aggregated activity-travel chains implicitly capture aspects of accessibility and hence of the impact of the quality of the underlying transport system, as well as the population's heterogeneity in the activity scheduling. Utility primitives are used in the framework of utility maximization estimation for groups of users.

Trip-primitives describe departure rates by activity type at the zonal level. These functions capture the aggregate quantity of users travelling by time of day, arriving to start, or departing having finished a given activity, hence giving rise to traffic flows across the area being modelled. Trip-primitives have previously been introduced in (Scheffer, Cantelmo, and Viti 2017) and in the previous chapter to describe the daily dynamics of activity-specific flows.

These three concepts are aligned with modelling activity-travel choices at three scales: micro, meso and macro. Utility-primitives are proposed as a mesoscopic instrument, used to estimate trip primitives within a departure choice model at the macroscopic level. However, the estimation procedure also generates distributions (of parameters of) marginal utilities, which can be meaningful at the microscopic level and could be used as a proxy for individual heterogeneity. The latter aspect is not developed in this thesis but constitutes an application opportunity of the proposed methodology in Synthetic Population (SP) creation for example.

Activity type: let A be the set of all potential activity types, with travelling included as one of the activity types $A = \{a_1, a_2, \dots, a_N\}$, with N being the number of considered activity types. For example, $A = \{a_{home}, a_{work}, a_{shopping}, a_{leisure}, a_{travel}\}$.

The primitives described in the following section are used to describe emerging behaviours for groups of people, considering a convolution of individual activity-trips sequences. The notation and scheduling process of an individual characterizing the parameters results in the representation of utility-primitives. The activity sequence representation of any individual resulting in emerging patterns adheres to the following notation and assumptions:

Activity sequence: we denote by $A^i = \left[A_1^i, A_2^i, \dots, A_{n^i}^i\right]$ the sequence of activities for the i-th individual, ignoring activity duration.

An example could be:

$A^i = \left[A_1^i, A_2^i, \dots, A_7^i\right] = \left[a_{home}, a_{travel}, a_{work}, a_{travel}, a_{shopping}, a_{travel}, a_{home}\right]$.

We do not seek to represent how individual activity chains are generated in this methodology.

Time of the day: the 24h day is discretised into time steps Δt so that time $t \in \{t_0, t_0 + \Delta t, t_0 + 2\Delta t, \dots, t_0 + (N-1)\Delta t\}$. The sequence of activities $[A_j^i]$ partitions the discretised day into n_i blocks of time, with $t^{i,j}$ the set of time points in the j-th block.

For convenience we also define activity start/end times and trip start/end times in a natural way.

Individual i begins activity A_j^i, at time $t_s^{i,j} = \min\{t^{i,j}\}$ and finishes at time $t_e^{i,j} = \max\{t^{i,j}\}$.

Travel time: in this chapter we assume that travel times are fixed throughout the model. The assumption of constant travel time can be easily relaxed and we can consider different constant trip times for incoming and outgoing trips, as well as considering travel times as demand- and time-dependent, as it was shown in (Cantelmo and Viti 2018). The inclusion of variable and mode-specific travel times will be shown in the next chapters.

V.2.2. Utility primitives

For individual i, the marginal utility for engaging in activity a at time t, which may depend on the start time t_s, is $u_a^i(t, t_s)$. For a given activity, the marginal utility is not zone-dependent but can differ between individuals. The total utility accrued by individual i performing activity a from time t_s to t_e is

$$U_a^i(t_s, t_e) = \sum_{t=t_s}^{t_e} u_a^i(t, t_s)\,\Delta\mathrm{t} \tag{5.1}$$

The marginal utility for travel activities will typically be negative, whereas other activities will likely generate positive utility, hence justifying the travelling activity. In this work, capturing the disutility of travel is most simply accomplished with a fixed cost per unit time: $u_{travel}(t, t_s) = u_t$ where $u_t \leq 0$. This is sufficient to distinguish between short trips (more attractive) and long trips (less attractive) to access an activity.

The utility accrued by individual i, when travelling from zone y to engage in activity a from time t_s to t_e is

$$U_a^i(t_s, t_e|y) = u_t tt_a^y + \sum_{t=t_s}^{t_e} u_a^i(t, t_s)\,\Delta t \tag{5.2}$$

The penalty associated with early and late arrival is only captured within equation (5.1). Let n be the number of activities performed by individual i. For readability, we omit that $n = n(i)$, and may simply denote activity using j rather than a_j^i. The j-th activity for individual i has (starting, ending) time denoted $(t_s^{i,j}, t_e^{i,j})$; these partitions the day. Each trip is included as an activity in the chain, and the associated u_j^i then represents the marginal (dis-)utility of travelling. We therefore have that individual i accrues total daily utility:

$$\boldsymbol{U}^i = \sum_{j=1}^{n} \sum_{t=t_s^{i,j}}^{t_e^{i,j}} u_j^i(t, t_s^{i,j})\,\Delta t \tag{5.3}$$

When studying emerging behaviour, the level of complexity that results from individual-specific choices becomes very high; hence to adopt the concept of utility-primitives, we aggregate the activity chain of each individual into just four components. When considering the primitive specific to activity a, we consider the marginal utility of any/all activities performed before a (denoted $a\,-$), the trip to access the activity (denoted $\rightarrow a$), the marginal utility specific to a, and the marginal utility of any/all activities performed after a (denoted $a\,+$). The activity chain for individual i is therefore $[A_{a-}^i, A_{\rightarrow a}^i, A_a^i, A_{a+}^i]$. Other inter-activity travel is subsumed into the before/after components. This simplification is one main assumption of the proposed approach. Listing all possible combinations and tour types would in fact unnecessarily increase the complexity of the system and the potential benefit would be minimal once aggregated at a macroscopic level, where such interdependencies would be lost. Thus, the proposed way of using utility theory is reduced to a very simple degree of precision and doesn't integrate the most advanced existing individual trip-based models. In line with this assumption, equation (5.4) expresses these three distinct activities, with times t_s and t_e representing the start and end time of activity a. Departure time is chosen to arrive at the desired start time: $t_d = t_s - tt_a^y$. Note that the trip components are assumed to be not individual specific. For individual i, with activity a starting and ending time (t_s, t_e) given a zone of origin y and a travel time tt_a^y. The start and end of the estimation period are t_0, t_N, respectively. This implicitly defines the before/after marginal utility functions for individual i:

$$\begin{aligned} \boldsymbol{U}_a^i(t_s, t_e|(y)) = \sum_{t=t_0}^{t_d-\Delta t} u_{a-}^i(t, t_0)\,\Delta t \quad + \quad u_t tt_{\to a}^y \quad + \quad \sum_{t=t_s}^{t_e} u_a^i(t, t_s)\,\Delta t \\ + \sum_{t=t_e+\Delta t}^{t_N} u_{a+}^i(t, t_e + \Delta t)\,\Delta t \end{aligned} \tag{5.4}$$

If we assume that the activity-specific marginal utility related to a certain time t, $u_a(t)$ does not depend on other parameters such as utility already accumulated or following utilities, a given time unit always yields to the same utility. The utility function becomes then separable meaning that treating each function independently does not introduce any further error (Cantelmo and Viti 2018). This leads to a more computationally efficient formulation. This notation could be extended to a more detailed chain of activities. Alternatively, longer tours could be separated into sub-components so that equation (5.4) would be applicable to study such sub-tours, focusing the estimation problem on one activity at a time.

V.2.3. Activity-travel Choice Model

The number of people based in each zone, y, with activity a in their activity chain is known and denoted Q_a^y. To determine the distribution over time, we use a choice model for activity starting and ending time, through a trade-off problem as illustrated in Figure V. 3 for $t_1 = 8AM$; $t_2 = 5{:}30PM$; $t_{t,a2} = 30\ min$. The shaded areas indicate the total accumulated utilities for the three activities; the red and blue lines correspond respectively to the start and end of trips.

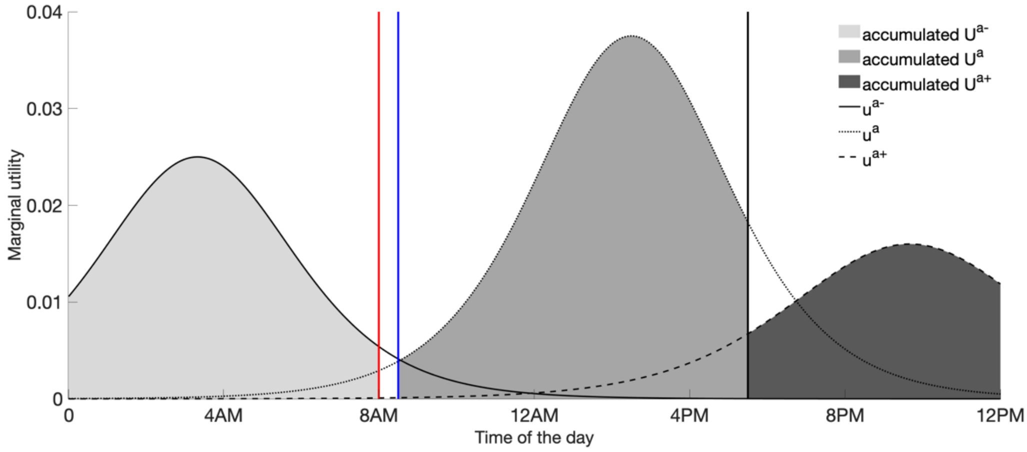


Figure V. 3 Accumulated Utility for a sequence of activities

For an individual in zone y who wishes to perform activity a in its activity chain, the total utility accrued from travelling to z to do a will be $U_a^{i,y}(t_s, t_e)$, if the activity (start, end) times are (t_s, t_e). This activity may be available in multiple zones, each with fixed access travel time and with an identical marginal utility. The probability of doing activity a between a starting time t_s and an ending time t_e, starting from zone y is:

$$P_a^y(t_s, t_e) = \frac{\exp\left((U_a^i(t_s, t_e) + u_t tt_a^y)/\sigma)\right)}{\sum_{t'_e > t_{s'}} \sum_{t_{s'}} \exp\left((U_a^i(t'_s, t'_e) + u_t tt_a^y)/\sigma\right)} \tag{5.5}$$

The scale coefficient is fixed in the following of the thesis with $\sigma = 1$.

Aggregating over all feasible end times gives the probability of departing zone y for activity a, starting in zone z at time t:

$$P_a^y(t) = \sum_{t_e > t} P_a^y(t, t_e) \tag{5.6}$$

The number of trips departing zone y at time t is therefore

$$T^{y\rightarrow}(t) = \sum_a \sum_{t_e > t} Q_a^y \, P_a^y(t, t_e) \tag{5.7}$$

There exist more sophisticated expressions for modelling this choice, and the independence axiom may be a limitation of the chosen formulation. However, the simplicity and ease of calculation of the model makes it convenient, interpretable, and efficient enough in the framework of this work. More complex random utility models can be used, but this would not allow to use a closed form.

The proposed transition from an individual-based formulation to the distribution of starting time at the population requires the assumption that the utility formulation can represent a group of individuals. While the meaning is not perfectly the same, this approach has been widely used when introducing utilities for population segments or considering a generic utility, for example in the conventional bottleneck model. To operate this transition from an individual to a group of users, some complexity of the trip chain is discarded. The usage of an artificial global accumulated utility before and after the relevant activity is selected to obtain the simplest possible formulation to consider trip chaining and so the lowest possible level of complexity required for the estimation purpose. In the resulting simplified form, individual heterogeneity across the population can be captured through having distributions of the model parameters in the case of a mesoscopic application of utility-primitives.

V.2.4. Marginal Utility Function

In order to apply equation (5.5) and estimate the total utility gain of performing activity a for each combination (t_s, t_e), a time-dependent functional form is sought that can fit multiple activity types. These marginal utility functions represent any individual and result in a macroscopic demand pattern in the form of activity-specific trips. When a utility function is separable (clock-based), users' behaviour depends only on the activity before and after (Adnan 2010; Cantelmo and Viti 2019). An assumption of the model is that the two secondary functions, from an aggregate perspective, represent any activity or combination of activities other than a such that every daily chain can be represented by a tour of three activities. We therefore do not specify which activities were performed before ($a -$) or after ($a +$) the activity whose parameters need to be estimated, but we can represent an aggregated function of all alternatives. Hence, these generic utility functions are mostly used for the estimation of $u_a(t, t_s)$. Recall that these utility-primitives will be combined with departure time choice models to give trip-primitives, which are used to determine aggregate trip rates. The representation of the marginal utility function is thus at a macroscopic level. Desirable features of the marginal utility formulation are:

- Peak at a certain time of the day (clock-based) or after a given duration (duration-based)
- Representation of fatigue effect, i.e. marginal utility may decrease while performing the activity for a period of time
- Low number of parameters to estimate

We tested a series of marginal utility function following these indicators, starting from (Yamamoto et al. 2000b) for which the distinction between the impact of time of day and decay linked to fatigue is difficult to be strictly established. The second formulation that has been tested is described in (Cantelmo and Viti 2019). Even though the behavioural representativeness is better, its impact on the synthetic MCMC examples, didn't prove an improvement of the likelihood functions at the level of generated trips. Furthermore, the complexity in terms of form and parameter description requires higher efforts for evaluating priors and especially computing accumulated utilities as their form depend both on starting and ending time.

We selected a model fulfilling these requirements in efficient way, without additional complexity: the marginal utility proposed by (Ettema and Timmermans 2003):

$$u_a(t, t_s) = \frac{\gamma_a \beta_a (U_a^{max})}{exp[\beta_a(t - (\alpha_a + t_a^s \tau_a))] \,.\, (1 + exp[-\beta_a(t - (\alpha_a + t_a^s \tau_a))])^{\gamma_a + 1}} \tag{5.8}$$

In order to evaluate $P_a(t_s, t_e)$ for each t_s and t_e element of T, the 5 parameters for each marginal utility (u_{a-}, u_a, u_{a+}) should be estimated for each activity a, in addition to D_a .

$$\Theta_a^{(\mathrm{u}=u_{a-}, u_a, u_{a+})} \ = \ (U_u^{max}, \alpha_u, \beta_u, \tau_u, \gamma_u) \tag{5.9}$$

For all parameter settings the marginal utility is unimodal. We define saturation point the maximum value and the increasing period before is called the warming up phase. Parameters α, γ and τ play a role on its position on the temporal axis. In particular, the behaviour of the function depends on each parameter value:

- α: if $\gamma = 1$ and $\tau = 1$ the saturation point is located at the α value and the function is symmetric.
- γ: if $\gamma > 1$, the saturation point is situated before α and the steepness of the left part is higher than the right part, $\gamma < 1$ represents instead a longer warm up phase.
- τ: controls whether saturation is reached at a fixed time of day, or is relative to activity duration. When τ is close to 0, utility is determined by time of day (regardless of activity start time). Whereas $\tau = 1$ describes a fully duration-based utility function; the utility function is translated in time to follow the activity start time.
- U^{max} : the parameter U^{max} represents the maximal possible accumulated utility for a certain activity, it impacts the magnitude of the marginal utility function. This parameter controls the gain in performing one activity relative to any other alternative. This means that by neglecting the travel costs, in this study the absolute values of U^{max} will be relative to the assumed trip times for each activity.

- β: determines dispersion around the saturation point. β <<1 results in a flat marginal utility, while a larger β gives more peaked marginal utility around the saturation point.

The possible shapes of function (11) in a 24h time-period, with respect to the 5 parameters is illustrated in Figure V. 4. Analysing the resulting shapes, we can provide a behavioural interpretation of the different components.

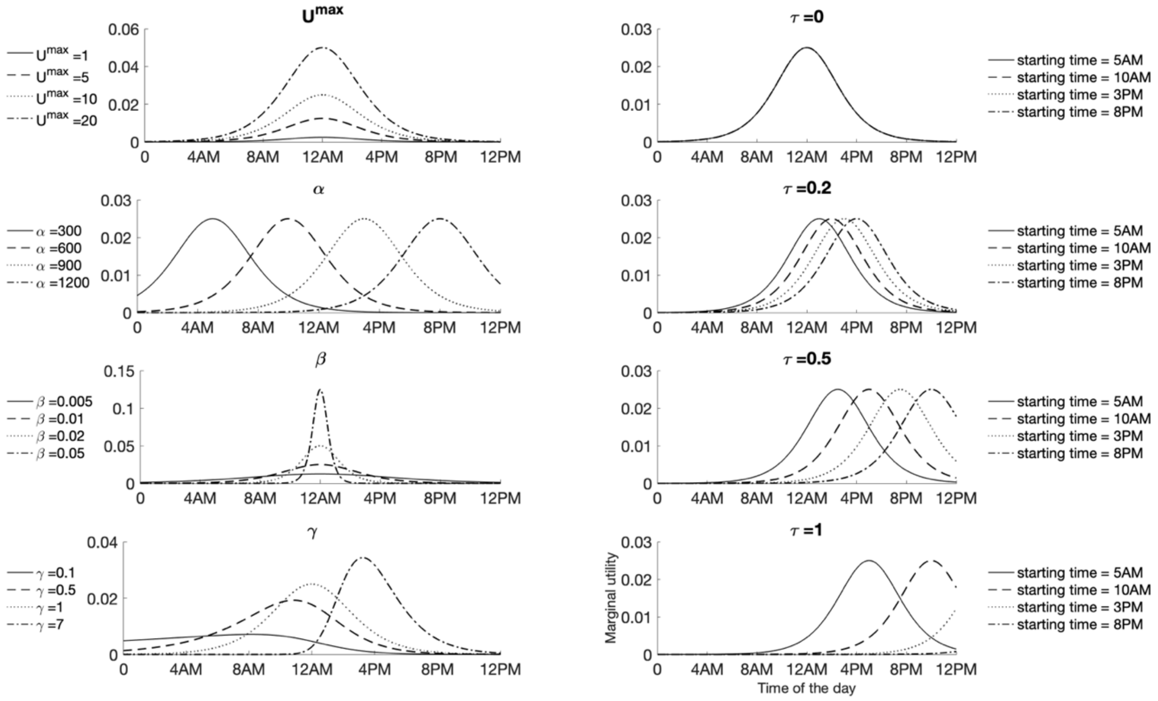


Figure V. 4 Impact of the marginal utility formulation's parameters (reference: U max=10; α=720; β=0.01; γ=1; τ=0)

The same functional form of marginal utility is used to calculate every utility U, though the parameters are activity-specific. This functional form has the flexibility to model, on the one hand, activities such as work, which is typically clock-time dependent (start times are usually concentrated in the morning and end times in the evening peak periods) and that is a resource which can potentially bring unlimited marginal utility, but it is subject to fatigue effects. On the other hand, activity types such as shopping are available rather uniformly throughout the day, hence less dependent on clock time, but are susceptible to stronger fatigue or satiation effects and are better described by their typical duration rather than the time at which they are performed. Individual heterogeneity is captured by probability distributions of model parameters, which emerge from the parameters' estimation process.

V.3. Model estimation and MCMC

The number of parameters to be estimated for deriving the total demand grows with the number of utility-primitives. Furthermore, the marginal utility functions are non-linear and the observations available for calibration are aggregated and can be limited in number. This encourages the usage of an extensive simulation-based approach for the estimation of parameters' values. Moreover, data such as traffic or passenger counts contain different sources of stochasticity, stemming from the inherent variability of the demand, counting errors, etc. These are reasons why the estimation of the parameters is done using a Bayesian procedure, and more specifically a Markov Chain Monte Carlo (MCMC) modelled through the Metropolis-Hastings algorithm (Metropolis et al. 1953).

As described in the previous section, plausible distributions from domain knowledge are used to initialise utility-primitive parameters. The estimation process then corrects these, based on available data. Given the aggregate nature of the estimation, this method is attractive because each variable is sampled without knowing its actual distribution but providing a posterior probability distribution as an output of the stochastic process. These parameter distributions are used to represent the heterogeneity of different users' marginal utility for each activity type. The expected value of each parameter is used as the basis for the utility primitive, which is then used in the trip-primitives estimation. Finally, the proposed method is flexible and can theoretically be applied to any kind of utility function and to different types of available data which makes it versatile.

V.3.1. <u>Process overview</u>

For each element of Θ (equation (5.9)), starting with initial values θ_0, the sampler proposes new values (equation (4.3)) for the parameters of each utility primitive (equation (5.8)). On one side, the proposed parameters are applied to the marginal utility function and used in the described departure time choice model (equation (5.5)) in order to calculate the estimated trips. These trips are then compared to the observed data, through the likelihood function computed according to equation (5.11). On the other side, the sampled parameters are evaluated in their plausibility, with respect to the prior, i.e. the probability of having the proposed value according to the modeller's selected density function equation. Plausibility and likelihood are used together for updating the proposed set of values in the posterior or continue to use the previous one (equation (4.7)). The posterior built after iterations is finally used to evaluate the primitives of the demand.

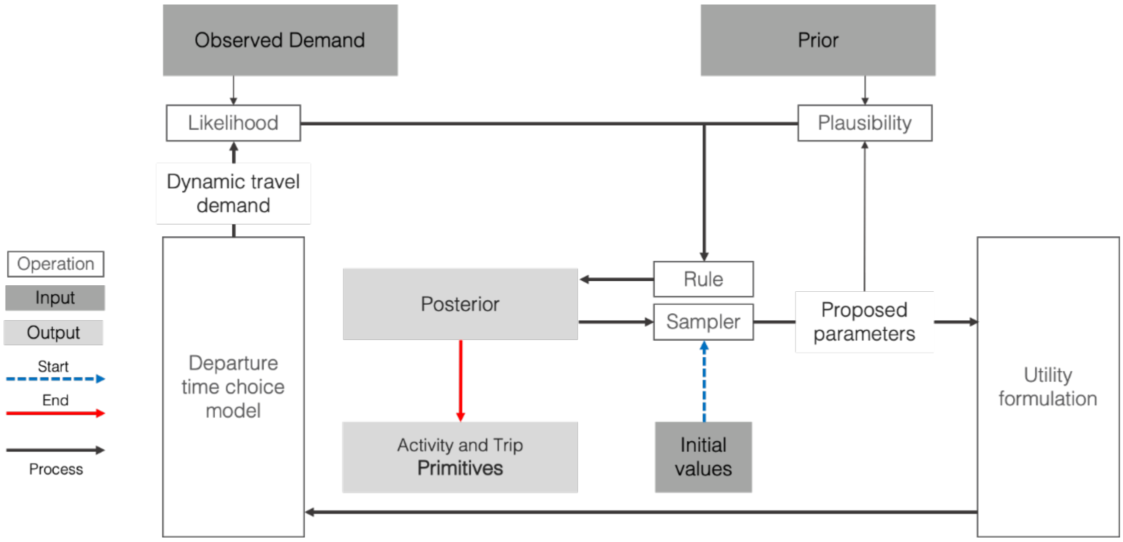


Figure V. 5 Methodological Framework

V.3.2. <u>Parameter update</u>

The utility formulation and the departure choice model described in the previous sections are the central elements as they build the model which reproduces the observed distribution (i.e., the observed demand data). The sampler and the rule are the two components of the MCMC which allow an efficient estimation process.

A conventional log-likelihood function is used to compare observed data to the model and a normal noise distribution is assumed, with no correlation, it is formulated as a simple function of residuals, at each iteration i:

$$\mathcal{L}_i = \sum_z \sum_t -\frac{1}{2}\left({r_{i,t\to z}}^2 + {r_{i,t,z\to}}^2\right) \tag{5.10}$$

with the residual calculated as the difference between simulated trips and observed starting trips to enter and leave the zone:

$$r_{i,t,\to z} = T^{\to z}(t) + \widehat{T_i^{\to z}(t)} \tag{5.11}$$

$$r_{i,t,z\to} = T^{z\to}(t) + \widehat{T_i^{z\to}(t)}$$

where $T^{\to z}(t)$ is the total number of trips ending in zone z at time t and $T^{z\to}(t)$ the total number of trips starting from zone z at time t and $\widehat{T_i^{z\to}(t)}$, $\widehat{T_i^{\to z}(t)}$ their estimation at the i^{th} iteration.

V.3.3. Application to activity-specific demand-generation

This methodology is applied in the remainder of the paper to the most aggregated possible information level, i.e. aggregate dynamic travel-demand generated at the level of a complete study area. Hence, starting from information on the observed aggregated time-dependent trips generated in a region, we want to distinguish which share of the said trips is done to perform a specific activity type. Although it is theoretically flexible enough to be applied at smaller scales, from OD pairs up to the most disaggregated agent-based models, this is the most delicate application level in terms of computation and the simplest in terms of data requirement.

In the case of estimating a complete study area, where no information is available at the zonal level some assumptions are required.

Since we do not give to the model any information which allow to distinguish if the trip is originating or ending at the activity location, there is no distinction in the observed $T^{\rightarrow z}(t)$ and $T^{z\rightarrow}(\mathrm{t})$ which are simplified to single a vector $\boldsymbol{T} = (\boldsymbol{T}_1, \boldsymbol{T}_2, \dots, \boldsymbol{T}_t)$. In the proposed study, it is used for a one-day estimation and with 30 minutes time interval it contains 48 elements.

This impacts the estimation process, equation (5.10) becoming:

$$\mathcal{L}_i = \sum_t -\frac{1}{2}\,({r_{i,t}}^2) \tag{5.12}$$

For this reason, the available information inside the estimation process is lower which increases the necessary number of iterations.

In the case of only intrazonal trips, the influence of travel time is negligeable because it doesn't impact the choice process. The period between t_1 and $t_1 + tt_a$ does not bring any positive utility and we neglect any disutility arising from these trips in this study. Again, the cost of the trips will be included in future research when we will extend the model to consider destination and mode choice. Here we simply consider that trips to a specific activity may be different, but they are fixed and constant by activity-type, during the day.

Finally, the home activity is not designated in the single zone framework, because we consider that home is the starting and ending activity for each daily activity-trip-chaining schedule.

This allows to apply the proposed formulation to the most extreme case where very few data is available. In order to show the potential of the proposed methodology, the following case studies are applied in this framework. Using more disaggregated data, and application opportunities to model more accurate estimates such as OD-specific and mode-specific trips will be subject of future research.

V.4. Case Study

To apply the proposed MCMC as described in the previous section, we set up two case studies. The first is a controlled synthetic experiment that allows to compare the results of the calibration in a controlled scenario, and perform a sensitivity analysis of the parameters, while the second is using real data collected from a multiday travel survey.

V.4.1. Case Study 1: Synthetic experiment

V.4.1.1. Single primitive

In order to test the estimation ability of the MCMC with a given number of datapoints and parameters, we created synthetic data representing one tour type only (i.e. one activity and one location), fixing the utility parameters of three functions and a demand of 5.000 people (table 1). The synthetic data have the same assumptions as the proposed model and are used to validate it in ideal conditions. With these three utility functions, we generate trip profiles using the probability choice model described in the methodological section, resulting in 10.000 trips. The synthetic dynamic trip counts are used as input of the MCMC without using further information about their formation in the process. This synthetic data is used as a benchmarking case where we know all the observed trips to engage on the given activity type. D_a^z is fixed in this specific optimization, while this number is estimated in the case of a more realistic experiment. Potential remaining trips can represent the demand which is not described by any tour.

Parameter	U1	U2	U3
α	250	725	1225
β	0.005	0.0075	0.005
γ	1	1	1
U_{max}	10	15	10
τ	0	0	0

Table V. 7 Parameter of the synthetic data

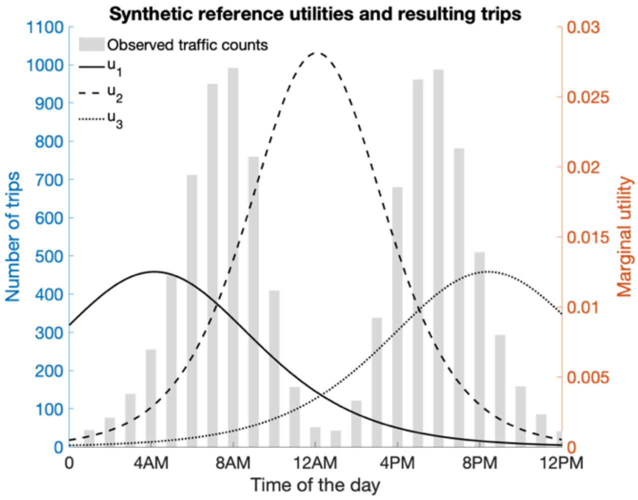


Figure V. 6 Synthetic experiment

Different scenarios are tested, where both the initial value of the parameters and their prior vary in order to see the impact of different prior types and available information to the estimation.

Scenario 1: the prior is normally distributed around the target value, used for generating the demand profile. This is an accurate and precise belief and represents a model with a lot of information available.

Scenario 2: the prior is slightly inaccurate with a positivity bias.

Scenario 3: the prior is uniform. This represents a model where no information is available apart from its feasible range.

Scenario 4: the prior is precise but not accurate. This represents a strong erroneous belief.

The four fitted profiles are compared to each other by means of the normalized root mean squared error with respect to the observed hourly traffic (24 data points corresponding to the histogram on Figure V. 6

Table V. 7) on one hand and with respect to the 13 target parameters on the other hand.

Table V. 8 Result of the 4 scenarios

	1	2	3	4

Prior	normal (μ, σ)	normal $(\mu + 10\%, 2\sigma)$ truncated for x<0	uniform $(0{,}01\mu, 100\mu)$	normal $(\mu + 80\%, \sigma)$
Initial value	$\mu + 50\%$	$\mu + 50\%$	$\mu + 50\%$	$\mu + 50\%$
Iterations	15.000	15.000	15.000	15.000
NRMSE (r2)				
traffic	0.05 (0.99)	0.193 (0.96)	0.578 (0.77)	0.217 (0.96)
parameters	0.04 (0.99)	0.163 (0.99)	0.320 (0.97)	0.738 (0.90)

This experiment allows to show the impact of prior types and step size for the different parameters, which will be used in the experiment using realistic data. For estimating the step size for each parameter, all parameters have been fixed but one, in turn, and step size changed for reaching better estimates, with the same settings. The results of this experiment show that in all cases the estimation is acceptable and that it is more interesting to have a uniformed prior instead of a wrong prior in terms of parameter estimation.

With a given low number of observations (24), a good estimation will depend on how accurate the initial guess is and a good prior but that the procedure is adequate for this problem.

V.4.1.2. Multiple primitives

Another controlled experiment was conducted for the case of multiple activity types. This synthetic experiment is a proof of concept where the observed traffic is artificially created using three types of activity and where the flows going to and returning from these activities are not overlapping. It is not the case in real observed data and this simplification will be relaxed later. This is used here as a numerical example to test the model with mixed types of priors and a greater number of parameters to estimate: 39, comprising 3 demand values and 4 parameters for each of the 3 utility functions of 3 activity types. τ is fixed, α, γ, U^{max} set as type 2 (from table 2) and only the β parameter is of type 3. Each of the three components is generated in the same manner as the former and their profiles are summed up as aggregated trips given as input, which gives a distinct profile from the previous case.

In Figure V. 7, the three components are compared one by one: the curve represents the estimated functions while the stems are the true underlying primitives used for generation we are seeking to estimate. As one can notice the MCMC identifies the shapes of each activity-specific trips distribution in a satisfactory way, despite relatively small deviations are observed, which can be attributed to the stochastic nature of the methodology, which keeps on randomly sampling and perturbing the estimation parameters until the stopping criterion is reached.

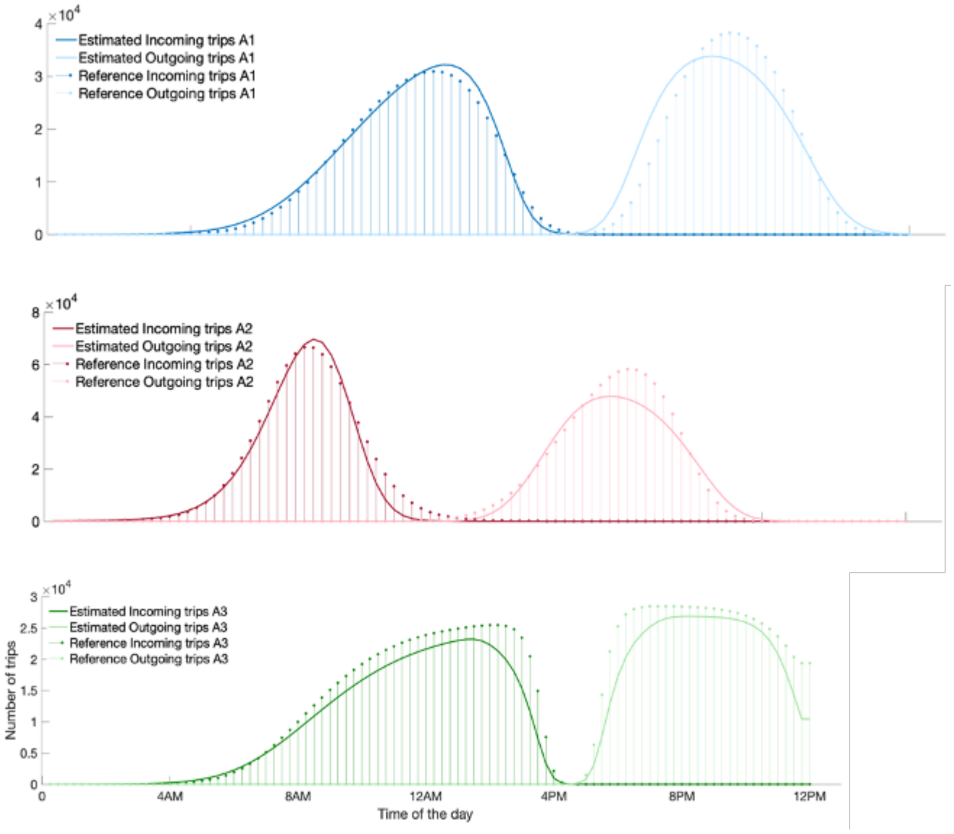


Figure V. 7 Estimated and reference trip-primitives by component

With this fully controlled experiment, a comparison parameter by parameter can be done and it shows an excellent fit. The estimation is very stable for most of the parameters and the estimated value close to the reference one. In the case of the β parameters the estimation is slightly worse; we can assume that this is due to the type of prior which is less informative.

V.4.2. Case Study 2: Ghent province trips

V.4.2.1. Database

The data used for the real case study results from a multiday travel survey collected in the area of Ghent in 2008 (Castaigne 2009). 707 individuals, randomly sampled with stratification criteria according to household size, gender and age, answered a 7-days travel diary. The database contains all their daily trips, 19.417 in total, described by origin, destination, starting, ending and travel time, modes, activity at origin and destination and their duration. 404 individuals were considered as «workers» as they described at least one trip for going to work. A description of the database and analysis of the variability of daily activity-travel pattern is available in (Raux, Ma, and Cornelis 2016).

Every observed trip performed during all working days has been considered for this experiment in order to increase the total number of observations for the MCMC, even though they include 7 different modes of transport. 12 specific activity types have been recorded in this multiday survey. In this paper, all survey categories apart from "home" are mapped to three modelled activity categories in the following manner:

- A1. "Work" are mandatory, repeated activities, for career or education purpose:
  work + school
  A1.1. Work – morning shift
  A1.2. Work – afternoon shift
  A1.3 Full day work (non-stop)
- A2. "Shopping" regroups necessary activities, mostly characterised by a relatively sharp duration constraint:
  long-term shopping + short-term shopping + drop off + personal business
- A3. "Leisure" are non-mandatory activities, for recreational purpose, and considered very flexible in their duration and scheduling constraints:
  leisure + eat out + walking/riding + visits + other

All starting times of the trips performed to reach those activities and the ending times of the activities have been considered summed over all days and counted as trips, used as input of the MCMC (Figure V. 8b). These aggregate trip numbers constructed from the survey data constitute the traffic data used in the estimation, at the scale of the complete study area.

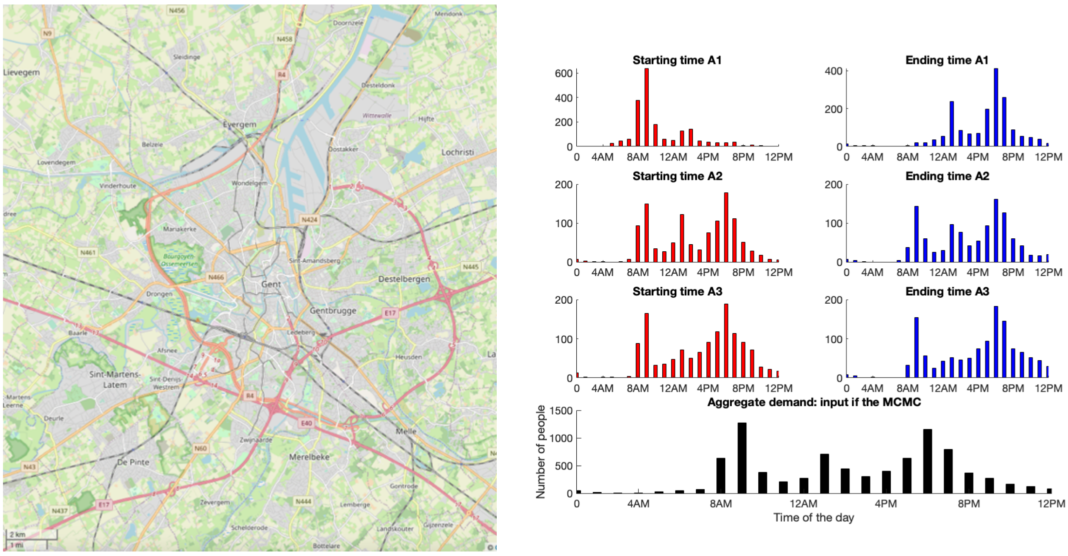


Figure V. 8 (a) Study Area (b) Observed demand by time of the day and activity type

V.4.2.2. Model Hypothesis and set up

In order to estimate activities with respect to the selected utility formulation (equation (5.8)), the following parameters are estimated:

$$\boldsymbol{\Theta} = \Theta_a^{\mathrm{u}} \text{ where } \mathrm{u} = u_{a-}, u_a, u_{a+};\ a = 1, \dots, 5 \tag{5.13}$$

As described above, the activity A1 is divided into three components to reproduce the observed midday lunch break and consider part-time workers. The demand D_{A1} is estimated for the total work demand and an additional parameter corresponding to the split between the three components is added to $\boldsymbol{\Theta}$. In order to include people working the full day but inserting a trip between the two working blocks, we have:

$$D_{A1.1} + D_{A1.2} + D_{A1.3} > D_{A1} \tag{5.14}$$

Conceptually, the posterior could result in a bimodal distribution for the activity A1 duration, but one of the goals is to find a representative average, that is why we have split A1 this way. For the same reason, it is difficult to model at an aggregate perspective both the universally available utility and the individual-specific utility that leads to fatigue effects. That is why the parameter τ_u is fixed to 0 or 1 depending on the activity type. For work ($A1$), we consider a fully clock-based utility with $\tau_u = 0$. For shopping ($A2$) and leisure ($A3$), $\tau_2 = 1$. As mentioned previously, fixing these parameters changes the meaning given to the parameter α_2 which, when $\tau_2 = 1$ corresponds to the time spent after which the marginal utility reaches its maximum value, it is thus coupled with a large β.

The time discretisation considered for calculation is 10 minutes. All feasible departure time combinations are evaluated for the proposed parameters $\boldsymbol{\Theta}$ and compared. A constant travel time is selected for each activity type, based on preliminary data analysis (Scheffer, Connors, and Viti 2021), partly presented in chapter 3:

$$tt_{A1} = 30\text{min};\ tt_{A2} = 10\text{min};\ tt_{A3} = 20\text{min}; \tag{5.15}$$

Another feasibility constraint is added for the shopping activity in order to account for opening hours. All starting times earlier than 7AM and ending times later than 9PM are excluded.

Because we focus on estimating expected values, priors are all assumed normal distributions. For parameters that we assume strictly positive, the selected prior is truncated from 0. Only the α_1 can be negative, and we assume negative values to represent a time of the day before midnight.

The mean values are selected based on the data analysis from (Castaigne 2009) and logical intuition. Because of the experimental setting, a large variance is selected that varies slightly among parameter type with respect to the modeller's confidence about the initial value. Preliminary experiments with synthetic data and order of magnitude impact this variance as well as the step size, which are defined by the following variance:

$$\Delta_{\beta,\gamma,U^{max},d} = 0.01\ \boldsymbol{\Theta}_{\text{initial}} \tag{5.16}$$

$$\Delta_{\alpha} = 5 \text{ if } \tau = 1$$

$$\Delta_{\alpha} = 20 \text{ if } \tau = 0$$

$$\Delta_{\tau} = 0$$

For every parameter, the initial values are set by the mean of the prior. The model has been implemented in MATLAB and run on a laptop with a 2.3 GHz Dual-Core and 8 GB memory. The computation time increases proportionately as the time interval decreases. Using a time interval of 30mins (10mins) takes 2700secs (8600secs) to complete 10 000 iterations. The computation time is drastically reduced when the assumed utility formulation is not dependent on activity starting time.

V.5. Results

V.5.1. Departure time and parameter estimation

After 150.000 iterations and considering a burn-in period of one-third we observe the total demand (red line) represented on Figure V. 9. The black line "initial estimation" represents the profile output from the first iteration, estimated with initial set of parameters. After the estimation process, the overall fit of the demand is significantly improved: the morning, afternoon and evening peak times are well estimated, and with the right order of magnitude.

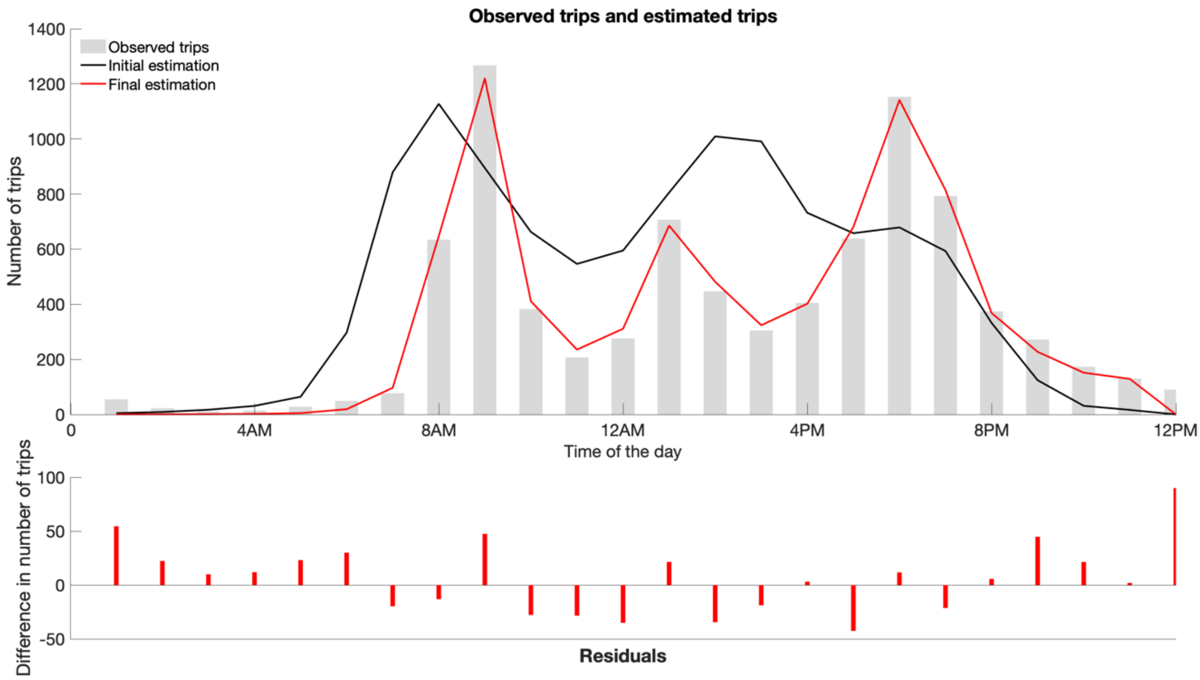


Figure V. 9 Aggregated demand estimation results

The evolution of the score can be seen on the top left of Figure V. 10. We can see that a convergence in terms of likelihood has been reached after less than 10 000 iterations. The grey part is not taken into account in the final estimation as it represents the burn-in period. The plot below represents the part of the score which refers to the prior. Its value drops fast as the parameters leave the initial region; this shows the ability of the model to move away from the prior belief. The example of activity's parameters $A1.3$ on the right shows this transition and the behaviour of the model. The estimation of γ and U^{max} do not show the same convergence stability as the other parameters. However, we can see that after the burn-in period they reach and remain within a given range. Furthermore, we refer to Figure V. 3 to support the fact that observed variations do not have a strong impact on the form of the function in the case of γ and in the case of U^{max}; the relative magnitude between the three U^{max} matters the most. The ratios of U^{max} are more stable with a standard deviation of 0.05.

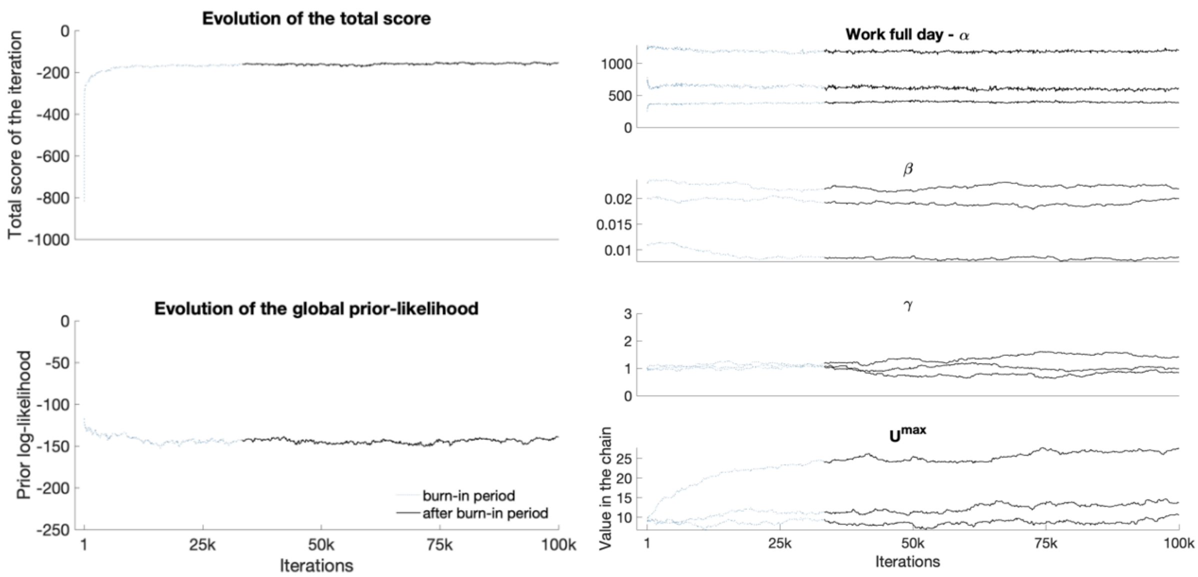


Figure V. 10 Evolution through 100 000 iterations

Even though the Metropolis process gives an acceptance rate of only 4%, we can see that the values continue to oscillate. Whereas for the parameters α and β, and to some higher extent U^{max}, the oscillations are relatively tight around a specific range, this is not the case for γ, which reveals considerable dynamics still after 100k iterations. This may be explained by the presence of γ in both the numerator, hence contributing to the overall magnitude of the utility, and as exponent in the denominator, hence contributing to the skewness of the marginal utility function.

As an example of the rule's output, we can see on Figure V. 11the evolution of the α parameter of u_1 for $A1.3$. The red line describes all the proposed values from the sampling and corresponds to the explored space, the blue line indicates the retained values i.e. the posterior.

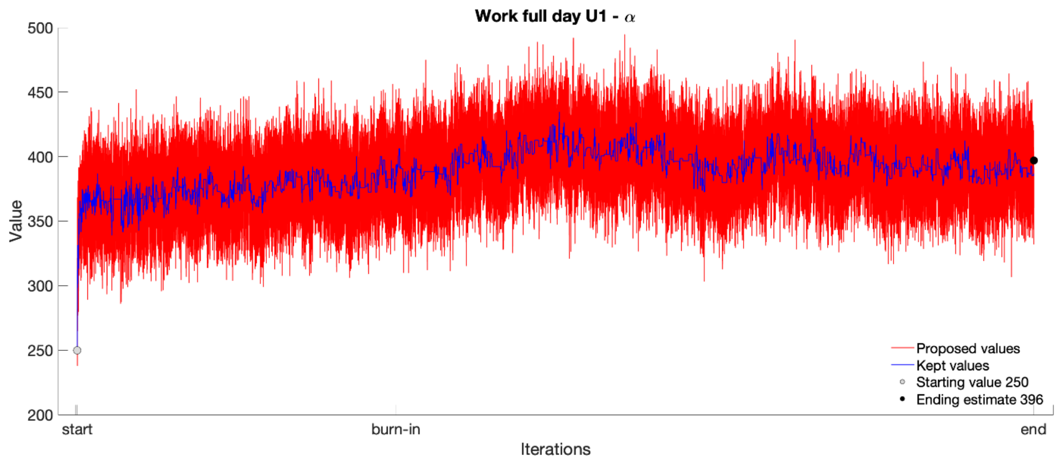


Figure V. 11 Example of a proposed and accepted values of the MCMC

Finally, the total estimated demand shown on Figure V. 9 is decomposed into the six utility-primitives representing the observed traffic shown in Figure V. 12. The Work activity is displayed as the sum of the three subcomponents (morning shift, afternoon shift and full day). It is sharply defined by the three peak periods of the day (around 8AM, 1PM, 7PM) while the Shopping and Leisure have a rather constant profile. These results confirm the intuition, i.e. working trips tend to be the main components of the morning, midday and afternoon peak period, whereas shopping and leisure activities do not have a sharp clock-time behaviour and are limited mainly by opening/closing times of the activities.

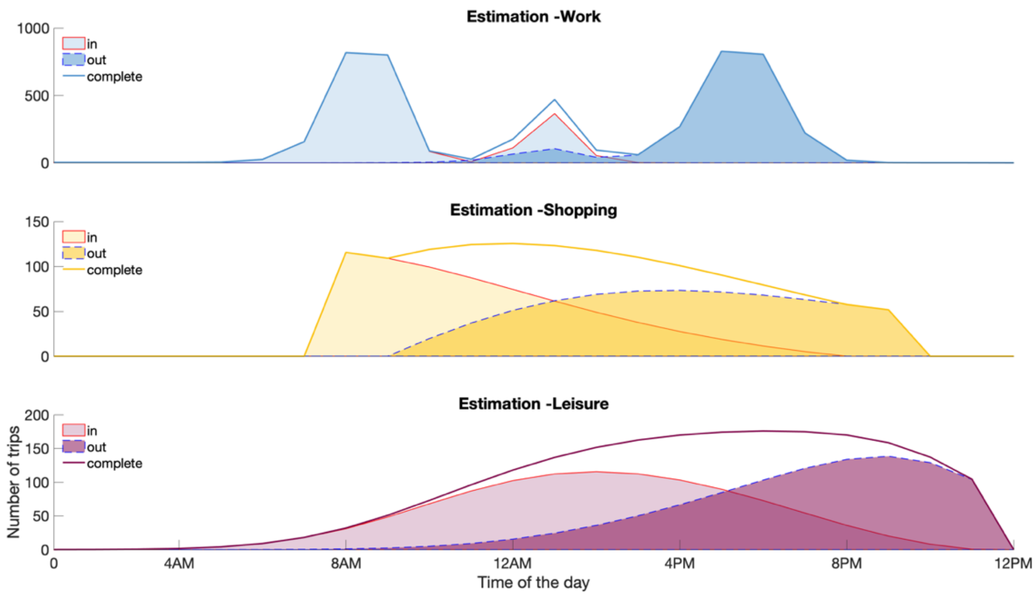


Figure V. 12 Six estimated trip-primitives

V.5.1.1. Analysis of estimated utility parameters

Differently from A1, in the case of $A2$ and $A3$ the utility varies with starting time and the accumulated utility depends on the ending times as well. In these two cases, the marginal utility has a very spiked profile with a very high U^{max} compared to the one of u_1 and u_3. This is a way to explain an activity which is still desirable despite the travelling loss but does not require a long duration time. In fact, the choice model focuses on starting and ending time knowing that a given number of people will be performing that activity at any moment of the day.

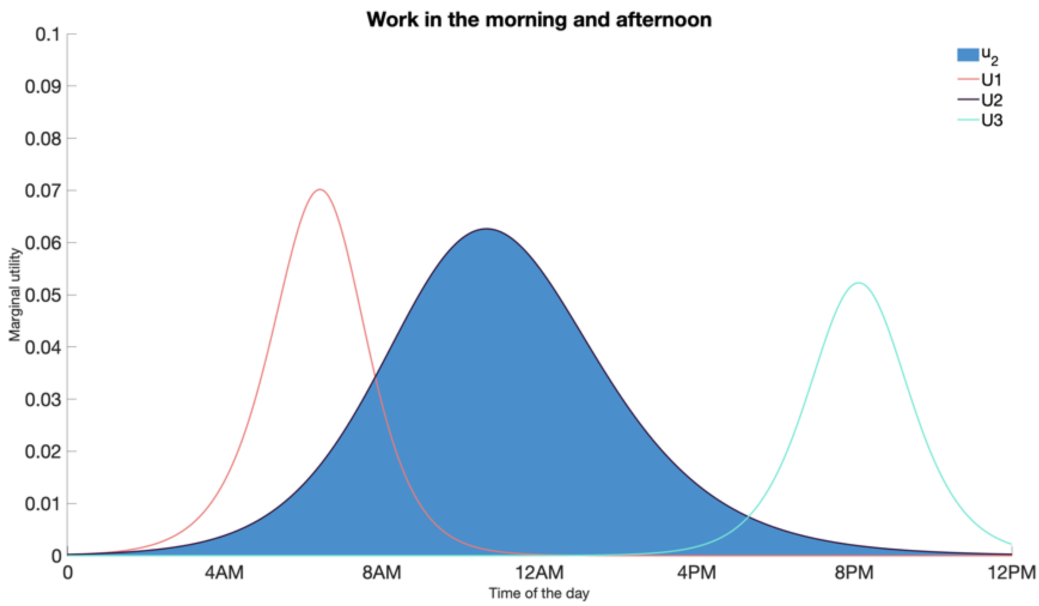


Figure V. 13 Estimated utility primitives for the work activity (full-day)

These functions are characterized by the final estimate of each parameter, which is represented by the average value of the posterior, after removing the burn-in period. At a mesoscopic level, the posterior of parameters can give information on their distribution in the population and their plausible values range. In most cases, parameters posteriors can be fitted with a normal distribution, the same shape as their prior. As an example, Figure V. 14 shows the histogram of the α parameter for five activity types and the three marginal utility functions used for departure time estimation.

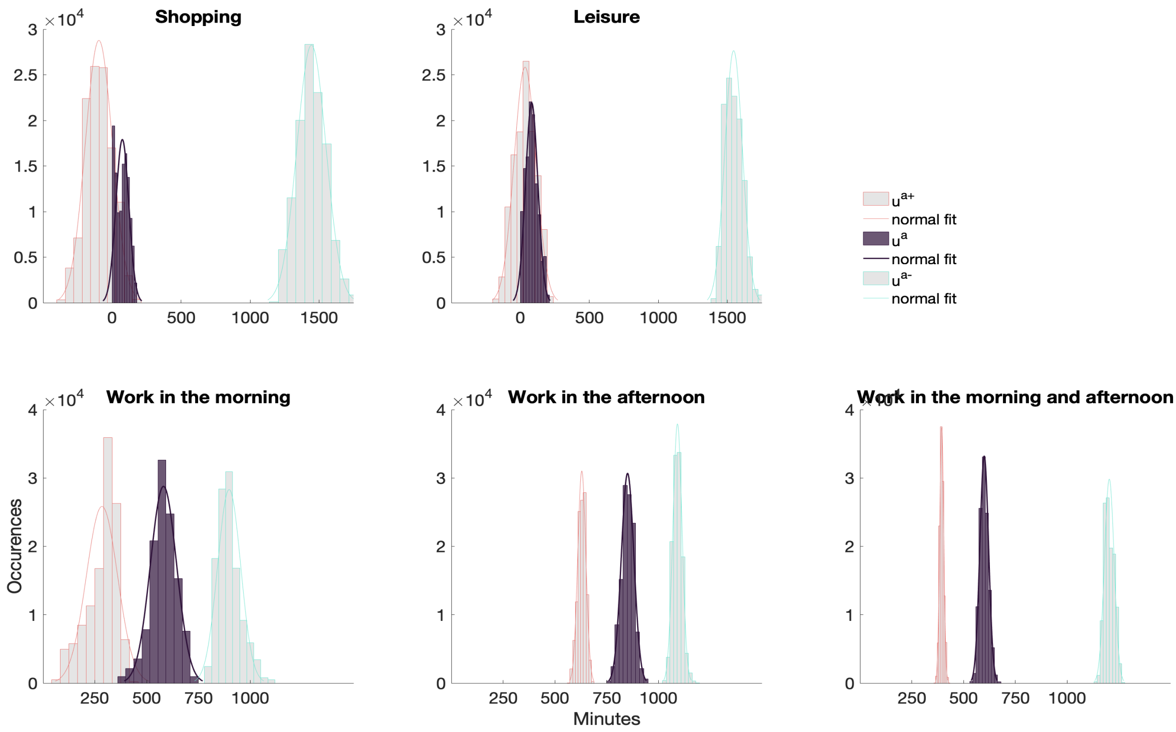


Figure V. 14 Posterior distribution of the α parameter [in minutes] for u_a

One can note here that in the case of A2 (shopping), the central α does follow this typical distribution (Figure V. 15). An explanation of this output is that the aggregation concerns individuals but also activity types. As we interpret α, in the case of $A3$, as the time before reaching the maximal utility, we can explain these two peaks as on the one hand "short-term, daily shopping" which resembles a very short-lasting activity and on the other hand "weekly grocery" or "shopping sessions" which have a longer duration.

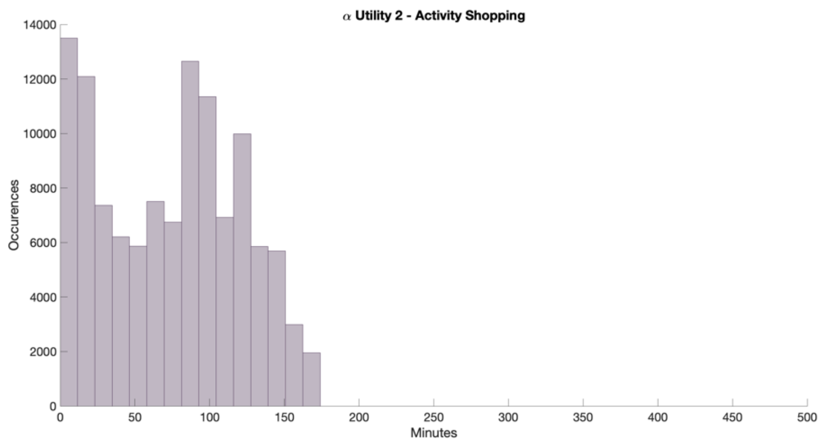


Figure V. 15 Example of non-normal posterior – heterogeneity

V.5.1.2. Evaluation of the results

In addition to actual trip data, we have additional information about the population's behaviour in the used dataset, which can be compared to the model's output. However, it is not possible to evaluate the veracity of the marginal utility parameters used as root for the trip estimation, since these are latent variables that cannot be observed or collected. Only a consistency-check is feasible with previous work which applied the same formulations. The output of the estimation procedure can be compared to ancillary indicators: for example, an index of the estimation plausibility is analysing the duration distribution of each activity. This indicator is an observed output based on the probability distribution for each time combinations corresponding to Δt.

$$N_{\Delta t} = \sum_{t_3-(t_2+tt)=\Delta t} P(t_{2,}t_3).D_a \tag{5.17}$$

Because it is not used as an input of the model, the comparison is fully independent from the estimation process.

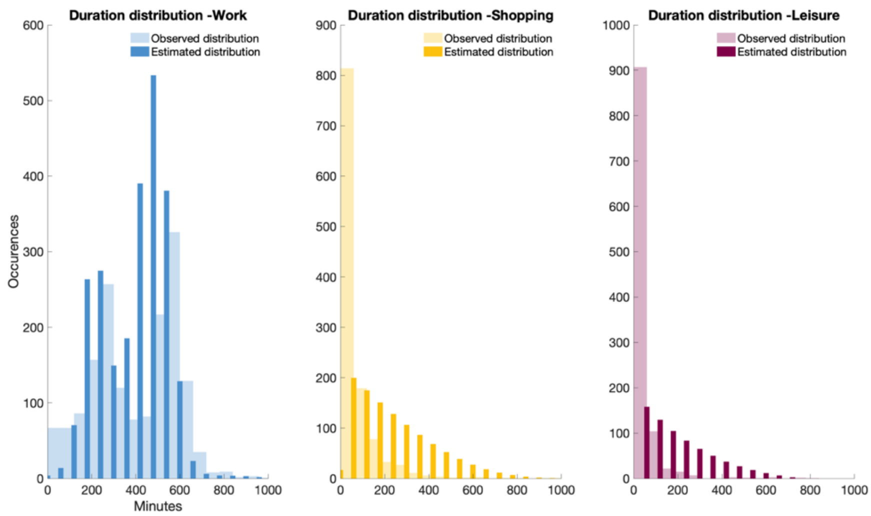


Figure V. 16 Activity duration distribution

Due to the form of the utility function, it is hard to represent very short activities at the aggregate level, given the peaks for U_1 and U_3 which are considered clock-based in all cases. The duration for activities $A2$ and $A3$ is rightly distributed with the highest value close to zero and descending, but overall overestimated. The actual mean duration is 41 minutes for shopping and 64 minutes for leisure. We have estimated $\alpha_1 = 18\text{min}$ and $\alpha_2 = 74\text{min}$ as the duration after which the utility reaches its maximum. If the value for leisure implies an overestimation, the order of magnitude supports the assumption on parameters τ and α.

In the case of the work activity, the bimodal distribution, resulting from the separation between part time and full day shifts is well represented with the two peaks pointing at the right locations (around 4 hours and 8 hours). However, the less observed durations (e.g. 6 hours) seem to be overestimated with our model. The duration distribution underlines that work is easier to be calibrated in comparison to secondary activities. Multiple reasons can explain this phenomenon. Firstly, the work activity is modelled in a more detailed way with three sub-activities, each of them having distinct specification. Secondly, the clock-based approach is easier to estimate with aggregated dynamic traffic counts as unique input information. Finally, the number of observations related to the work demand is larger than the secondary activities, which makes it easier to be distinct in the aggregate counts. Additionally, overestimation of the activity duration shopping and leisure may also be due to the assumption of fixed travel time, since short duration activities are likely performed after shorter travels (e.g. buying bread at a local bakery shop).

Lastly, we can compare, in the case of work during the full day (A1.3), the estimated final values with the empirical estimation done by (Ettema and Timmermans 2003) using a genetic algorithm approach predicting the best parameters based on 79 observed individual travel times, work organization and personal characteristics. This procedure uses input data specific to the traditional home-work-other tour type (i.e., work trips starting between 6AM and 10AM). We can so assess the consistency of the MCMC by comparing these outputs, the "model" value in table (3) refers to these estimation results.

Table V. 9 comparison of parameter estimation

	U1			U2			U3		
Parameter	Ettema 2003	Initial MCMC	Final MCMC	Ettema 2003	Initial MCMC	Final MCMC	Ettema 2003	Initial MCMC	Final MCMC
α	335	250	369	753	725	665	1184	1225	1193
β	0.023	0.005	0.024	0.011	0.050	0.010	0.020	0.005	0.021
U^{max}	9.77	10	9.97	9.78	15	21.83	8.35	10	8.35
γ	1	1	1.1	1	1	1.2	1	1	0.9

The estimation procedure manages to get off non-optimal initial values and get closer to the ones estimated by (Ettema and Timmermans 2003). The most different parameters are referring to the activity work itself (U2); in particular, the maximal utility is higher and at an earlier time (11AM).

This could be explained by the fact that we separated the work activity in more activity types and so that we refer only to the full-time aspect. It can also be due to settings and environmental aspect of the experiment which is based on observed data.

V.6. Discussion

Because we used real data, we can compare the obtained results to the ground truth in terms of activity-specific trips. Validation is therefore based on the observed activity-specific trips.

Firstly, we can compare the share of trips, by activity type and time of the day. The estimation loses in accuracy at such level of detail, in opposition to the aggregate estimation by time of the day which provides an $r^2 = 0.99$. The trips related to work demand has the best estimation which is satisfactory as it constrains traffic the most with a large number of generated trips and so impact more traffic conditions. Given the level of detail of input data, an $r^2 = 0.76$ for work-related trips is considered satisfactory and we expect higher level of accuracy with different kind of data. In addition, the differentiation between incoming and outgoing trips is even more difficult to predict, in particular for this case where only intrazonal trips are considered and the problem is underdetermined. The time periods when those activity-specific trips are inexistant, we observe a clear overestimation by the model, due to the probabilistic estimation of generated trips that always give a non-zero value.

Starting and ending trips are considered together and not using for example an origin-destination matrix to distinguish them. Therefore, the model's division exaggerates the peak in the morning to start an activity and even more the peak in the afternoon to stop an activity. Almost all the demand falls in this category after 4PM, while in reality trips to start leisure or shopping are substantial during this period. The activity share is however coherent with real data. These results show the current model limitations when used in such aggregated form and it highlights the possibilities in order to further improve it. We believe that in the multiple OD case, using input data differentiating incoming and outgoing trips as well as distribution of travel times, these results can only get better. These current limitations are inherent to the assumptions and set up of this specific case study and will be studied in future research, giving also the possibility to include other decisions such as mode and destination choice estimation.

V.7. Conclusion

In this chapter we described the core of a utility-based model relying on advanced sampling methods to determine activity-specific demand based on traffic data through the calibration of a departure time choice model. The concept of trip chaining is handled by first dividing the tour into separated sub tours and then estimating the two-parameters departure time choice. The proposed MCMC is shown to be able to distinguish activity-specific flows from the aggregated demand and the utility-based probabilities prove to be adequate for reproducing a whole day traffic pattern. Inserting strong constraints on the probability form allows to have a better interpretation of the results, however, these constraints make the model unable to reproduce distributions being away from their inherent form. Nonetheless, as the probability curves are calculated with the current model, the results when combined with the actual dynamic OD matrices, can give a useful interpretation to the flows. The errors in the estimation can be explained by constant and equal trip costs as well as the stochasticity and non-linearity of the calibration process. The trip cost is a factor that surely impacts the departure time choice as it varies throughout the day and by origin (or destination). A first measure has been taken, using different trips durations by activity types, however, there is still strong heterogeneity in clustered activities, a finer distinction of activity types could lead to a better estimation. At this stage, the proposed approach treats highly complex model output with the lowest possible dataset level of details, such that the points used for calibration are fewer than the number of parameters to calibrate and an many solutions can fit them. If more data is available, the process could further be guided towards real values. Despite simplifications in term of trip costs and trip chaining and the complexity of the calibration process, the comparison with real data shows already very good results. The unique framework proposes multi-scale utility specifications which offers promising prospects at three different modelling scales.

Chapter 6

Highlights of the chapter

1. The mesoscopic activity-based model is extended for destination choice by explicitly accounting for OD-dependent trip distances
2. A tour-based macroscopic estimation of trips distribution is proposed using the MCMC framework
3. Application to OD generation in 2 case studies is showcased

In order to further develop a complete activity- and mode-specific demand model, and to continue the analogy with the four-step model, it is natural to focus in this chapter on the choice of destination where to perform an activity. While the method has so far proved adequate for the estimation of demand generation in chapter 5, with very accurate results when comparing the total number of generated trips in a whole region, when comparing the trip primitives with microscopic data room for improvement is suggested. This relatively poorer estimation result is expected to be due to the choice of using a constant trip time for each activity, and a single set of parameters for the utility primitives. In this chapter we relax these two assumptions, but we also resort to more data or additional assumptions to be able to extend the modelling approach. The challenge here is to include in a formulation based essentially on the positive part of the accumulated utility a realistic distribution of trips, and therefore a more realistic disutility component.

The spatial framework is therefore extended to traffic analysis zones and the temporal aspect remains the same with a one-day horizon. Two essential components are used. Firstly, we should have a reliable estimate of the number of people performing a given type of activity in each of the zones of the study area. This allows an additional variable to be approximated within the estimation to the marginal utility formulation. This is a factor applied to the maximum potential cumulative utility, which varies according to the destination area. The degree of complexity increases, resulting in higher computation times, as the accumulated utility for each pair of departure times has to be calculated for each potential zone. The second aspect used to estimate the distribution is included in the trip cost. Until now, the disutility of the trip was only reflected in the interruption of the positive utility accumulation during the trip time. To consider variable trip times, but still keeping a simplified approach to include travelling costs, we incorporate a formulation of the disutility associated with the trip, as a function of the typical travel time between two zones and within the same zone. This time is estimated from the database used in the practical application but more generally it can be approximated by the distance and size of the zone pairs in question, or obtained from other available sources (e.g. Google API, TomTom).

To be able to estimate this new component more correctly, a second component of the plausibility formula has been added, in addition to the total number of people attracted per area and per activity. This is a conservation control criterion. Indeed, although we model only one trip of the tour related to an activity, we also estimate the end time of the activity, in the chosen zone. Thus we can compare the number of trips starting in a zone at a certain time of day with the number of activities ending there.

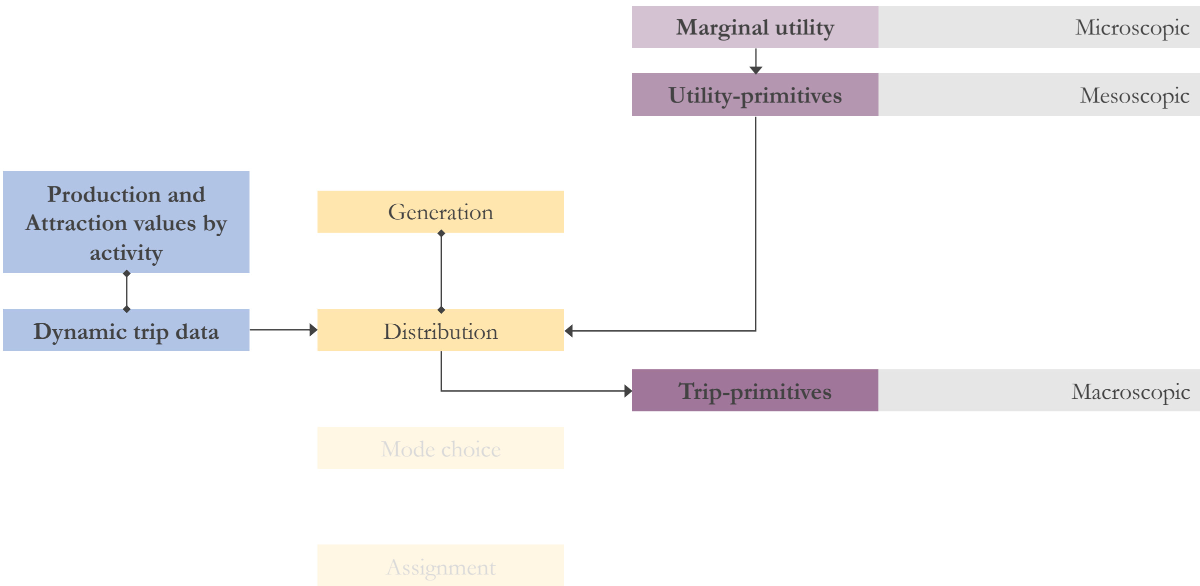


Figure VI. 1 Thesis framework chapter 6

The work presented in this chapter has been described in the following paper:

<u>"Estimation of macroscopic activity-travel demand: A utility maximization approach"</u>

Poster at Scientific congress:
Transportation Research Board TRB Annual Meeting 2020

And

The 25th International Conference of Hong Kong Society for Transportation Studies 9-10 December 2021

VI. DESTINATION CHOICE

VI.1. Introduction

In this chapter, we apply the model described in the previous chapters and propose a methodology to estimate the daily profiles of each identified activity components and the resulting Origin-Destination matrices. Relying on the same general principle of utility maximization, this paradigm is applied at the population level like in the previous chapter to jointly estimate the starting times, destination, and activity at destination of all trips. All components of marginal utility functions are estimated through an *MCMC*. The principles of the *MCMC* are the same as described in chapter IV.2.2 and V.3 but some settings are adapted to the specific application and choice formulation employed. In the proposed form, trips resulting from the choice model are compared at each iteration to a new set of indicators. This model is tested on the same dataset as in the previous case studies, collected in the city of Ghent (2008), containing 15397 trips and four main activity types. We show that the proposed model can properly generate purpose-specific dynamic demand at zonal level and estimate daily origin-destination activity-travel demand. The result of this approach can therefore be applied to estimate activity-specific OD matrices that can for instance be used as input for dynamic origin-destination flow estimation from traffic data.

VI.2. Methodology

This chapter extends the methodology proposed in the previous chapter and includes to the choice facet the destination for every trip. To do so, in this chapter a new parameter is introduced in the marginal utility function, and travel times vary by origin and destination. In this chapter, travel time is not differentiated by time and mode, which is an assumption that will be relaxed in the next chapter. The notations and assumptions described in the previous chapter are used also for the methodology proposed here.

VI.2.1. Utility accumulation

An activity schedule performed by individual i is represented by a triple of vectors: $(\boldsymbol{a}^i, \boldsymbol{o}^i, \boldsymbol{d}^i)$. The activity vector $\boldsymbol{a}^i = [a_1^i, \dots, a_N^i]$ shows the activity type engaged in at every discrete time point. Location is specified using two zone vectors: $\boldsymbol{o}^i = [o_1^i, \dots, o_N^i]$ and $\boldsymbol{d}^i = [d_1^i, \dots, d_N^i]$, with N being the time periods. When engaging in an activity at time t in zone z: $o_t^i = d_t^i = z$. When travelling from zone y to zone z, $o_t^i = y$ and $d_t^i = z$ for the duration of the trip. For example, an individual begins the day at home in zone 8, travelling to zone 4 to work, then shopping in zone 3, returning home to zone 8 (Figure VI. 2). Elements corresponding to travel are highlighted in red.

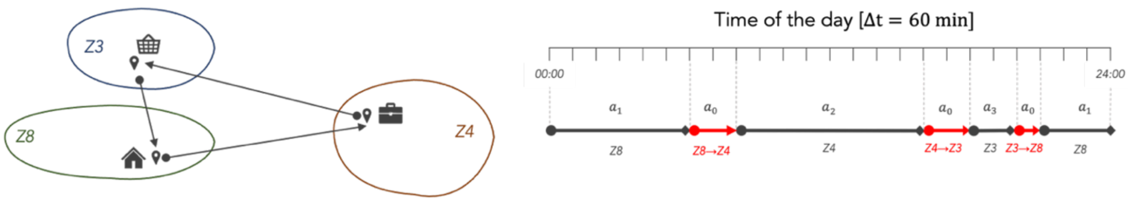


Figure VI. 2 Example of an activity-travel chain for individual i

We discretise the time considering $\Delta t = 60mins$ and encode each activity type as follows: 0 = travel, 1 = home, 2 = work, 3 = shopping, 4 = leisure. Hence, in this instance the vector triple corresponding to $a^i = [a_1^i, a_2^i, \dots, a_7^i] = [1,0,2,0,3,0,1]$ could be represented as:

$$a^i = [1,1,1,1,1,1,0,0,2,2,2,2,2,2,2,2,0,0,3,3,0,1,1,1]$$

$$o^i = [8,8,8,8,8,8,8,8,4,4,4,4,4,4,4,4,4,4,3,3,3,8,8,8]$$

$$d^i = [8,8,8,8,8,8,4,4,4,4,4,4,4,4,4,4,3,3,3,3,8,8,8,8]$$

With a slight abuse of notation, travel activity start/end times will also be denoted using the zone indices: individual i departs zone y at time $t_s^{i,yz}$ to travel to zone z, arriving at $t_e^{i,yz}$.

Let the travel time from zone y to zone z to engage in activity a be tt^{yz}; this variable is not considered here individual-specific but depends on time of day, and varies only by origin and destination. Intrazonal travel is denoted similarly, tt^{yy}. Someone in zone y who wishes to engage in activity a in zone z, starting that activity at time t' therefore needs to depart from zone y at $t' - tt^{yz}$.

The utility U_a^i accrued by individual i, obtained by engaging in activity a from time t_s to t_e is then:

$$U_a^i(t_s, t_e, \mathrm{z}) = \sum_{t=t_s}^{t_e} u_{a,z}^i(t, t_s)\,\Delta t \tag{6.1}$$

This function has the same form as equation (5.1) but varies with the destination zone z as $u^{i}_{a,z}$ is the marginal utility function related to activity a for individual i in a specific zone.

In the case of the "trip" activity, the accumulated utility is typically negative (disutility) and depends on the origin y, destination z, and departure time $t_d = t_s - \ tt^{yz}(t_d)$. The cost associated to this travel time is calculated as follows:

$$c_t(tt^{yz}) = tt^{yz}(t_d)^2 \tag{6.2}$$

This form has been chosen in order to reflect in a very simple formulation the difference between actual travel time and the perceived travel time and in particular in order to support the attractiveness of very short trips while ensuring $c_t(0) = 0$. With this form, very short trips have a lower cost than with a linear cost function while longer trips are fast penalized. This form can be adapted and updated to other attribute-specific functions, such as activity or mode specific cost, remaining consistent with basic concepts of perception (Clark 1982; Stevens 1957). The other costs related to travelling, such as access or waiting time, are ignored.

With the assumption on activity chain simplification, this results in the following utility accumulated when studying activity a. The difference with the formulation proposed in the previous chapter (5.4) relates to the specificities introduced above, i.e. a marginal utility depending on the destination as well as a quadratic formulation of the cost.

$$\begin{aligned} \boldsymbol{U}^{i}_{a}(t_s, t_e | (y, z)) \\ &= \sum_{t=t_0}^{t_d-\Delta t} u^{i}_{a-}(t, t_0)\,\Delta t \quad + \quad u_t tt^{yz}(\mathrm{t}_d)^2 \; + \; \sum_{t=t_s}^{t_e} u^{i}_{a,z}(t, t_s)\,\Delta t \\ &+ \sum_{t=t_e+\Delta t}^{t_N} u^{i}_{a+}(t, t_e + \Delta t)\,\Delta t \end{aligned} \tag{6.3}$$

We assume individual(s) i to be based in zone y and hence incurring travel time tt^{yz} when going to zone z to engage in activity a. Without regard to the details of the previous activity-travel chain A^{i}_{a-} and characteristics of the person, the "base" is defined as origin zone of the trip to a. The trips after activity a is included in the activity $a+$ which could be located in any zone. If compared to equation (5.4), equation (6.3) does include a more specific travel cost and a zone-specific utility accumulation for activity a.

VI.2.2. Marginal utility formulation

The form chosen for the marginal utility formulation is derived from (Ettema and Timmermans 2003) like in the previous chapter (equation (5.8)). However, an additional multiplicative factor has been added with respect to the original formulation in order to reflect that different zones may offer different marginal utility for a given activity type (e.g. some zone may offer better paid jobs or may be more attractive for leisure activities). All activity types, except travelling, bring positive utility in the following manner:

$$u_{a,z}(t,t_s) = \frac{\gamma_a \beta_a (v_{a,z} \, . \, U_a^{max})}{exp[\beta_a(t-(\alpha_a + t_s\tau_a))] \, . \, (1+exp[-\beta_a(t-(\alpha_a + t_s\tau_a))])^{\gamma_a+1}} \qquad (6.4)$$

where the model parameters are defined as in section V.2.4 and additionally:

- $v_{a,z}$: is a new factor affecting U^{max}. It reproduces the relative attraction of a zone z for the given activity type a.

Adding this term instead of defining $U_{a,z}^{max}$ has been decided in order to keep the exact same base structure for different zones, this means that the calibration process can be done in different phase. This could allow for example to apply the same methodology to different zoning or insert easily changes in attractiveness without calibrating any of the other parameters. Having a common variable for different zones also simplifies the approach for defining priors. The factor $v_{a,z}$ can indeed be derived any available data and relative attractiveness of zones for specific activities.

This formulation can still be used in two different formal types following the same rationale explained in the previous chapter. When τ is equal to 0, the utility is determined by time of day, regardless of activity start time. This represents activity types such as work. Whereas $\tau = 1$ describes a fully duration-based utility function: the utility function is offset in time in function of the activity starting time. In this case the saturation point describes an optimal duration instead of a time of the day. Here, for each activity a, the parameter τ is fixed to be $\tau = 0$ or $\tau = 1$ and does not vary in the calibration phase.

VI.2.3. Choice model

Given the formulations (6.1)-(6.4), and in line with the previous chapter, a logit model is used in order to determine the probabilities (and by extension the distribution over the population) of choosing an alternative for the following aspects:

- Starting time of activity a
- Ending time of activity a
- **Zone where to perform activity a**

The destination choice is in this way dependent on the disutility of travelling and in particular on the expected travel time as well as the expected accumulated utility in a destination zone. The relative attractiveness varies thus with time of the day and trip purpose.

In addition, to the number of people (based) in each zone, y, with activity a in their activity chain be Q_a^y, we now Let the total number of people doing each activity in each zone per day be D_a^z and. For all individuals in y pursuing activity a, the attractiveness of different destination zones is distinguished by the relevant travel time. If travelling from zone y to zone z, they will need to depart at $t_s - tt_a^{yz}$. Then the probability of choosing to perform activity a in zone z departing from zone y starting at time t_s will be:

$$P_a^{yz}(t_s, t_e) = \frac{\exp\left((U_a^i(t_s, t_e) + u_t tt_a^{yz})/\sigma)\right)}{\sum_w \sum_{t_e' > t_s'} \sum_{t_s'} \exp\left((U_a^i(t_s', t_e') + u_t tt_a^{yw})/\sigma\right)} \tag{6.5}$$

With respect to the model presented in the previous chapter, here the lower sum includes only those zones (including the current zone) where activity a is available.

Aggregating over all feasible end times gives the probability of departing zone y to go to zone z for activity a, starting in zone z at time t:

$$P_a^{yz}(t) = \sum_{t_e > t} P_a^{yz}(t, t_e) \tag{6.6}$$

We associate the incoming trips to zone z at time t, including intrazonal trips, with the number of people engaging on activity a in zone z at that time:

$$T^{\to z}(t) = \sum_y \sum_a Q_a^y \, P_a^{yz}(t) = \sum_y \sum_a \sum_{t_e > t} Q_a^y \, P_a^{yz}(t, t_e) \tag{6.7}$$

Similarly, the number of trips departing (ending activity) zone z at time t is

$$T^{z \to}(t) = \sum_y \sum_a \sum_{t_s < t} Q_a^y \, P_a^{yz}(t_s, t) \tag{6.8}$$

We assume that the total number of people per day engaging in activity a in zone z is known:

$$D_a^z = \sum_t \sum_y \sum_a \sum_{t_e > t} Q_a^y \, P_a^{yz}(t, t_e) \tag{6.9}$$

Finally, for each activity, the total number of participants matches the total demand summed across all zones.

$$\sum_z D_a^z = \sum_y Q_a^y \tag{6.10}$$

VI.2.4. Utility parameter distribution and calibration

Marginal utility functions typically describe the possible utility accumulated by individuals, whereas we recall that in this thesis we use it to represent an aggregated variable. Each marginal utility function, including those composing the utility primitives, may have different values from individual to individual. This heterogeneity can be described through the distribution of parameters.

VI.2.5. Process

As in the previous chapters, each variable is sampled without knowing its actual distribution but providing a posterior probability distribution as an output of the stochastic process. These distributions can well represent the heterogeneity of different users and the marginal utility of different activity types and components of the utility-primitives.

VI.2.6. Scoring

The computed score is:

$$S_i = \mathcal{L}_i + \sum_{\mathbf{\Theta}} \log\left(\pi(\mathbf{\Theta}_i)\right) \tag{6.11}$$

In order to include multiple indicative aspects, output of the model, the likelihood function includes the following different components, all in terms of trip numbers:

$$\begin{aligned} \mathcal{L}_i = \rho_1 * \sum_z \sum_t -\frac{1}{2}\left(r_{1,i}(z,t)^2\right) + \rho_2 \sum_z \sum_a -\frac{1}{2}\left(r_{2,i}(z,a)^2\right) \\ + \rho_3 \sum_z \sum_t -\frac{1}{2}\left(r_{3,i}(z,t)^2\right) \end{aligned} \tag{6.12}$$

where

- ρ_n is a scale factor applied to the n^{th} component of the likelihood;
- $r_{1,i}(z,t)$ is the difference for generated demand by zone and time of the day;
- $r_{2,i}(z,a)$ is the difference for total attracted demand by zone and by activity type;
- $r_{3,i}(z,t)$ is the conservation between attracted demand and generated demand at the zonal level.

These residuals are calculated according to equations (6.13), as the difference between the observed data and the output of the utility-based choice model for departure time, destination, and mode.

$$
\begin{aligned}
r_{1,i}(z,t) &= T^{z\rightarrow}(t) - \widehat{T'^{z\rightarrow}}(t) \\
r_{2,i}(z,a) &= D_a^z - \widehat{D_a^z} \\
r_{3,i}(z,t) &= \widehat{T^{z\rightarrow}}(t) - \widehat{T'^{z\rightarrow}}(t)
\end{aligned}
\tag{6.13}
$$

The values of the scale factors are selected in order to give more or less importance to the different results and keep an order of magnitude comparable with the prior component. They are fixed for the full process and chosen based on the goal of the simulation as well as the number of datapoints included.

VI.3. Case study

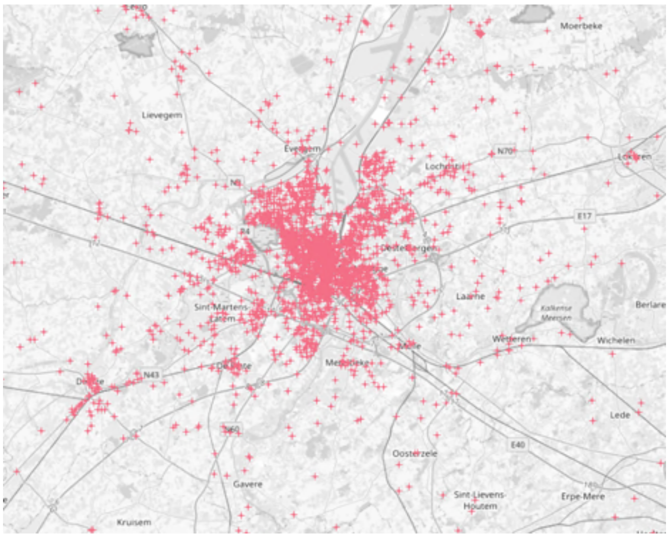
VI.3.1. <u>Dataset</u>

The proposed methodology has been applied to a dataset collected in the area of Ghent in 2008 (Castaigne 2009). For this case study, the area of Ghent has been divided in 10 zones (Figure VI. 3c). A total of 15397 trips is used in this application, for which we know the

- Origin
- Destination
- Starting time
- Ending time
- Travel time
- Modes
- Activity at origin

- Activity at destination

They result from a multi-day survey collected on 707 individuals, randomly sampled with stratification criteria according to household size, gender and age. A description of the database and analysis of the variability of daily activity-travel pattern is available in (Raux, Ma, and Cornelis 2016).



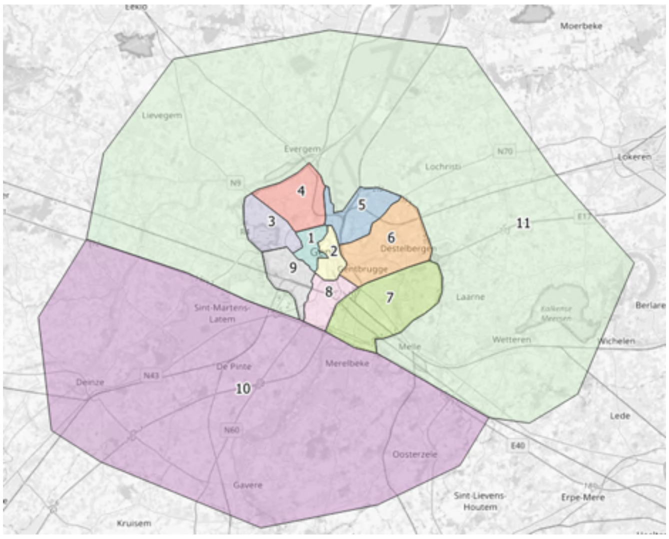


Figure VI. 3 Study area (a) points visited (b) and zoning (c) in Ghent, Belgium

The survey includes 12 activity types, clustered in the following categories:

- Home
- Work
- Shopping & other mandatory activities
- Leisure & other secondary activities

The generated demand for different activities is similar from one zone to the other while the attracted demand varies more, in particular for home and work trips (Figure VI. 4).

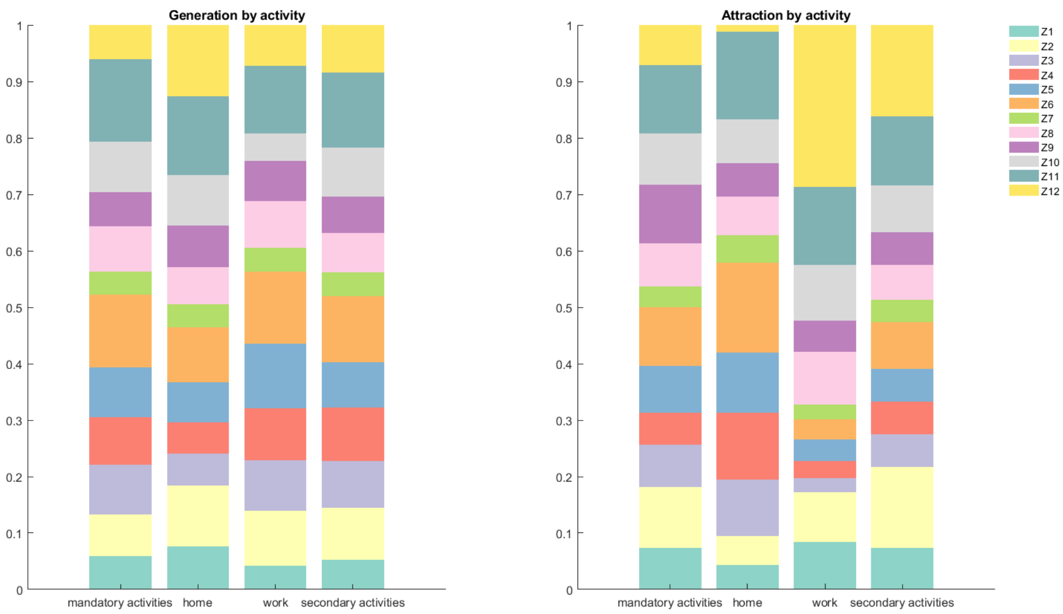


Figure VI. 4 Proportion if total generated and attracted demand, by activity type, for each zone

VI.3.2. Input to the MCMC

In order to apply the Markov Chain Monte Carlo estimation process using the available dataset, needed input data are to start the procedure:

- the total number of people willing to perform a certain activity from each zone $Q_a^z \ \forall \, z, a$
- for the likelihood estimation:
- the number of trips starting by zone and time of the day: $T^{z\rightarrow}(t) \ \forall \, z, t$
- the total number of people doing a certain activity in each zone $D_a^z \ \forall \, z, a$
- and for calculating the trip duration and so the related disutility of the trip:
- the average travel time for each OD pair $d^{yz} \ \forall \, y, z$

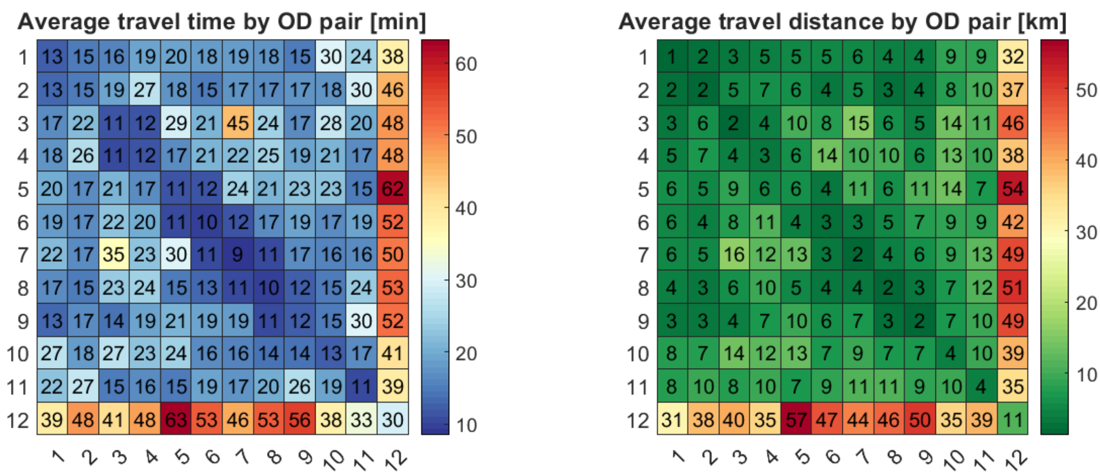


Figure VI. 5 average by OD pair for (a) distance travelled (b) travel time

VI.3.2.1. Parameters to estimate

In order to estimate the demand related to the 4 activities considered, assumptions are made to better reproduce typical behaviours. The home and work activities are separated into different subcategories having diverse characteristics.

Table VI. 10 Activity type definition

	Activity type	Type of utility u_a	Total observed demand
1	Shopping	Duration-based	3655
2	Leisure		4280

3	Work AM	Clock-based	
4	Work PM		
5	Work AMPM		
-	Work	-	1687
6	Home AMPM	Duration-based	
7	Home PM	Clock-based	
-	Home	-	5775

The parameter τ_n is fixed to $\tau = 0$ for clock-based utilities and $\tau = 1$ for duration-based utilities and all u_{a+} and u_{a-}. For each of these seven activities, 13 parameters are thus estimated in the MCMC, corresponding to the parameters of equation (6.4) and the proportions of the whole demand corresponding to the given work-type and home-type:

- $\Theta_a^{\mathrm{u}} = \left(U_{u,a}^{max}, \alpha_{u,a}, \beta_{u,a}, \gamma_{u,a}\right)$ where $\mathrm{u} = u_{a-}, u_a, u_{a+};\ a = 1, \dots, 7$
- $\Theta_a^{\mathrm{z}} = \upsilon_{\mathrm{a,z}}$ where $z = 1, \dots, 12;\ a = 1, \dots, 7$
- $\Theta_{\mathrm{share(W)}}^{(w=1:3)} = P(W_w)$ with $\sum_{w=1:3} P(W_w) = 1$
  with $w = 1\ work\ AM, w = 2\ work\ PM$ and $w = 3\ work\ AMPM$
- $\Theta_{\mathrm{share(H)}}^{(h=1:2)} = P(H_h)$ with $\sum_{h=1:2} P(H_h) = 1$
  with $h = 1\ home\ PM$ and $h = 2\ home\ AMPM$

$$\mathbf{\Theta} = \sum_u \sum_a \Theta_a^{\mathrm{u}} + \sum_z \sum_a \Theta_a^{\mathrm{z}} + \sum_w \Theta_{share(W)}^{\mathrm{w}} + \sum_h \Theta_{share(H)}^{\mathrm{h}}$$

In total $\mathbf{\Theta}$ contains 173 elements to estimate. For each of these parameters, a prior function is defined according to table 2. The initial value is set to the mean of the prior for every parameter apart from $U_{max} = 9$ for all utilities in order to start from a neutral balance between alternatives.

Table VI. 11 Utility priors

Shopping					
Parameter	Shape	Truncated [0;+inf]	Mu	Sigma	Step size
α	Normal	Yes	[480;15;1080]	[50;50;50]	[20;5;20]
γ	-	-	0	0.25	0.1/d

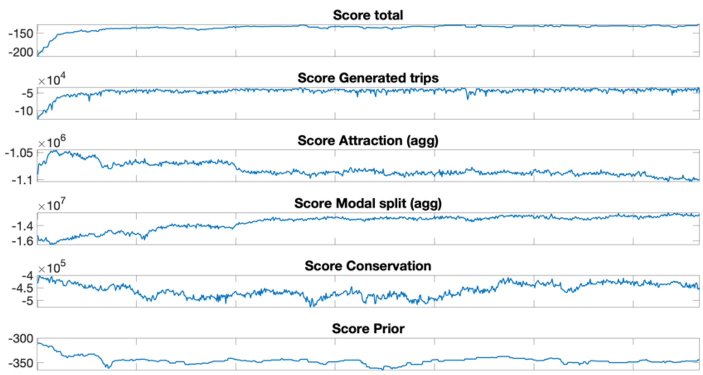
τ	-	-	[0;1;0]	-	0
U^{max}	-	-	[10;15;10]	5	mu/d
β	-	-	[0.008;0.05;0.008]	0.2	0.01/d
Leisure					
α	Normal	Yes	[600;40;1440]	[50;50;50]	[20;5;20]
γ	-	-	0	0.25	0.1/d
τ	-	-	[0;1;0]	-	0
U^{max}	-	-	[10;15;10]	5	mu/d
β	-	-	[0.008;0.05;0.008]	0.2	0.01/d
Work AM					
α	Normal	Yes	[250;600;960]	[50;50;50]	[5;5;5]
γ	-	-	0	0.25	0.1/d
τ	-	-	[0;0;0]	-	0
U^{max}	-	-	[10;15;10]	5	mu/d
β	-	-	[0.008;0.05;0.008]	0.2	0.01/d
Work PM					
α	Normal	Yes	[600;900;1225]	[50;50;50]	[5;5;5]
γ	-	-	0	0.25	0.1/d
τ	-	-	[0;0;0]	-	0
U^{max}	-	-	[10;15;10]	5	mu/d
β	-	-	[0.008;0.05;0.008]	0.2	0.01/d
Work AMPM					
α	Normal	Yes	[360;780;1225]	[50;50;50]	[20;20;20]
γ	-	-	0	0.25	0.1/d
τ	-	-	[0;0;0]	-	0
U^{max}	-	-	[10;15;10]	5	mu/d
β	-	-	[0.008;0.05;0.008]	0.2	0.01/d
Home AMPM					
α	Normal	Yes	[0;60;1080]	[100;50;50]	[5;5;5]
γ	-	-	0	0.25	0.1/d
τ	-	-	[0;1;0]	-	0
U^{max}	-	-	[10;15;10]	5	mu/d
β	-	-	[0.008;0.05;0.008]	0.2	0.01/d
Home PM					
α	Normal	Yes	[720;1380;1440]	[100;100;10]	[20;20;0]
γ	-	-	0	0.25	0.1/d
τ	-	-	[0;0;0]	-	0
U^{max}	-	-	[10;15;0]	5	0
β	-	-	[0.008;0.05;0.008]	0.2	0.01/d

The prior for parameters linked to the destination zone is selected based on the value of D_{α}^{z} and because this value is known the variance of this parameter is smaller than for the others.

VI.4. Results

The presented results have been produced with the settings presented above and 3000 iterations of the MCMC. A burn-in period of 300 iterations has been chosen, i.e. all the values from iteration 1 to 300 are not included in the posterior distribution.

The iterative process shows a good convergence from the overall scoring function (Figure VI. 6a). The prior-related score is decreasing, this is because the initial values are in most cases the mean of the normal prior, such that the initial score is the highest score possible. This behaviour shows the ability of the MCMC to rectify the prediction while keeping plausible values. The score related to attraction is decreasing as well because its impact on the estimated choice model is very weak, in addition the factor ρ_1 is higher than for the other components.



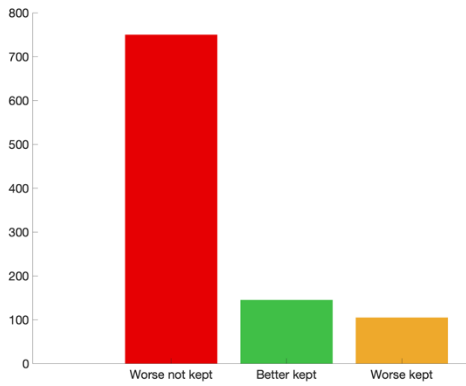


Figure VI. 6 (a) Score's components for 3000 iterations (2) acceptance rate of sampled parameters

One of the features of the adopted Metropolis algorithm is that a set of sampled parameters resulting in a lower score can be selected based on the acceptance criterion (equation (4.7)). Figure VI. 6b shows the proportion of parameters' set increasing the score, decreasing the score but accepted and finally decreasing the score but rejected. The observed ratio of these three components is satisfactory because we can see that the score improves but that there is possibility of leaving local optima and search better the feasible range.

VI.4.1. Generated demand: trip-primitives

The total aggregated demand at the zonal level is reproduced with a very high accuracy ($r^2 = 0.9$) (Figure VI. 7). This complete demand profile is not used in the likelihood computation, the fitting is instead governed by the demand by time of the day and zone of origin.

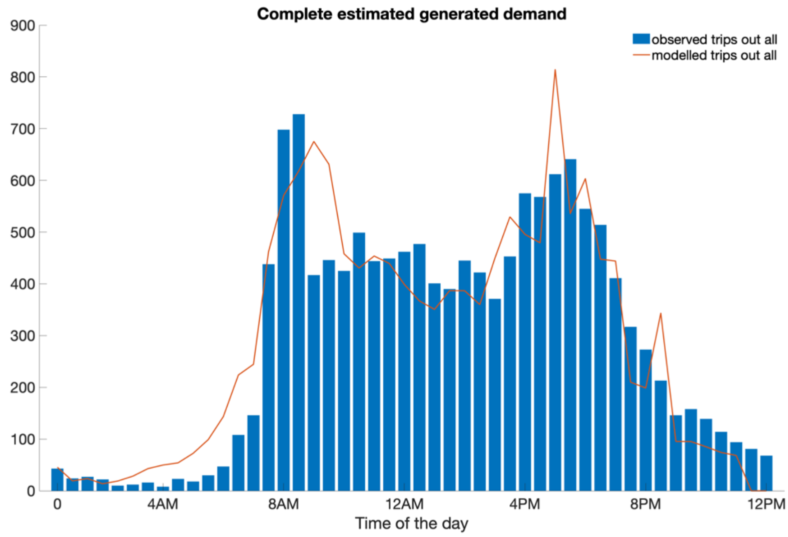


Figure VI. 7 Generated demand by time of the day for the complete study area

Figure VI. 8 shows the zone-specific results in terms of generated demand for all kind of activities in the case of different zone types. The procedure is slightly more efficient for bigger zones and this affects the final scoring. This result is not considered as a strong disadvantage of the methodology because a higher generated demand results also in contributing in stronger way to congestion, so it is more important to estimate well those components. We see however on Figure VI. 8 that atypical demand profiles are modelled properly as well. For example, zone 1 where the peak in the morning is almost not apparent. Despite a common utility formulation, as the examples in Figure VI. 8 show, the profiles can vary a lot from one zone to the other.

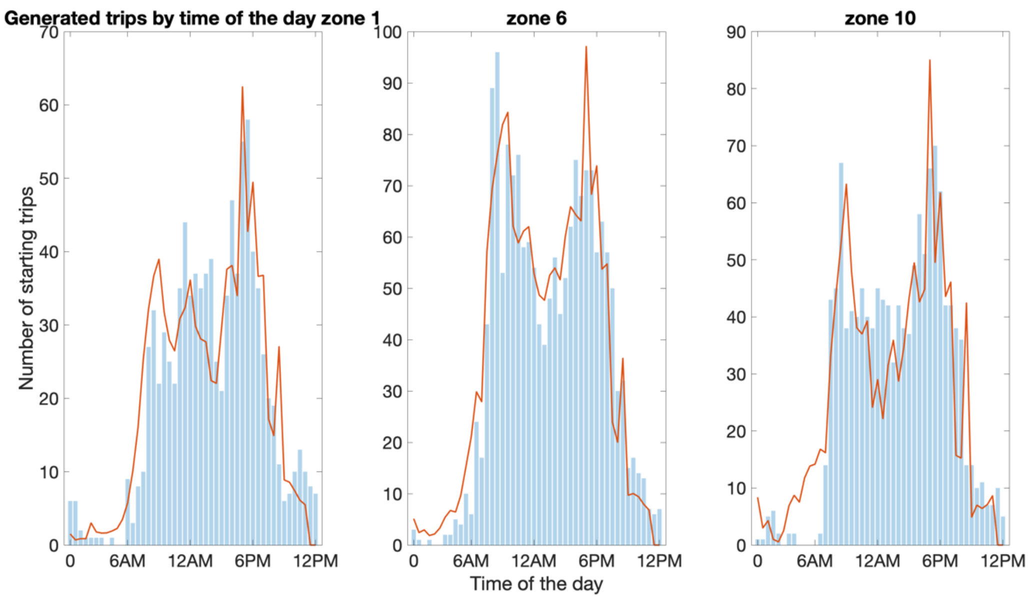


Figure VI. 8 Modelled and observed demand for three zones

We can also describe in detail the different components of the full demand. Figure VI. 9 shows the trip-primitives estimated by the model for zone 9 in comparison with the reference values. We can see that the magnitude by activity type is broadly in line with the observations and that the overall daily dynamics are well reproduced. In particular, the peaks for working activity which corresponds well to the clock-based utility function and the more uniform demand for activities having a duration-based utility function. The different sub-activity types for home and work allow to reproduce very well the daily variations of these activities. The relative importance of the peaks depends on the proportions of activity sub-types. Similarly, the time limits for activity Shopping and Leisure could be included, to reflect opening hours and so represent better the starting and ending time of the uniform profile.

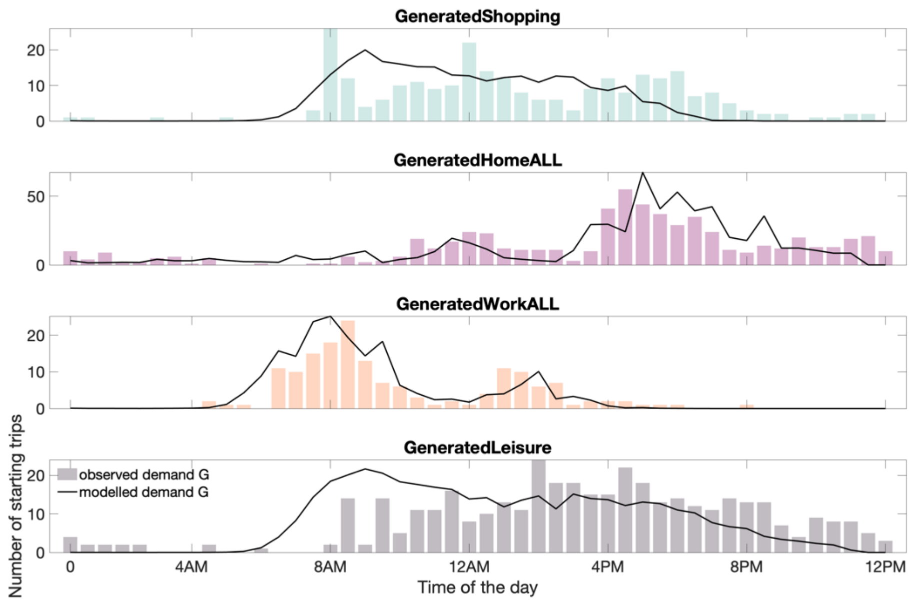


Figure VI. 9 Trip-primitives for one zonem during one day

Overall, the estimation of generated trips by activity type and time of the day is satisfactory, given the level of detail of the output. The estimation's drawback is for times of the days, activity and zones which do not count many observations. In particular when the observed demand is close to 0, the shape of the proposed utility function is unsuitable. In addition, when the demand is not following the usual form, expected for any zone, it is more difficult for the model to differentiate the relative components of the demand.

VI.4.2. Attracted demand

The attracted demand results from the destination choice model, coupled to the departure time choice model. The estimation of these attracted trips depends on the difference in travel time and the attractiveness factor. However, the difference in travel time has a relatively minimal impact as the difference in travel time from one zone to the other is rarely higher than the considered time bins. Making the calculation with smaller time intervals increases drastically the duration of each iteration to detect differences, the average travel time differences between zones being less than 5 minutes, we would ideally need a one-minute simulation. This can be done theoretically, but it would preferably require a level of input data (dynamic trip counts) which has the same level of precision and is in practice very hard to collect. The highest difference in the average travel time is 40 minutes (zone 5 internal is 7 minutes and from zone 5 to zone 10 is 47 minutes). Attracted demand is not controlled by dynamic data inside the likelihood function but only through the aggregated number of attracted trips by activity D_a^z. For this reason, the daily profiles of attracted demand are not as accurate as the ones of the generated demand. However, the attracted demand by zone and activity type arrives to a close result. The estimated OD matrix is presented on Figure VI. 10a and is close to the actual one $(\rho^2 = 0{,}72)$. Overall daily OD trips are modelled with a lower accuracy than the generated trips by zone and time of the day $(\rho^2 = 0{,}85)$ but a higher one than activity specific OD demand $(\rho^2 = 0{,}6)$.

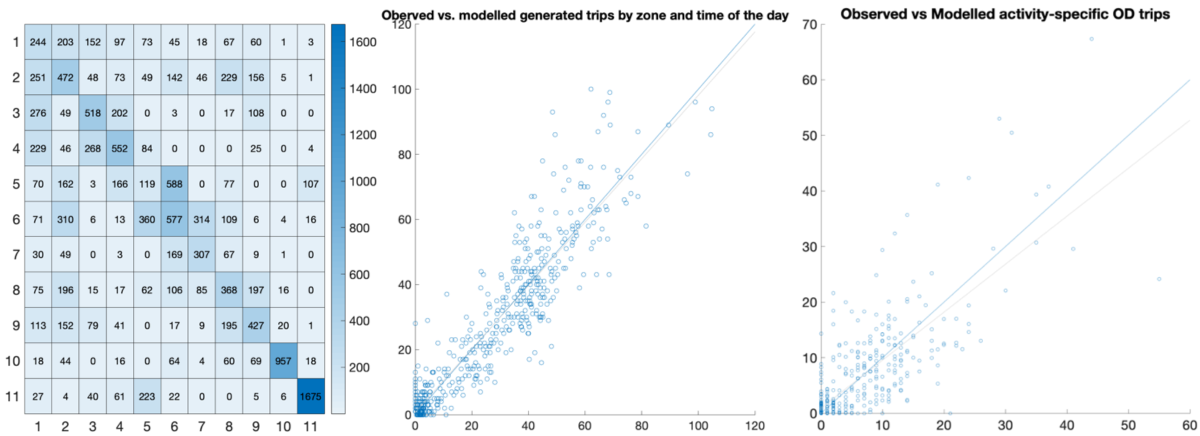


Figure VI. 10 (a) Estimated OD matrix and fit of (b) generated trips (c) activity-specific OD trips

VI.4.3. Utility-primitives

The utility-primitives are the main component resulting in the modelled demand. On Figure VI. 11, we see the resulting utility functions for a selection of activities. The parameter values are the average value of the posterior vector Π_θ, (except the burn-in period). This estimation process is done individually for each activity type so the relative order of magnitude of these curves is relevant only within the three components of utility-primitives and among zones but not among distinct activity types.

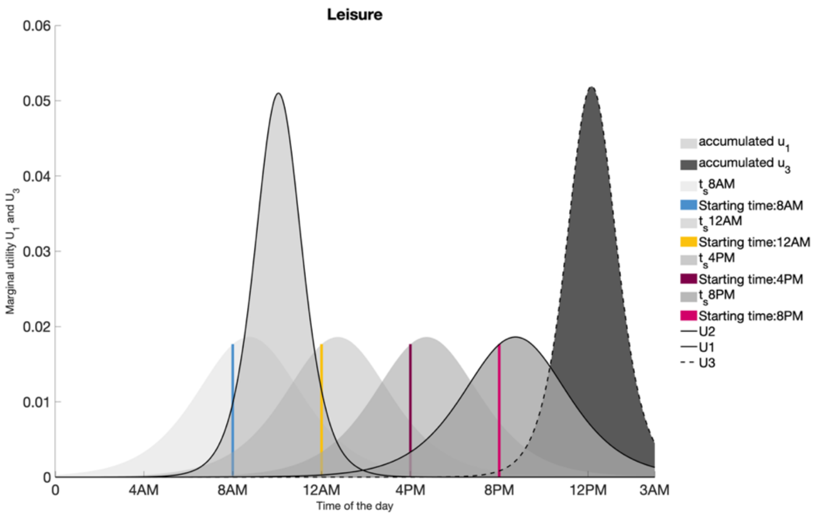


Figure VI. 11 Utility primitives of Work and Leisure

VI.4.4. Supplementary indicators

In addition to the aggregate travel, the choice model results in estimating the distribution of the duration of an activity which differs for each activity type and by extension it can provide an indication of the number of people staying in a zone (zonal occupancy). This occupancy can be calculated by the difference between the cumulative number of arriving trips and leaving trips by time of the day:

$$O^{z,a}(t) = \sum_{t_s<t} T_a^{\rightarrow z}(t_s) - \sum_{te<t} T_a^{z\rightarrow}(t_e) \tag{6.14}$$

For the home activity, all trips at the end of the activity (returning) happen before 12PM because the studied time frame is from 00:00AM to 12:00PM.

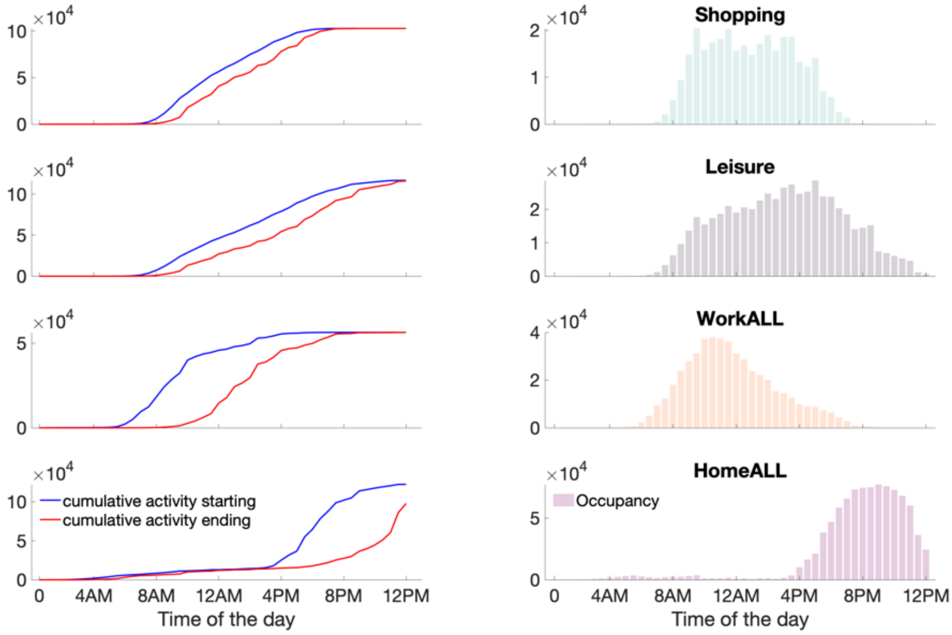


Figure VI. 12 Total occupancy (study area) by activity type

The occupancy is an interesting indicator for the assessment of management solutions and can be used on the side of the activity duration. For instance, it can be used to indicate the demand for parking in an area.

However, the duration for every activity is mostly overestimated by our model, in particular for duration-based marginal utilities. We can explain this by the fact that u_{a-} and u_{a+} are assumed clock based. This means that no utility is gained from returning from that activity before a certain time of the day so the duration of this activity can be longer than expected. Having a duration-based marginal utility formulation u_{a-} and u_{a+} would increase a lot the computational time and wouldn't result in a better estimation of the resulting traffic. On the contrary, these formulation act as an upper and lower limit for activity participation. For the home activity, we can see well the two components which results in two peaks of duration. Finally, work is slightly overestimated, this can be due to a wrong estimation of the share between full time and part time work. This could easily get better if the share is used as input of the MCMC.

VI.5. Conclusion

In this chapter, we presented a model for recognizing activity-specific trips components in observed travel demand. Its application to the proposed case study shows the potential to model accurate demand at the origin-destination level with a very low necessary level of detail for input data. In addition, the MCMC approach has the advantage of delivering information on the heterogeneity of the population and on the aggregation of activity types.

While it impossible to observe actual utility values and so to calibrate marginal utility functions, we proposed here an alternative approach using trip counts deriving from these functions to describe accumulated utility for various activity types. Some assumptions needed to be made on those utility functions to be applied at the population level, with an aggregate perspective. Yet existing formulations designed for the individual scale can fit the proposed scope.

The results of the case study show that the efficiency of the method is higher for large travel demand. Despite the parsimonious aspect of the method, it should be noted that the quality of the results increases with more detailed available information. Several aspects, in addition to the departure time choice model are integrated in the proposed model to include destination choice aspect. The potential applications of this additional level are however many with for example the creation of realistic seed matrices for OD estimation. It is also possible to understand the correlation between trips observed at different times of the day and between different zones and so to forecast better dynamic demand.

Chapter 7

Highlights of the chapter

1. Introduction and joint calibration of an activity-mode specific trip cost formulation
2. Two approaches to activity mode choice approximation are presented and compared
3. A two-step MCMC procedure to estimating OD specific modal split is proposed

The last aspect of demand modelling that we include in the methodological contribution of this thesis naturally concerns modal choice. First, we describe a simple expansion of the model described in the previous chapter and then a version more adapted to the modelling of this aspect of demand estimation. These two approaches, according to the available data, could be combined in application.

In order to integrate modal split and therefore generate activity-mode specific origin-destination trips, the simplest estimation consists in combining and integrating the choices of departure time, destination and mode within the utility maximisation formulation. The different modes are here distinguished by travel time. On the one hand, we use the hourly speed, in order to represent the level of service, affected by the frequency of public transport or the presence of traffic jams. On the other hand, the distance varies between the two zones in question. This provides an approximation of the travel time by mode, OD pair and time of the day and thus it creates a dynamic modal split.

But unlike the choice of destination, it turns out in this section that this aspect can hardly be integrated through only regarding the marginal utility formulated at destination, and the journey costs then become an essential component. Thus, we propose a formulation of the cost depending on both the mode and the activity at destination. We incorporate a mode-specific system access cost that can be calibrated for each area to reflect both the size and the infrastructure and services available within that area. In addition, we assume that a journey time is not perceived in the same way depending on the activity at destination. This variation in the value of time is transcribed through a quadratic cost factor that allows us to model a higher acceptance for a commute than for a leisure trip, for example.

Thanks to the variation in utility between each destination zone, the different generation profiles of the origin zones, and a distinct activity split for each pair, we can with this last method also obtain a modal split that varies according to the time of day.

In order to model this cost aspect as well as possible, we propose a two-step approach to calibration. A first quick MCMC allows us to fix the parameters related to the marginal utility and the positive share while in a second step, a second MCMC focuses on the different parameters related to the cost function.

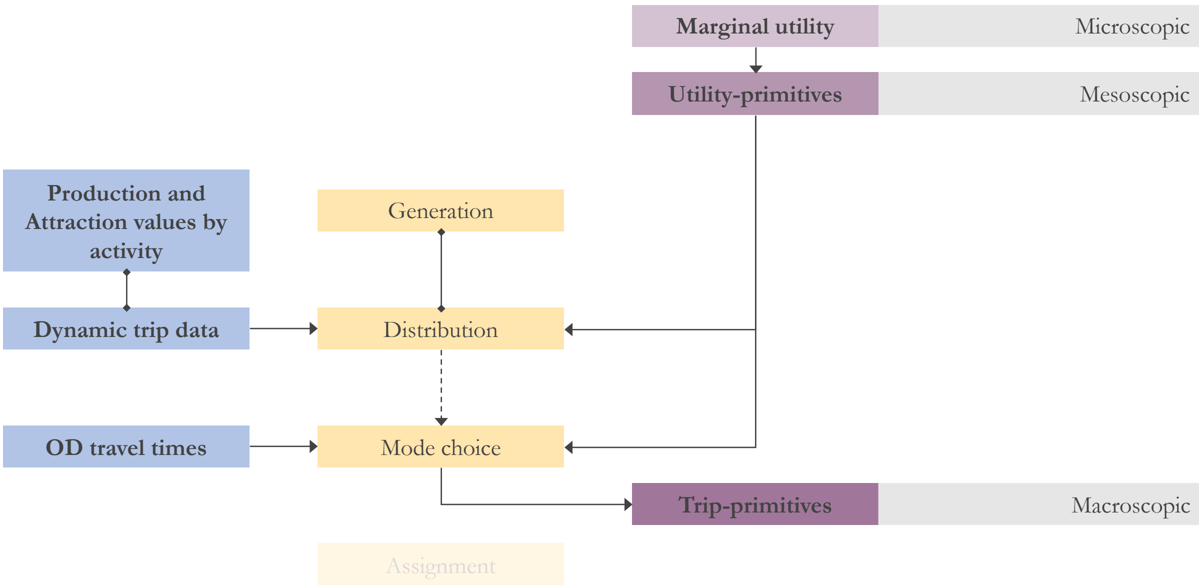


Figure VII. 1 Thesis framework chapter 7

The work presented in this chapter has been described in the following paper:

“Dynamic Modal Split Incorporating Trip Chaining: A Parsimonious Approach to Mode-Specific Demand Estimation”

Transportation Research Procedia
24th EURO Working Group on Transportation Meeting, EWGT 2021, 8-10 September 2021

VII. MODE CHOICE

VII.1.1. Introduction

Dynamic mode choice is essential to understand the potential effectiveness of policies aiming to achieve desirable modal split targets or to manage the demand for resource-limited systems such as shared mobility services. In this chapter, we propose two distinct approaches to modal split estimation. First, an estimation of dynamic modal split for work-related trips, including mode- and time-specific costs. To obtain an accurate profile while remaining at an aggregate level, three types of work activities are described (full time, morning and afternoon shift). The MCMC procedure is used to evaluate the marginal utility function parameters which are used in a joint departure time and mode choice evaluation.

The second approach is less demanding in terms of dynamic input data but includes stronger assumptions at the modelling level. A function of the disutility of travelling including mode and activity specific components is proposed. The dynamic variations of modal split are in that case a consequence of variations in trip-purpose share. The estimation procedure is slightly different as the departure time and destination are estimated jointly over a first phase with an expected travel time while the mode choice is handled over a second phase. The two phases have different level of detail and so this reduces considerably the execution time. Indeed, the generation does not require to compute marginal utilities for each different zone while introducing the destination specific factor increases the computation time. Furthermore, in that application, we showcase a possible utilization of the MCMC calibration process where different parameters cannot or don't need to be jointly estimated.

The estimated modal split concerns motorized vehicles, soft modes but also train and urban public transport. Based on utility maximization principles, the accumulated utility is formulated within the departure time choice model described in the previous chapter.

VII.2. Method I

Mode specific travel speed for each time of the day is used to estimate also travelled distance distribution per mode. The methodology is applied and tested, using data collected in Ghent in 2008, as in the previous chapters. 16.749 work-related trips have been considered in a simplified estimation where two successive trips are constrained to be done with the same mode. This assumption is in line with the conclusions of the empirical analysis of chapter 3. The proposed method is easy to implement using only dynamic trip counts, without the need for simulation or traffic assignment.

VII.2.1. Methodology

In this part, we assume to have information, or a reliable estimation, of the total daily demand for work-related trips within a study area, D_w. Then, thanks to the proposed model, the time- and mode-dependent demand $D_w^m(t)$ is calculated according to departure time probabilities calculated based on utility maximization principles. Individuals are assumed to optimise their schedule to maximise their utility and the aggregated trips, resulting from a probability model which produces emerging traffic flows. We estimate their choices as a utility maximization problem with time-dependent travel times which are mode-specific but not function of the estimated flows.

Following the general framework proposed in (Yamamoto et al. 2000b), we define the overall net utility accumulated during the reference time period $\boldsymbol{U}$ as the sum of the disutility/cost of travelling and the positive utility of performing one or more activities (equation (5.3)).

In this section, we do not include destination-specific utilities, however, equation (7.1) can easily be extended to capture zone-dependent marginal utilities simply by labelling activities in different zones as distinct activity types offering different marginal utilities. We believe anyhow that considering marginal utilities to be zone-independent is acceptable for this specific study since our focus is on generating dynamic modal split. Jointly estimating destination and mode choice would be a straightforward extension of the model proposed in the previous chapter.

Additionally, in this study we focus on the work activity only, again for the sake of simplicity but the same approach can be used for estimating any activity type. In the same manner as we considered the notation for activity a, a^-, a^+ in the previous chapters, for sake of clarity in this special application case, we use the following notation:

- w^- relates to all activities which took place before the trip to go to work
- w^+ relates to all activities after leaving the work activity, including any trips made to access them.

We formulate these two before/after blocks as singular activity types. We assume that the disutility of travelling has constant marginal cost, $c_t < 0$ per unit time. Travel time to the working activity, from zone y to zone z by mode m, $tt_{work}^{yz,m}(t_d)$, is OD- and mode-specific and depends on the departure time, t_d. Departure time is chosen to arrive at the desired start time: $t_d = t_s -$

$tt_{work}^{yz,m}(t_d)$. Note that the trip components are assumed to be not individual specific. For individual i, commuting using mode m, with work starting and ending time (t_s, t_e) given a zone of origin y and destination z and hence travel time $tt_{work}^{yz,m}(t_d)$, the total daily utility gained is

$$\begin{aligned} \boldsymbol{U}_{work}^{i}(t_s, t_e, m|(y,z)) & \\ &= \sum_{t=t_0}^{t_d-\Delta t} u_{w-}^{i}(t, t_0)\,\Delta t \quad + \quad c_t tt_{work}^{yz,m}(t_d) \quad + \quad \sum_{t=t_s}^{t_e} u_{work}^{i}(t, t_s)\,\Delta t \\ &+ \sum_{t=t_e+\Delta t}^{t_N} u_{w+}^{i}(t, t_e+\Delta t)\,\Delta t \end{aligned} \tag{7.1}$$

The total accumulated utility (2) includes four components. The first and last represent all trips and activities performed before and after work; the central components are the work activity and the trip to reach it. This formulation could easily be generalized to include dependence of the marginal utilities on the location where activities are performed, as shown in the previous chapter. This would result in estimating OD-specific parameters. However, location choice is not explicitly studied in this part of the study and hence marginal utilities are considered zone-independent. Similarly, without information at the zonal level, travel time is approximated based on the type of activity (at the destination) and is mode-specific, i.e. $tt_{work}^{m}(t_d)$. These elements are the necessary components of the utility maximization process.

VII.2.1.1. Marginal utility formulation

All marginal utilities follow this functional form, including those capturing the 'activities' w^- and w^+. In this model we use equation (7.1) to express the expected marginal utility at an aggregated level, i.e. at a zone level. Individual heterogeneity is captured by probability distributions of model parameters, which emerge from the parameters' estimation process. For a given activity, this collection of model parameters across individuals is denoted $\boldsymbol{\theta}$.

VII.2.1.2. Trip duration formulation

At the aggregate level, a new assumption needs to be done in terms of experienced travel time. For each time of the day, various data sources can be used to get the expected travelling speed by mode $v_m(t_d)$, which is assumed to vary by time of day to consider congestion, varying service frequencies and quality levels. This value is considered as given and fixed while the travelled distance by mode is a distribution d_m, used to compute the expected travel time by mode and time of the day:

$$tt_{work}^{m}(t_d) = v_m(t_d) * E[d_m] \quad (7.2)$$

VII.2.1.3. Estimation process

Based on formulations (2)-(4), the departure time is expressed as a discrete choice process using a multinomial logit model like in the previous chapters. The probability of choosing the pair of starting and ending times (t_s, t_e) and mode m is computed as follows:

$$P_{work}(t_s, t_e, m) = \frac{\exp\left(U(t_s, t_e, m) + c_t tt_{work}^{m}\right)}{\sum_m \sum_{t'_e > t_s'} \sum_{t_s'} \exp\left(U_a^i(t'_s, t'_e) + c_t tt_{work}^{m}\right)} \quad (7.3)$$

The time allocation probability is used to estimate the distribution of departure times of the complete population. The proposed framework is used to estimate all work-related trips inside a study area. This includes trips for which the activity at destination is work (6a) and the trips starting after the work activity (6b) which results in 2 trips for each worker (6c):

$$T_{\rightarrow work}(t) = D_{work} \sum_{m} \sum_{t_e > t_s} P_{work}(t_s = t + tt_{work}^{m}, t_e, m) \quad (7.4)$$

$$T_{work\rightarrow}(t) = D_{work} \sum_{m} \sum_{t_s} P_{work}(t_s, t_e = t, m) \quad (7.5)$$

$$T_{work}(t) = T_{\rightarrow work}(t) + T_{work\rightarrow}(t) \quad (7.6)$$

The score S_k at iteration k is reflecting both the plausibility with respect to the defined prior (i.e. the a priori knowledge available on parameter's values) and with respect to the observed data.

$$S_k = \sum_{\Theta} \log\left(prior(\Theta_k)\right) + \mathcal{L}_k(T) + \mathcal{L}_k(D) \quad (7.7)$$

The sampled values for the marginal utility functions are used in order to generate trip distributions during the day with respect with choice probability (equation (7.3)) and total demand D_{work}. This output of trip generation model is compared to observed daily demand. In this application to mode choice estimation, the likelihood is made of two elements: $\mathcal{L}_k(T)$ relates to the observed total generated demand $T_{work}(t)$ for each time interval Δt calculated in equation (7.6) and $\mathcal{L}_k(D)$ relates to the known distribution of travelled distances for working purpose in the study area, regardless of the mode.

VII.2.2. <u>Case study</u>

To test the methodology, we use the same database used in the previous chapters, obtained from a multiday travel survey collected in the province of Ghent in 2008, including multiple users, days, and tour types (Castaigne 2009). All trips to and from work are considered for this analysis. Since only substantial travel time differences will be influential, time resolution of 5 mins intervals for the observed demand is sufficient.

Modes are grouped into the following categories, based on their travel time distribution: motorized modes, train, urban public transport, and soft modes (Figure VII. 2a). For each of these modes, we estimated the travelled distances directly from survey data (for all kind of trip purposes) and fitted this distribution as starting point for the estimation. One of the goals of the estimation is to estimate those distributions for the work activities.

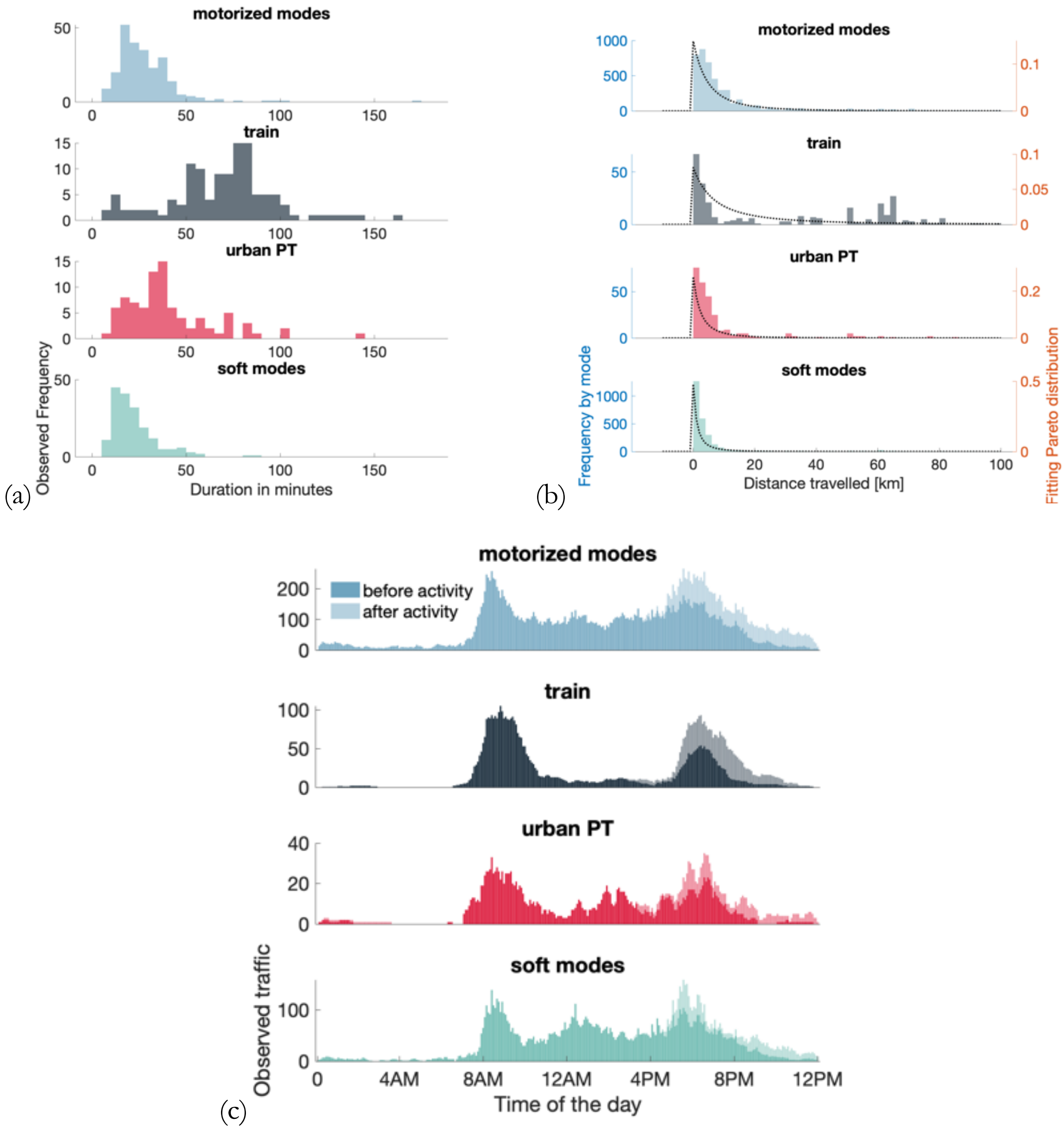
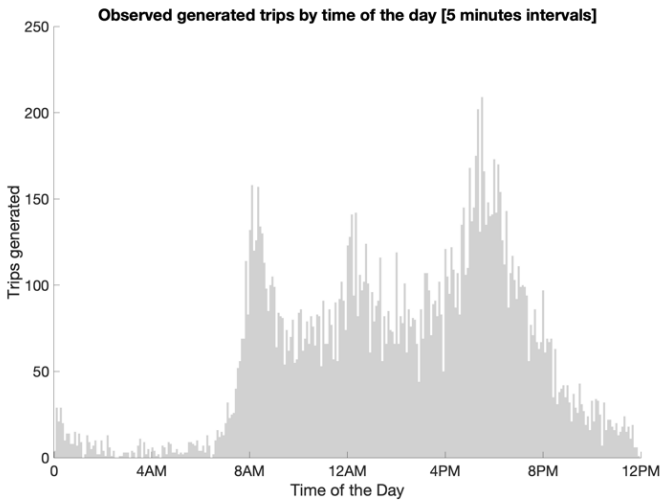


Figure VII. 2 Observed (by mode) (a) trip duration frequency (b) travelled distance distribution (c) traffic by time of the day (5 minutes interval)

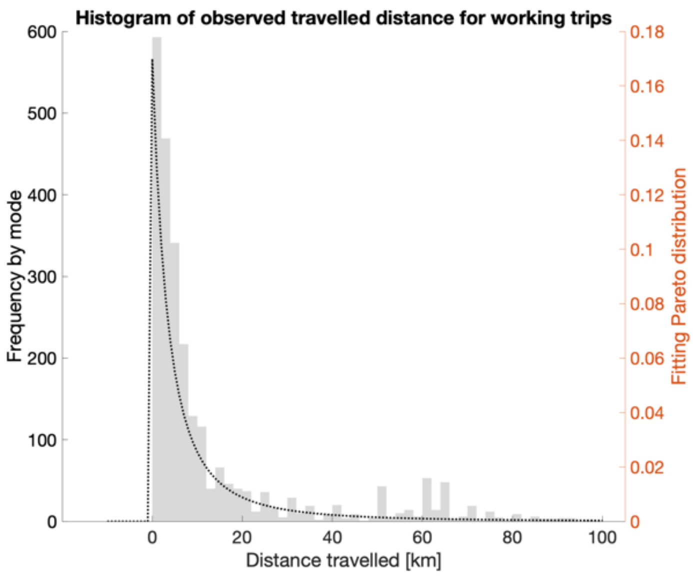
The input of the MCMC are

- The generated total demand by time of the day
- the total travelled distance distribution, and,
- the modal speed by time of the day (Figure VII. 2).

The first two are used for assessing the estimation quality and calibrate the parameters while the last is used inside the derived estimation of the travel time distribution. To estimate realistic travel times with respect to the observed traffic, we calculate a truncated average of observed travel speed using survey answers for each time period. Despite the large number of observations, there was not much data for every mode in every time interval. Missing data was added via linear interpolation of neighbouring data.



(a)



(b)

Figure VII. 3 Input: (a) work-related generated demand by time of the day (b) Total travelled distance distribution

To reproduce daily dynamics, particularly the lunch-time peak, work activity has been separated into three sub-components: $w_{j=1}$ is "work in the morning", $w_{j=2}$ is "work in the afternoon" and $w_{j=3}$ is "full-day work". Each component has distinct marginal utilities for all three components of the estimation w_j^-, w_j, w_j^+.

The parameter τ_n is fixed to $\tau = 0$, meaning that all marginal utilities are fully clock-based. For each of these three activities, 13 parameters are thus estimated in the MCMC, corresponding to the parameters the marginal utility function and the proportion of the whole demand corresponding to the given work-type. Two additional parameters are estimated for each mode in order to evaluate the distribution of travelled distance with a Pareto form:

- $\theta_{\text{utility}}^{(n=1:3)} = (U_n^{max}, \alpha_n, \beta_n, \tau_n, \gamma_n)$ for all work-activity subtype j
  with $n = 1\ is\ w^-, n = 2\ \ is\ work$ and $n = 3\ is\ w^+$
- $\theta_{\text{share}}^{(j=1:3)} = P(w_j)$ with $\sum_{j=1:3} P(w_j) = 1$
  with $j = 1\ work\ in\ the\ morning, j = 2\ work\ in\ the\ afternoon$ and $j = 3\ full - day\ work$
- $\theta_{distance}^{(m=1:4)} = (scale_m, shape_m)$

$$\boldsymbol{\Theta} = \sum_j (\theta^j + \sum_n \theta^{n,j}) + \sum_m \theta^m \tag{7.8}$$

In total $\boldsymbol{\Theta}$ contains 48 elements to estimate. The total number of trips to be distributed is considered as fixed and known, since they represent the expected number of users working daily in a certain area, which is an information often available. Preliminary analysis showed that the number of successive trips performed by the same mode is higher than 80% and that this number is even higher for owned resources such as bike and car (Scheffer, Connors, and Viti 2021). For this reason and to accelerate the estimation procedure, we assume in the proposed case study that consecutive trips are done using the same mode.

VII.2.2.1. Results

The estimation procedure has been run for 15.000 iterations. The resulting final demand profile is rather accurate with respect to observed generated demand, with $r^2 = 0{,}82$. The main result of the estimation is the posterior distribution of estimated parameters. To estimate final marginal utilities, a burn-in period of 200 iterations is selected and all remaining values of the posteriors are used to compute the average of each parameter. Figure VII. 4 shows the three estimated marginal utility functions. They are the primary component for the estimation of departure time distributions. The estimation of these three components is not correlated because the proportion is estimated independently. For this reason, the difference in the values of U^{max} is relevant within a sub-activity only.

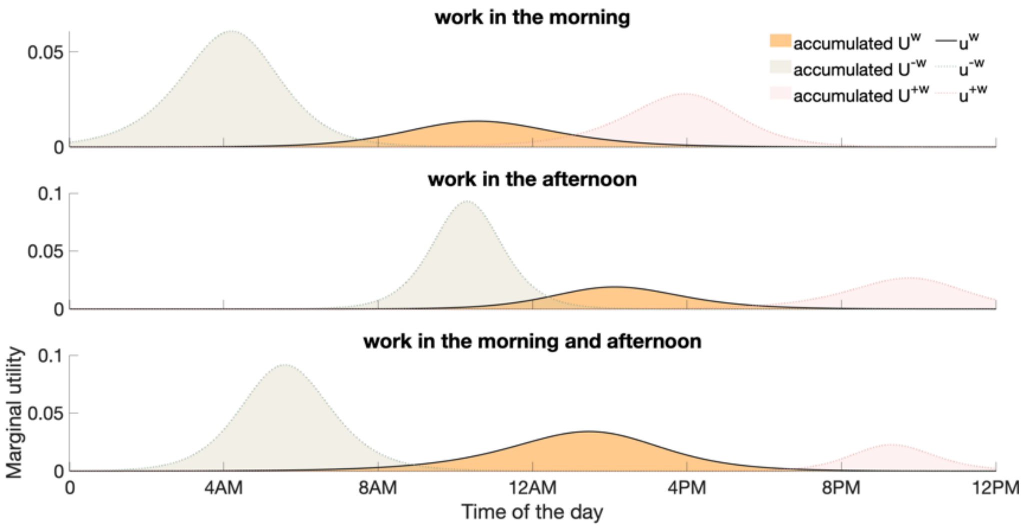
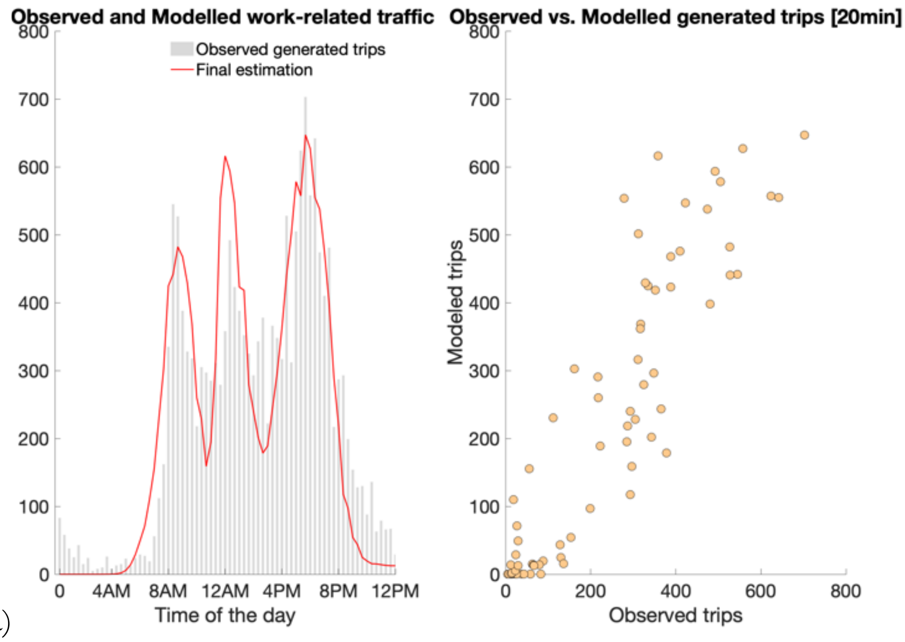
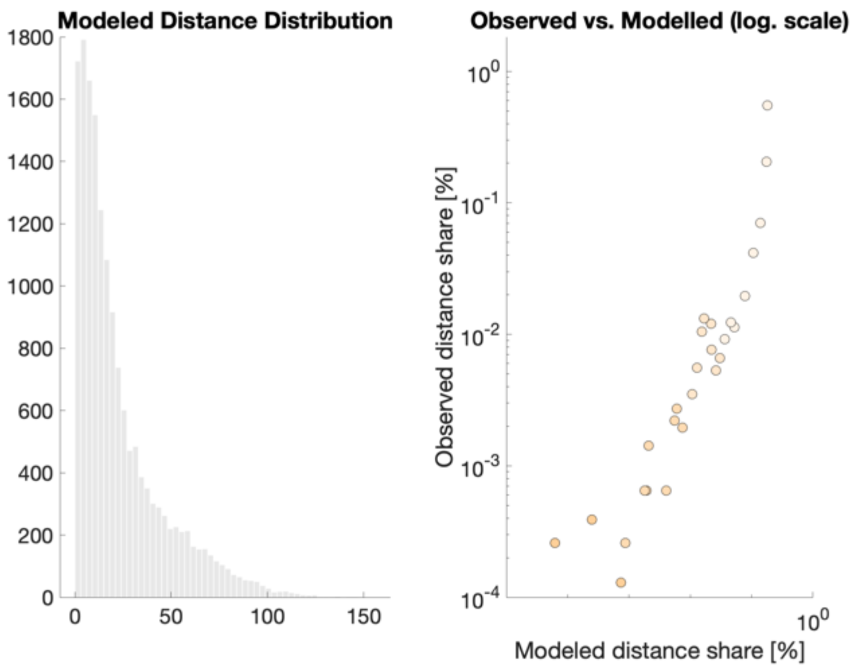


Figure VII. 4 Estimated marginal utility functions of the three components

Figure VII. 5a shows the final estimation for each time of the day. The separation of the work activity into three components makes it possible to better reproduce the mobility dynamics, with the three observed peaks which are well captured. The other derived output is the total travelled distance. In order to compare distance distributions, a number of distances values corresponding to the number of mode-specific trips is sampled from the estimated corresponding distribution. Figure VII. 5b shows the frequency of all modes together and the comparison of share for each interval of 5km (the intensity of the colour represents the length of the trip). Travelled distances are overall overestimated compared to the observed data. The relatively poorer fitting with respect to the demand is because the number of data points to fit and the order of magnitude is smaller, and also because the correlation between the estimated parameters and the output of each simulation is lower.



(a)



(b)

Figure VII. 5 Estimated (a) traffic for each time of the day (b) travelled distance distribution

Including a mode component in the choice set results in a dynamic modal split estimation (Figure VII. 6a), which shows the ability of the model to produce such output even without resorting to advanced cost functions but only based on the positive component of the accumulated utility and a probability distribution of travelled distances. The estimation was done for a 5-minute interval, but the modal split results are shown for a 20-minute time interval. This avoids skewing the output with missing or outlying data, in terms of modal speed for example. The mode-specific departure time profiles (Figure VII. 6b) indicate a good representation of large-scale temporal dynamics. For example, around mid-day, the peak is more visible for soft modes and car users and almost no train users appear. The estimated work-related trips can be compared to Figure VII. 3 that represent the real values for all kind of trip purpose.

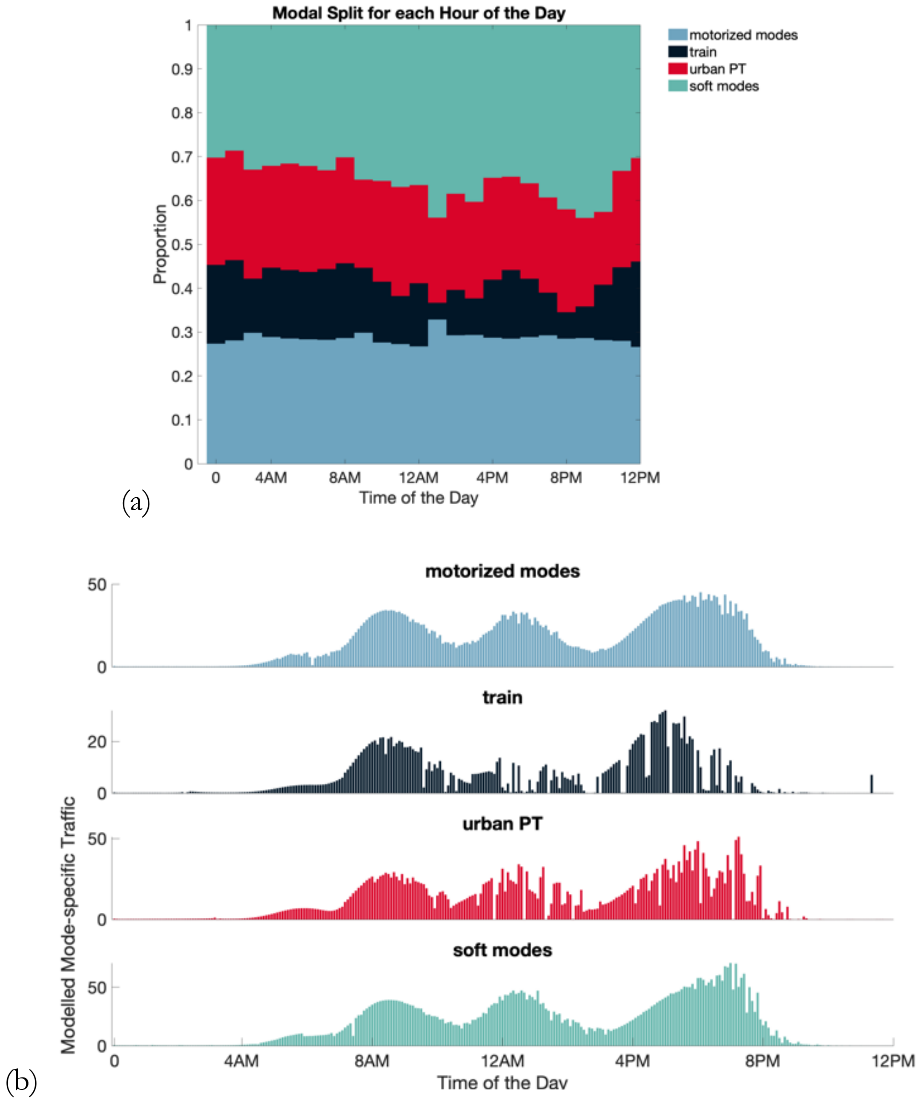


Figure VII. 6 Estimated (a) dynamic modal split (b) mode specific demand profiles

However, the scale is slightly misrepresented because the total number of mode-specific users is not known or controlled in any manner, and there is no other component than the travel time, and the travelled distance distribution is not dependent on the time of the day. The peak of soft modes around 12AM can normally be explained by shorter trips associated to the shopping or eating activity. In general, there is an overrepresentation of modes for which trip duration is typically short, such as soft modes.

VII.3. Method II

This correlation between mode choice and activity types can lead to temporal variations as well in an indirect manner. For the second section of modal choice modelling, the approach is different for two main reasons. The first reason is related to a practical issue of computation time. As soon as the choice model incorporates a destination for which the marginal utility function as a function of time varies from one area to another the computation time increases considerably. The second reason is related to the conclusions derived from the empirical analyses. By including modal choice and consequently a calibration of the travel cost function, the number of parameters increases, and we propose a simplification of the estimation procedure.

VII.3.1. Rationale

In order to apply the observations linking activity type to trip-characteristics, the objective in this section is to develop a cost function that reflects the variation in perception and disutility based on activity-chain characteristics, in a simple way. The two most important conclusions we draw from the database analysis are the following:

- The disutility associated with a trip is perceived differently depending on the activity at the destination and,
- Private transport modes such as bicycle or car have a high chance of being used in consecutive trips (Figure VII. 7).

The first argument is built into the model within the trip cost function while the second allows for a penalty within the conservation function used in the likelihood formula, less constraining than in the previous section where all successive trips were performed with the same mode.

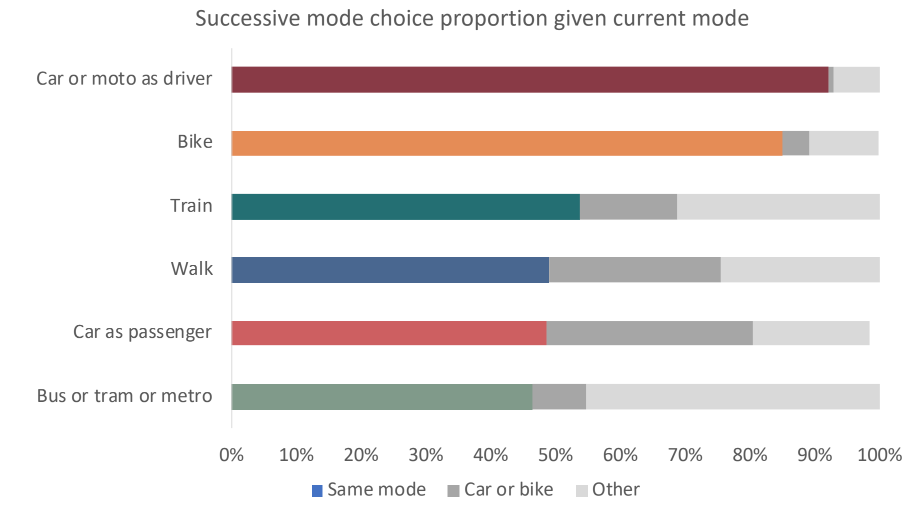


Figure VII. 7 Successive mode proportion with respect to current mode

On this figure, we see that car drivers and bike users have the tendency (at more than 85%) to use again this mode of transport in their trip. While this was used in the previous method as argument to constrain the usage of the same mode for successive trips, we believe that it could be used as a softer indicator of the likelihood of modelled trips.

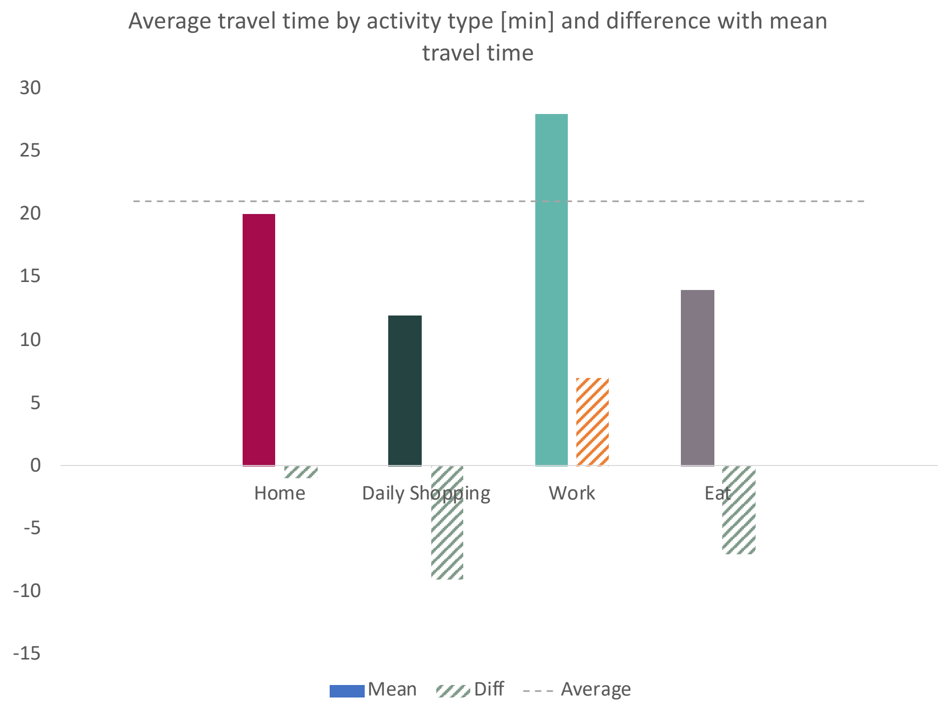


Figure VII. 8 Average travel time by activity type

Figure VII. 7 shows the average travel time by activity type. This suggests that a traveller does not have the same willingness to travel for any kind of activity. In particular, the work activity is the only one which exceeds the observed average travel time. For daily shopping, the average travel time is more than 5 minutes lower than for another kind of trip. This can be due to many different reasons, for example the choice set for destination of this kind of activity or the level of importance of performing this activity or not. For both these examples, the impact can be translated in the utility maximization formulation we are using and so it justifies a particular form of the disutility based on the activity at destination.

VII.3.2. Methodology

In this section, the calibration process used is the same as in the previous chapters, however, the reference parameters and indicators are separated into two categories. On the one hand, we have those that concern generation (i.e. almost all parameters of the utility function) and on the other hand those that concern distribution and modal choice. This distinction is made in order to increase the computational capacity and to use a two-step estimation.

VII.3.2.1. Disutility of travelling:

In order to estimate the modal split and distribution of modelled trips generated by time of the day, we focus in this section on the difference in trip cost. Because the estimation is done for an aggregate group, the selected function is simple in its form and contains the following characteristics:

- It varies with respect to the chosen mode of transport,
- It varies with respect to the origin destination zone,
- It can consider the accessibility of different modes of transport which can vary by time of the day and zones
- Depending on the activity type, it can be perceived smaller or longer than the actual travel time until a certain threshold

$$c_{tt}(t_s) = p_a tt_{yz}^2 + c_m^{yz}(t_s) \qquad (7.9)$$

with $p_a < 1$

as in the previous chapter, we select a quadratic function of the typical travel time between two zones (6.2), multiplied by an activity-specific parameter p_a. The component $c_m^{yz}(t_s)$ can be interpreted as an access cost by mode of transport, which could depend on different factors in order to reflect their level of service by time of the day and OD.

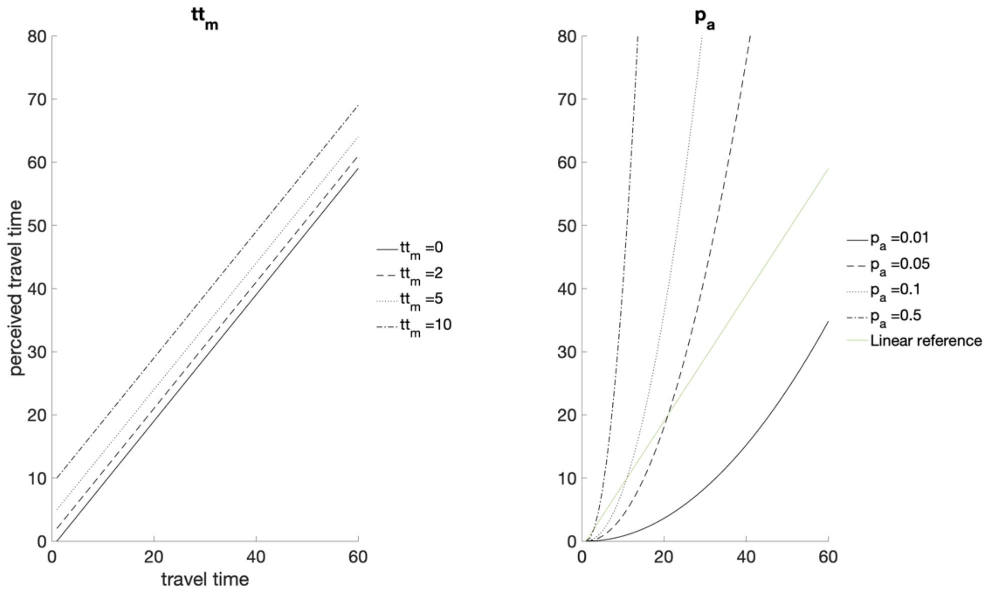


Figure VII. 9 Components of the travelling cost

VII.3.2.2. Estimation process

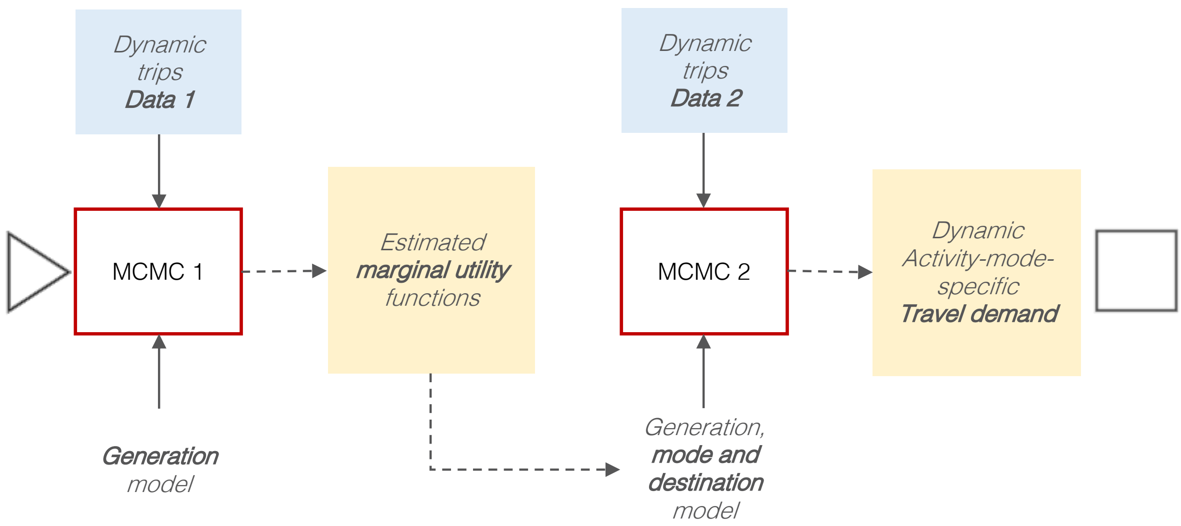


Figure VII. 10 Estimation process

In Figure VII. 10 we indicate the two phases of the estimation: *MCMC1* and *MCMC2*. The destination choice model, when including $v_{a,z}$ requires calculating the possible marginal utility at each time of the day for each different destination zone at every iteration. For this reason, the time needed for completing an iteration is more than nz times longer than the otherwise. nz is number of zones in the study area and so this issue becomes bigger with larger number of zones.

A possibility to overcome this issue is to simplify this indicator and define it based on some zonal characteristics in order to make it common to groups of zones. However, given the relative independence of some parameters on certain indicators, we decided here to maintain the $\upsilon_{a,z}$ parameter and estimate the other marginal utility parameters over a first phase. In the light of previous chapters focusing on the generation part, it is reasonable to consider the parameters which are common to all the modelled zones to be estimated with enough accuracy during the generation process only, i.e. $\alpha_a, \gamma_a, \tau_a, U_a^{max}, \beta_a$. The output if the *MCMC1* is used in order to be included in the choice model estimated in the *MCMC2*. The expected value for the 60 calibrated parameters is fixed, given the estimated prior and used to calculate the marginal utilities of every utility-primitives.

The number of parameters to be calibrated is lower but there are more likelihood components, however each of these components contains less elements to be fitted.

VII.3.3. Case study

In order to apply this methodology to a realistic case study, we consider the city of Ghent and the same database used in the Method I. However, we include this time the following activity types:

- Home
- Work
- Shopping & other mandatory activities
- Leisure & other secondary activities

And modes have been grouped in the same four categories:

- Private motorized mode
- Soft
- Urban public transport
- Train

The zoning is the same as the one used in the destination choice model for which we have the modal split shown in Figure VII. 11.

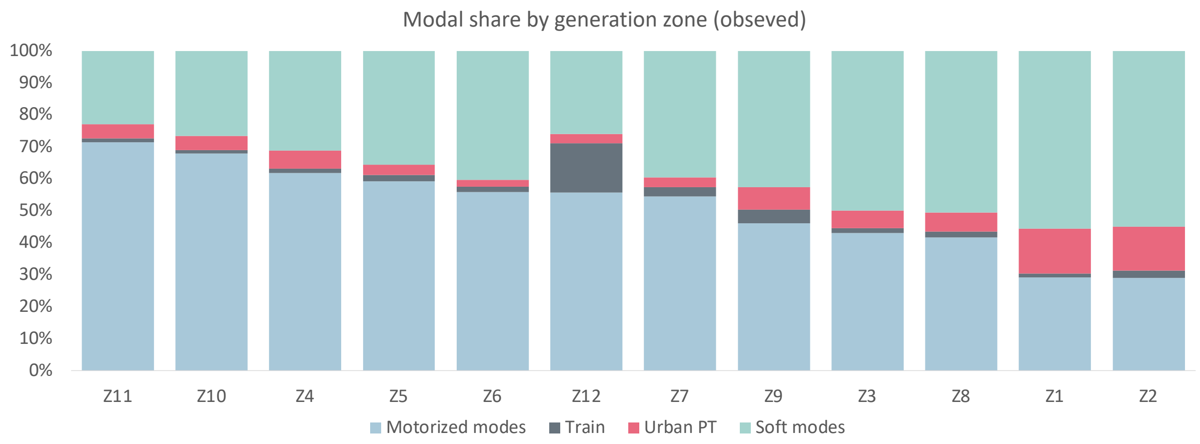


Figure VII. 11 Modal split by zone of generation

In this case study, an assumption is made in terms of travelling disutility: $c_m^{yz}(t_s)$ becomes c_m^y. This means that we do not include a variable access cost by time of the day neither do we include the impact of destination on that cost. This is due to the data availability for this case study. In this case, the parameter is calibrated because no information is available on the level of service of the different zones and time of the day, in particular in year 2007. If more data is available, this value can be averaged and fixed. In the case the data can be approximated, this approximation can be used as a prior for the calibration of more specific parameters.

The two calibration models used are the same, but the parameters are different:

Table VII. 12 Properties of the two MCMCs

	MCMC1	**MCMC2**
Choice model	$P_a^y(t_s, t_e)$	$P_a^{yz}(t_s, t_e, m)$
Likelihood components	$T^{z\rightarrow}(t)$	$T^{z\rightarrow}(t)$ $\widehat{T'^{z\rightarrow}}(t)$ D_a^z M^y
Types of calibrated parameters	$\alpha_a, \gamma_a, \tau_a, U_a^{max}, \beta_a$	$\upsilon_{a,z}, p_a^{tt}, c_m^y$
Fixed parameters	tt_a^{yz}	$\alpha_a, \gamma_a, \tau_a, U_a^{max}, \beta_a$
Number of parameters to be calibrated	60	56
Iterations	25000	10000

Computation time	$ca. 20\ hours$	$ca. 20\ hours$
Output	$T^{\mathrm{y}\to\mathrm{a}}(t)$	$T_m^{yz\to\mathrm{a}}(t)$

The likelihood function is calculated in the following manner for the two MCMC:

$$\mathcal{L}_{i,1} = \rho_{1,1} * \sum_z \sum_t -\frac{1}{2}\left(r_{1,i}(z,t)^2\right) + \rho_{2,1} \sum_z \sum_a -\frac{1}{2}\left(r_{2,i}(z,a)^2\right) + \rho_{3,1} \sum_z \sum_t -\frac{1}{2}\left(r_{3,i}(z,t)^2\right) \tag{7.10}$$

$$\mathcal{L}_{i,2} = \rho_{1,2} * \sum_z \sum_t -\frac{1}{2}\left(r_{1,i}(z,t)^2\right) + \rho_{2,2} \sum_z \sum_a -\frac{1}{2}\left(r_{2,i}(z,a)^2\right) + \rho_{3,2} \sum_z \sum_t -\frac{1}{2}\left(r_{3,i}(z,t)^2\right) + \rho_{4,2} \sum_z \sum_m -\frac{1}{2}\left(r_{4,i}(z,m)^2\right)$$

where

- $\rho_{n,L}$ is a scale factor applied to the n^{th} component of the likelihood for the $MCMC_L$;
- $r_{1,i}(z,t)$ is the difference for generated demand by zone and time of the day;
- $r_{2,i}(z,a)$ is the difference for total attracted demand by zone and by activity type;
- $r_{3,i}(z,t)$ is the conservation between attracted demand and generated demand at the zonal level.
- $r_{4,i}(z,m)$ is the different in modal usage at the zonal level.

All these indicators' unit is [trips].

VII.3.4. Results

The result section of this modelling part focuses on the second level of the MCMC estimation. Figure VII. 12 shows the overall trip generation estimation after the *MCMC2* and the comparison with the initial estimated demand at the end of *MCMC1*. With the set of parameters varying in the second estimation, we can see an improvement with respect to the first level, in particular the height of the peak which is closer to reality. This justifies the introduction of the likelihood component $r_{1,i}(z,t)$ in the score as the impact of cost-related parameters is non negligible on the time profile of generated trips. To assess the impact of the *MCMC1* output, three different marginal utility parameters values have been fixed in order to start the *MCMC2* with the same other configuration.

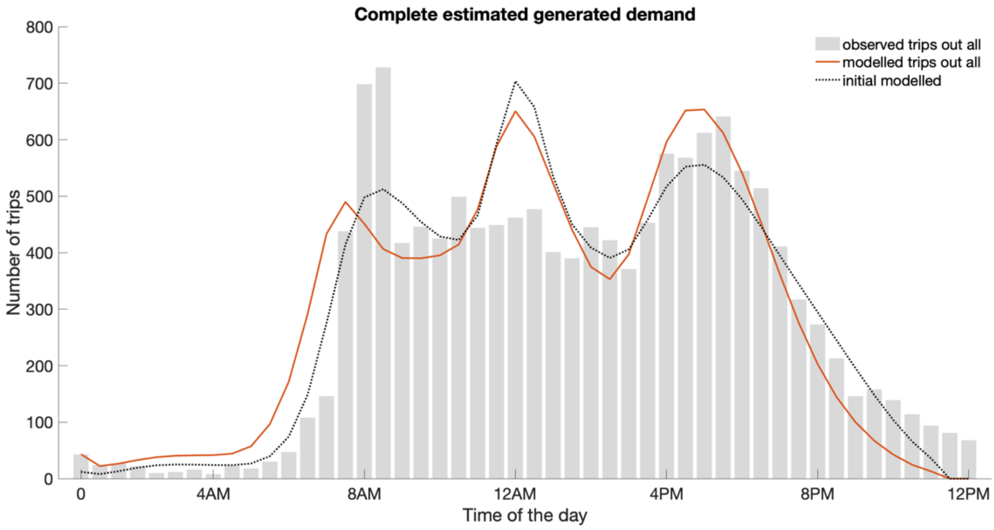


Figure VII. 12 Complete estimated demand *MCMC1* and *MCMC2*

The different score components are shown in Figure VII. 13. The generation by zone and time of the day is the most well reproduced component. This is due to the impact of *MCMC1*, to the number of points to be estimated in each component and to the ρ_n factors. The issue linked to the relative magnitude of individual value also results, as described in the previous chapters, to a weaker representation of the lowest values. This can be seen in particular for mode choice where the differences between car-related trips and other modes is very high. In this case, values representing bus and train usage are counting for less than 10% of the total number of trips.

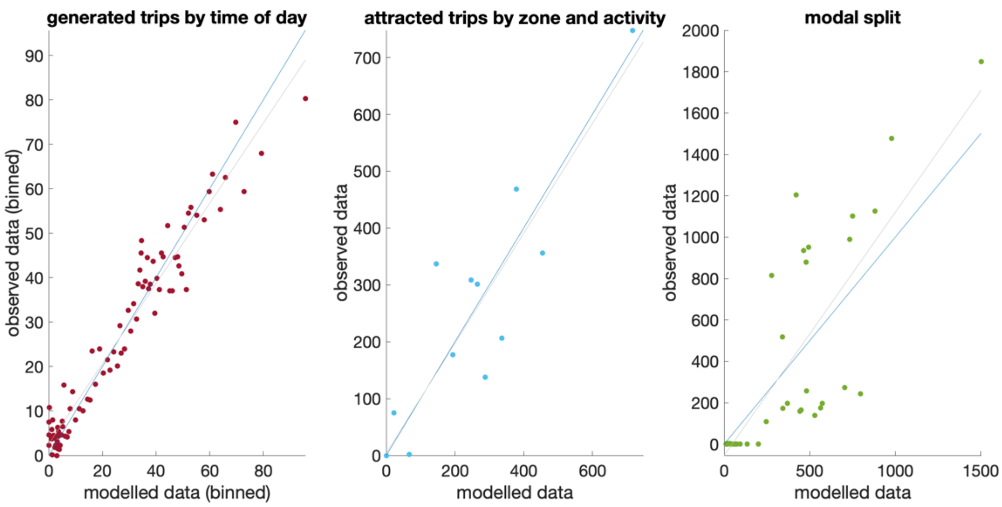


Figure VII. 13 Scatter plot of three likelihood components

On the modelling side, this results in a higher estimated value for access cost of those mode Figure VII. 14(a). The estimation of access cost seems realistic: it is fixed to 0 for cars. For soft modes the value is almost null as well (<1 minute) and higher for public transport. The highest value for train (8 minutes) reflects the density of train station which is lower than the one of bus station for which the access time is estimated at 5 minutes. In this case study, the study area covers the city of Ghent which can explain low values for this component.

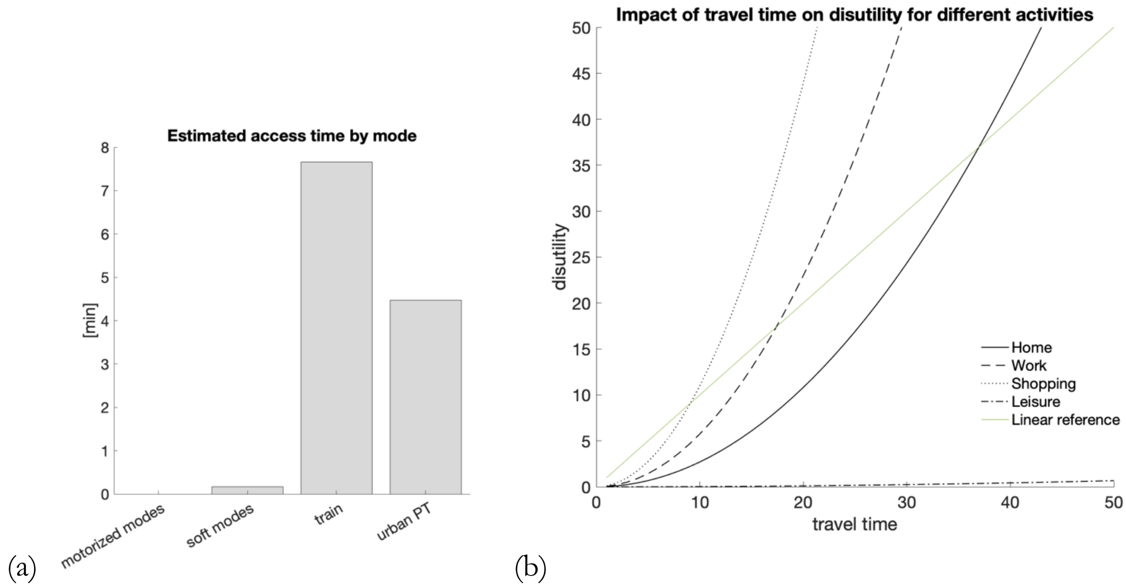


Figure VII. 14 Components of the disutility by (a) mode (b) activity

The activity-specific component of the disutility is seen on figure. While the value for leisure is overestimated, we can see for home, work and shopping that the relative order of magnitude is realistic: returning home is usually bringing the highest utility and shopping the lowest. While the starting value and prior was the same for all the activities, we can see that for work and shopping, the estimated value is in line with the observed data (Figure VII. 8). The point at which the function becomes higher than the green line is at a travel time of 9 minutes for shopping and 18 minutes for work.

These estimated function result in variations of modal split by activity type and time of the day. As forseen on figure Figure VII. 13, the estimation for train and bus use, on another level than zonal, is lower than the actual distribution and the distribution is done mostly between soft modes and car. In this case study, the travel time tt_{yz} does not depend on the mode chosen, in opposition to method 1. The impact of the mode is only included inside the access cost. For example, the cost for accessing train service is not gained by experiencing a faster journey. This results in an overestimation of soft modes for some trip type and by extension at some times of the day. For this reason, Figure VII. 15 shows only the car usage rate by time of the day and activity type.

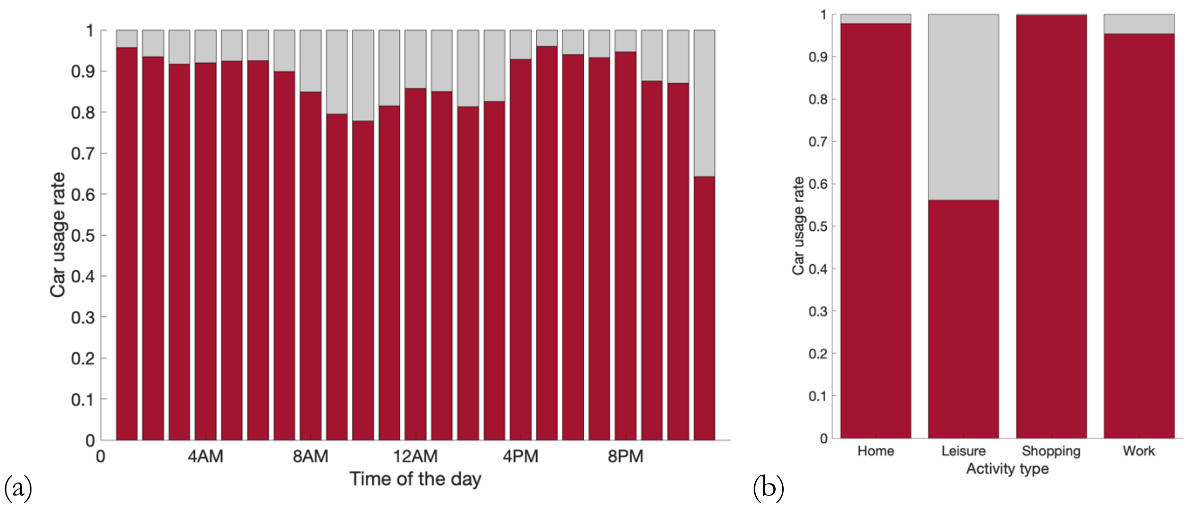


Figure VII. 15 Car usage rate by (a) time of the day (b) activity type

The final indicator of this case study concerns the duration of activities. The introduction of a more detailed travel cost has an impact on the estimation of activity durations. Compared to the durations estimated in the generation model, the model described in this section shows a better estimation for short activities like shopping, where the decrease is faster than in previous estimations. Even though the model allows to mimic activities chains continuing the next day, with continuous marginal utility functions after midnight, the duration is estimated until the end of the day only, this explains that the durations for home activity are lower than one could expect. Concerning the work activity, the two peaks are still visible and corresponding better to the actual value. In particular for the full day of work, we can observe a peak for a 9-hour activity which is in line with observed data and longer than the value estimated before. Introducing the trip disutility, perceived differently by activity, allows to model better the trade-off process and its results on activity characteristics.

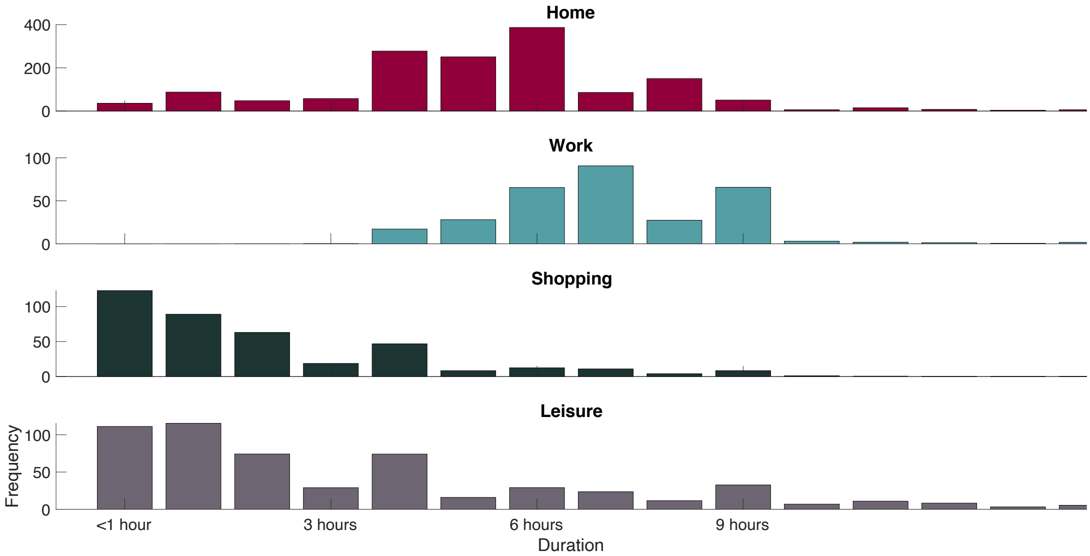


Figure VII. 16 Duration distribution by activity

VII.4. Conclusions

In this chapter we proposed a Bayesian approach to estimate utility parameters of marginal utility and travel cost functions. The aspect of mode choice is estimated using dynamic speed and estimated distance travelled such that the total accumulated utility in a day, given mode choice and trip timings results in travel demand daily distributions. The application of the proposed methodology shows a relatively good estimation, in particular considering the low input data requirements. It also underlines the possibility to combine many information sources relating to diverse aspects of travel decisions. It is feasible with the proposed model to include the possibility of using different modes however this increases the computational time significantly and lacks to describe the strong existing correlation with the proposed, simplified choice model. Both methods show strengths and weaknesses but can be applied for different purpose and with different available of information. A better overall estimation can be reached if combining those methods if a modeller is able to access both speed variations by mode and time of the day and apply a more detailed cost function. However, this can also be done using probability distributions and combining the demand model with traffic simulators. Finally, this chapter shows the possibility to carry out an MCMC in multiple levels, which can be interesting in terms of computational time and transferability of the model.

Chapter 8

Highlights of the chapter

1. 24h demand model for Luxembourg nation
2. Combination of a novel demand approach with commercial network loading models
3. Description of possible application for real network situation

In the last chapter of the thesis, an application of the method to another database and modelling environment is proposed. This shows the feasible integration and possible benefit of combining traffic modelling and the proposed demand model. The network of Luxembourg nation is modelled in a commercial software, and a 24h activity-based demand is generated for the country, based on a dataset collected in Luxembourg in 2017. The integration of this models outlines also the possible applications in terms of management and forecasting when the described demand model is applied within a real network.

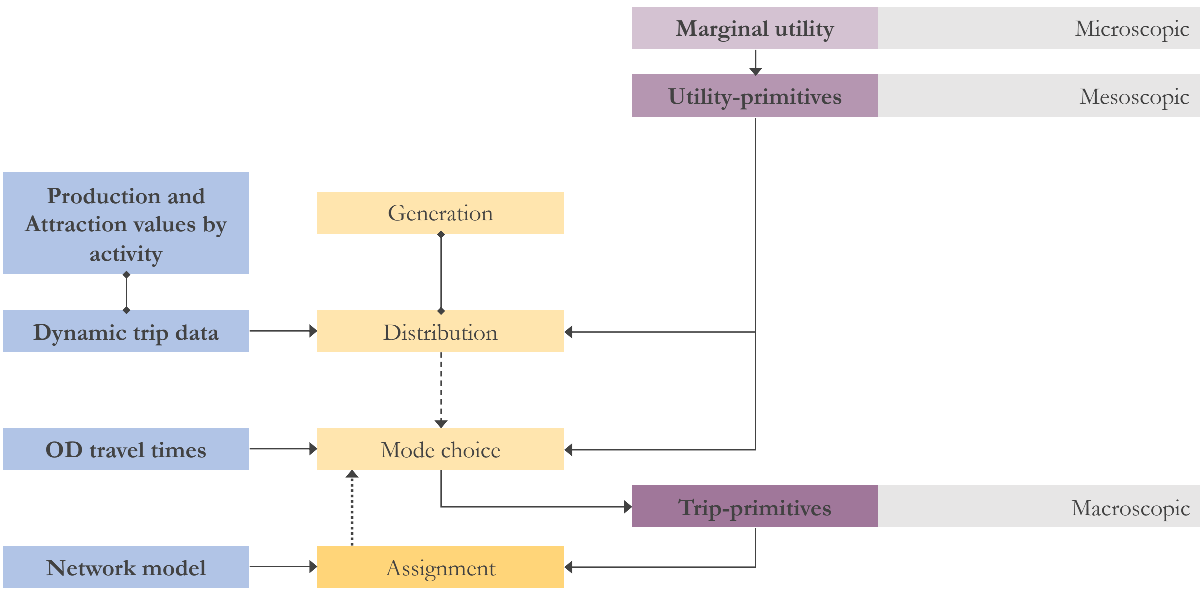


Figure VIII. 1 Thesis framework chapter 8

VIII. A CASE STUDY OF LUXEMBOURG

The implementation of the model to be applied for the country of Luxembourg needs to generate two aspects of the full model. On one side, the network of Luxembourg is modelled using the VISUM software and open-source data including the road network as well as the public transport services. On the other side, the demand model described in the previous chapters is applied to data collected in Luxembourg to be consistent with the environment in question. To do so, a survey has been used as starting point. The input data is generated with available datasets in order to have a vision of the full population and trips of the study area. This chapter, rather than a validation of the model is thought to showcase application opportunities.

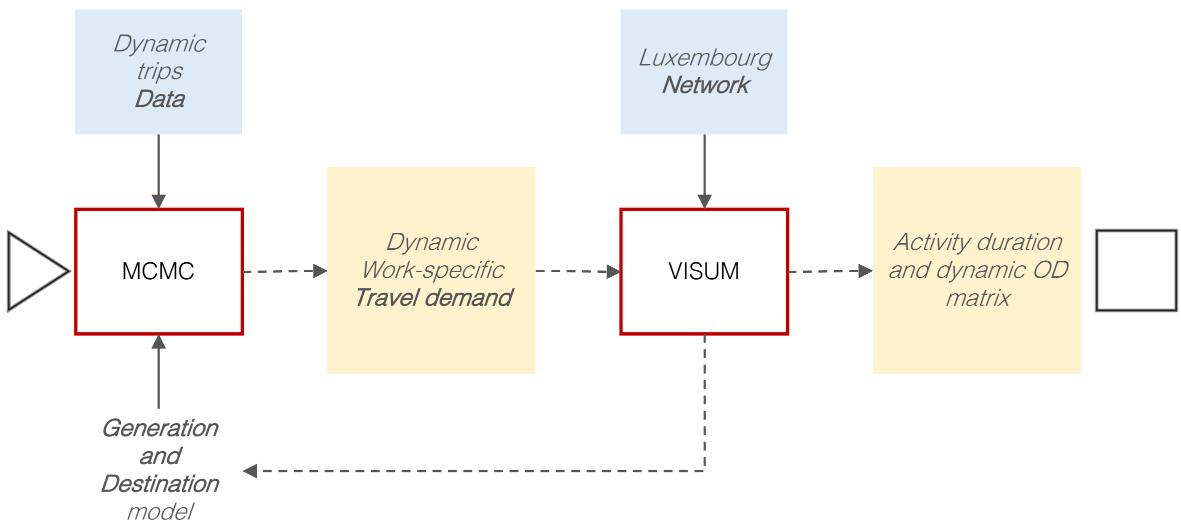


Figure VIII. 2 Estimation framework - traffic model integration

On Figure VIII. 2, the *MCMC* is connected to the VISUM model in an iterative manner. This is not applied in this way in the rest of the chapter but illustrates a possible way of integrating both systems.

VIII.1. Data

In this section, the demand model has been developed using a dataset collected in Luxembourg in 2017. The LuxMobil travel survey (Ministère du Développement durable et des Infrastructures

2017) has been conducted by the ministry of sustainable development and infrastructure in order to describe the travel characteristics of residents and cross-borders of Luxembourg and support governmental decision in terms of mobility and transport. 37.500 respondents have described a typical day and the derived trips performed. This dataset contains personal, household and trips characteristics. For more information concerning the outcome of this survey, we refer to the report published by the government as well as publications focusing on various aspects of the survey or using its results, as for example (Ma and Xie 2021; Lambotte, Jean-Marc 2021; Ministère du Développement durable et des Infrastructures 2017).

Concerning our study, variables of interest are related to the characteristics of activity and travel chains. In reference to the *chapter 1*, we describe here some aggregated indicators connected to tested hypothesis, in order to first support our conclusions and secondly to justify the usage of Ghent case study throughout this thesis, by showing divergence and similarities.

Modal split changes over the day and is statistically correlated with activity choice dynamics

The profile usage of 5 referenced modes is compared between the Luxmobil and BMW database. The Figure VIII. 3 shows these two profiles and their difference. The usage profile of these modes is strongly similar for the two dataset which confirm the typicity of different modes of transport and their usage by time of the day. In particular, the train which is mostly used for commuting.

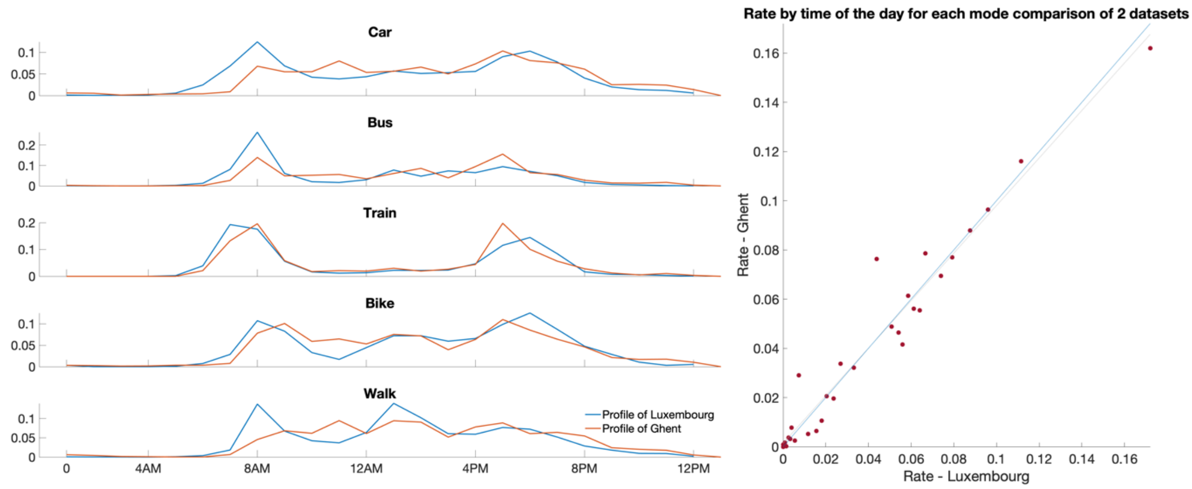


Figure VIII. 3 Daily profile of mode use of LuxMobil compared to Ghent database

In order to confirm a connection between the observed daily profiles with the activities, we calculated the actual use of a certain mode of transport for an activity, compared to the expected utilization rate, as explained for Ghent for Figure III. 8.

This shows the same over- and under-representation of certain modes for given activities. We see for example that walking is a dominant mode for going to eat and soft modes in general for personal business. Train is mostly used for commuting while other modes (car, bus) have a more regular type of usage, they are the most "basic" modes of transport.

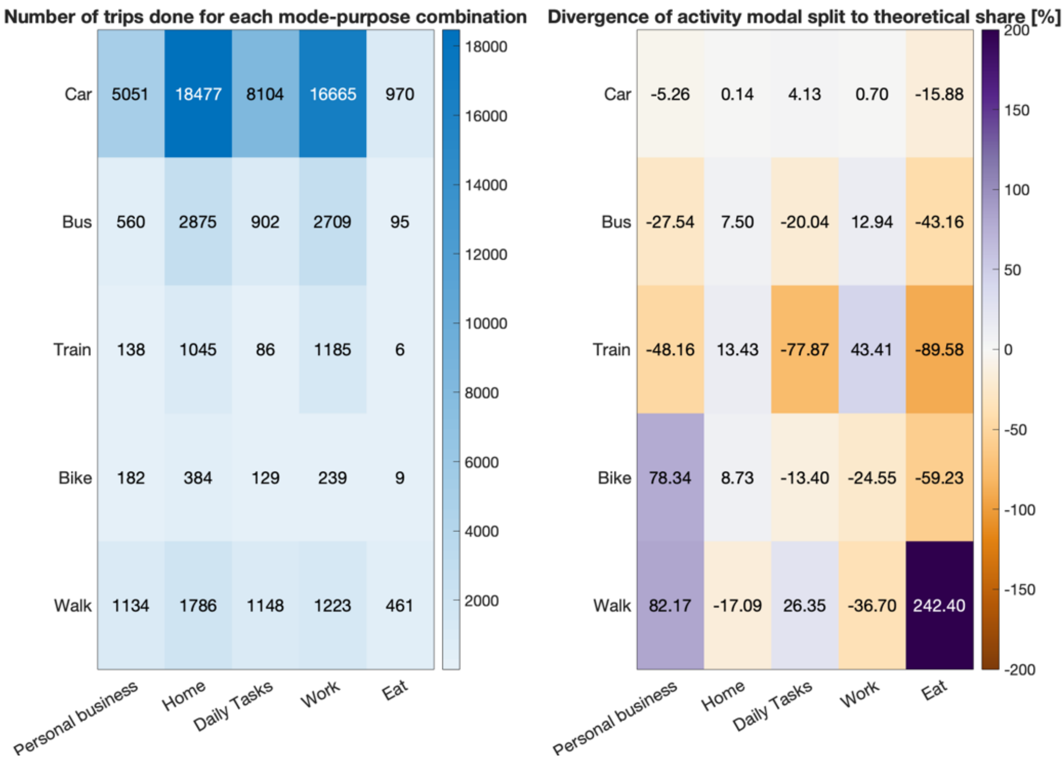


Figure VIII. 4 Correlation mode choices

The mode chosen for a trip strongly depends on the mode chosen at an earlier time of the day and owned resources such as bike and car increase this dependency

For this second hypothesis, we reproduced also the matrix generated for Ghent, showing the transition from one mode to the other. The comparison reference here is Figure III. 9.

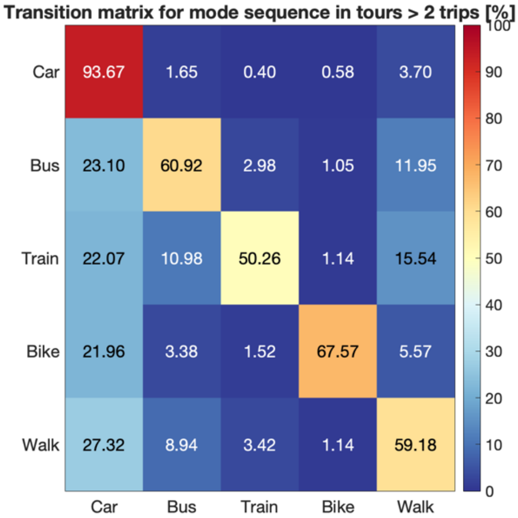


Figure VIII. 5 Transition Matrix modes

As in Ghent, we can see that the probability of using twice the same mode is particularly high for cars. The value for bike is here the second highest but is lower than inside the BMW. Two reasons can be mentioned to explain this, first of all usage rate of bike is overall lower in Luxembourg and is less used than in Ghent. This can be explained by the fact the whole country is included and not only the city centre but it can also be explained by for cultural reasons. This means that we believe in Luxembourg, bike is used more for leisure than for commute for example. Furthermore, the city provides a service of bike sharing (which was not the case in Ghent in 2007). The usage of bike sharing service does not bring the same constraint as using its own vehicle for successive trips. Unfortunately, the dataset does not differentiate bikes provided by the Vel'oh service and private ones.

Hypothesis concerning AS of single users is not handled in this section as we analysed only emerging behaviours at the population level. The data collected concerns a large set of individuals, but they described their travel behaviour only for one day. The estimation of their individual activity space from collected data cannot be done with only this level of information.

VIII.2. Network

Luxembourg, officially the Grand Duchy of Luxembourg, is a country situated in western Europe bordering France, Belgium, and Germany. Main center of gravity of the Great Region, Luxembourg attracts a high number cross-border workers (STATEC 2021). The Grand Duchy has a multimodal connection system consisting of several motorways (Figure VIII. 6), railways and bus lines (Figure VIII. 7) and a recently built tram line.

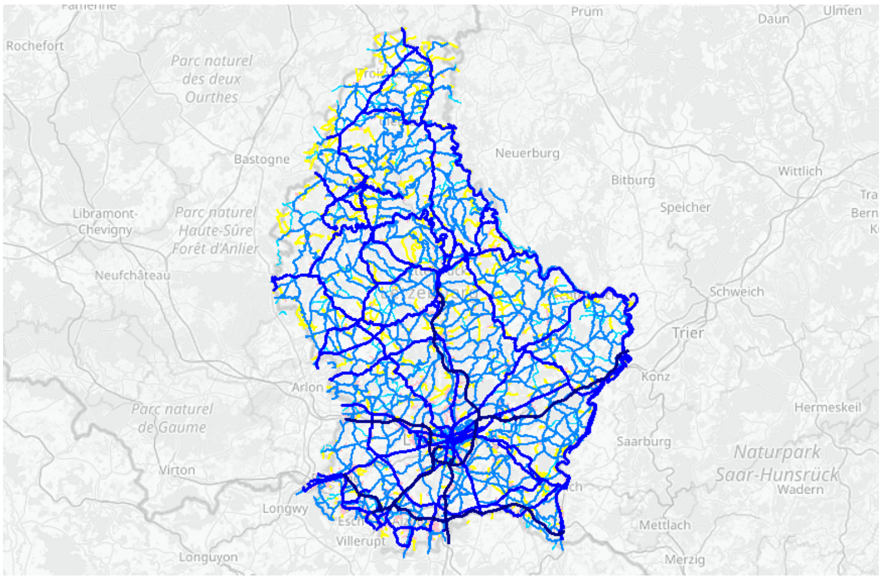


Figure VIII. 6 Private Network of Luxembourg

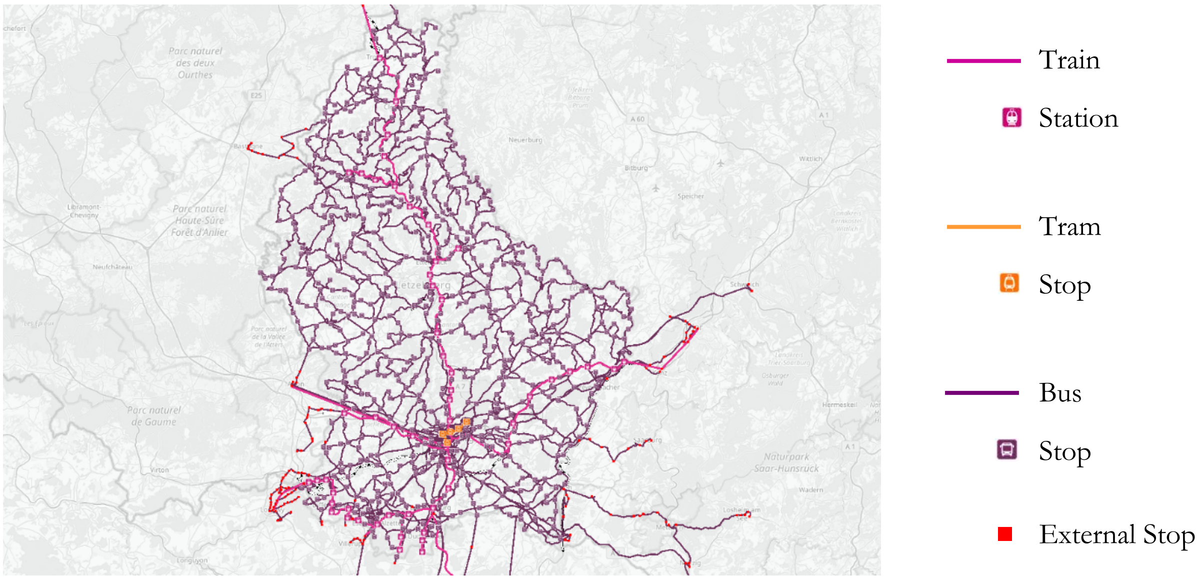


Figure VIII. 7 Public Network of Luxembourg

Following the zoning created ad hoc as part of the Luxmobil investigation, the area in question was divided into 153 zones of which 147 internal and 6 external. While the inland zones broadly represent the municipal boundaries and sometimes their sub-divisions, the outer areas represent the three neighboring countries of France, Belgium and Germany (Figure VIII. 8).

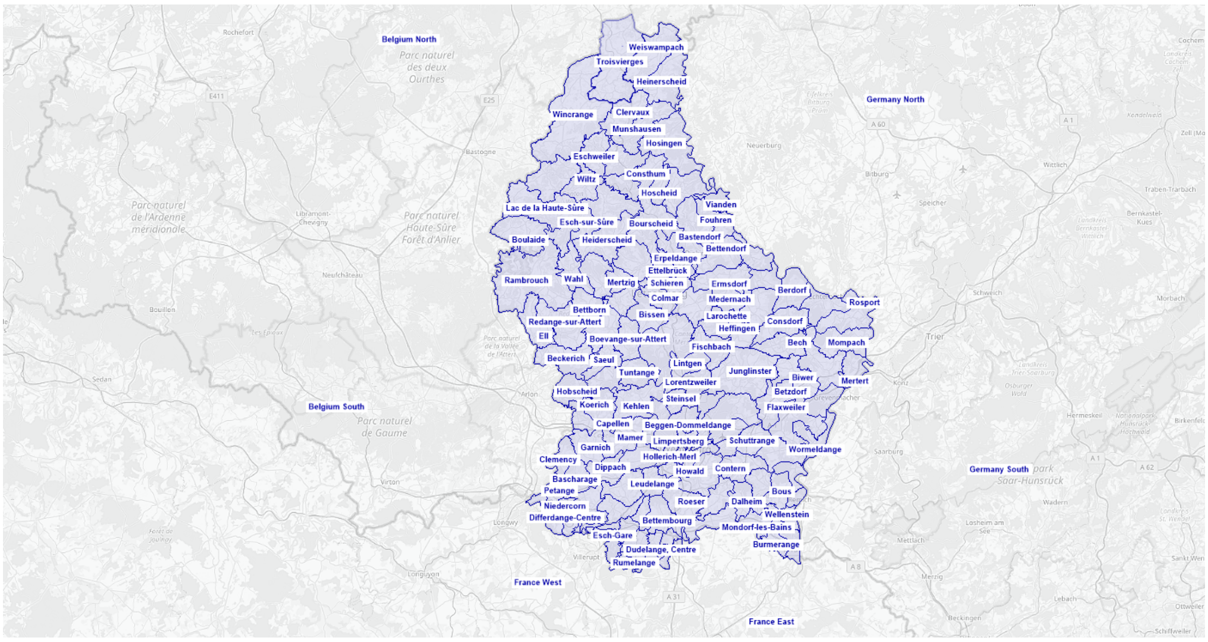


Figure VIII. 8 Zoning

VIII.3. Demand model

In this section, we used the Luxmobil dataset in order to generate the data needed for the calibration model described in the previous chapters. The input used for calculation is the total number of generated and attracted trips by zone. From the survey's answers, each zone is given a percentage for each activity type. This factor is multiplied with the total number of trips in order to have the values in terms of trip number.

The other series of data needed in the model is the reference data used for estimating the likelihood. This is dynamic trip data which in this case contains number of trips generated by zone and time of the day. To create these profiles, the data from the survey has been scaled up using the number of employments by zone and the population, for the internal zones of Luxembourg country. For the cross-border trips, the total number of workers living abroad and working in Luxembourg has been used. An example of the profile for three kind of zones is shown on Figure VIII. 9.

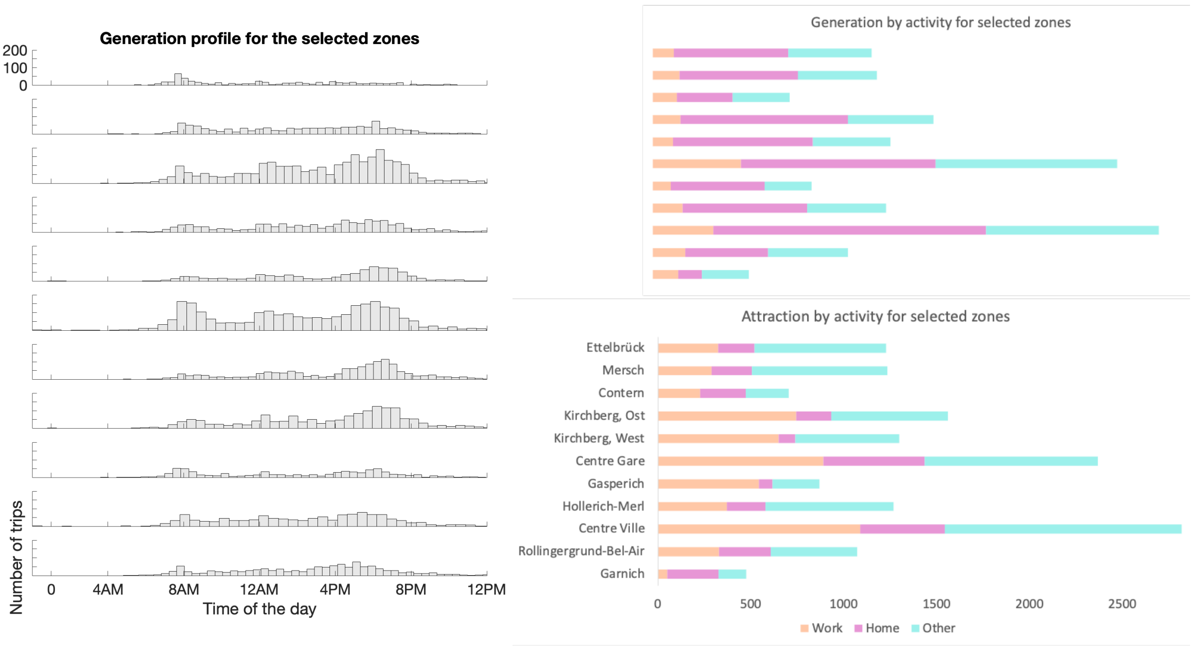


Figure VIII. 9 Input data MCMC - LuxMobil

The other aspect that we extracted from the Luxmobil dataset are the desirable output values of the *MCMC*. The profiles by activity at the aggregated level, which we try to reproduce are the following:

Figure VIII. 10 Primitives from the observed data

The OD matrix derived from the survey is shown on Figure VIII. 11. Even though the number of answers does not allow to compare the *MCMC* estimation with it, it gives insights on the most important OD pairs and trip generators/ attractors. Black cells are showing a value of 0. This also underlines the complexity of using the Luxmobil dataset in a straightforward scale up estimation process.

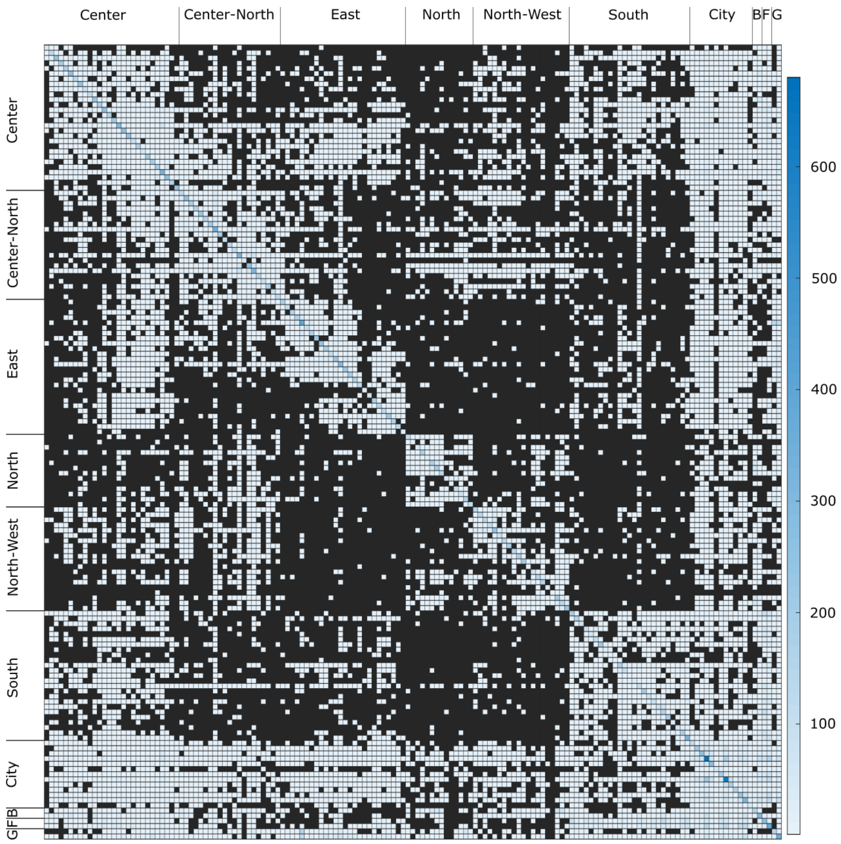


Figure VIII. 11 OD matrix from database

VIII.4. Applications opportunities

VIII.4.1. <u>Output of the model</u>

The first kind of application opportunity of the model, as exemplified with this application to Luxembourg, relates to the direct output of the *MCMC* calibration. First, the results described in this section suggest that the *MCMC* model is applicable, even though it has been developed using synthetic data and the BMW database, to other real setting. On Figure VIII. 12, we can see that the MCMC gives a good results for the three indicators, but does not reach the level of precision of generated trips shown in the previous chapters. This may be due to the fact that the travel time has not been taken into consideration but only the destination choice model presented in chapter V. Another factor explaining the level of precision is that the model has not been tuned to be applied for Luxembourg and for example we can see on Figure VIII. 10 that the work-related profiles does not have two strong peaks like observed in Ghent. This suggests that the model can be applied but that it needs some specifications related to the area.

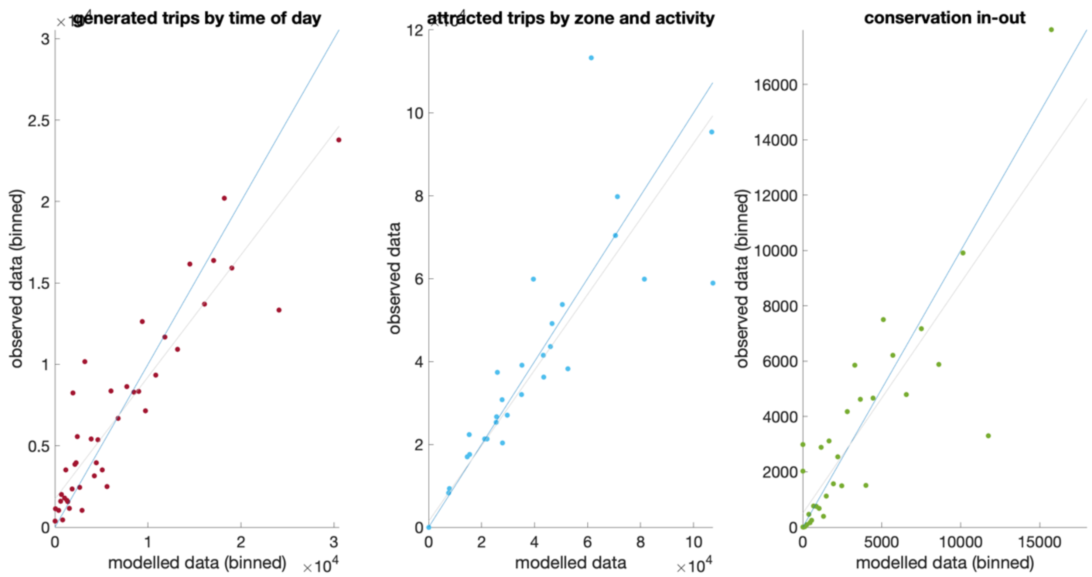


Figure VIII. 12 Scatter plot of likelihood components

An interesting output of the Luxembourg application is the comparison between the city centre of Luxembourg and other zones of the country. On Figure VIII. 13, we can see their demand profiles. It can be seen that the generated trips outside peak hours are very low for other areas than Luxembourg city and that the major demand results at the morning peak hour.

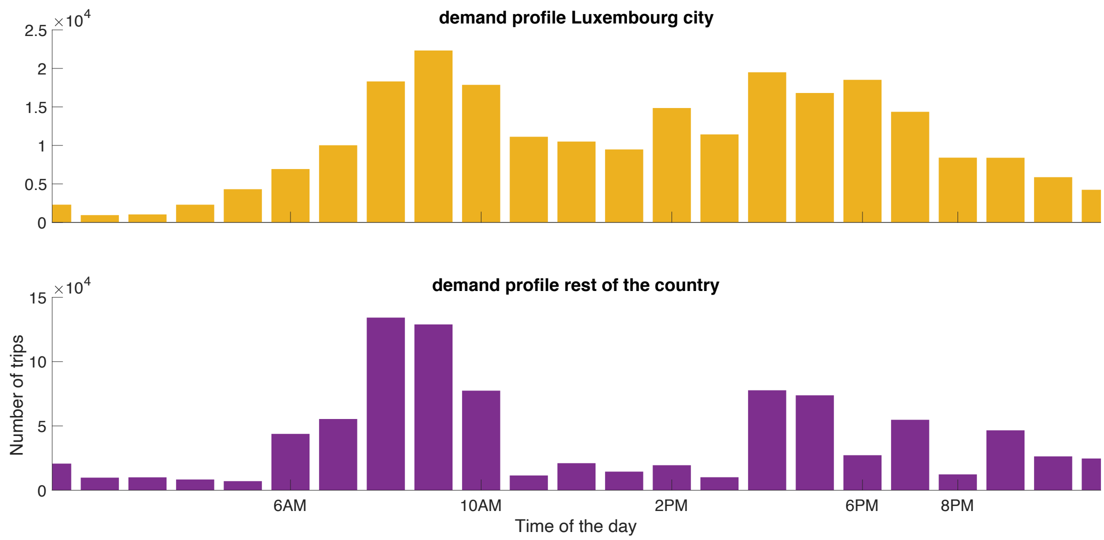


Figure VIII. 13 Demand profile Luxembourg city and the rest of the country by time of the day

This can be explained when observing trips attracted towards these two zones by time of the day (Figure VIII. 14).

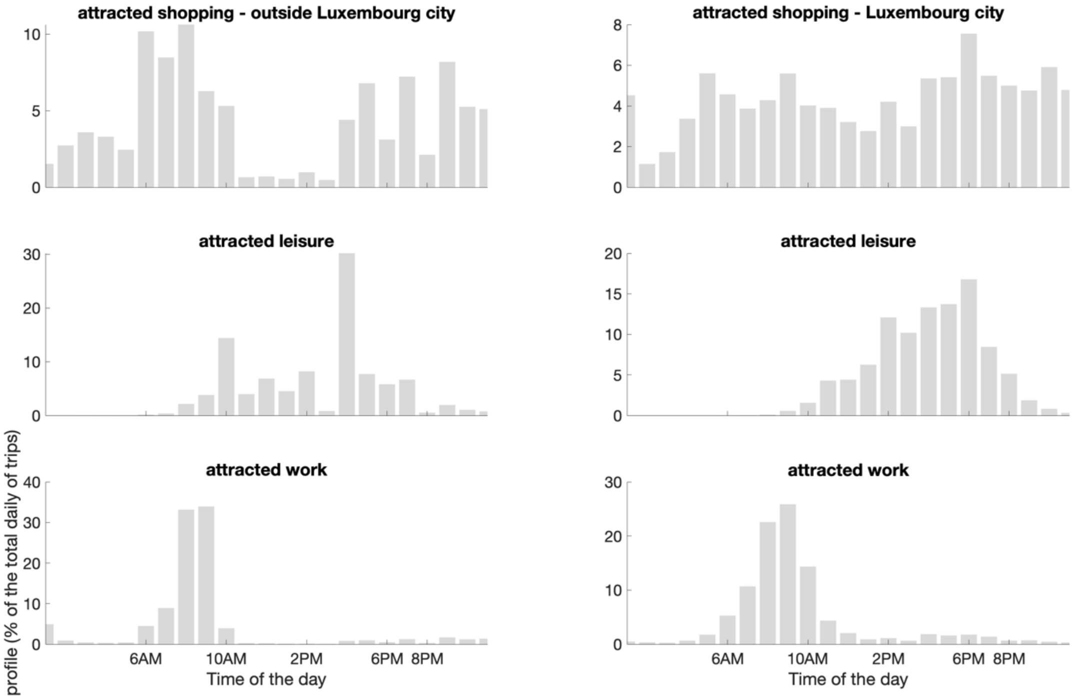


Figure VIII. 14 Profile of attracted trips by time of the day for two types of zones

It is interesting to see that the work profile does not differ much between the two kind of zones while shopping is uniform inside Luxembourg city and outside working hours for villages around the capital. For leisure, we also have a realistic output: the peak for leisure activities outside Luxembourg city is after the end of the working period while it is increasing with time and lasting until later for the city. It is important to note here that the marginal utilities related to shopping and leisure do not depend in any manner on the time of the day in their functional form.

These preliminary results show that the model in its simple form can be applied to Luxembourg and that introducing a more detailed estimation of travel time we can obtain a good estimation of other indicators as presented in the rest of the thesis. For example, the occupancy by zone and the duration of activities by zone and time of the day can be used to model the dynamic demand for parking or charging stations.

VIII.4.2. Network assignment

A second possible utilization of the model's output is its integration within a macroscopic traffic simulator. The model produces matrices at any level of detail, higher than the modelled time intervals. With the approach presented in this thesis, the reduction of the time bin's size does not increase the number of parameters to be calibrated. It only increases the computational time due to the resolution of calculation to be done. On the contrary commonly used methods to compute seed matrices usually calibrate individually values at each time interval, which raises the complexity and increases the potential errors. Potential errors also lie in the fact that the correlation between successive time-specific matrices is sometimes hard to be defined and integrated. In this model, each matrix results from a continuous in time demand profile which ensures stability and consistency.

A series of 24 matrices (one for each hours of the day) has been produced for the country of Luxembourg with the zoning defined in section VIII.2. The estimation's results show an underestimation of cross-borders demand. This is due to the fact that this demand has a very particular profile. It is reasonable to believe the utility of these transport users is different from the ones living in Luxembourg and that a large part of their trips is done outside of the study area. However, it is very easy to estimate, outside of the MCMC trips related to this section of travellers. On figure we can see the network state in terms of $\frac{volume}{capacity}$ for two different hours of the day, using the matrices generated by the *MCMC* and cross-border trips as input to a user equilibrium assignment.

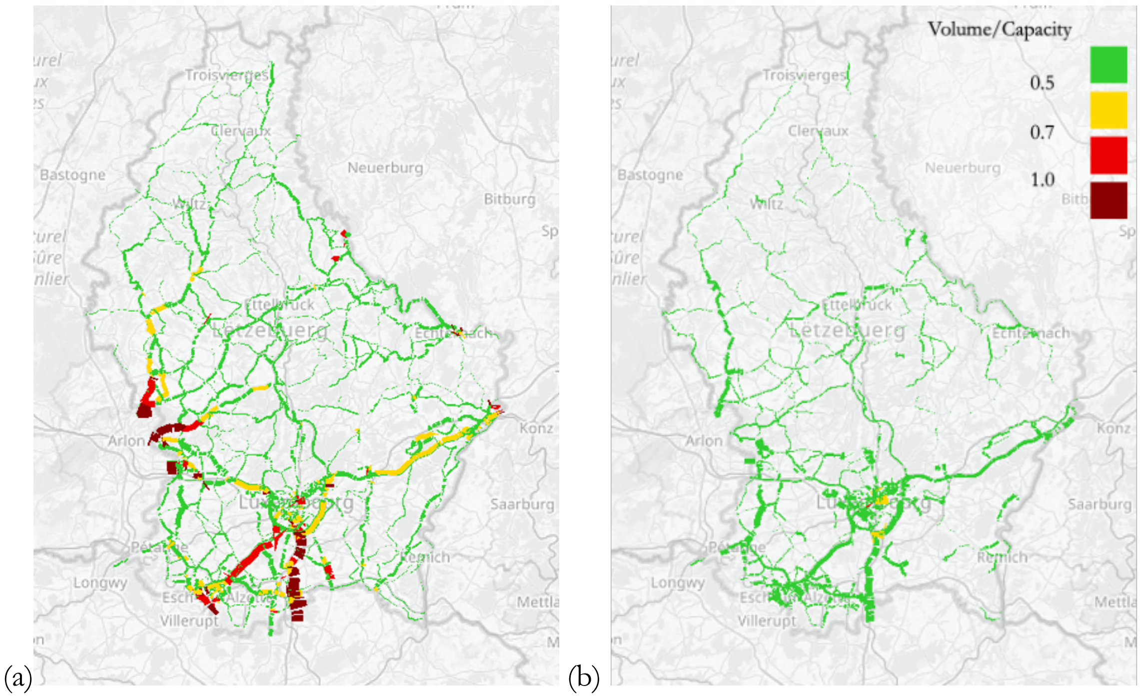
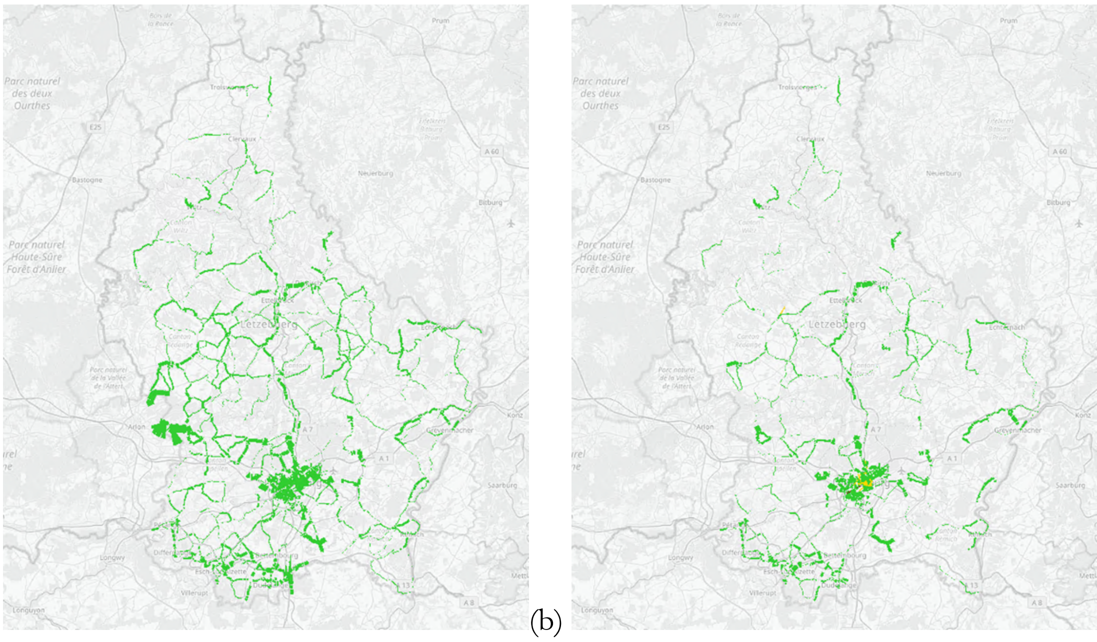


Figure VIII. 15 State of Luxembourg network at (a) 7-8AM (b) 1-2PM

Another advantage of the MCMC is that the matrices can be generated by activity type. For example, if we take the same time interval as on Figure VIII. 15 but without work-related trips, we obtain the results depicted on Figure VIII. 16.



(a) (b)

Figure VIII. 16 State of Luxembourg network non-work trips (a) 7-8AM (b) 1-2PM

Comparing Figure VIII. 16 with Figure VIII. 15 we can see that in both cases trips are reduced between the morning period and afternoon period but that this difference is stronger difference if we consider work trips. In the case of non-work-related trips, the reduction is smaller which reflects a more homogeneous demand through the day. The distribution of non-work-related trips suggest that those trips contribute only slightly to the congestion in the morning but has consequences on the one appearing in the city centre at lunch time.

These results, apart from showing the generation of a consistent seed matrix for a 24h calibration, shows potential for application for different days of the week and off-peak hours for example. Such information can be valuable for certain policy or planning strategies that can be modelled either within the MCMC if they impact the input values used for the estimation, such as Q_a^y or D_a^z or within with the network model, like pricing strategies. In such cases, an iterative process could benefit from the modelling of the supply side integrated inside the *MCMC* (dis)utility functions variables and parameters. In an iterative setting, the MCMC can also include in its scoring function OD matrices produced over a first phase and (re-)calibrated through their application in the traffic simulator. This would allow to include a higher level of information to better estimate a set of parameters linked both to utility and trip costs.

VIII.4.3. Combination with other models

The final application opportunity of the demand model and the output described above is its combination with other models. As described in section IV.2, the calibration process itself can be applied to any kind of generation model, and section VII.3 also described the possibility of including different models and formulation in a given stage of the MCMC procedure. A particular illustration of this would be the inclusion of probabilistic ellipses described in section III.3.2 within the destination choice part. For example, using the two-level approach, we could estimate over a first phase the distribution of home-work related trips. For those trips, data is usually the easiest to collect at a macroscopic scale, (e.g. employment and number of workers by zone). Secondary activities can be estimated using the aggregate activity space, knowing information related to the estimated home-work trips. The combination of the MCMC with other models can also refer to the form of the MCMC itself. As described in the combination with VISUM, the MCMC can be used with different forms of components for the score calculation and likelihood estimation. This means for example, that if generated trips by time of the day are not available, other kind of dynamic information can replace those in the calibration phase, such as dynamic OD matrices or even the number of people present in a zone or doing a given activity in that zone, by time of the day (e.g. using a by-product of *google popular times*).

Yet, even in the current application context, it can be used as a pure calibration method for any kind of marginal utility formulations and trip cost functions.

Finally, given the general output of the MCMC can help to obtain a very precise input for models that can benefit from modal usage and their variations by time of the day and zones such as emissions model. More detailed output like the posterior of utility parameters can be used for generating agents with their own individual preferences taken from the obtained distributions.

IX. CONCLUSION

The purpose of this thesis was to present a novel approach to demand choice estimation which would have the flexibility to be applied at different modelling levels with a focus on macroscopic activity-travel modelling. We also remind the focus on mode choice aspect, present in the main research question studied in this thesis:

"Is it possible to capture spatial and temporal distribution of activity- and mode-specific flows over a day from aggregate data?"

IX.1. Answer to the research questions:

The overall goal of this model in the perspective of this thesis is to generate mode-specific flows at a macroscopic level. Travel flows are characterized by their spatial and temporal dynamics for which we want to ensure consistency with individual behaviour. To do so, and starting from the common activity-based paradigm, we introduced the concepts of "trip-primitives" and "utility-primitives" to apply activity-specific choice models to the different milestones of a traditional demand model. These concepts are also tools used to translate methods used in individual travel modelling to heterogeneous groups of people.

The research question has been also broken drown in four ancillary problems which have been answered through the manuscript.

RQ1: Can we quantify how much commuting mode choice has an impact on other activities or trips?

The first research question has been addressed in the second and third chapters of this thesis. First, the review of the literature shows that the mode choice in general, is modelled with a higher and adequate level of detail in an activity-based approach.

The review of the literature also showed that the flows are modelled accurately with a trip based macroscopic approach. At this level, spatial and temporal dynamics are the essential emerging behaviour to be observed. In order to select more in detail, the characteristics of the tour-based approach which can result in variations at this level, chapter three used a multiday travel survey for a set of analysis. These permitted to make a selection of possible input and desirable outputs of a tour-based approach for explaining mode-related flows. The validated hypotheses encourage the usage of an activity-based approach as mode choice varies depending on activity chain characteristics which can be detected at both spatial and temporal level, due to inherent correlation between choices. Indeed, daily variations of modal split can be explained by daily variations of activity share at destination, related distance travelled and visited locations. However, some indicators and aggregate statistics allowed to detect acceptable assumptions at the macroscopic level, keeping some of the tour-based approach.

It is always delicate to use a single dataset in order to quantify relationships between a set of observed variables. In this thesis, we used multiple indicators in order to outline a qualitative relationship between characteristics of the home-work trip and the following trips. The usage of activity space at the level of groups of people sharing common characteristics is a way of quantifying the impact of commuting mode choice in a probabilistic manner. The ellipses described in chapter 3 are defined by the home and work location and can be characterized also by the used mode of transport. It is clear that the definition of the typical ellipse would vary from one environment to another, but it can be used as a proxy for delimiting the plausible activity area of destinations visited by commuters. In this way, commuting mode choice's impact on other activity trip can affect the location, mode used but also number of activities performed during the day.

RQ2: Can the trade-off governing individual activity scheduling explain emerging travel flows?

This question has been answered in the fourth chapter with the introduction of the concept of trip-primitives. Trip-primitives are the most straightforward dynamic description of activity-specific travel flows. We showed that their form can vary according to the generation model used as basis for their generation. Even in the simplest case of a probabilistic distribution, governed by a parametrized density law, the concept of trip-primitives does include underlying trade-off component as it allows to model dynamic variations through the day. However, to model this trade-off better, we also formally defined the concept of utility-primitives. This concept is a key contribution of this thesis as it allows to reproduce the classical choice mechanisms at the population level. The interpretation of utility-primitives used in this thesis allows to include heterogeneity and model variations in individual responses. We showed in chapter five, that the trade-off governing activity scheduling can explain and even allow to model emerging travel flows. "Primitives" are also a key element to the novelty proposed in this thesis, which is the definition of a multiscale approach where the balance between complexity of modelling and required data on one side and behavioural interpretation and consistency on the other side.

RQ3: Can we model mode dynamics considering explicitly their correlation with activity-travel chaining?

Chapters six and seven are the last two steps of the step-by-step approach presented through this thesis for modelling activity-mode-specific travel flows considering correlations with activity travel chaining. These chapters introduce another level of "primitives" used in the generation model, to a more complex two-step integrated model. The model shows that a utility-maximization approach can be used for modelling mode dynamics. Even though a higher level of information on travel times would be required for modelling mode choice with accuracy, we showed the potential of our approach for obtaining spatial and temporal variations of modal split which are consistent with activity chaining.

RQ4: How can an activity-based aggregated model estimate and predict the effect of disruptive policies and new services?

A tour-based multimodal model, because of its level of detail can by definition be used to observe effects of policies and infrastructure on different aspects. Because it is multimodal, the interactions between modes and how they are impacted by new services such as car-sharing can be observed at the zonal level. Emerging behaviours linked to specific activity types can also be useful in the case of large-scale developments of the study areas. The connection of successive trips has proven to be fundamental in the decision process of singular travellers. However, modelling every agent of a population can be very accurate to detect these choices but is sometimes problematic in particular in forecasting disruptive services or new infrastructure. The model proposed in this thesis, because of its formulation can be used to include many relevant parameters, for example in the travel cost function and would allow to estimate the joint distributions of modes and activity starting time for each zone.

IX.2. Main findings

The main findings and practical contributions of this thesis can be separated in three aspects. The first one is related to the empirical analysis and its conclusions while the other two are more methodological.

Empirical analysis:

The main findings of the empirical analysis are listed below, the first half of the list has been confirmed by the usage of two different datasets, while the ones related to activity space are based on the BMW dataset only:

- Modal split changes over the day

- These dynamics are statistically correlated with activity dynamics
- The mode chosen for a trip strongly depends on the mode chosen at an earlier time of the day
- Owned resources such as bike and car increase this dependency
- Activity Space varies with the most frequently mode used
- The centre of an Ellipse AS can be approximated by the centre of the home-work segment
- The orientation of an Ellipse AS can be approximated by the orientation of the home-work segment
- Ellipse AS can be adapted to groups of people in a probabilistic manner

Primitives:

The second part concerns the introduction and formal definition of *primitives* concepts. These modelling tools are essential elements to answer the modelling challenge presented in this work, i.e. using individual-based principles and apply them to estimate emerging travel behaviours. The flexibility of their usage, with different formulations and different models, makes them an important instrument for multi-level estimation mechanism. We proved that they can be applied at all the stages of the traditional four-step model and the assumptions we used allow to give them a more universal value. This is the case in terms of distributions for example.

Markov Chain Monte Carlo:

This leads us to the third contribution of the thesis, which lies in the usage of a Markov Chain Monte Carlo to estimate and calibrate parameters of the proposed generation model. Utility-based models are often complicated to be calibrated as their values are not directly declarable or detectable. Furthermore, the proposed model has the flexibility to decide the degree at which the correlations between choice are explicit or not. The adaptability of the model depends obviously on the level of available information to be included in the objective function. This ranks from aggregate zonal characteristics to dynamic OD matrices or even socio-demographic distributions.

IX.3. Future research

The different chapters described in this thesis have all potential value for future research. First of all, the proposed method can be applied for various level of input data and with different modelling elements. This means that every choice model included within the MCMC estimator can be tuned for different application purposes. This includes for example, inside the destination choice, to integrate the model outlined in chapter 3 with the Ellipses definition. Future work could focus on exploiting those implications in dynamic demand modelling and include the revealed regularities both in mode choice and location choice models. Observed outcomes on the individual AS will be included in the calibration of potential areas for secondary activities locations using the gaussian fit. Furthermore, limitations have shown to be linked to the neglect of level-of-service and land use. They can be included in the estimation of activity spaces to make them more accurate and adaptable in order to potentially apply it to planning and new services implementation. This includes the definition of a more accurate trip cost function which would allow to represent better the mode dynamics in particular. Chapter 7 also showed the possibility of decomposing the calibration of the different parameters within multiple MCMC simulations. This can be done as well by modelling commuting trip characteristics and distributions over a first phase and secondary trips at a later stage. This is made possible by the simple tour-based approach used inside the generation model that links two successive trips one to the other.

Another major advancement in the potential research would be to benefit of the probabilistic output of the MCMC to integrate inside the calibration a set of distributions. For example, the travel time distribution by mode and OD pair could be modelled in order to estimate better the mode choice dynamics described in chapter 7.

Other potential future research include the application of estimated distributions in synthetic population construction.

Finally, a more comprehensive integration of travel cost could allow to apply this model to different scenarios and be used for actual planning and management purposes.

Chapter 9 References

Abdelghany, Ahmed F., and Hani S. Mahmassani. 2003. "Temporal-Spatial Microassignment and Sequencing of Travel Demand with Activity-Trip Chains." *Transportation Research Record: Journal of the Transportation Research Board* 1831 (1): 89–97. https://doi.org/10.3141/1831-10.

Adnan, Muhammad. 2010. "Linking Macro-Level Dynamic Network Loading Models with Scheduling of Individual's Daily Activity–Travel Pattern." In *Chapters*. Edward Elgar Publishing. https://ideas.repec.org/h/elg/eechap/13831_13.html.

Antoniou, Constantinos, Jaume Barceló, Martijn Breen, Manuel Bullejos, Jordi Casas, Ernesto Cipriani, Biagio Ciuffo, et al. 2016. "Towards a Generic Benchmarking Platform for Origin–Destination Flows Estimation/Updating Algorithms: Design, Demonstration and Validation." *Transportation Research Part C: Emerging Technologies*, Advanced Network Traffic Management: From dynamic state estimation to traffic control, 66 (May): 79–98. https://doi.org/10.1016/j.trc.2015.08.009.

Antoniou, Constantinos, Carlos Lima Azevedo, Lu Lu, Francisco Pereira, and Moshe Ben-Akiva. 2015. "W-SPSA in Practice: Approximation of Weight Matrices and Calibration of Traffic Simulation Models." *Transportation Research Part C: Emerging Technologies*, Special Issue on International Symposium on Transportation and Traffic Theory, 59 (October): 129–46. https://doi.org/10.1016/j.trc.2015.04.030.

Arentze, Theo A, and Harry J. P Timmermans. 2004a. "A Learning-Based Transportation Oriented Simulation System." *Transportation Research Part B: Methodological* 38 (7): 613–33. https://doi.org/10.1016/j.trb.2002.10.001.

Arentze, Theo A, and Harry J.P Timmermans. 2004b. "A Learning-Based Transportation Oriented Simulation System." *Transportation Research Part B: Methodological* 38 (7): 613–33. https://doi.org/10.1016/j.trb.2002.10.001.

Arentze, Theo, Harry Timmermans, and Frank Hofman. 2007. "Creating Synthetic Household Populations: Problems and Approach." *Transportation Research Record: Journal of the Transportation Research Board* 2014 (December): 85–91. https://doi.org/10.3141/2014-11.

Axhausen, Kay, Andrea Zimmermann, Stefan Schönfelder, Guido Rindsfüser, and Thomas Haupt. 2002. "Observing the Rhythms of Daily Life: A Six-Week Travel Diary." *Transportation* 29 (May): 95–124. https://doi.org/10.1023/A:1014247822322.

Ballis, Haris, and Loukas Dimitriou. 2020a. "Revealing Personal Activities Schedules from Synthesizing Multi-Period Origin-Destination Matrices." *Transportation Research Part B: Methodological* 139 (September): 224–58. https://doi.org/10.1016/j.trb.2020.06.007.

———. 2020b. "Deriving Daily Activity Schedules from Dynamic, Purpose-Dependent Origin–Destination Matrices." *Transportation Research Record: Journal of the Transportation Research Board* 2674 (10): 375–86. https://doi.org/10.1177/0361198120939094.

Barthelemy, Johan, and Philippe L. Toint. 2013. "Synthetic Population Generation Without a Sample." *Transportation Science* 47 (2): 266–79. https://doi.org/10.1287/trsc.1120.0408.

Becker, Gary S. 1965. "A Theory of the Allocation of Time." *The Economic Journal* 75 (299): 493–517. https://doi.org/10.2307/2228949.

Becker, Henrik, Milos Balac, Francesco Ciari, and Kay Axhausen. 2019. "Assessing the Welfare Impacts of Shared Mobility and Mobility as a Service (MaaS)." *Transportation Research Part A: Policy and Practice* 131 (September). https://doi.org/10.1016/j.tra.2019.09.027.

Beckman, Richard J., Keith A. Baggerly, and Michael D. McKay. 1996. "Creating Synthetic Baseline Populations." *Transportation Research Part A: Policy and Practice* 30 (6): 415–29. https://doi.org/10.1016/0965-8564(96)00004-3.

Ben-Akiva, Moshe E., and John L. Bowman. 1998. "Activity Based Travel Demand Model Systems." In *Equilibrium and Advanced Transportation Modelling*, edited by Patrice Marcotte and Sang Nguyen, 27–46. Boston, MA: Springer US. https://doi.org/10.1007/978-1-4615-5757-9_2.

Bernardin, Vincent L., Frank Koppelman, and David Boyce. 2009. "Enhanced Destination Choice Models Incorporating Agglomeration Related to Trip Chaining While Controlling for Spatial Competition."

Transportation Research Record: Journal of the Transportation Research Board 2132 (1): 143–51. https://doi.org/10.3141/2132-16.

Bhat, Chandra, Ajay Govindarajan, and Vamsi Pulugurta. 1998. "Disaggregate Attraction-End Choice Modeling: Formulation and Empirical Analysis." *Transportation Research Record: Journal of the Transportation Research Board* 1645 (1): 60–68. https://doi.org/10.3141/1645-08.

Bhat, Chandra R. 1997. "Work Travel Mode Choice and Number of Non-Work Commute Stops." *Transportation Research Part B: Methodological* 31 (1): 41–54. https://doi.org/10.1016/S0191-2615(96)00016-1.

Bhat, C.R., and R. Misra. 1999. "Discretionary Activity Time Allocation of Individuals between In-Home and out-of-Home and between Weekdays and Weekends." *Transportation* 26 (2): 193–229. https://doi.org/10.1023/A:1005192230485.

Bowman, J. L, and M. E Ben-Akiva. 2001a. "Activity-Based Disaggregate Travel Demand Model System with Activity Schedules." *Transportation Research Part A: Policy and Practice* 35 (1): 1–28. https://doi.org/10.1016/S0965-8564(99)00043-9.

Bowman, J.L, and M.E Ben-Akiva. 2001b. "Activity-Based Disaggregate Travel Demand Model System with Activity Schedules." *Transportation Research Part A: Policy and Practice* 35 (1): 1–28. https://doi.org/10.1016/S0965-8564(99)00043-9.

Boyce, D E. 1986. "Integration of Supply and Demand Models in Transportation and Location: Problem Formulations and Research Questions." *Environment and Planning A: Economy and Space* 18 (4): 485–89. https://doi.org/10.1068/a180485.

Bradley, Mark, John L. Bowman, and Bruce Griesenbeck. 2010. "SACSIM: An Applied Activity-Based Model System with Fine-Level Spatial and Temporal Resolution." *Journal of Choice Modelling* 3 (1): 5–31. https://doi.org/10.1016/S1755-5345(13)70027-7.

Cantelmo, Guido, and Francesco Viti. 2018. "Incorporating Activity Duration and Scheduling Utility into Equilibrium-Based Dynamic Traffic Assignment." *Transportation Research Part B: Methodological*, August. https://doi.org/10.1016/j.trb.2018.08.006.

———. 2019. "Incorporating Activity Duration and Scheduling Utility into Equilibrium-Based Dynamic Traffic Assignment." *Transportation Research Part B: Methodological* 126 (August): 365–90. https://doi.org/10.1016/j.trb.2018.08.006.

Cantelmo, Guido, Francesco Viti, Ernesto Cipriani, and Marialisa Nigro. 2018. "A Utility-Based Dynamic Demand Estimation Model That Explicitly Accounts for Activity Scheduling and Duration." *Transportation Research Part A: Policy and Practice*, Transportation and Traffic Theory: Behavior and Planning Applications, 114 (August): 303–20. https://doi.org/10.1016/j.tra.2018.01.039.

Carlier, Kristof, Tno Inro, Stella Fiorenzo-catalano, Charles Lindveld, Piet Bovy, Kristof Carlier, Stella Fiorenzo-catalano, Charles Lindveld, and Piet Bovy. 2003. "A Supernetwork Approach towards Multimodal Travel Modeling." In *Proceedings of the 81st Transportation Research Board Annual Meeting, Washington DC.*

Cascetta, Ennio, Domenico Inaudi, and Gérald Marquis. 1993. "Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts." *Transportation Science* 27 (4): 363–73. https://doi.org/10.1287/trsc.27.4.363.

Castaigne, M. 2009. "Behaviour and Mobility during the Week 'BMW.'" Belgian Science Policy 2009 – 120 p. (Research Programme Science for a Sustainable Development).

Castaigne, M., E. Cornelis, R. Frederix, C. Tampere, P. Toint, F. Viti, and F. Walle. 2009. "Behaviour and Mobility within the Week BMW." *Final Report. Belgian Science Policy, Research Programme Science for a Sustainable Development, Brussels*, 120.

Charypar, D., and K. Nagel. 2005. "Generating Complete All-Day Activity Plans with Genetic Algorithms." *Transportation* 32 (4): 369–97. https://doi.org/10.1007/s11116-004-8287-y.

Chow, Joseph Y.J., and Adel E. Nurumbetova. 2015. "A Multi-Day Activity-Based Inventory Routing Model with Space–Time–Needs Constraints." *Transportmetrica A: Transport Science* 11 (3): 243–69. https://doi.org/10.1080/23249935.2014.958120.

Cirillo, C., and K. W. Axhausen. 2002. "MODE CHOICE OF COMPLEX TOURS." In *Publication of: Association for European Transport*. https://trid.trb.org/view/728061.

Cirillo, Cinzia, and Kay W. Axhausen. 2010. "Dynamic Model of Activity-Type Choice and Scheduling." *Transportation* 37 (1): 15–38. https://doi.org/10.1007/s11116-009-9218-8.

Clark, James E. 1982. "Modeling Travelers' Perceptions of Travel Time." *Transportation Research Record*, 5.

Currie, Graham, and Alexa Delbosc. 2011. "Exploring the Trip Chaining Behaviour of Public Transport Users in Melbourne." *Transport Policy* 18 (1): 204–10. https://doi.org/10.1016/j.tranpol.2010.08.003.

De Witte, Astrid, Joachim Hollevoet, Frédéric Dobruszkes, Michel Hubert, and Cathy Macharis. 2013. "Linking Modal Choice to Motility: A Comprehensive Review." *Transportation Research Part A: Policy and Practice* 49 (March): 329–41. https://doi.org/10.1016/j.tra.2013.01.009.

Duguay, Gerald, Woo Jung, and Daniel McFadden. 1976. *SYNSAM: A Methodology for Synthesizing Household Transportation Survey Data*. Urban Travel Demand Forecasting Project, Institute of Transportation Studies.

Ettema, Bastin, Polak, and Ashiru. 2007. "Modelling the Joint Choice of Activity Timing and Duration." *Transportation Research Part A: Policy and Practice* 41 (9 SPEC. ISS.): 827–41. https://doi.org/10.1016/j.tra.2007.03.001.

Ettema, and Timmermans. 2003. "Modeling Departure Time Choice in the Context of Activity Scheduling Behavior." *Transportation Research Record: Journal of the Transportation Research Board* 1831 (January): 39–46. https://doi.org/10.3141/1831-05.

Farooq, Bilal, Michel Bierlaire, Ricardo Hurtubia, and Gunnar Flötteröd. 2013. "Simulation Based Population Synthesis." *Transportation Research Part B: Methodological* 58 (December): 243–63. https://doi.org/10.1016/j.trb.2013.09.012.

Flötteröd, Gunnar. 2009a. "Cadyts - a Free Calibration Tool for Dynamic Traffic Simulations." In . https://infoscience.epfl.ch/record/152443.

———. 2009b. "Cadyts a Free Calibration Tool for Dynamic Traffic Simulations." Infoscience. 2009. https://infoscience.epfl.ch/record/152358.

Fosgerau, M., and L. Engelson. 2011. "The Value of Travel Time Variance." *Transportation Research Part B: Methodological* 45 (1): 1–8. https://doi.org/10.1016/j.trb.2010.06.001.

Fu, Xiao, and William H. K. Lam. 2018. "Modelling Joint Activity-Travel Pattern Scheduling Problem in Multi-Modal Transit Networks." *Transportation* 45 (1): 23–49. https://doi.org/10.1007/s11116-016-9720-8.

Gao, Wenli, Michael Balmer, and Eric Miller. 2010. "Comparison of MATSim and EMME/2 on Greater Toronto and Hamilton Area Network, Canada." *Transportation Research Record: Journal of the Transportation Research Board* 2197 (December): 118–28. https://doi.org/10.3141/2197-14.

Gargiulo, Floriana, Sônia Ternes, Sylvie Huet, and Guillaume Deffuant. 2010. "An Iterative Approach for Generating Statistically Realistic Populations of Households." *PLOS ONE* 5 (1): e8828. https://doi.org/10.1371/journal.pone.0008828.

Hägerstrand, Torsten. 1970. "What about People in Regional Science?" *Papers of the Regional Science Association* 24 (1): 6–21. https://doi.org/10.1007/BF01936872.

Halat, Hooram, Ali Zockaie, Hani S. Mahmassani, Xiang Xu, and Omer Verbas. 2016. "Dynamic Network Equilibrium for Daily Activity-Trip Chains of Heterogeneous Travelers: Application to Large-Scale Networks." *Transportation* 43 (6): 1041–59. https://doi.org/10.1007/s11116-016-9724-4.

Harding, Christopher, Zachary Patterson, and Luis Miranda-Moreno. 2013. "Activity Space Geometry and Its Effect on Mode Choice | Semantic Scholar." In . https://www.semanticscholar.org/paper/Activity-Space-Geometry-and-Its-Effect-on-Mode-Harding-Patterson/f4b0840ff973696baa798b1d5454b48bef1fa070#related-papers.

Hensher, David A., and April J. Reyes. 2000. "Trip Chaining as a Barrier to the Propensity to Use Public Transport." *Transportation* 27 (4): 341–61. https://doi.org/10.1023/A:1005246916731.

Hesam Hafezi, Mohammad, Naznin Sultana Daisy, Hugh Millward, and Lei Liu. 2021. "Framework for Development of the Scheduler for Activities, Locations, and Travel (SALT) Model." *Transportmetrica A: Transport Science*, May, 1–33. https://doi.org/10.1080/23249935.2021.1921879.

Horni, Andreas, Kai Nagel, and Kay W. Axhausen. 2011. "High-Resolution Destination Choice in Agent-Based Demand Models." Application/pdf, 23 p. https://doi.org/10.3929/ETHZ-A-006686309.

Horni, Andreas, Darren M. Scott, Michael Balmer, and Kay W. Axhausen. 2009. "Location Choice Modeling for Shopping and Leisure Activities with MATSim: Combining Micro-Simulation and Time Geography." Application/pdf, 25 p. https://doi.org/10.3929/ETHZ-A-005871421.

Jonnalagadda, Nageswar, Joel Freedman, William Davidson, and J. Hunt. 2001. "Development of Microsimulation Activity-Based Model for San Francisco: Destination and Mode Choice Models." *Transportation Research Record: Journal of the Transportation Research Board* 1777 (January): 25–35. https://doi.org/10.3141/1777-03.

Khan, Nazmul Arefin, and Muhammad Ahsanul Habib. 2021. "Modeling Activity-Based Tour Shared Travel Choices, Tour-Level Activity Participation and Time Allocation." *Transportmetrica A: Transport Science*, February, 1–38. https://doi.org/10.1080/23249935.2021.1876178.

Kitamura, Ryuichi. 1984a. "Incorporating Trip Chaining into Analysis of Destination Choice." *Transportation Research Part B: Methodological* 18 (1): 67–81. https://doi.org/10.1016/0191-2615(84)90007-9.

———. 1984b. "A Model of Daily Time Allocation to Discretionary Out-of-Home Activities and Trips." *Transportation Research Part B: Methodological* 18 (3): 255–66. https://doi.org/10.1016/0191-2615(84)90036-5.

Krygsman, Stephan, Theo Arentze, and Harry Timmermans. 2007. "Capturing Tour Mode and Activity Choice Interdependencies: A Co-Evolutionary Logit Modelling Approach." *Transportation Research Part A: Policy and Practice* 41 (10): 913–33. https://doi.org/10.1016/j.tra.2006.03.006.

Lam, William H. K., and Yafeng Yin. 2001. "An Activity-Based Time-Dependent Traffic Assignment Model." *Transportation Research Part B: Methodological* 35 (6): 549–74. https://doi.org/10.1016/S0191-2615(00)00010-2.

Lambotte, Jean-Marc. 2021. "Distribution Spatiale de l'emploi et Des Modes de Transport Des Travailleurs Actifs Au Grand-Duché de Luxembourg Analyse Des Données de l'enquête Luxmobil 2017." Application/pdf. https://doi.org/10.25353/UBTR-XXXX-78AE-A574.

Lenntorp, Bo. 1976. *Paths in Space-Time Environments: A Time-Geographic Study of Movement Possibilities of Individuals*. Royal University of Lund, Department of Geography.

Lenormand, Maxime, and Guillaume Deffuant. 2012. "Generating a Synthetic Population of Individuals in Households: Sample-Free vs Sample-Based Methods." *ArXiv:1208.6403 [Math, Stat]*, August. http://arxiv.org/abs/1208.6403.

Li, Ran, and Daoqin Tong. 2016. "Constructing Human Activity Spaces: A New Approach Incorporating Complex Urban Activity-Travel." *Journal of Transport Geography* 56 (October): 23–35. https://doi.org/10.1016/j.jtrangeo.2016.08.013.

Liao, Feixiong, Theo Arentze, and Harry Timmermans. 2013. "Incorporating Space–Time Constraints and Activity-Travel Time Profiles in a Multi-State Supernetwork Approach to Individual Activity-Travel Scheduling." *Transportation Research Part B: Methodological* 55 (September): 41–58. https://doi.org/10.1016/j.trb.2013.05.002.

Lim, Yongtaek, and Hyunmyung Kim. 2016. "A Combined Model of Trip Distribution and Route Choice Problem." *Transportmetrica A: Transport Science* 12 (8): 721–35. https://doi.org/10.1080/23249935.2016.1166171.

Lin, Dung-Ying, Naveen Eluru, S. Travis Waller, and Chandra R. Bhat. 2008. "Integration of Activity-Based Modeling and Dynamic Traffic Assignment." *Transportation Research Record: Journal of the Transportation Research Board* 2076 (1): 52–61. https://doi.org/10.3141/2076-06.

Lindveld, C. D. R. 2003. "Dynamic O-D Matrix Estimation: A Behavioural Approach." http://resolver.tudelft.nl/uuid:fbe10ed4-0c17-4d2f-a272-30ce17399174.

Liu, Peng, Feixiong Liao, Hai-Jun Huang, and Harry Timmermans. 2016. "Dynamic Activity-Travel Assignment in Multi-State Supernetworks under Transport and Location Capacity Constraints." *Transportmetrica A: Transport Science* 12 (7): 572–90. https://doi.org/10.1080/23249935.2016.1189739.

Llorca, Carlos, Sasan Amini, Ana Tsui Moreno, and Rolf Moeckel. 2019. "Validation of an Agent-Based Travel Demand Model with Floating Car Data." *Transportation Research Procedia* 37: 242–49. https://doi.org/10.1016/j.trpro.2018.12.189.

Ma, Tai-Yu, and Simin Xie. 2021. "Optimal Fast Charging Station Locations for Electric Ridesharing with Vehicle-Charging Station Assignment." *Transportation Research Part D: Transport and Environment* 90 (January): 102682. https://doi.org/10.1016/j.trd.2020.102682.

Maronna, Ricardo A., and Ruben H. Zamar. 2002. "Robust Estimates of Location and Dispersion for High-Dimensional Datasets." *Technometrics* 44 (4): 307–17.

McNally, Michael G. 2007. "The Four-Step Model." In *Handbook of Transport Modelling*, 1:35–53. Handbooks in Transport 1. Emerald Group Publishing Limited. https://doi.org/10.1108/9780857245670-003.

Metropolis, Nicholas, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. 1953. "Equation of State Calculations by Fast Computing Machines." *The Journal of Chemical Physics* 21 (6): 1087–92. https://doi.org/10.1063/1.1699114.

Miller, Eric J., Matthew J. Roorda, and Juan Antonio Carrasco. 2005. "A Tour-Based Model of Travel Mode Choice." *Transportation* 32 (4): 399–422. https://doi.org/10.1007/s11116-004-7962-3.

Ministère du Développement durable et des Infrastructures. 2017. "Enquête Luxmobil 2017." December 7, 2017. http://transports.public.lu/fr/publications/situation-actuelle/enquete-luxmobil-2017.html.

Moeckel, Rolf, Leta Huntsinger, and Rick Donnelly. 2017. "From Macro to Microscopic Trip Generation: Representing Heterogeneous Travel Behavior." *The Open Transportation Journal* 11 (1): 31–43. https://doi.org/10.2174/1874447801711010031.

Mueller, Kirill, and Kay W. Axhausen. 2011. "Hierarchical IPF: Generating a Synthetic Population for Switzerland." In . Louvain-la-Neuve: European Regional Science Association (ERSA). https://www.econstor.eu/handle/10419/119994.

Nurul Habib, Khandker M., and Eric J. Miller. 2009. "Modelling Activity Generation: A Utility-Based Model for Activity-Agenda Formation." *Transportmetrica A* 5 (1): 3–23. https://doi.org/10.1080/18128600802591384.

Ortuzar, J. de D., and Luis G. Willumsen. 2011. *Modelling Transport*. 4th ed. John Wiley & Sons, Ltd. https://www.wiley.com/en-us/Modelling+Transport%2C+4th+Edition-p-9780470760390.

Patterson, Zachary, and Steven Farber. 2015. "Potential Path Areas and Activity Spaces in Application: A Review." *Transport Reviews* 35 (6): 679–700. https://doi.org/10.1080/01441647.2015.1042944.

Patwary, A. U. Z., Wei Huang, and Hong K. Lo. 2021. "Metamodel-Based Calibration of Large-Scale Multimodal Microscopic Traffic Simulation." *Transportation Research Part C: Emerging Technologies* 124 (March): 102859. https://doi.org/10.1016/j.trc.2020.102859.

Pendyala, R. M., R. Kitamura, and D. V. G. Prasuna Reddy. 1998. "Application of an Activitybased Travel Demand Model Incorporating a Rule-Based Algorithm", Environment and Planning B." *B*, 25–753.

Perchoux, Camille, Yan Kestens, Frédérique Thomas, Andraea Van Hulst, Benoit Thierry, and Basile Chaix. 2014. "Assessing Patterns of Spatial Behavior in Health Studies: Their Socio-Demographic Determinants and Associations with Transportation Modes (the RECORD Cohort Study)." *Social Science & Medicine* 119 (October): 64–73. https://doi.org/10.1016/j.socscimed.2014.07.026.

Polak, J., P. Jones, and P. Vythoulkas. 1993. "Travellers Choice of Time of Travel under Road Pricing." 1993.

Pucher, John, and John L. Renne. 2003. "Socioeconomics of Urban Travel: Evidence from the 2001 NHTS." 2003. /paper/Socioeconomics-of-Urban-Travel%3A-Evidence-from-the-Pucher-Renne/3cd62e2331d2d65ba8318aae53fee401f6233af7.

Rai, R. K., Michael Balmer, Marcel Rieser, V. S. Vaze, Stefan Schönfelder, and Kay W. Axhausen. 2007. "Capturing Human Activity Spaces: New Geometries." *Transportation Research Record: Journal of the Transportation Research Board* 2021 (1): 70–80. https://doi.org/10.3141/2021-09.

Raux, Charles, Tai-Yu Ma, and Eric Cornelis. 2016. "Variability in Daily Activity-Travel Patterns: The Case of a One-Week Travel Diary." *European Transport Research Review* 8 (4): 26. https://doi.org/10.1007/s12544-016-0213-9.

Saadi, Ismaïl, Ahmed Mustafa, Jacques Teller, and Mario Cools. 2016. "Forecasting Travel Behavior Using Markov Chains-Based Approaches." *Transportation Research Part C: Emerging Technologies* 69 (August): 402–17. https://doi.org/10.1016/j.trc.2016.06.020.

Scheffer, Ariane, Guido Cantelmo, and Francesco Viti. 2017. "Generating Macroscopic, Purpose-Dependent Trips through Monte Carlo Sampling Techniques." *Transportation Research Procedia* 27: 585–92. https://doi.org/10.1016/j.trpro.2017.12.111.

Scheffer, Ariane, Richard Connors, and Francesco Viti. 2021. "Trip Chaining Impact on Within-Day Mode Choice Dynamics: Evidences from a Multi-Day Travel Survey." In *Transportation Research Procedia*. Elsevier.

Schönfelder, Stefan, and Kay W Axhausen. 2003. "Activity Spaces: Measures of Social Exclusion?" *Transport Policy* 10 (4): 273–86. https://doi.org/10.1016/j.tranpol.2003.07.002.

Small, Kenneth A. 1982. "The Scheduling of Consumer Activities: Work Trips." *The American Economic Review* 72 (3): 467–79.

STATEC. 2021. "Travailleurs Frontaliers Occupés Au Luxembourg Selon La Résidence et La Nationalité (En 1 000 Personnes) 1974 - 2020." 2021. https://statistiques.public.lu/stat/TableViewer/tableViewHTML.aspx?ReportId=12928&IF_Language=fra&MainTheme=2&FldrName=3&RFPath=92.

Stevens, S S. 1957. "On the Psychophysical Law," 29.

Supernak, Janusz. 1992. "Temporal Utility Profiles of Activities and Travel: Uncertainty and Decision Making." *Transportation Research Part B: Methodological* 26 (1): 60–76. https://doi.org/10.1016/0191-2615(92)90020-W.

Timmermans, Harry, Theo Arentze, and Chang-Hyeon Joh. 2002. "Analysing Space-Time Behaviour: New Approaches to Old Problems." *Progress in Human Geography* 26 (2): 175–90. https://doi.org/10.1191/0309132502ph363ra.

Toole, Jameson L., Serdar Colak, Bradley Sturt, Lauren P. Alexander, Alexandre Evsukoff, and Marta C. González. 2015. "The Path Most Traveled: Travel Demand Estimation Using Big Data Resources." *Transportation Research Part C: Emerging Technologies*, Big Data in Transportation and Traffic Engineering, 58, Part B (September): 162–77. https://doi.org/10.1016/j.trc.2015.04.022.

Tseng, Y.-Y., and E.T. Verhoef. 2008. “Value of Time by Time of Day: A Stated-Preference Study.” *Transportation Research Part B: Methodological* 42 (7–8): 607–18. https://doi.org/10.1016/j.trb.2007.12.001.

Tsoleridis, Panagiotis, Charisma F. Choudhury, and Stephane Hess. 2022. “Utilising Activity Space Concepts to Sampling of Alternatives for Mode and Destination Choice Modelling of Discretionary Activities.” *Journal of Choice Modelling* 42 (March): 100336. https://doi.org/10.1016/j.jocm.2021.100336.

Tyrinopoulos, Yannis, and Constantinos Antoniou. 2013. “Factors Affecting Modal Choice in Urban Mobility.” *European Transport Research Review* 5 (March): 27–39. https://doi.org/10.1007/s12544-012-0088-3.

Vickrey, William S. 1969. “Congestion Theory and Transport Investment.” *The American Economic Review* 59 (2): 251–60.

Voas, David, and Paul Williamson. 2000. “An Evaluation of the Combinatorial Optimisation Approach to the Creation of Synthetic Microdata.” *International Journal of Population Geography* 6 (5): 349–66. https://doi.org/10.1002/1099-1220(200009/10)6:5<349::AID-IJPG196>3.0.CO;2-5.

Voorhees, Alan M. 2013. “A General Theory of Traffic Movement: The 1955 ITE Past Presidents’ Award Paper.” *Transportation* 40 (6): 1105–16. https://doi.org/10.1007/s11116-013-9487-0.

Vovsha, Peter, and Mark Bradley. 2004. “Hybrid Discrete Choice Departure-Time and Duration Model for Scheduling Travel Tours.” *Transportation Research Record: Journal of the Transportation Research Board* 1894 (1): 46–56. https://doi.org/10.3141/1894-06.

Vovsha, Peter, James E. Hicks, Gaurav Vyas, Vladimir Livshits, Kyunghwi Jeon, Rebekah Anderson, and Gregory Giaimo. 2017. “Combinatorial Tour Mode Choice.” In . https://trid.trb.org/view/1439366.

Vrtic, M., P. Fröhlich, N. Schüssler, K. W. Axhausen, D. Lohse, C. Schiller, and H. Teichert. 2007. “Two-Dimensionally Constrained Disaggregate Trip Generation, Distribution and Mode Choice Model: Theory and Application for a Swiss National Model.” *Transportation Research Part A: Policy and Practice*, Selected Papers on Applications of Discrete Choice Models Presented at the European Regional Science Conference, Amsterdam, August 2005, 41 (9): 857–73. https://doi.org/10.1016/j.tra.2006.10.003.

Wang, James Jixian. 1996. “Timing Utility of Daily Activities and Its Impact on Travel.” *Transportation Research Part A: Policy and Practice* 30 (3): 189–206. https://doi.org/10.1016/0965-8564(95)00023-2.

Yamamoto, Toshiyuki, Satoshi Fujii, Ryuichi Kitamura, and Hiroshi Yoshida. 2000a. “Analysis of Time Allocation, Departure Time, and Route Choice Behavior Under Congestion Pricing.” *Transportation Research Record: Journal of the Transportation Research Board* 1725 (1): 95–101. https://doi.org/10.3141/1725-13.

———. 2000b. “Analysis of Time Allocation, Departure Time, and Route Choice Behavior Under Congestion Pricing.” *Transportation Research Record: Journal of the Transportation Research Board* 1725 (January): 95–101. https://doi.org/10.3141/1725-13.

Yasmin, Farhana, Catherine Morency, and Matthew J. Roorda. 2017. “Macro-, Meso-, and Micro-Level Validation of an Activity-Based Travel Demand Model.” *Transportmetrica A: Transport Science* 13 (3): 222–49. https://doi.org/10.1080/23249935.2016.1249437.

Ye, Xin, Karthik Charan Konduri, Ram M. Pendyala, Bhargava Sana, and Paul Waddell. 2009. “Methodology to Match Distributions of Both Household and Person Attributes in Generation of Synthetic Populations.” In . https://trid.trb.org/view.aspx?id=881554.

Ye, Xin, Ram M. Pendyala, and Giovanni Gottardi. 2007. “An Exploration of the Relationship between Mode Choice and Complexity of Trip Chaining Patterns.” *Transportation Research Part B: Methodological* 41 (1): 96–113. https://doi.org/10.1016/j.trb.2006.03.004.

Zhang, Xiaoning, Hai Yang, Hai-Jun Huang, and H. Michael Zhang. 2005. “Integrated Scheduling of Daily Work Activities and Morning–Evening Commutes with Bottleneck Congestion.” *Transportation Research Part A: Policy and Practice* 39 (1): 41–60. https://doi.org/10.1016/j.tra.2004.04.005.