UNIVERSITÉ DU LUXEMBOURG

PhD-FSTC-2018-09
The Faculty of Sciences, Technology and Communication

DISSERTATION

Defence held on 29/01/2018 in Luxembourg

to obtain the degree of

DOCTEUR DE L'UNIVERSITÉ DU LUXEMBOURG

EN SCIENCES DE L'INGÉNIEUR

by

Guido Cantelmo
Born on 24 March 1987 in Rome, (Italy)

DYNAMIC ORIGIN-DESTINATION MATRIX ESTIMATION WITH INTERACTING DEMAND PATTERNS

Dissertation defence committee

Dr Francesco Viti, dissertation supervisor
Professor, Université du Luxembourg

Dr Constantinos Antoniou
Professor, Technische Universität München

Dr Thomas Engel, Chairman
Professor, Université du Luxembourg

Dr Chris Tampère
Professor, Katholieke Universiteit Leuven

Dr Ernesto Cipriani, Vice Chairman
Professor, Università Degli Studi Roma Tre

The only way to have real success in science, the field I'm familiar with, is to describe the evidence very carefully without regard to the way you feel it should be. If you have a theory, you must try to explain what's good and what's bad about it equally. In science, you learn a kind of standard integrity and honesty.

Richard P. Feynman

Acknowledgments

These four years of Ph.D. have been long, intense and full of contingencies. It seems to me that I have been living four lives in one. I can hardly find any word to describe them without diminishing this experience or, even worse, diminishing my appreciation for all those who supported and encouraged me during these years. Nevertheless, I will try my best.

I would first like to acknowledge my Supervisor, Prof. Dr. Francesco Viti. I remember that, when I started my Ph.D., at every conference there was some professor telling me how lucky I was to have him as a supervisor. I can definitely say that I have been extremely lucky, and I am grateful to Francesco for the opportunity he gave me. He has not only been a great mentor, but also a good friend. He taught me to always question my results, specifically when they match with my expectations. Although frustrating in the beginning, this really helped me out during the last years of my research.

I am particularly grateful to both members of my supervisory committee, Prof. Dr. Ernesto Cipriani and Prof. Dr. Thomas Engel. They constantly motivated and advised me to keep digging into my research. A special thank goes to Ernesto and Marialisa. They have been the first ones to have faith in my scientific capabilities and to suggest me to look for a Ph.D. position. I would have probably never achieved this goal without you.

I would also like to express my sincere gratitude to the defence committee members, Prof. Dr. Chris Tampère and Prof. Dr. Constantinos Antoniou, for taking the time to go through my dissertation and provide valuable advice.

I am grateful to my fellow colleagues of the MobiLab group, who definitely added some colour to this journey. When I started, back in 2014, we were only three grad students and one Professor. In only a few years, the group more than tripled in size. The merit of this success clearly goes to Francesco and his everlasting will for running new projects, attending all conferences and recruiting new students. But even more important, he inspired us and made everybody in the team contribute to this success. Marco, you deserve a special mention for helping me out when I was struggling with those insecurities many young researchers face at the beginning of their career. Francois, Georgios, Thierry, it has been a pleasure to attend so many conferences together; I will see you at the piano bar.

I also want to acknowledge my family in Luxembourg. The “Dommeldange Fraternity” (I still believe that the University of Luxembourg should officially recognize our community) and the “Amigos del Vicindario”. I would not have been able to survive all these years without you cheering me up. You are far too many to mention, but you know how much I value your friendship and how important it has been for me. I am not going to forget our adventures, and I am quite confident we will have more in the future. Eva, Omar, Nastya, Chris and Giulio, it is specifically because of you that I managed to feel home while being so far away from Rome.

At last, I would like to give a warm hug to my family. Among all of the beautiful things Italy has to offer, my family is what I missed the most. Roberta and Silvio, my parents, and my brother Francesco. These years have been extremely difficult for all of us and I really wanted to spend more time with you. You guys always told me to enjoy my life and to never walk back. There are no words to express how much I missed you, how much I love you and how much you mean to me.

I would like to thank you all, and I give my apologies if other people, certainly important, were not mentioned. This is the end of a beautiful period of my life. These years have not always been perfect, I had many ups and downs, but you have always been there for me. Now it is time to start a new Chapter, and I am sure most of you will still be part of it.

Preface

It has become very fashionable to talk about *Mobility as a Service*, multimodal transport networks, electrified and green vehicles, and sustainable transportation in general. Nowadays, the transportation field is exploring new angles to solve mobility issues, applying concepts such as using machine learning techniques to profile user behaviour. While for many years “traffic pressure” and “congestion phenomena” were the most established keywords, there is now a widespread body of research pointing out how new technologies alone will solve most of these issues.

One of the main reasons for this change of direction is that earlier approaches have been proven to be more “fair” than “effective” in tackling mobility issues. The main limitation was probably to rely on simple assumptions, such as in-elastic mobility travel demand (car users will stick to their choice), when modelling travel behaviour. However, while these assumptions were questionable twenty years ago, they simply do not hold in today's society. While it is still true that high-income people usually own a car, the concept of urban mobility evolved. First, new generations are likely to buy a car ten-twenty years later than their parents. Second, in many cases, users can choose options that are more effective by combining different transport modes. Wealthy people might decide to live next to their working place or to the city centre, rather than to buy a car. Thus, it becomes clear that to understand the evolution of the mobility demand we need to question some of these assumptions.

While data can help in understanding this societal transformation, we argue in this dissertation that they cannot be considered as the sole source of information for the decision maker. Although data have been there for many years, congestion levels are increasing, meaning that data alone cannot solve the problem. Although successful in many case studies, data driven approaches have the limitation of being capable of modelling only what they observed in the past. If there is no record of a specific event, then the model will simply provide a biased information. In this manuscript we point out that both elements – data and model – are equally relevant to represent the evolution of a transport system, and specifically how important is to consider the heterogeneity of the mobility demand within the modelling framework in order to fully exploit the available data.

In this manuscript, we focus on the so-called Dynamic Demand Estimation Problem (DODE), which is the problem of estimating the mobility demand patterns that are more likely to best fit all the available traffic data. While this dissertation still focuses on car-users, we stress that the activity based structure of the demand needs to be explicitly represented in order to capture the evolution of a transport system. While data show a picture of the reality, such as how many people are travelling on a certain road segment or even along a certain path, this information represents a coarse aggregation of different individuals sharing a common resource (i.e. the infrastructure). However, the traffic flow is composed of different users with different trip purposes, meaning they react differently to a certain event. If we shut down a road from one day to another, commuting and not commuting demand will react in a different way. The same concept holds when dealing with different weather conditions, which also lead to a different demand pattern with respect to the typical one. This dissertation presents different frameworks to solve the DODE, which explicitly focus on the estimation of the mobility demand when dealing with typical and atypical user behaviour. Although the approach still focuses on a single mode of transport (car-users), the proposed formulation includes the generalized travel cost within the optimization framework. This key element allows accounting for the departure time choice and, in principle, it can be extended to the mode choice in future work.

The methodologies presented in this thesis have been tested with a “state of the practice” dynamic traffic assignment model. Results suggest that the models can be used for real-life networks, but also that more efficient algorithm should be considered for practical implementations in order to unleash the full potential of this new approach.

Index

Notation

Acronyms:

AB-DTA = Activity Based Dynamic Traffic Assignment

AB-DODE = Assignment Based Dynamic OD Estimation

DNL = Dynamic Network Loading

DTA = Dynamic Traffic Assignment

DODE = Dynamic Origin-Destination demand Estimation

DSA = Daily Systematic Activity

DUE = Dynamic User Equilibrium

FDSA = Finite Difference Stochastic Approximation

GDP = Gross Domestic Product

MU = Marginal Utility

NSA = Not Systematic Activity

OD = Origin-Destination

ODE = Origin-Destination demand Estimation

SB-DTA =Schedule-Based Dynamic Traffic Assignment

SBODE = Sensitivity Based OS Estimation

SPSA = Simultaneous Perturbation Stochastic Approximation

TA = Traffic Assignment

TB-DTA = Trip-Based Dynamic Traffic Assignment

TC-DTA = Trip-Chain-based Dynamic Traffic Assignment

UB-DODE = Utility Based Dynamic OD Estimaiton

UMT = Utility Maximization Theory

WSA = Within week Systematic Activity

Notations:

Irregular Demand Patterns:

N= Number of Users

c^i= Numerical perturbation at Iteration i.

α^i= Step-size at Iteration *i*.

μ= Mean;

σ= Variance

$\boldsymbol{G}^i$ =Gradient at the iteration *i*

$\boldsymbol{J}$= Jacobian Matrix

$\boldsymbol{I}$= Identity Matrix

$\boldsymbol{b}$ = Vector of binary variables indicating whether a link flow is on the corrected branch or not

$\boldsymbol{d}^*$= Estimated OD flows

$\widehat{\boldsymbol{d}}$ = Vector of the historical OD flows

$\widehat{\boldsymbol{g}}_n$= n-th Stochastic Gradient approximation

$\boldsymbol{l}$ = Vector of the simulated link flows

$\hat{\boldsymbol{l}}$ = Vector of the observed link flows

$\boldsymbol{p}$ = Descent direction for the SBODE algorithm;

$\boldsymbol{q}^*$= Vector of the simulated node measures

$\widehat{\boldsymbol{q}}$ = Vector of the observed node measures

$\boldsymbol{r}^*$= Vector of the simulated route measures

$\hat{\boldsymbol{r}}$ = Vector of the observed route measures

$\boldsymbol{x}$ = Vector of the OD flows

$\boldsymbol{z}$ = Estimator for the error

Δ^n= SPSA random perturbation vector with {-1;+1} elements.

$\boldsymbol{\theta}$= Vector of generic variables to be updated

$\boldsymbol{\xi}^n$= Vector with all-zeros values except for the variable *n* to be optimized

Regular Demand Patterns:

U = Total net utility

U^t = Disutility of travelling for a specific trip in the system

U^a = Utility related to performing a certain activity

U^T = Overall disutility of travelling

U^A= Overall utility of performing activities

n = Subscript for the user

N = Number of users

p = Subscript for the number of activities/purposes

P = Number of activities/purposes

s = Subscript for the trip

S = Number of trips

m = Subscript for the mode of transport

M = Set of feasible modes of transports

r = Subscript for the route

R = Set of feasible routes

t = Analysis time interval

t^0= Preferred arrival time

t^{d-o}= Departure time for which user n arrives on time at the destination

t_p^s = Actual starting time for a specific activity *p*

t_p^e= Actual ending time for a specific activity *p*

EA =Scheduling delay - Early Arrival

LA =Scheduling delay - Late Arrival

VoT= Value of Time of unit duration for the travel time

VoE= Value of Time of unit duration for arriving early

VoL= Value of Time of unit duration for arriving late

U^{MAX} is the maximum accumulated utility for an activity *p*

$\beta_P, \alpha_P, \gamma_P$ parameters for the clock-based MU for an activity *p*

τ_P= Parameter considering the flexibility of an activity *p*

η_P= Parameter for the duration-based MU for an activity *p*

T = Travel Time

T^F= Free flow travel time

T^b= Time spent at the bottleneck

D(t) = length of the queue during time interval *t*

S = Capacity of the bottleneck

ε time interval in which congestion is observed on the network

G= parameters accounting for the fatigue effect

1

Introduction

Demand Estimation is the process of inferring the origin-destination demand flows from the available traffic data. This problem is considered extremely important for any application in Transport Engineering, as inaccurate demand flows lead to inaccurate traffic predictions, thus economic and social losses.

This dissertation focuses on the structure of the Origin-Destination demand flows. The addition of a parametric representation of the demand results in including activity patterns and increasing the overall reliability of the estimation process.

This Chapter introduces the context of this thesis, the Demand Estimation problem, the contribution of this dissertation and its outline.

1.1 Context and background

1.1.1 Context

Transportation is a key sector of any country's economy, as it contributes to improve both the economic growth and the quality of life for the citizens. In the European Union alone, this sector sustains over 10 million jobs and contributes approximately for 4.5% of the overall Gross Domestic Product (GDP) (European Commission 2014). Nevertheless, mobility also represents a major cost for our society. As the demand for mobility services is increasing with the economic welfare, pressure on existing infrastructures has reached its limits. Even without considering environmental and social externalities, in Europe, congestion alone costs about 1% of the total GDP (European Commission 2014). To provide an example, according to TomTom data, Luxembourg Ville is the 32th most congested city in Europe and is ranked 78th worldwide (TomTom 2017). Statistics show that the extra travel time during the rush hour is 40 minutes per day, meaning that, over a year, each user spends 1.7% of his life stuck in the queue. The challenge for the public authorities is therefore twofold: on one hand, to expand transport facilities in order to support the economic growth, on the other hand to reduce costs and externalities related to the transport sector.

Traditionally, response measures to this phenomena can be classified into three main domains: strategic planning, tactical planning and operation management programs (Cascetta 2009). Strategic and tactical planning involves long and medium-term decisions, such as expanding or building new infrastructures (e.g. roads, logistic hubs), promoting emerging technologies (e.g. electric and autonomous vehicles) or boosting cooperative or on-demand mobility (e.g. car sharing, ride sharing). Operation management programs focus instead on short-term solutions to optimize the usage of existing transport facilities. Ramp-metering control strategies, traffic signal design and transit timetable optimization are some examples of management solutions. Regardless the adopted response measures, public authorities need support tools to quantify their effect. One solution involves using data driven techniques to analyse the current situation and to predict the evolution of the system. However, these approaches are often leading to a biased estimation, as they do not account explicitly for user behaviour or the future evolution of the system (which could substantially differ from the current one). Concerning road users, a preferable option is to deploy Traffic Assignment (TA) models, which for many years have been successfully applied as supporting decision tools for the evaluation of traffic planning and managing solutions. TA models not only offer the opportunity to estimate and predict the traffic state on the transport network but also to implement the proposed solution within the model itself, in order to quantify the gain of each alternative and compare it with the do-nothing scenario – i.e. the current situation or its projection in the future. These models are composed of two main components: the choice model and the propagation model (Corthout 2012). The former deals with the behavioural aspect of the user that decides to travel from one place to another, and can account for route, location, and - in a dynamic context - departure time choice. The propagation model deals instead with the physical relations between vehicles and infrastructure. Conventional TA models assume that both the infrastructure characteristics (called *supply*) and the mobility demand are constant over a certain reference time period (*within-period stationary* case). Although this assumption is acceptable to estimate average traffic conditions on the network, congestion is underrepresented. Thus, last decades witnessed to an intensive research effort in developing new models, in which these characteristic are modelled as time dependent (*within-period dynamics* case) (Cascetta 2009). Time dependent - or Dynamic - Traffic Assignment Models (referred to as DTA in the rest of this thesis) represent the state of the art in modelling transport systems, as they provide a realistic representation of the congestion and a wide range of time-varying outputs, such as length of the queue, route costs and travel times. These output measures are also extremely relevant in order to perform an ex-post analysis of the results and compute additional performance measures analysis or a cost-benefit analysis.

One of the essential inputs for (stationary or dynamic) TA is the mobility demand, as it results that a biased demand pattern will likely lead to a biased congestion pattern. Typically, mobility demand is represented as an Origin-Destination (OD) demand matrix, where each cell of the matrix represents the number of trips from one traffic zone to another, for a certain trip purpose and mode of transport. In the dynamic case, time is usually discretised in a specified number of intervals, and a dedicated OD matrix is then associated to each of them (Figure 1.1).

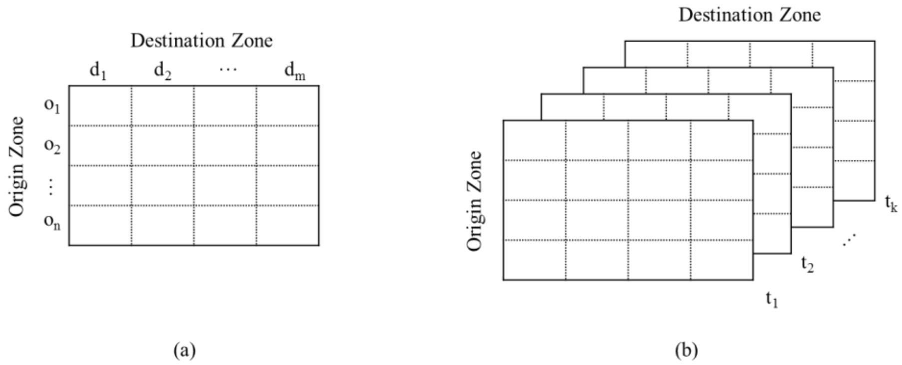


Fig.1.1. OD Demand Matrix for the within-period stationary (a) and dynamic (b) case;

The main problem is that state of the art measurement systems are not sufficient for estimating a realistic OD demand matrix. Conventional monitoring systems, such as loop detectors, measure the effect of the demand on the network rather than the demand itself. More advanced techniques, such as GPS coordinates and plate number recognition, provide more insights into the number of trips between a certain origin and destination. However, they usually capture only a sample of the whole demand, whose representativeness is related to the penetration rate of the adopted technology. As a consequence, practitioners usually turn to demand generation models in order to approximate the demand. The most widely adopted demand model is the so-called Four-Step model (McNally 2007). This model is based on aggregate demographic data and, by means of a series of probability functions, estimates the OD matrix. This representation of the demand is normally called *macroscopic* since the demand is represented at a strongly aggregated level. The main limitation of this approach is that demographic variables are static, which means that the result will be a stationary OD matrix. However, DTA models need as input a dynamic OD matrix rather than a static one. Moreover, in reality, mobility demand is activity rather than trip based. Each user chooses his/her departure time based on the activities he/she scheduled along the day.

To take into account this element, an alternative approach is to use Activity-Based Models (ABM) and, specifically, Activity-Based demand generation models (Ben-Akiva and Bowman 1998), to represent the demand. In this case, a population is generated by census data, generating Activity Plans which describe the entire daily activity pattern for each user on the network. The main advantages using this representation are two. First, since the Activity Plan includes departing/arrival time at the destination, the resulting demand is dynamic in nature, which is a desirable property for DTA models. Furthermore, this plan includes different activities, which allows the model to consider a tour of activities rather than simple trips. While the relevance of the trip chain has been already investigated in the literature, normally this element is still missing in the macroscopic models.

The OD matrix generated through demand generation models is usually not accurate enough to reproduce the expected traffic patterns when coupled with a TA model. Thus, it is required a calibration phase where this demand is corrected in order to reproduce the available historical traffic information. This problem, which is briefly discussed in the next sub-session, is indeed the main focus of this thesis and it is well known in the literature as Origin-Destination demand Estimation (ODE) problem.

1.1.2 Origin-Destination Demand Estimation using Traffic Data

The OD Estimation (ODE) is the process of improving an existing available OD matrix by combining the demand generation models described in the previous section with the available traffic data. This process can be considered as an *indirect* estimation of the demand, as it relies on a model – the traffic assignment – in order to ensure consistency between the network performances (e.g. link flows, speeds, travel time) and the demand, which is the target of the estimation. As this process takes traffic data as an input in order to provide the mobility demand as an output, the literature often refers to it as the reverse assignment problem (Figure 1.2) (Cascetta 2009).

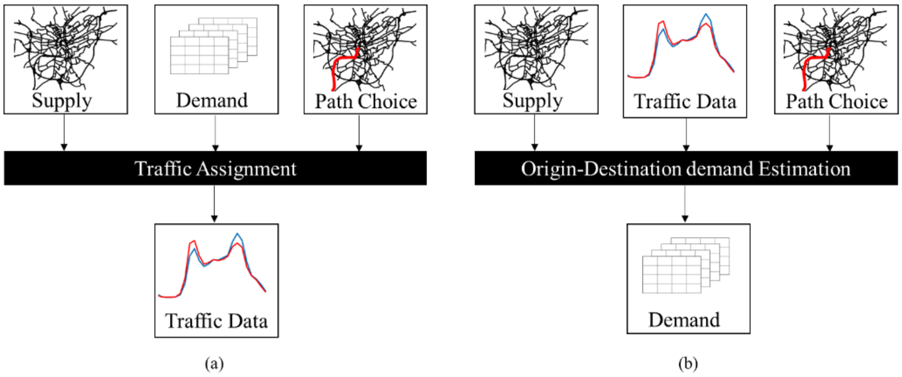


Fig.1.2. Comparison between the Traffic Assignment (a) and the OD Estimation (b) problem;

While an extensive overview of the problem is out of the scope of this chapter, this section aims at providing the general definition of the ODE problem and at highlighting the contribution of the current thesis. Demand estimation has been initially applied for the calibration of *within-day stationary* TA models and, without loss of generality, can be formulated as an optimization problem. The goal is to update the current values of the OD demand matrix in order to minimize the error between simulated and observed traffic data. By considering only the link and demand flows, for the stationary case, the objective function can be formulated as in Equation (1.1).

$$\mathbf{d}^* = \arg\max_x \left[z_1(\mathbf{x}, \hat{\mathbf{d}}) + z_2(\mathbf{l}, \hat{\mathbf{l}}) \right] \tag{1.1a}$$

Subject to

$$\mathbf{l} = \mathbf{Mx} \tag{1.1b}$$

Where z_1 and z_2 are distance functions measuring the error between current and reference values, $\boldsymbol{x}$ is the vector containing the OD flows to be updated, $\hat{\boldsymbol{d}}$ the one with the historical values of the OD flows obtained through the demand model, $\hat{\boldsymbol{l}}$ the vector with the available traffic counts and $\boldsymbol{l}$ the one with the simulated link flow. Finally, $\boldsymbol{d}^*$ is the vector with the calibrated demand flows, which minimises the objective function. The optimization problem is constraint to Equation (1.1b) in order to ensure consistency between demand flows and simulated traffic measures. This done through the assignment matrix **M**, which maps the demand flows to the link flows. Equation (1.1) can be directly extended to the dynamic OD demand Estimation (DODE), with the main difference that both OD flows and traffic measurements will be time dependent. Based on this formulation, existing research on the DODE problem can be divided into five main categories:

1) *Application domain:* A first classification is between *online* and *offline* approaches. In the first case, the model uses real time traffic data to update the existing OD-matrix, repeating the process when new data are available. This model is quite common for performing real time traffic predictions. Concerning the second category, *offline* models do not require real time data but calibration is based on a database which represents "typical" dynamics during the study period, such as the congestion during the morning peak. These models usually work together, as one generates the input for the other (Antoniou 2004).
2) *Objective functions*: The objective function is a key building block of the DODE as measures the magnitude of the error related to each variable. Typical research on this direction focuses on including additional information. While this was proven to be a good solution for motorways or small networks, it is not a sufficient condition when dealing with urban or big sized networks, as the enormous number of OD pairs makes the problem extremely complex. To avoid this issue, researchers are developing new solutions for decreasing the number of variables (Djukic et al. 2014).
3) *DTA model*: The Dynamic Traffic Assignment model is also extremely relevant during the DODE process. An unrealistic DTA will always lead to a poor representation of the demand, as, even when a perfect estimation of the demand is available, the model is not able to replicate the real traffic data. At the same time, if the DTA is too complicated, the computational time might become cumbersome. (Frederix 2012);
4) *The solution algorithm:* On this point, a main classification can be done between *analytical* and *DTA-based* models. The former models assume that an analytical relation between OD and Link flows exists and it is explicitly used to update the current solution. The *DTA-based* are more general, as they do not depend on any explicit relation. The advantage is twofold. First, many commonly adopted models do not provide this explicit relation, meaning that analytical models are not an option. Second, *DTA-based* models are more flexible as they can deal with many different data sources, such as GPS coordinates and even mobile phone network data, without assuming any analytical relation with the decision variables. By contrast, this flexibility comes with significant costs in terms of computational time (Barceló et al. 2012).
5) *Stochastic or deterministic OD flows:* Traditional OD estimation techniques look for the OD flows that are more likely to fit the available data. The result is a deterministic number of trips for each OD pair. However, the mobility demand is not deterministic by nature, meaning that OD flows are likely to change from one day to another. Thus, a possibility is to consider the OD flows as stochastic variables to consider the probability of observing a certain trip (Shao et al. 2015).

This thesis focuses on the *offline* case using a *DTA-based* approach. The work presented in the next chapters investigates the opportunities of extending both the objective function and the DTA model in order (i) to produce more reliable estimations with respect to the base case and (ii) to incorporate user behaviour in both the demand and DTA model, in order to account for heterogeneous mobility patterns. Finally, the current manuscript still estimates deterministic OD flows, as considering stochastic variables would further increases the complexity of the problem.

1.2 Objective and scope

1.2.1 Objective

The ideal situation would always be to directly observe the mobility demand and to feed the model with this direct information. Although this solution is possible from a technological point of view, it is too expensive in practice. Even when these data are available, their penetration rate is not enough to capture the entire population. For this reason, other data sources such as link counts and speeds, are always considered in state of the art models. Figure (1.3) shows the relation between data and demand generation/estimation models. The diagonal of the table shows the direct[1] measurement systems (we observe what we model), while other data sources can be classified as "redundant" when the information is more accurate than the model and "incomplete" when it is more aggregate. As direct observations are rarely available, modellers showed that combining "redundant" and "incomplete" data sources within the objective function leads to significant improvements of the OD estimation, as the redundancy corrects some of the issues related to using an incomplete – or indirect – measure (Antoniou et al. 2016).

While this assumption has been investigated for conventional data sources, such as GPS trajectories and plate number recognition, the main objective of this thesis is to include activity information within the DODE problem. Although Census Data are an incomplete and highly inaccurate source of information, they capture the behavioural nature of the demand.

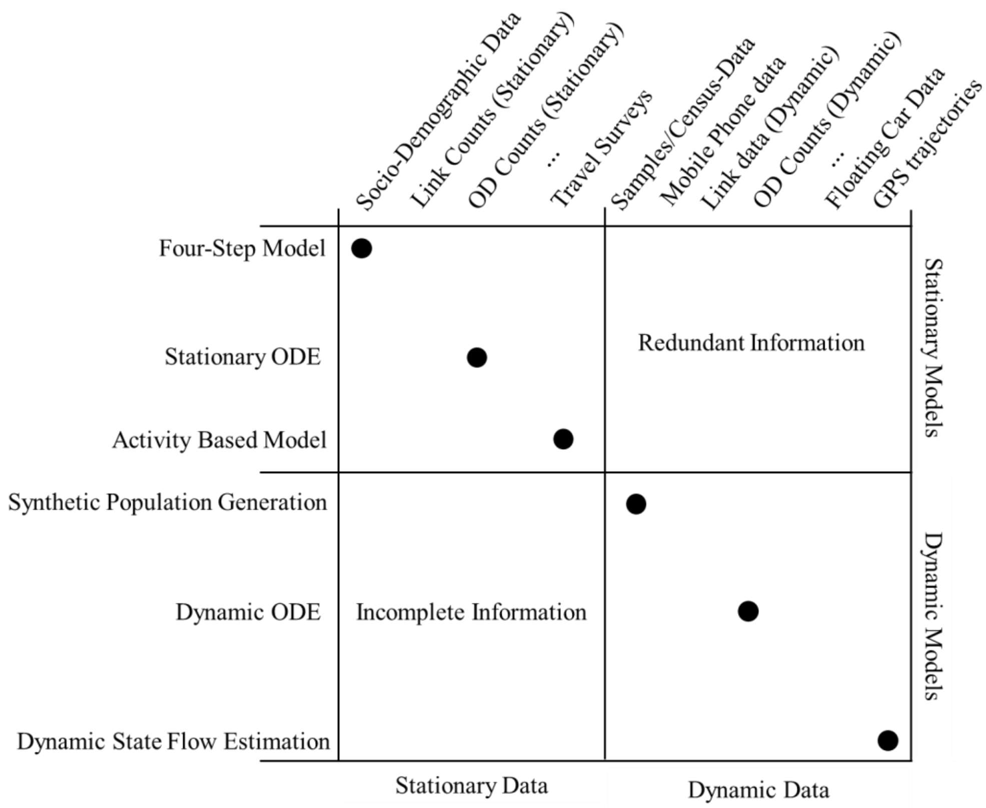


Fig.1.3. Relation between demand models and data sources;

In order to create a structure of the demand that accounts for a realistic user behaviour, the following research questions can be then formulated:

- *R.Q.1: How activity information can be included within the DODE problem?*

  Mobility Demand, which has been always modelled as activity based in theory, follows a trip-based representation in most of the state of the art DODE approaches. This work supports the idea that a correlation between an aggregated representation of the demand and the activity based user behaviour is needed in order to use macroscopic DTA models in a more efficient way. This research assumes that the OD dynamic demand matrix is a convolution of different

[1] More specifically, direct information represents in this thesis the ideal source of data for a certain demand model. To make an example, Socio-Demographic data do not always contains information at the OD level, but still represent the ideal data source for the Four-Step approach.

activity patterns. Thus, based on partial information at the activity level, a structure for the OD matrix can be created and updated by considering conventional measurement systems.

- *R.Q.2: How can we estimate purpose dependent OD flows without increasing the number of parameters?*

  While including activity information within a demand model seems reasonable, to simply account for activity based OD flows within Equation (1.1) is not an option. Because of the large number of parameters involved in the process, DODE is well-known for being a complex and highly nonlinear problem. In essence, if the number of variables increases in order to account for activity based OD flows, the DODE will very likely over fit the available traffic data. Thus, to reduce the problem dimension is a fundamental step in order to achieve our goal. This problem becomes even more relevant if considering that activity plans are tour-based, thus combining DTA models based on Dynamic OD matrices with a tour of activities is one of the main challenges of this work.

- *R.Q.3: How much we can increase the reliability of the OD estimation by improving the quality of the initial matrix?*

  The DODE problem, as specified in the previous sub-section, estimates the OD flows by updating an existing dynamic OD matrix. However, there is not an established approach for generating a dynamic OD matrix to use as initial input. State of the practice models rely on the assumption that a historical matrix exists, while in reality practitioner often over impose a temporal profile to an existing static OD matrix. Although researchers agree that the initial matrix is a fundamental aspect, research effort in this direction is very limited. A major challenge of this research is to improve the reliability of the DODE when an existing reliable OD matrix is not available, which is a major issue for practical implementations of any DODE framework.

1.2.2 Scope

The research presented in the next chapters deals with the DODE problem in the case of car users. As this thesis focuses on the objective function and the behavioural aspects of the underlying DTA model, other aspects such as the efficiency of the solution algorithm, introducing new data sources and dynamic network loading will not be discussed in detail, but only when necessary.

The methodologies illustrated are general and applicable to any DTA model which is based on dynamic OD matrices. As proposed in (Zhou and Mahmassani 2007) the OD demand is considered in this thesis as the convolution of three functional functions: *Regular Demand Pattern*, *Structural Deviations* and *Random Fluctuations* (Zhou and Mahmassani 2007). Regular Demand Patterns represent the typical demand profile, Structural Deviations take into account those phenomena, such as weather condition, that the analyst can model, while random fluctuation considers those deviations that cannot be explained. This classification results useful for identifying those conditions where activity information should be included within the DODE.

Finally, *online* and *real-time* applications are out of the scope of this thesis.

1.3 Thesis Contribution

This thesis analyses the problem of extending the Dynamic OD demand Estimation problem, bringing the following practical and scientific contribution.

1.3.1 Practical Contribution

- Mobility Patterns: The mobility demand is heterogeneous in nature, as it is composed of different users with different preferences. Part of this heterogeneity is captured within the DTA model. This research extends this concept by showing that the overall demand can be subdivided into multiple demand segments, each of them representing a different activity plan, which can be approximated by a simplified function. These functions can be used for representing activities at a macroscopic level.
- Non-commuting demand matters: Trip based approaches works quite well when the goal is forecasting the impact of major transportation infrastructures but are usually inadequate in analysing complex transport policies (McNally, Michael G. 2007). This because the conventional four step approach does not link trips and activities but only considers simplified temporal and spatial dependencies between zones (McNally, Michael G. 2007). This assumption might still hold for the commuting demand, which can be approximated through a gravity model. However, as we will show in the next section, commuting demand is about 20% of the overall mobility demand. The remaining share of the mobility demand will likely be underestimated if trip purpose and travel behaviour are not included within the demand model. This thesis introduces a methodology for clustering activity patterns and including them in the demand estimation process, reducing this phenomenon.
- Dynamic Structure of the Demand: The temporal profile of the demand changes not only over time but over space too. First, the work presented in this manuscript shows that the temporal profile extrapolated from traffic loops can be more or less representative of the demand, depending on the road type. Specifically, while traffic flow on a motorway is very aggregate, primary roads are a better proxy of the temporal profile of the demand. Second, a departure time choice model is introduced within the DTA, which creates a different temporal profile for each traffic zone based on information such as the preferred departure time for the morning and evening commute. This allows to reduce the error within the starting point, as more realistic demand profiles are generated.

1.3.2 Scientific Contributions

- Utility-Based Departure Time Choice Model: In this thesis, a departure time choice model based on complex utility functions is integrated with the traffic assignment module. Thus, the choice model of the DTA is extended in order to jointly work on route, departure time choice and activity location. The main contribution in this phase is to investigate the error related to use different utility functions within the departure time choice model. We demonstrate that, while in some cases introducing a simple – or even constant – value of the utility can lead to substantial errors, in other cases this simplification can be accepted, speeding up significantly the process.
- Utility-Based DODE: A Utility-Based model for the Dynamic OD estimation is presented. By using the parameters of the departure time choice model as decision variables, the number of variables strongly decreases with respect to the base case. The key contribution of this model is to partially fill the gap between trip based and activity based models, as it accounts for multiple activity types and to explicitly model activity tours, under the condition that the underlying DTA model accounts for tours as well.
- Reliability issues: A new framework called "Two-Step approach" is also presented in this thesis. This model deals with the classical DODE problem, as it does not account for activities and tours, and it divides the DODE in two sub optimization problem: the first step looks for a reliable time dependent OD matrix to use as initial solution for the second optimization problem, which adjusts the OD flows in order to minimize the error with respect to the available traffic data. Results show that this model provides more reliable results, as it reduces the negative impact of having a low quality initial matrix.

- Regular and irregular demand patterns: This thesis highlights that two classes of models should be considered within the OD estimation. Many DODE models assume that a good initial OD matrix is available, meaning that these models mostly work on random and structural deviations of the demand while keeping the systematic component constant. However, if the error on the systematic component – i.e. the *regular demand pattern* – is relevant, then a model that explicitly accounts for activity patterns is needed.

1.4 Thesis Outline

This dissertation is based on a collection of publications. Except for Chapter 2, each chapter introduces an original contribution, therefore it contains dedicated introduction and literature review sections. The manuscript is structured as follows:

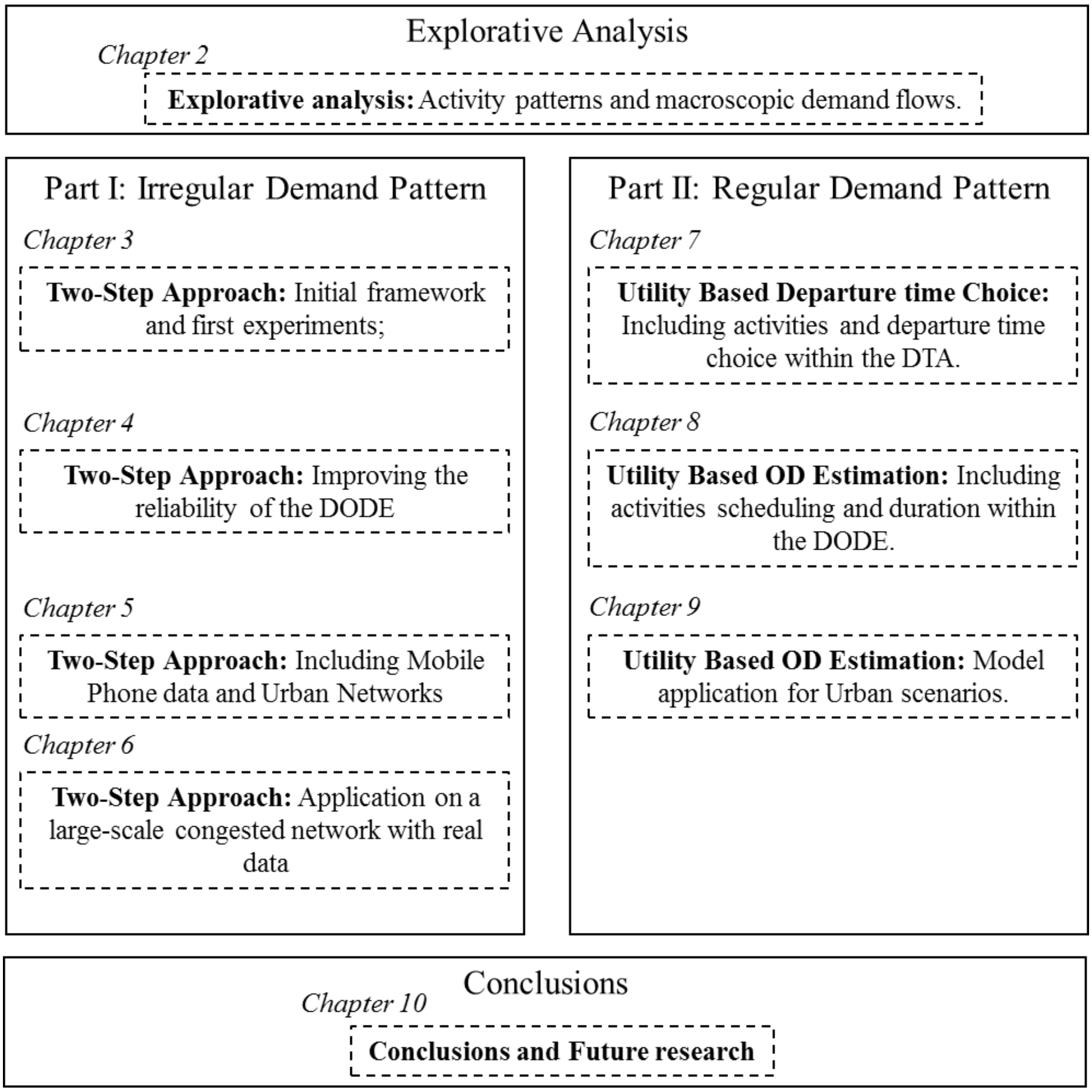


Fig.1.4. Thesis outline;

While it might seems formally more correct to introduce the *Regular Demand Pattern* before the *Irregular* ones, this could make this manuscript complicated to read and hard to access. The reason is that reading Chapters 7-8, which contains the main contributions of this dissertation, requires an advance knowledge of the OD estimation problem. Before moving to these chapter, the reader should already be aware of the limitations for conventional approaches, including the quality of the initial demand matrix, the underdetermines of the problem and the possibility of including additional traffic data within the objective function, just to mention a few. This topics are discussed within the first part of this manuscript.

1.4.1 Explorative Analysis

Before moving to the theory, this section analyses empirical data in order to verify the assumption that a relation between activities and OD demand flows exist. Empirical evidence supports the idea that a different modelling approach should be considered when dealing with regular or irregular demand patterns.

1.4.2 Part I: Irregular Demand Pattern

The first part of this thesis focuses on the estimation of irregular mobility patterns. The input is an existing dynamic OD matrix, while the goal is to look for the OD matrix that best fits the available traffic data. As this process results in just the adjustment based on empirical observations, it is likely to capture fluctuations from the systematic component of the demand – such as the user behaviour during a raining day – while is less likely to account for user preferences, such as the preferred departure time.

1.4.3 Part II: Regular Demand Pattern

In this case, the assumption is that a good dynamic OD matrix is not available or that we want to update it in order to account for new activity patterns. A departure time choice model and a parametric approach to the dynamic OD Estimation model are presented. The target, in this case, is not to fit the traffic data, as the parametric function does not allow the same precision as the model presented in Part I, but to identify the activity-based structure of the demand, thus its systematic component.

1.5 References

Antoniou, Constantinos. 2004. "On-Line Calibration for Dynamic Traffic Assignment." Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/28639.

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. doi:10.1016/j.trc.2015.08.009.

Barceló, J., L. Montero, M. Bullejos, O. Serch, and C. Carmona. 2012. "Dynamic OD Matrix Estimation Exploiting Bluetooth Data in Urban Networks." In *Proceedings of the 14th International Conference on Automatic Control, Modelling & Simulation, and Proceedings of the 11th International Conference on Microelectronics, Nanoelectronics, Optoelectronics*, 116–121. ACMIN'12. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS). http://dl.acm.org/citation.cfm?id=2230549.2230572.

Ben-Akiva, Moshe E., and John L. Bowman. 1998. "Activity Based Travel Demand Model Systems." In *Equilibrium and Advanced Transportation Modelling*, 27–46. Springer. http://link.springer.com/content/pdf/10.1007/978-1-4615-5757-9_2.pdf.

Cascetta, Ennio. 2009. *Transportation Systems Analysis*. 2nd ed. Vol. 29. Springer. http://www.springer.com/us/book/9780387758565.

Corthout, Ruben. 2012. "Intersection Modelling and Marginal Simulation in Macroscopic Dynamic Network Loading." PhD Thesis. https://lirias.kuleuven.be/handle/123456789/344247.

Djukic, Tamara, S. P Hoogendoorn, Hans van Lint, and TRAIL Onderzoekschool. 2014. "Dynamic OD Demand Estimation and Prediction for Dynamic Traffic Management." Delft: TRAIL Research School.

European Commission, ed. 2014. *Transport: Connecting Europe's Citizens and Business*. Ms. updated in November 2014. The European Union Explained. Luxembourg: Publ. Off. of the Europ. Union.

Frederix, Rodric. 2012. "Dynamic Origin-Destination Matrix Estimation in Large-Scale Congested Networks." PhD Thesis. https://lirias.kuleuven.be/handle/123456789/344248.

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

Shao, Hu, William H.K. Lam, Agachai Sumalee, and Martin L. Hazelton. 2015. "Estimation of Mean and Covariance of Stochastic Multi-Class OD Demands from Classified Traffic Counts." *Transportation Research Part C: Emerging Technologies*, 59 (October): 92–110 doi.org/10.1016/j.trc.2015.05.017

TomTom. 2017. "TomTom Traffic Index 2017." Accessed August 25. https://www.tomtom.com/en_gb/trafficindex/city/luxembourg.

Zhou, Xuesong, and Hani S. Mahmassani. 2007. "A Structural State Space Model for Real-Time Traffic Origin–destination Demand Estimation and Prediction in a Day-to-Day Learning Framework." *Transportation Research Part B: Methodological* 41 (8): 823–40. doi:10.1016/j.trb.2007.02.004.

2

Empirical analysis of daily demand patterns

This Chapter aims to show empirical support to some of the main assumptions behind the theory presented in this thesis. The first section introduces the concept of splitting the mobility demand in several "primitives", each of those representing a certain activity. Next, an analysis of how much information on the day-to-day evolution of the demand can be obtained by loop detectors is presented. Then, the analysis focuses on the activity-based structure of the mobility demand. The discussion supports the idea that users with the same trip purpose can be grouped in homogeneous classes and that a limited number of classes represents a large share of the overall demand. Finally, the author stress that this analysis presents simple empirical findings, which cannot be generalized to all cases but represent a requirement for adopting the methodologies presented in the next chapters.

The Content of this chapter has been partially presented at the following conferences and its content is unpublished to date:

Cantelmo, Guido, and Francesco Viti. 2015. "Activity Demand: An Empirical Analysis on the Influence of the Activities on the Traffic within-Day Demand Profile." *In 4rd HEART Conferenc*e.

Cantelmo, Guido, Francesco Viti, and Chris MJ Tampere. 2014. "Exploiting the Relation between Activity Data and Traffic Data within the Dynamic Demand Estimation Problem." *In 3rd HEART Conference*.

2.1 Introduction

In this chapter, we explore the opportunity of clustering different activities in order to observe user behaviour at a macroscopic level. The main idea behind the methodology proposed in this thesis is that a few parameters can represent a homogeneous class of users. Simply stated, it should be possible to model users travelling from a certain origin to a certain destination for a certain purpose through a few parameters, such as the preferred departure time and its variance. However, before moving to the main body of this research, which is presented in the next chapters, this assumption must be properly verified through empirical data.

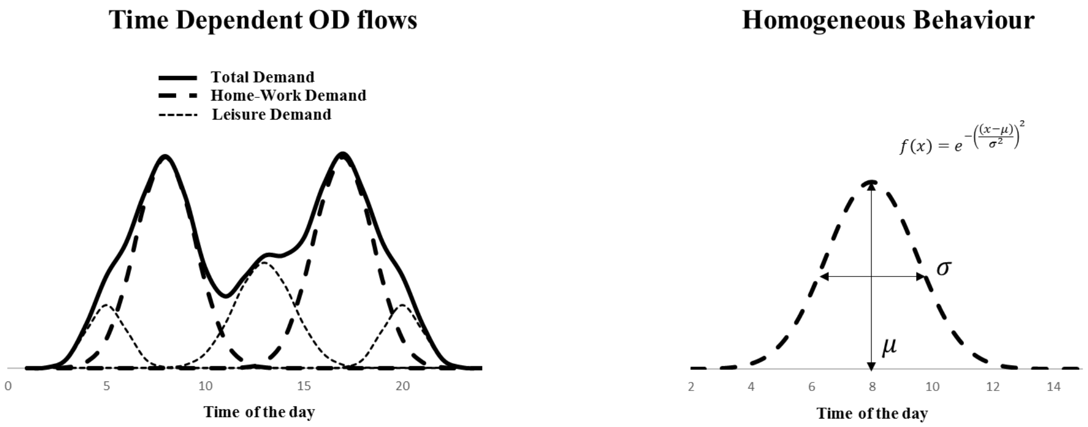


Fig.2.1 Mobility Demand as a convolution of different activities;

While many DODE formulations have been proven to successfully estimate the demand flows when a trustworthy dynamic OD matrix is available, in practice, historical demand flows are often derived from static models and their temporal profile is extrapolated from the traffic data. When large-urban networks are involved in the DODE process, the problem becomes even more complex, as this error increases with network sizes. For example, commuters travelling from different origins to the same common destination will clearly have a different departure time. On this point, the analysis presented in the next sections focus on two main aspects. Section (2.1) investigates the correlation between traffic data and temporal demand profile, while Section (2.2) proposes to use probability functions for modelling activity patterns and generates a more realistic profile, as shown in Figure (2.1).

The BMW (Behavioural and Mobility within the Week) database is the main source of information used to validate our assumptions. The dataset, collected in the region of Ghent-Belgium, contains information from 717 different individuals in the form of Travel Diaries. For a week, each individual provided information over:

1) The Departure Time
2) The Arrival Time
3) The starting location of the trip
4) The final location of the trip
5) The kind of activities performed
6) The number/sequence of activities
7) The mode choice

8) The household composition (not used in this study)

The travel survey started on the 08 September 2008 and finished on 07 December 2008, hence covering a period of three months. Each user has been monitored for one week period and traffic counts are available for the same period of time. For further information, see the full BMW report (Castaigne et al. 2010)[2].

Next sections introduce some insights and conclusions about the aggregate user behaviour.

2.1.1 Temporal and Spatial structure of the demand

This section analyses the discrepancy between mobility demand and traffic profile. Thus, the first step is to compare the temporal trend of the observations obtained from traffic data with the trend of the aggregate Activity demand estimated from the travel diaries. Main differences between these types of information are:

- *Traffic Data:* Direct observation about all the demand, cheap but strongly aggregated.
- *Activity Data:* Sample information (user level), expensive but disaggregate. Statistically biased if the sample size is not sufficiently large.

In Figure 2.2 is possible to observe the two temporal profiles, for two different days (Wednesday and Thursday).

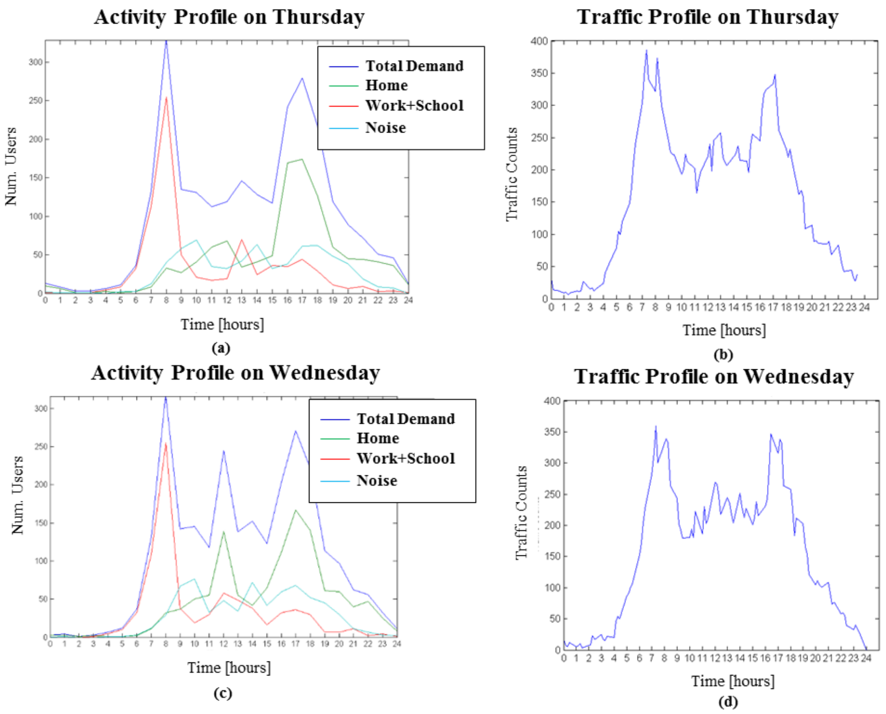
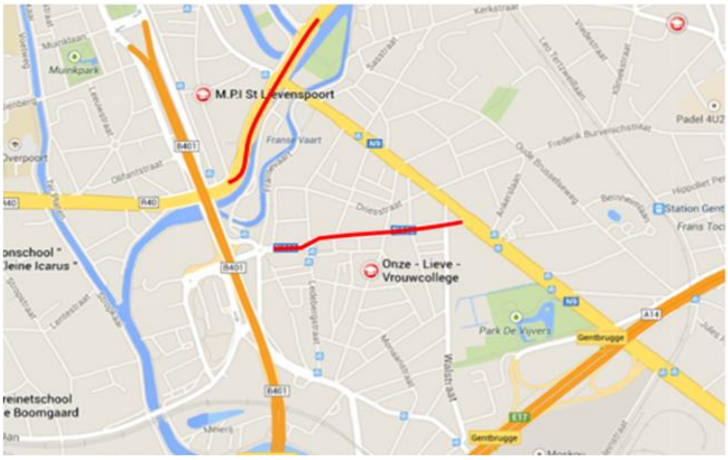


Fig.2.2: (a) Aggregate plot of the observed activities for Thursday; (b) Observed traffic Flow on a detector. Thursday; (c) Aggregate plot of the observed activities for Wednesday; (d) Observed traffic Flow on the same detector;

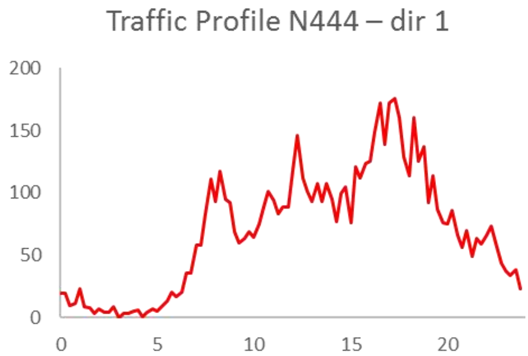
[2] Available at this link: http://www.belspo.be/belspo/SSD/science/Reports/BMW_FinRep.ML.pdf

Figures 2.2 (a-b) and (c-d) show that the aggregate temporal profile on Thursday results similar when using Activity Data or Traffic Counts, while this is not the case on Wednesday. Wednesday is a specific day in Belgium, as schools close earlier and people often go home after lunch in order to stay with their family. This behaviour, although systematic, is not captured by most of the detectors on the network suggesting that, if this specific detector is used for creating the time-dependent demand profile, day-to-day dynamics are likely to be underestimated within the generated dynamic OD matrix. While this observation seems obvious, the question is: which detectors should be used to create the time-dependent OD matrix and, even more important, which detectors are more likely to capture the different mobility pattern showed in Figure (2.2c)?

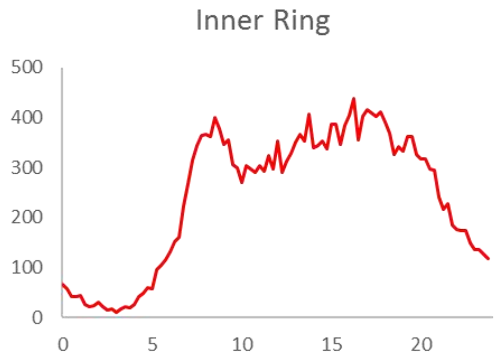
On this point, we can assume traffic flows on a local road to be mostly generated by the incoming/outgoing users for that specific traffic zone, while regional roads and highways are composed by a more aggregated traffic flow. Consider the city map of Ghent showed in Figure 2.3a. Traffic profile on two sections is shown, for a local road (Figure 2.3b), very close to a school, and for the inner ring (Figure 2.3c).



(a)



(b)



(c)

Fig.2.3: (a) Network of Ghent (b) Observed traffic Flow on a Local Road; (c) Observed traffic Flow on the Inner Ring;

While temporal profile on the Local Road shows similarity with the one depicted in Figure 2.2c, the time profile on the inner Ring is extremely aggregated. The reason is that, while the Local Road is mainly used by users arriving at their destination– the place in which the activity will be consumed –

the ring is shared by different users, with different purposes, different departure time and different travel time.

In order to investigate this effect at a network level, the time-dependent profile obtained through the travel survey and traffic counts have been compared for all available detectors in the area of Ghent. Each observation – both travel surveys and traffic counts - has been separated into three components: Morning Peak, Evening Peak and afternoon, and each time period has been normalized. The function obtained through this procedure has been then compared with the one obtained from the travel surveys. Results are summarized in Table 2.1., which shows the correlation in terms of r-squared between road detector and activity-based demand profile for each day of the week.

Table 2.1: Relation Between Traffic Counts and Activity-Based temporal profile

	Monday r2	Tuesday r2	Wednesday r2	Thursday r2	Friday r2	Saturday r2	Sunday r2
Regional Roads	0.35	0.34	0.37	0.4	0.32	0.38	0.13
Provincial Roads	0.5	0.45	0.43	0.47	0.45	0.38	0.13
Local Roads	0.51	0.49	0.47	0.52	0.48	0.37	0.12
Inner Ring	0.42	0.33	0.38	0.41	0.37	0.39	0.14
External Ring	0.71	0.65	0.69	0.74	0.73	0.38	0.14
Link Road	0.79	0.53	0.65	0.77	0.58	0.4	0.08

In Table 2.1, r2 represents the average "r-squared" over all the detectors belonging to the same road category. For instance, 0.51 is the average r-squared for all traffic counts of category "*Local Roads"* on *Monday*. However, Table 2.1 shows this value can change a lot from one detector to the other. Some local roads present a very high correlation, like in 2.2, while other do not. Thus, if the profile of the demand is known, it becomes quite easy to identify between different detectors the one providing a realistic demand profile for the neighbouring traffic zones. If we focus on the highways, the *Inner Ring* presents a lower value of r2 with respect to the *External Link* or *Link Road*. This is reasonable because the *External link* is a large motorway running around the city mainly used by commuters that live and work outside the city, while the *Link Road* is a segment of the highways entering directly in the city centre, which has only a few detectors. This allows us to observe the inflow/outflow related to all those activities located in the city centre.

2.1.2 Activity-Based structure of the demand

2.1.2.1 Activity Patterns and Activity Components

In this section, we investigate the aggregate relation between Activity-based OD flows and traffic states. Specifically, we assume that the mobility demand is a convolution of different *Activity Patterns*. Some of them are rigid, like the home-work activity, while others are more flexible. It is intuitive to realize that rigid activities determine the network condition, while the flexible ones are influenced by the given traffic state – i.e. users can reschedule their activities if the cost of reaching the destination is too high. Thus, we can classify the mobility demand according to the definition of *rigid* and *flexible Activity Demand Components*. In general, we can identify at least three groups of *Activity Components*:

I. *Within-Day-Systematic Activities (DSA):* These are *rigid* activities, in which arrival and/or departing time is not usually flexible (i.e. going to work, returning home)

II. *Within-Week-Systematic Activities (WSA):* These are *flexible* activities, which are not systematic within the day, but recur regularly, e.g. every week (i.e. swimming pool, weekly shopping).

III. *Not-Systematic Activities (NSA):* These *flexible* activities represent extraordinary events with respect to the usual user activity scheduling (i.e. visiting the doctor is an example).

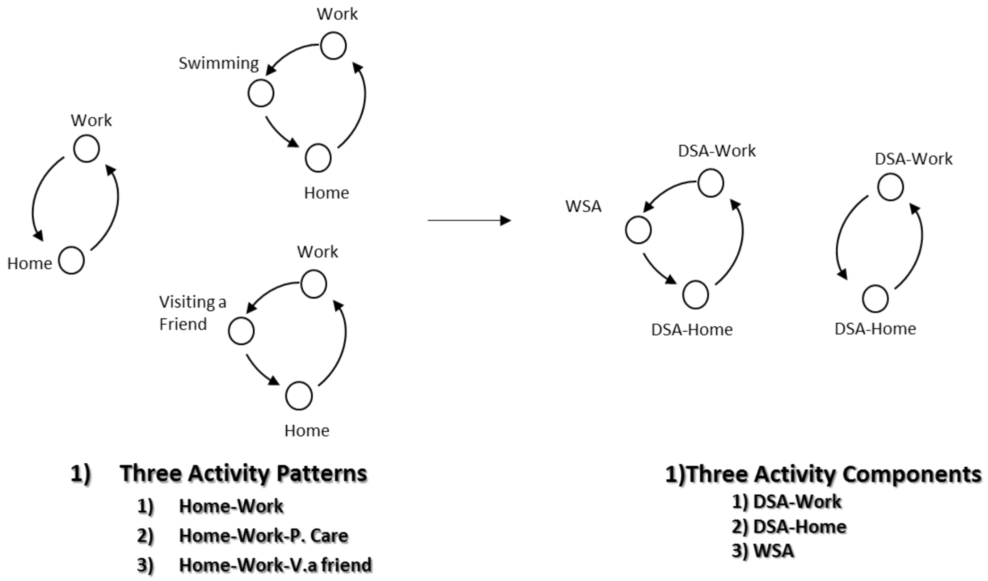


Fig.2.4: From Tours to Activity Component

The classification is done on the basis of the concept that, for the single user, the agenda during the month will be composed of DSA activities, like work, WSA activities, like the swimming pool or the weekly shopping, and NSA activities, like visiting the doctor. All the activities reported within the BMW database have been classified according to the above three groups of Activity Components by mean of a hierarchical cluster analysis, as shown in Figure 2.4. Four artificial functions, representing *within day systematic* (Home and Work), *within week systematic* and *not systematic activities* have been generated based on the aforementioned definition. The cluster analysis exploits the following rules:

- The first activity to be grouped in a cluster is the most similar, by definition, with the artificial function (DSA, WSA, NSA). For instance, DSA – Home function has to be grouped, as the first step, with Work Purpose.
- If two artificial functions are grouped, the procedure should be stopped
- According to the general rule of hierarchical clustering, when by grouping two clusters too much information is lost, the procedure should be stopped.

Point 2 and 3 stress that this procedure can generate no more than four activity components, the number of artificial functions used in this problem. For details on this procedure, and specifically how the artificial functions have been created, we refer the interested reader to Appendix A. Figure 2.5 reports the list of activities described within the BMW database and how these can be classified according to these four *Activity Components*. We can now define an *Activity Pattern* as the combination of two or more *Activity Components.* The main different between *Activity Pattern* and *Activity Component* in this thesis is that the former considers generally a series of *Activity Components* and their location, while the latter represents an aggregate purpose dependent demand flow for a specific traffic zone.

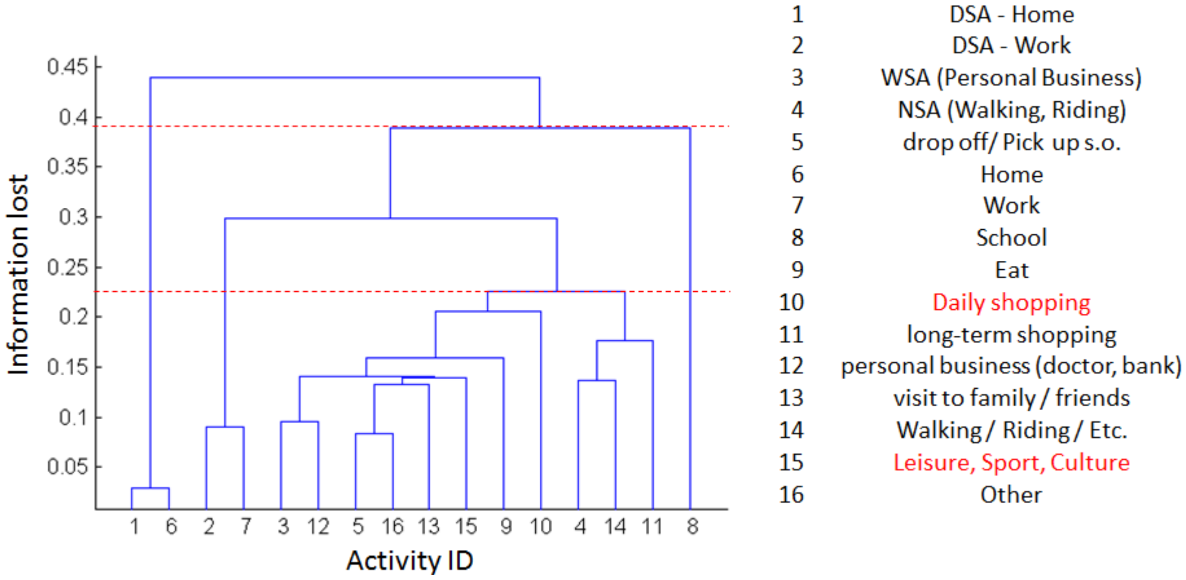


Fig.2.5: Hierarchical Cluster Analysis results

Under these assumptions, the *rigid* component represents a relevant share of the total demand (Figure 2.6). The activity pattern (Home-Work) counts 12% of the total demand, while the (Home-School) the 6%. If we consider all the *point-to-point* movements - which means not more than two trips during the day – the percentage rises up to 35%. These percentages, which are similar to other in the literature (Bowman and Ben-Akiva 2001), show how to consider trip chains and daily patterns is fundamental to capture the overall mobility demand. Another observation to report is the relevance of the commuting trips with respect to total amount of the demand. If disaggregate purposes are used to represent the demand, home-work trip represents the most important tour, according to the number of observations. However, if WSA activities are grouped, the *Activity Pattern* (Home-Leisure) becomes the second most relevant in terms of percentage, pointing out that the commuting based demand represents only a small component of the demand.

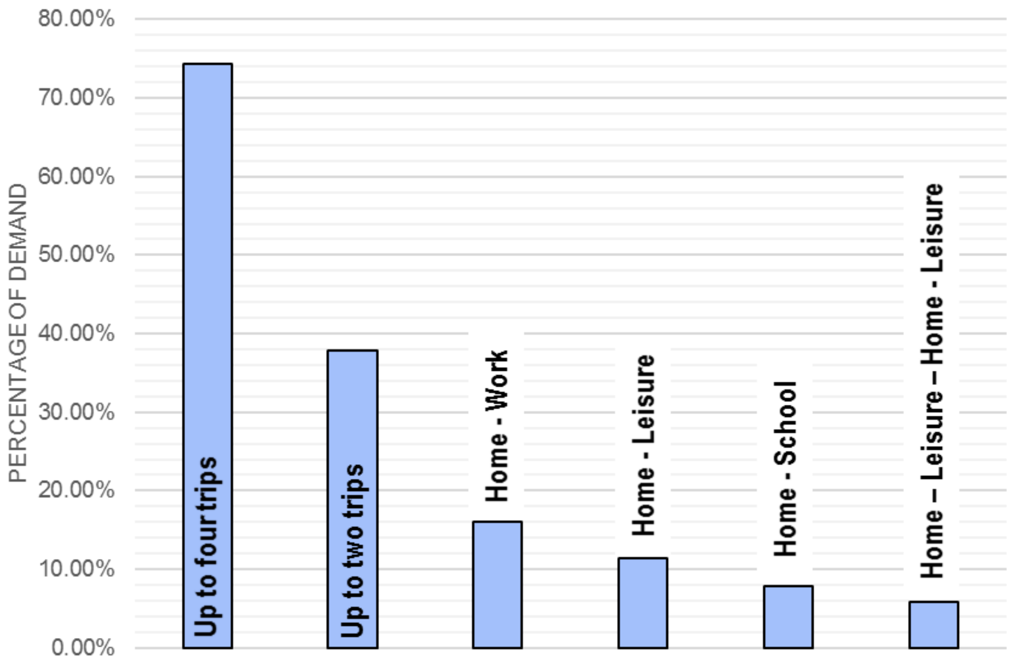


Fig.2.6: Percentage demand for Activity Pattern

After the cluster analysis, 375 *Activity Patterns* have been identified within the BMW database, but only four of them, showed in Figure 2.6, are able to capture 42% of the total demand. These *Activity Patterns* are: (Home-Work); (Home-Leisure) (Home-School) (Home-Leisure-Home-Leisure). If we focus on the number of trips, 75% of the monitored users perform between 2 and four trips during the day. This suggests that the demand model should be able to represent at least four *Activity Patterns* and till four trips for each user.

2.1.2.2 Activity Scheduling and Duration

To analyse at an aggregate level the *Activity Pattern*, three of the most important parameters of the activity chain have been evaluated: *Travel Time distribution*, *Departing Time Distribution* and *Activity Duration*. This information will be used in this section to approximate an *Activity Probability Function.*

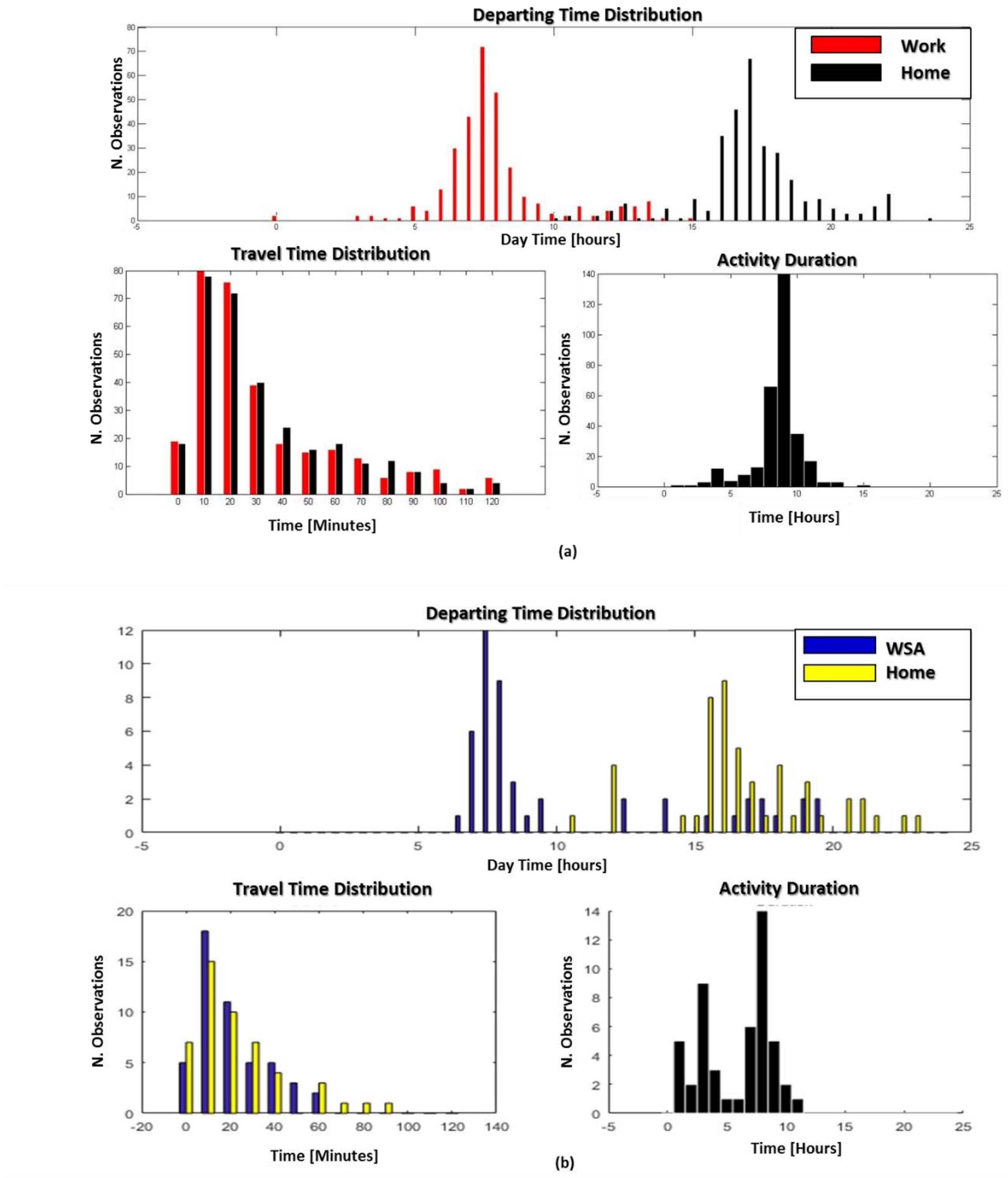


Fig.2.7: Departure time, Travel Time and Activity duration distribution;

In Figure 2.7a, departure time, travel time and the activity duration for the Home-Work commute are presented, while in Figure 2.7b the same parameters are reported for the *WSA*-Home activity pattern. First, we can observe that the *flexible* component of the demand enters into the network after the morning peak when the *rigid* component leaves the system. Second, we can observe a different behaviour at the level of activity duration. While the *rigid* demand component presents a distribution around one single value (9 hours), for the *flexible* one we have two (9 and 2 hours), showing that the duration needs to be properly included in the model in order to account for the intra-cluster heterogeneity of the demand.

Figure 2.7 also shows that the duration of the activity and the departing time can be represented by a probability distribution. We define *Activity Probability Function* the probability function which describes a certain *Activity Pattern*. Specifically, the *mean value* of this distribution is the average departing time for users within one *Activity Pattern,* while the *covariance* term shows the dispersion of the departing time with respect to this value.

For the (Work-Home) tour, the chosen *Activity Function* in this explorative analysis is the Gaussian distribution. This choice is mainly related to the fact that this probability function provides a good approximation with respect to the available data. The equation of the Gaussian function is the following:

$$f(x) = N \cdot e^{-\left(\frac{(x-\mu)}{\sigma^2}\right)^2} \quad (2.1)$$

Where *N* is the number of users belonging to that specific *Activity Component*, μ is the average departure time and σ the variance. The following table contains the fitting parameters obtained by fitting departing time distribution and activity duration for the two *Activity Component:*

Table 2.2: Fitting Parameters Work-Home Activity Components

	N	μ	σ^2
Departure Time from Home in the Morning	31.54	7.521	1.139
Departure Time from Work in the Evening	29.51	16.95	1.126
Duration of the activity “Work”	36.28	8..89	1.015

It is relevant to stress that every component of Equation 2.1, when used to represent the departure time, has a physical meaning, as it represents a different component of the demand. Specifically, the *N* term represents the volume of demand, μ the average departing time and σ^2 its variance.

Here the first possibility to properly predict the traffic state in the afternoon exploring rigid demand properties appears. If we estimate the (correct) activity function for the morning, a good approximation of the evening rush hour can be performed. We can estimate the Probability of departing from Work in the evening by assuming that the average departure time is equal to the average departure time in the morning plus the duration of the activity. So we get μ= 8.89+7.521= 16.41 (instead of the real one, which is 16.95), while we assume all the other parameters to be constant. By comparing this values with the real *Activity Function* the discrepancy between the parameters is extremely low.

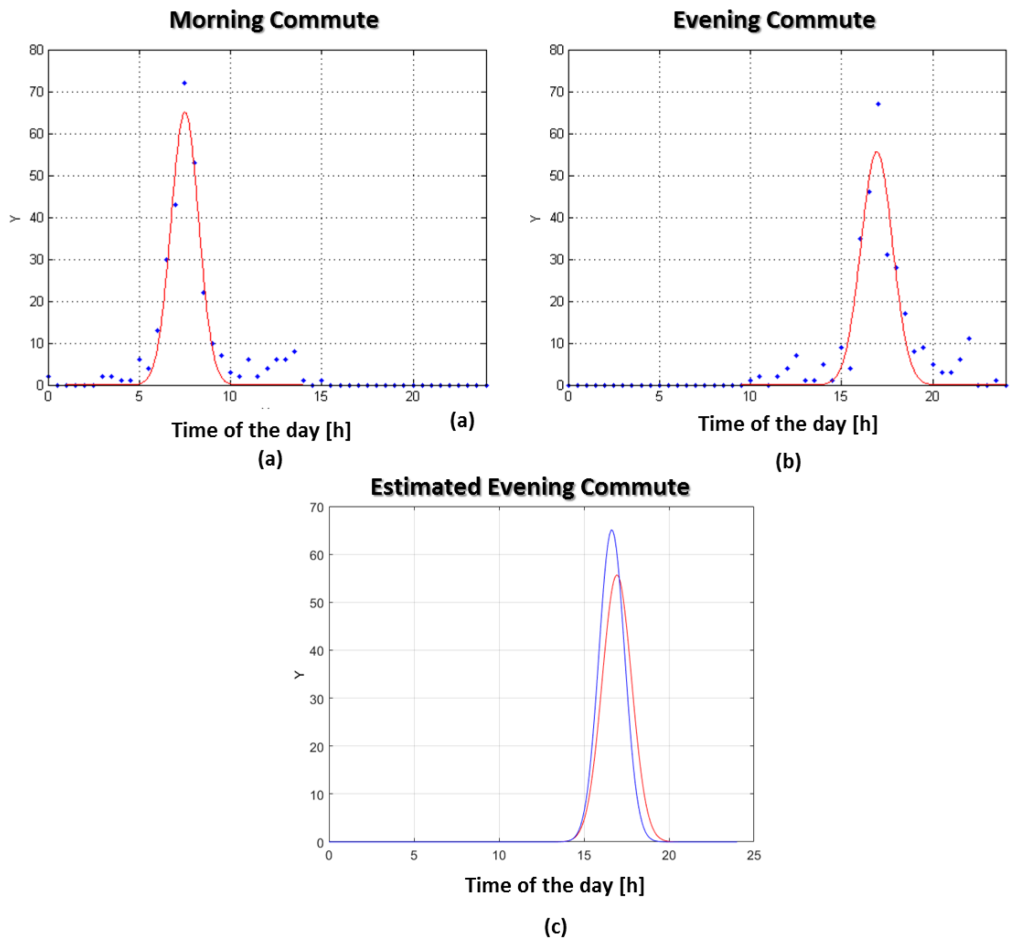


Fig.2.8: (a) Activity Function of the DSA-W demand component; (b) Activity Function of the DSA-H demand component; (c) Comparison between the real DSA-H Activity Function (red), and the estimated one (blue), based on the average duration of the activities;

2.1.3 Conclusions on the Explorative analysis.

In this Chapter, an analysis on the structure of the mobility demand has been proposed. First, we analysed the correlation between traffic data and the temporal distribution of the demand. As only traffic counts were available, our analysis mostly focused on comparing the traffic profile on the detectors with the one obtained from the travel surveys. We observed that the time-dependent demand profile strongly changes over space and that some detectors are more likely to capture this difference. Second, we showed that only detectors located at the entrance of a traffic zone are likely to capture atypical mobility patterns and day-to-day evolution of the demand.

This suggests that two options can be considered to explicitly account for this issue. One way it to tackle this issue from the data point of view. A possibility is to formulate the problem as a sensor location problem, i.e. search for those links that capture these trends or to use new data sources (Viti et al. 2014). A second option is to explore the possibility to use a model for generating the dynamic OD matrix that explicitly accounts for such dynamics. In this dissertation, we explore this second option.

The second part of this chapter analysed the opportunity of breaking down the mobility demand from one unique irregular flow to several Activity-based demand patterns – called *Activity Components.* After analysing the available data, the conclusion is twofold.

1) The demand can be considered as the convolution of different activities, and each activity can be modelled as a simple function.
2) Simple Functions, such as the Gaussian distribution, might be useful for estimating the user behaviour at an aggregate level. However, results show that, although the probability function captures the regular behaviour, the fitting with respect to the actual data is poor (Figure 2.8a-b).

Based on these conclusions, two different models for solving the DODE should be considered. As proposed in (Zhou and Mahmassani 2007) the OD demand is considered as the convolution of three functional functions: *Regular Demand Pattern*, *Structural Deviations* and *Random Fluctuations*. *Regular Demand Patterns* represent the typical demand profile, *Structural Deviations* take into account those phenomena, such as weather condition, that the analyst can model, while random fluctuation considers that deviation that cannot be explained.

Many DODE approaches have been proven to be efficient when a "good" Dynamic OD matrix exists. However, this assumption holds only if the underlying Activity-Based structure of the demand is correct, thus the *Regular Demand Pattern* is unbiased. Under this assumption, we propose a "conventional" framework for the DODE, which aims at correcting the *Structural* and *Random* fluctuations of the demand. The goal is to achieve a good fit between simulated and observed Traffic Data while at the same time increase the reliability of the state of the art methodologies. This model is presented in the first part of this manuscript, which deals indeed with estimating dynamic OD flows with Irregular mobility Patterns.

The second part of this thesis deals instead with the complementary problem, thus to update the dynamic OD flows when the *Regular Demand Pattern* contains substantial errors. In this case, we propose to subdivide the demand into several *Activity Components* and to use probability functions to estimate their shape. Rather than fitting the data, the goal is in this case to generate a realistic dynamic OD matrix. This model is presented in the second part of this manuscript, which deals with the estimation of Regular Mobility Patterns.

2.2 References

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

Castaigne M., Cornelis E., Walle F., Frederix R., , Chris MJ Tampere, Toint P., Viti, F. 2010. "Analyzing Weekly Activity-Travel Behaviour from Behavioural Survey and Traffic Data." 12th WCTR, July 11-15, 2010 – Lisbon, Portugal.

Viti, F., M. Rinaldi, F. Corman, and C.M.J. Tampere. 2014. "Assessing Partial Observability in Network Sensor Location Problems." T*ransportation Research Part B: Methodological* 70:65–89. https://doi.org/10.1016/j.trb.2014.08.002.

Zhou, Xuesong, and Hani S. Mahmassani. 2007. "A Structural State Space Model for Real-Time Traffic Origin–destination Demand Estimation and Prediction in a Day-to-Day Learning Framework." *Transportation Research Part B: Methodological* 41 (8): 823–40. doi:10.1016/j.trb.2007.02.004.

PART I

Irregular Mobility Patterns

3

The Two-Step Approach

As discussed in the previous Chapters, most state of the art DODE models have difficulties in reproducing the correct traffic regime if the initial dynamic OD matrix is not sufficiently close to the real one. In this chapter, a new and intuitive procedure to specify an opportune starting seed matrix is proposed: it is a two-step procedure based on the concept of dividing the problem into small-size problems, focusing at each step on specific set of OD pairs.

The first step focuses on the optimization of a subset of OD variables. In the second step, the framework works on all the OD pairs, using as starting matrix the matrix derived from the first step. In this way is possible to use a more performing optimization method for each step, improving the performance of the method and the quality of the result with respect to the classical approach. In this chapter, the procedure is tested on the real network of Antwerp, showing that the proposed model can avoid systematic errors due to adopting a certain "good" initial matrix and improve the overall performances of the model.

The Content of this chapter has been presented in the following works:

Cantelmo, Guido, Francesco Viti, Chris Tampère, Ernesto Cipriani, and Marialisa Nigro. 2014. "Two-Step Approach for Correction of Seed Matrix in Dynamic Demand Estimation." *Transportation Research Record: Journal of the Transportation Research Board* (December): 125–33. doi:10.3141/2466-14.

Cantelmo, Guido, Francesco Viti, Chris Tampère, Ernesto Cipriani, and Marialisa Nigro. 2014. "Two-Step Approach for Correction of Seed Matrix in Dynamic Demand Estimation." *Transportation Research Board 2014*

3.1 Introduction and Literature Review

Traffic congestion, especially in urban networks, is nowadays a relevant societal problem, and of primal interest in traffic engineering. Typically, congestion phenomena are due to bottlenecks that cause the propagation of congestion on the network, making very difficult to trace back its real causes. A correct representation of the spread of congestion, which is essential for the proper evaluation of management operations, requires tools capable of simulating and predicting time-dependent network traffic conditions.

The dynamic demand estimation (or the demand adjustment, if we start from a known OD matrix usually derived by a combination of surveys and mathematical models) searches for temporal OD matrices that best fit link measurements as traffic counts. The problem is well-known in both the off-line (medium-long term planning and design) and in the on-line (real-time management) context. Cascetta et al. (1993) proposed to face the problem using a sequential or a simultaneous approach: the first makes the demand estimation for each single time slice, holding constant the others. In the simultaneous approach the matrices of every time slice are perturbed simultaneously to guarantee full consistency between estimation periods. This approach is virtually more correct than the sequential one, as it takes into account the relationship among different OD pairs. On the other hand, the computational time is higher, so it is commonly adopted only for the off-line context.

Different approaches and solution algorithms have been developed in the last years for both off-line and on-line dynamic OD estimation; it is possible to distinguish between formulating the estimation as a single level optimization problem (Zhou, Lu, and Zhang 2012) , or as a bi-level optimization problem, as in (Cipriani, Gemma, and Nigro 2013; Yang 1995); another classification distinguishes approaches explicitly using the assignment matrix as a link between traffic counts and demand (Cascetta, Inaudi, and Marquis 1993), or approaches using a linear approximation of the assignment matrix (R. Frederix et al. 2011; Toledo and Kolechkina 2013), or assignment-free approaches (Cremer and Keller 1984).

About the solution algorithms, it is well known the effectiveness of Kalman filtering, especially for capturing day-to-day dynamics (Zhou and Mahmassani 2007) or for on-line estimation (Ashok 1996; Ashok and Ben-Akiva 2000); however, also studies on the Kalman filter for the off-line context are known (Balakrishna, Koutsopoulos, and Ben-Akiva 2005). New stochastic solution approaches have been recently proposed by (Antoniou et al. 2009) and (E. Cipriani et al. 2011).

Different authors focused on the problem of increasing the amount of information required by the estimation including in the objective function of the problem adding further measures compared to the traditional traffic counts, which are not able alone to discriminate between the congested or uncongested state of the network: for example, link speed and occupancy measurements have been proposed by (Balakrishna 2006) and (R. Frederix, Viti, and Tampère 2010), probe data from vehicle equipped by AVI tags by (Dixon and Rilett 2002; Zhou 2004; Eisenman and List 2004; Caceres, Wideberg, and Benitez 2007; Barceló et al. 2012; Mitsakis et al. 2013), aggregate demand data such as traffic emissions and attractions by zones by (Iannò and Postorino 2002; E. Cipriani et al. 2011), turning movements by (Choi et al. 2009), trajectories data and scheduling (Kim and Jayakrishnan 2010).

The majority of the approaches reported in literature focus on the estimation of the dynamic OD matrix from the assumption that a good starting matrix (here called seed matrix) is available. This is not always possible, although the quality of the seed matrix can deeply influence the estimation result (Bierlaire and Crittin 2004; Ernesto Cipriani et al. 2013).

Starting from these remarks, this study aims at proposing a method which, based on state-of-the-art Dynamic OD demand Estimation (DODE) procedures, allows to build a proper dynamic seed matrix to be used as input in the estimation problem. Therefore, we start by presenting and testing different deterministic and stochastic optimization methods to solve the estimation problem; once verified the difficulties of these methods in obtaining a demand able to reproduce the correct traffic regime on the network, as discussed in (Rodric Frederix, Viti, and Tampère 2013), a Two-Step procedure is proposed in order to improve the quality of the seed matrix.

3.2 Methodology

The dynamic demand estimation problem is generally solved as an optimization problem. Its formulation requires the specification of the objective function, also known as goal function, its variables, elements of the OD demand matrix to be estimated, and its constraints related to feasibility and routing conditions. Specifically, the aim of the estimation problem is to find the matrix that minimizes the distances with respect to both the traffic measurements and the seed matrix. Cascetta and Nguyen (1988) formalized the problem in the static case as:

$$\boldsymbol{d}^* = argmin\left[z_1\left(\boldsymbol{x},\widehat{\boldsymbol{d}}\right) + z_2\left(\boldsymbol{v}(\boldsymbol{x}),\hat{\boldsymbol{l}}\right)\right] \tag{3.1}$$

Where:

v(x)=***Mx***, with ***M***= Static Assignment Matrix

The corrected-estimated matrix **d*** is the one that minimizes its inaccuracy in replicating measurements $\hat{\boldsymbol{l}}$, once assigned on the network, while trying not to move away from the seed-starting matrix $\widehat{\boldsymbol{d}}$. The functions z_1 and z_2 are estimators of these measures. Among others, these functions are defined according to the maximum likelihood or generalized mean square error (GLS) theory.

The most common traffic measurements adopted for the Dynamic Demand Estimation are flow, density and speed observations collected on the roads using different types of sources. Being the problem underdetermined (more unknowns than observations), especially when only link measurements are available, multiple matrices could generate the correct regime on the network. In order to overcome this issue, additional a priori information on demand matrix must be added in the problem: this is the reason why the distance with respect to the seed matrix is usually included in the goal function. The generic goal function, using a simultaneous approach on the variables, has the following form (Toledo and Kolechkina 2013):

$$(\boldsymbol{d}_1^*, \dots, \boldsymbol{d}_n^*) = argmin \begin{bmatrix} z_1\left(\boldsymbol{l}_1, \dots, \boldsymbol{l}_n, \widehat{\boldsymbol{l}_1}, \dots, \widehat{\boldsymbol{l}_n}\right) + \\ +z_2(\boldsymbol{q}_1, \dots, \boldsymbol{q}_n, \widehat{\boldsymbol{q}_1}, \dots, \widehat{\boldsymbol{q}_n}) + \\ +z_3\left(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n, \widehat{\boldsymbol{d}_1}, \dots, \widehat{\boldsymbol{d}_n}\right) + \\ +z_4(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n, \widehat{\boldsymbol{r}_1}, \dots, \widehat{\boldsymbol{r}_n}) + \end{bmatrix} \tag{3.2}$$

Where

- $\mathbf{l}/\hat{\mathbf{l}}$ are the simulated values/measurements on the links;
- $\mathbf{q}/\hat{\mathbf{q}}$ are the simulated values/measurements on the nodes;
- $\mathbf{x}/\hat{\mathbf{d}}$ are the estimated/starting value of the demand;
- $\mathbf{r}/\hat{\mathbf{r}}$ are the simulated values/measurements on the route.
- $\boldsymbol{d}_{\boldsymbol{n}}^*$ estimated demand matrix for time interval n;
- $\boldsymbol{z}$ is the estimator

The dependence between simulated information in Equation (3.2) and the estimated demand is obtained directly performing a user equilibrium dynamic traffic assignment (DTA), so that:

$$(\boldsymbol{l}_1, \dots, \boldsymbol{l}_n, \boldsymbol{q}_1, \dots, \boldsymbol{q}_n, \boldsymbol{r}_1, \dots, \boldsymbol{r}_n) = \mathbf{F}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$$

With **F** = user equilibrium Dynamic Traffic Assignment (DTA).

The solution of the problem (3.2) requires the definition of an optimization method and the adoption of an updating rule of the solution during each iteration. Different solution algorithms have been proposed in the past. For a detailed overview we refer to (Lindveld 2003; Balakrishna 2006). Concerning the optimization method, in this study, three path-search methods have been used as reference: the Finite Difference Stochastic Approximation (FDSA), the Simultaneous Perturbation Stochastic Approximation (SPSA) and the Sensitivity-Based OD Estimation (SBODE) method. These are here briefly introduced.

3.2.1 Finite Difference Stochastic Approximation (FDSA)

The FDSA (Finite Difference Stochastic Approximation (Kiefer and Wolfowitz 1952)) is a method usually adopted when dealing with stochastic measurements. It obtains the descent direction perturbing every variable (OD pair) in the matrix as in Equation (3.3):

$$\boldsymbol{\theta}^{i+1} = \boldsymbol{\theta}^{i} - \alpha^{i}\boldsymbol{G}^{i} \tag{3.3}$$

where **θ** is the matrix for the iteration *i*, α is the step length and $\mathbf{G}_i$ is the gradient. The gradient is obtained as follows:

$$\boldsymbol{G}^{i}(\boldsymbol{\theta}^{i}) = \begin{bmatrix} \dfrac{z(\boldsymbol{\theta}^{i} + c^{i}\boldsymbol{\xi}^{1}) - z(\boldsymbol{\theta}^{i})}{c^{i}} \\ \vdots \\ \dfrac{z(\boldsymbol{\theta}^{i} + c^{i}\boldsymbol{\xi}^{r}) - z(\boldsymbol{\theta}^{i})}{c^{i}} \end{bmatrix} \tag{3.4}$$

where **ξ** is the vector with zeros, except for the OD pair to be perturbed, and c^i is the step. In this method every OD pair is perturbed independently, so the number of simulations required for computing the gradient in any iteration is equal to the number of the OD pairs plus the value of *z* in the starting point.

3.2.2 Simultaneous Perturbation Stochastic Approximation (SPSA)

The Simultaneous Perturbation Stochastic Approximation (SPSA (Spall 2012)) is a path search optimization method, where an approximation of the gradient is computed based on a simultaneous perturbation of all the variables. In the SPSA, the equation to update the matrix is the standard formulation reported in Eqaution (3.3). The gradient **G** is obtained in this model with a numeric perturbation of the matrix **θ,** as follows:

$$\widehat{\boldsymbol{g}}_{k}(\boldsymbol{\theta}^{i}) = \frac{z(\boldsymbol{\theta}^{i} + c^{i}\Delta^{k}) - z(\boldsymbol{\theta}^{i})}{c^{i}} \begin{bmatrix} (\Delta_{1}^{k}) \\ \vdots \\ (\Delta_{r}^{k}) \end{bmatrix} \tag{3.5}$$

$$\boldsymbol{G^{i}} = \overline{\boldsymbol{g}}(\boldsymbol{\theta^{i}}) = \frac{\sum_{\boldsymbol{k=1}}^{\boldsymbol{Grad\_rep}} \widehat{\boldsymbol{g}}_{k}(\boldsymbol{\theta}^{i})}{Grad\_rep} \tag{3.6}$$

With c^i the perturbation step. *Grad_rep* is the number of the gradient replications.

With respect to the FDSA, the gradient has a stochastic component, but the computational time to obtain the descent direction is smaller. In the equation above, the formulation of the SPSA model is presented with the asymmetric perturbation. The model formulated in this way takes the name SPSA-AD (Asymmetric Design (E. Cipriani et al. 2011)). The advantage of using this formulation is that the number of assignment needed to compute the gradient is reduced of the 50% with respect to the basic SPSA with symmetric design (SD). Both these variants (SPSA, SPSA AD) will be tested on the case study.

3.2.3 Sensitivity-Based OD Estimation (SBODE)

The last method that has been considered in this study is the Sensitivity-Based OD Estimation model (SBODE (Rodric Frederix et al. 2011)). The SBODE model is based on the idea of perturbing every OD pair as in the FDSA method. The first step, after the initialization of the variables, is the simulation of the starting matrix, to obtain the goal function value and the link flows on the network. Then, the Jacobian is obtained from the starting matrix, perturbing every OD pair. In this case, the higher the dimension of the OD matrix is, the longer will be the computational time (the algorithm requires one simulation for every OD pair perturbed). The SBODE model starts from the standard Gauss-Newton method to update the solution at the *i*-th iteration:

$$\boldsymbol{\theta}^{i+1} = \boldsymbol{\theta}^{i} + \boldsymbol{\alpha}^{\boldsymbol{i}}\boldsymbol{p}^{\boldsymbol{i}} \tag{3.7}$$

$$\boldsymbol{p}^{\boldsymbol{i}} = -(\boldsymbol{J}^T\boldsymbol{J})^{-\mathbf{1}}(\boldsymbol{J}^T\boldsymbol{F}(\boldsymbol{x}_{i-1})) \tag{3.8}$$

Where $\mathbf{F}(\mathbf{x}_{i-1}.)$ is the deviation between the measured and simulated link flows acquired by assigning $\mathbf{x}_{i-1}$, while $\mathbf{J}$ is the Jacobian of $\mathbf{F}(\mathbf{x}_{i-1}.)$. It contains the sensitivity of every link flow deviation to each OD flow. In this model it is possible to include also the deviation from the a priori matrix as regularization term, leading to the following formulation:

$$\boldsymbol{p}^{\boldsymbol{i}} = -(\boldsymbol{J}^T\boldsymbol{J} + \varepsilon\boldsymbol{I})^{-\mathbf{1}}(\boldsymbol{J}^T\boldsymbol{F}(\boldsymbol{x}_{i-1}) - \varepsilon(\boldsymbol{x}_{i-1} - \widetilde{\boldsymbol{x}})) \tag{3.9}$$

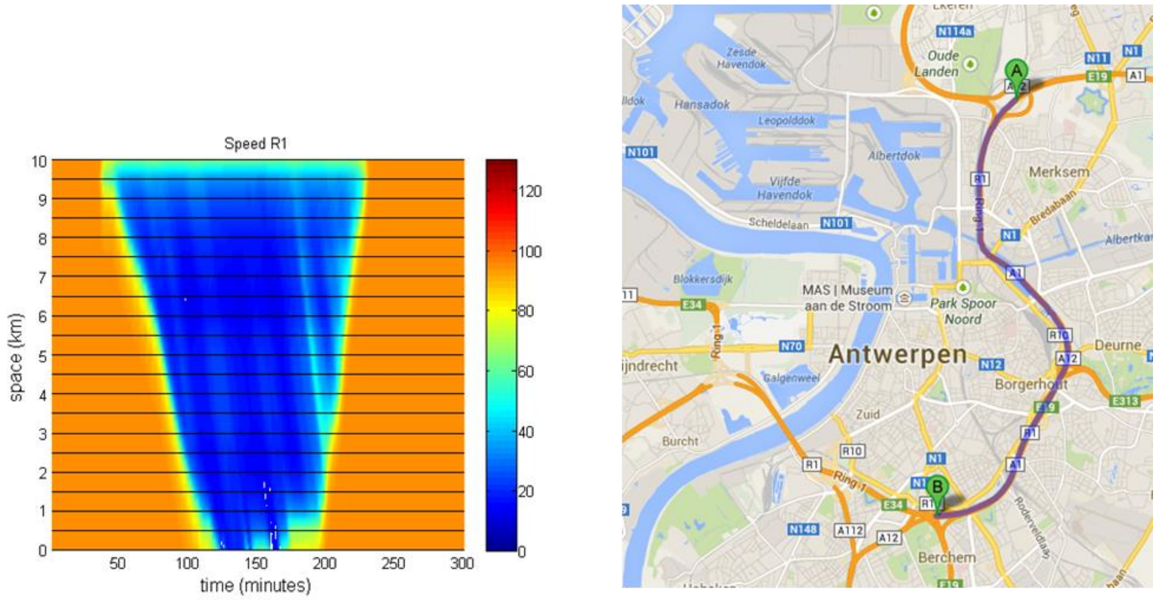
with ε the weight of the regularization term $\boldsymbol{p}^{\boldsymbol{i}}$ is the update vector. It specifies both the direction and the size of the update. This update vector is multiplied with step size parameter $\boldsymbol{\alpha}$ that is determined via a Line Search**.** So the SBODE model uses the Gauss-Newton to obtain the direction, and then uses a Line Search along the direction of $\boldsymbol{p}^{\boldsymbol{i}}$ to find the optimal step. In this Line Search a different goal function can be used, with a Boolean term to check whether the new solution is still in the correct regime.

$$\boldsymbol{G}^k(\boldsymbol{x}) = \frac{\left\|\hat{\boldsymbol{l}} - \boldsymbol{l}(\boldsymbol{x}_k)\right\|_{\mathbf{2}}^{\mathbf{2}}}{\left\|\hat{\boldsymbol{l}} - \boldsymbol{l}(\boldsymbol{x}_{k-1})\right\|_{\mathbf{2}}^{\mathbf{2}}} + \frac{A}{k}\frac{\|\boldsymbol{b} - \boldsymbol{b}(\boldsymbol{x}_k)\|_{\mathbf{2}}^{\mathbf{2}}}{\boldsymbol{I}} \tag{3.10}$$

With $\boldsymbol{l}(\boldsymbol{x}_k)$ and $\hat{\boldsymbol{l}}$ the simulated and measured flows on each link of the network equipped with sensors. Here **b** and b(**x**) are vectors of binary variables indicating whether a link flow is on the corrected branch of the fundamental diagram or not.

3.3 Test of the different estimation approaches

The test case study is related to the inner ring-way around Antwerp, Belgium. The network includes 56 links, 39 nodes, with 46 OD pairs, all mainly connecting the different entry and exit points of this stretch of motorway, making rerouting options not likely. The considered morning peak period occurs between 05:30 and 10:30. The field data – speeds and flows – were available every 5 minutes. The detectors are located at the on and off-ramps and on some intermediate sections. The OD flows have been estimated for 15-minutes departure intervals, so the dynamic matrix contains 966 OD pairs; the seed matrix, that amounts to 202,200 trips, is derived from an existing static OD matrix by superimposing a time profile. Flows of a selection of OD pairs have been increased so obtaining a congestion pattern similar to the actual one. In doing so, the seed matrix implies the correct traffic regime.



(a) (b)

Fig.3.1: (a) x,t plot of the measured speeds on the network, (b) Ring of Antwerp;

For that reason, only flows have been included in the GLS goal function that assumes the following expression:

$$\min f\left(\boldsymbol{\theta}^{i}\right) = \left[h(\boldsymbol{l}-\hat{\boldsymbol{l}})\right] = \sum\nolimits_{l \in D} (\boldsymbol{l}-\hat{\boldsymbol{l}})^{2} \quad (3.11)$$

With $\boldsymbol{l}$ and $\hat{\boldsymbol{l}}$ the simulated and measured flows on each link and D the subset of network links with sensors.

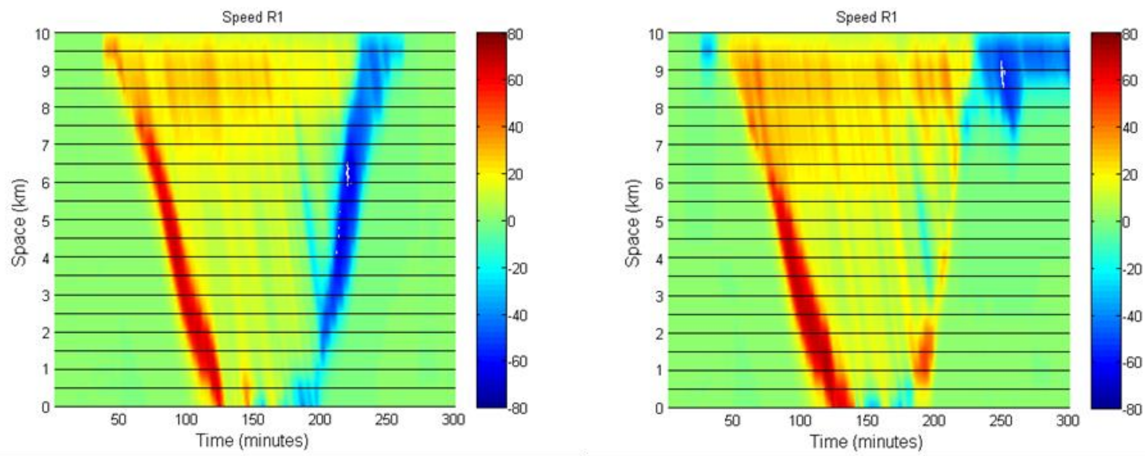
The speed measurements have been used only for validation, since it is expected that if the initial traffic regime is accurately represented on any link, then the new estimated matrix reduces link errors related to flows while preserving such correct traffic regime.

SPSA, SPSA AD and SBODE methods have been used to solve Equation (3.11). In the application of the SPSA, the step c_k is a percentage of the OD pair itself. In this way it is possible to obtain a more representative value of c_k, taking into account the different dimension of the OD pairs, as already reported by other authors (Balakrishna 2006; Cipriani, Gemma, and Nigro 2013; Frederix 2012). Table 3.1 shows the results found with the different methods. Regarding the computational time, these tests have been obtained on computer that, for every iteration, sends the results to a server. This has increased the computational times, as reported in the table. Computational speed is however not a main concern of this study. More comprehensive information about the computational efficiency of this method can be found in others' works (Cipriani, Gemma, and Nigro 2013, 200; Frederix et al. 2011; Cipriani et al. 2011; Cipriani et al. 2013; Frederix 2012; Frederix, Viti, and Tampère 2013). In this study, the computational time is used only to compare the performances of the different solutions, so it is regarded as only a metric.

Table 3.1: Results for each Approach

	SBODE	SPSA (ck=0.01; Grad_rep=50)	SPSA (ck=0.01; Grad_rep=1)	SPSA AD (ck=0.01; Grad_rep=50)	SPSA AD (ck=0.01; Grad_rep=1)
Final O.F. value	6.10E+07	4.29E+08	6.55E+08	3.28E+08	3.52E+08
O.F. improvement [%]	97.08	79.50	68.66	84.29	83.17
Final RMSE on link flows	237.73	628.29	795.44	552.03	570.67
*Final RMSN on link flows[*100]*	7.35	19.44	24.61	17.08	17.65
Final RMSE on link speeds	18.64	17.93	27.72	20.35	20.61
*Final RMSN on link speeds [*100]*	28.93	27.83	38.94	31.58	31.98
N°of iterations	40	93	1,000	273	929
Comp.time for one iteration [min]	420	41	1	21	0.5
Tot comp.time [h]	280.00	63.55	16.67	95.55	7.74
Tot comp.time [days]	11.67	2.65	0.69	3.98	0.32

It is possible to observe that the SBODE model obtains the best improvement of the goal function, but at the same time is implies the greatest computational time. The SPSA-AD is more effective than the basic model. In the following tests the version of the SPSA-AD with c_k equal to 1% and Grad_rep=50 – about the 5% of the Demand Matrix dimension - is used. The SPSA AD with Grad_rep=1 obtained good results, close to the version with *Grad_rep*=50. Anyway by using 50 gradient replications the method results more stable with respect to the basic one.



(a) (b)

Fig.3.2: (a) x,t plot of the differences between simulated and measured speeds on the network for the solution of the SPSA AD, (b) x,t plot of the differences between simulated and measured speeds on the network for the solution of the SBODE;

Concerning the results, it is important to highlight that a congestion pattern very close to the real one (Figure 3.1a) has been obtained with all the tested methods. At the same time, all the methods present an offset in the congestion pattern. This offset is shown in Figure (3.2). In this figure the time-space plots of the vector of the differences between simulated and measured speeds are presented. The red zone on the left, representing an overestimation of the speeds, implies that congestion is estimated to begin later in time with respect to the actual congestion pattern. On the other hand, the blue zone on the right, representing a significant underestimation of the speeds, implies longer recovery time with respect to actual one. This error is present in both the models, deterministic SBODE and stochastic SPSA AD, although they differ from each other significantly, mainly in the congestion recovery part. If the offset is clearly defined in the SPSA AD, this difference is less evident in the SBODE.

3.4 The Two-Step Approach

Both the deterministic SBODE and the stochastic SPSA AD procedures underline the same problem at the end of the estimation: an offset in the representation of the congestion pattern. This result leads to think that the final error is not related to the model adopted, but to the specific case study and in particular of the specific seed matrix, so highlighting the importance of a proper starting point.

As a consequence, a new approach for solving the problem is proposed: this approach – called “Two-Step Approach”- aims at improving the quality of the seed matrix, performing a prior step preceding the optimization process for the estimation of the final matrix. The basic idea is to divide the problem in two small-sized problems, and solve them separately. The procedure is generic and thus applicable to both the SPSA AD and the SBODE method, so gathering information about its general properties. The approach works as follows:

- FIRST STEP: The first step is focused on the optimization of a subset *N* of the OD pairs. The result of this step is to obtain a new matrix with respect to the starting seed matrix (in the remaining of this Chapter called “wrong seed matrix”) to be used as input in the second step. The matrix obtained solving the first step is called in the remaining of the Chapter “correct seed matrix”.
- SECOND STEP: In the second step, the usual estimation procedure of the OD matrix is performed (the same used to obtain the results presented in Table 1), starting from the new “correct seed matrix” obtained from the first step.

Before carrying on the experiments, it is necessary to define the subset *N* of variables. Two ways have been explored in this work. One based on the *data analysis* (Approach 1) and another one more generic (Approach 2) based on the *network analysis*.

3.4.1 Approach 1

In this approach the subset *N* is defined as the subset of OD pairs that generate the greater link flows on the network. The goal function is the same presented in (3.11), so the subset *N* is expected to contain OD pairs related to the most important descent directions for the starting seed matrix. Sometimes a great error is present on the links with high flows. In this situation a wrong matrix – worse with respect to the seed – that generate a correct flow only on the links with the greater link flows could aim to a reduction of the goal function value. This reduction is due to an irregularity of the goal function that generates a local minimum in correspondence to a matrix strongly different from the real one. So it is possible to select the OD pairs that have more influence on this links, correct them and reduce the number of minimum points generating a more clear descent direction to the global minimum of the function. Furthermore, in the specific case study, the greater errors of the wrong seed matrix were on the higher links flows, so by doing so, we focus on the part of the goal function that contributes to the largest gain. An analogous approach was recently proposed by (Djukic et al. 2012), where Principal Component Analysis is proposed for lowering the number of OD demand variables, as it is proposed in this study.

In the first step, 126 OD pairs out of 966 have been selected to be included in the optimization method. The selected OD pairs are responsible of the flows higher than 8000 veh/h, the highest on the network. This OD pairs are an important quote of the wrong seed matrix, generating 40,480 trips on 202,000.

Taking into account the smaller number of variables, and the will to obtain a good gradient to correct the wrong seed matrix along the main descent direction, the method chosen for the optimization is the deterministic gradient FDSA. Starting from the results of the FDSA, both the SPSA AD and the SBODE are then applied. The matrix obtained by the optimization of the second step is called in the rest of the article "final matrix". The results are presented in Table 3.2.

Table 3.2: Results for the Two-Step Approach 1

Optimization Method	final O.F. value	O.F. Improvement [%]	Final RMSE on link flows	Final RMSN on link flows [*100]	Final RMSE on link speeds	Final RMSN on link speeds [*100]	N. Iterations
Step 1 (FDSA)	5.80E+08	72.26	730.385	22.6	18.47	28.67	53
Step 2 (SPSA AD)	4.49E+08	78.52	644.82	19.95	13.67	21.16	508
Step 2 (SBODE)	7.85E+07	96.24	269.61	8.34	16.76	26	7

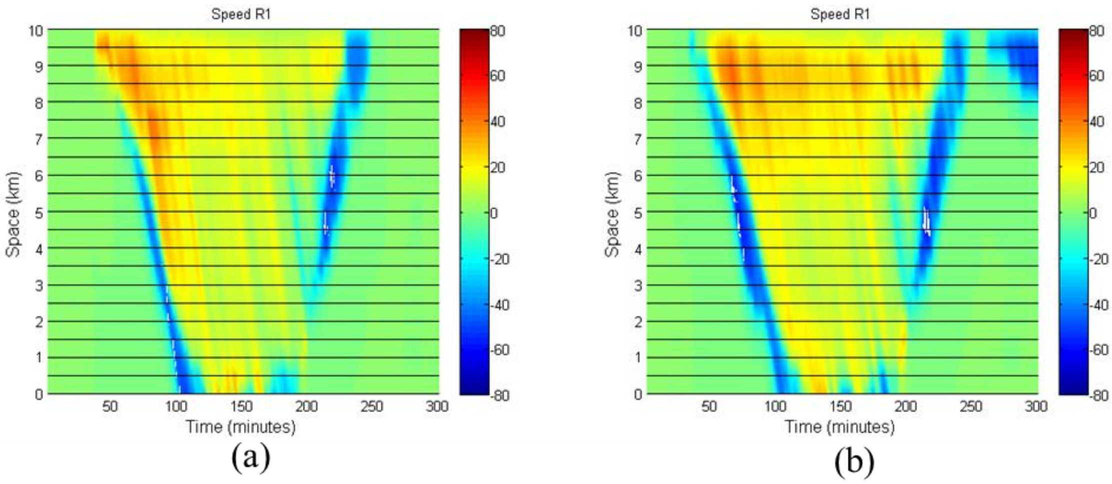


Fig.3.3: x,t plot of the differences between simulated and measured speeds on the network for the solution of the two-step approach for the SPSA AD (a) and the SBODE (b);

The results of the Two-Step approach using Approach 1 can be summarized as follows:

1) Using SPSA-AD:
   a. The final deviation is greater with respect to the basic one step SPSA-AD.
   b. The best congestion pattern is obtained, as shown in Figure 3.3b.
   c. The speeds RMSE is equal to 13.67, which is lower than the basic SBODE and all the others models.
   d. The absolute distance between the final matrix and the wrong seed matrix is equal to 6.29E+04; the absolute distance between the wrong and the correct seed matrix was equal to 5.10E+04. In the basic SPSA-AD the distance between the final matrix and the wrong matrix was equal to 9.18E+06, so the result obtained by the two step approach is closer to the seed matrix.
   e. The congestion pattern has a longer duration than the real one as shown in Figure 3.3a.
2) Using SBODE:
   a. Also in this case the final deviation of the goal function is greater with respect to the basic one step version.
   b. The congestion pattern is better than the basic SBODE, as shown in the Fig. 3.3b, and by the RMSE of the link speed, equal to 16.76.

c. The final matrix is closer to the wrong seed matrix. The absolute distance between the final matrix and the wrong seed matrix is equal to 1.75E+05 travellers; The distance between the wrong and the correct seed was equal to 5.10E+04. For the “one step” approach the distance between the SBODE final matrix and the wrong seed matrix was equal to 1.89E+05, so in the second step the algorithm is closer to the seed matrix.
d. The congestion pattern has a longer duration than the real one and the offset disappeares (Fig.3.3b).

In both situations, using the SBODE or the SPSA AD in the second step, the offset disappears, but the error on the congestion patter is again on the boundary of the congestion period. In the Two-Step case the congestion is slightly longer with respect to the real one, anyway the error is smaller than the one obtained with the basic approach, as demonstrated by the values of RMSE/RMSN of the speeds.

For both the SPSA AD and the SBODE method, the Two-Step results present a smaller error both on the speeds both on the distance from the wrong seed matrix. Typically these terms are included in the goal function, while in this experiment are used only to verify the quality of the results. It is important to observe as this two measurements, not directly included in the goal function, have been reduced using a “two-step” approach and correcting the wrong seed matrix.

Using SBODE in the second step, the computational time is lower than the computational time for the SBODE in the one step approach. The adoption of the two-step approach allows a lower improvement of the goal function (96.24% against 97.08% with the basic SBODE) with a significant reduction of the computational time (from 11.67 to 3.88 days).

However, we have to underline that it is not possible to completely solve the problem with simply the introduction of the two-step with respect to the “one-step” approach, since we have adopted a first-order network loading model, which has inherent simplifications in the way traffic back propagates and congestion discharges. Furthermore it also to observe how the error in the boundary will be ever the highest; to delete this error it is necessary that the matrix reproduces exactly the real speeds on the network.

3.4.2 Approach 2

The Approach 1 relies on the basic assumption that bigger OD demand flows, corrected in the first step, are mainly accountable of the descent direction. Despite this approach is faster than the basic method, it is not however proven that the main descent directions imply a final matrix closer to the real one. In order to generalize such property, another subset N of OD pairs can be proposed. The idea is to obtain the correct regime on the bottleneck in the first step, and to use the second step to obtain the final demand estimation. Specifically, in the first step, the only variables considered in the optimization procedure are the OD demand flows passing on the links where bottlenecks are located. In the second step, as in the previous case, the global optimization is performed. In summary, in this second approach 630 OD pairs out of 966 are perturbed with the FDSA in the first step, while in the second step all the 966 OD pairs are included in the optimization, using both the SPSA AD and the SBODE.

The main problem of this Approach 2 is the FDSA itself, since the number of variables to be optimized is very high and, differently from SBODE, the model uses a constant step (not line search). For this reason the computational time is very high and equals to 9 days. The value of the goal function for the matrix obtained from FDSA in the first step is similar to the value obtained in the basic SPSA AD, while the error on the speeds is smaller. The only optimization with a better RMSE/RMSN of the link speeds is the optimization obtained in the Approach 1 using the SPSA AD in the second step (Table 3.2).

The absolute distance of the correct seed matrix obtained with the Approach 2 from the wrong seed matrix is equal to 6.24E+04, so it is smaller than the distance obtained with both the basic one step models. Starting from the result of the FDSA, the second-step optimization is obtained with both the SPSA AD and the SBODE.

Table 3.3: Results for the Two-Step Approach 2

Optimization Method	final O.F. value	O.F. Improvement [%]	Final RMSE on link flows	Final RMSN on link flows [*100]	Final RMSE on link speeds	Final RMSN on link speeds [*100]	N. Iterations
Step 1 (FDSA)	5.80E+08	72.26	730.385	22.6	18.47	28.67	53
Step 2 (SPSA AD)	4.49E+08	78.52	644.82	19.95	13.67	21.16	508
Step 2 (SBODE)	7.85E+07	96.24	269.61	8.34	16.76	26	7

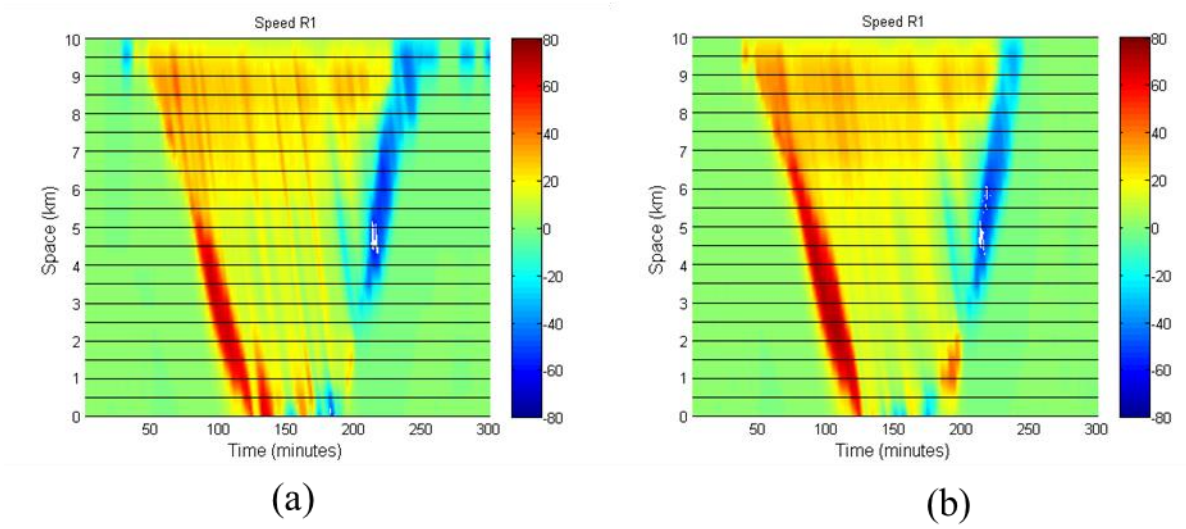


Fig.3.4: x,t plot of the differences between simulated and measured speeds on the network for the solution of the two-step approach for the SPSA AD (a) and the SBODE (b);

Figure 3.4b shows the x-t plots of the differences between simulated and measured speeds for the solution of the second step, obtained using the SPSA AD as optimization method. The error of the speed is very high and the offset in the congestion pattern is greater than in the solution of the FDSA. Table 3.3 shows the results for the first and the second step.

With respect to the basic SPSA AD the final value of the goal function is smaller. The distance between the wrong seed matrix and the final matrix is the highest and is equal to 1.15E+05. As it is possible to observe in Table 3, the error in the speeds and the offset in the congestion pattern is greater than in the previous cases. As for the SPSA AD, also the SBODE has obtained the best value of the goal function, but the computational effort is quite high (14 days for the SBODE while 24 days for the SPSA AD). It is necessary to highlight as the convergence criterion for the SPSA AD is stronger than the convergence criterion for the SBODE. For the SBODE the convergence was obtained fixing the maximum distance between iterations equal to 1. Using this criterion the SPSA AD arrived to the convergence after few iteration obtaining a not satisfactory results, so the maximum distance in the SPSA AD was set to 0.1. Anyway is possible to set the stop criterion as a maximum number of iterations to solve the problem. After 250 iteration the final value of the goal function is 1.96E+08. The high computational time of this approach is related to the first step, the FDSA. One way to solve this problem is to use the SBODE itself in the first step.

3.4.3 Summary of the results

With reference to the results above, the following considerations and conclusions can be done:

Approach 1:

- It is possible to improve the effectiveness and efficiency of the classical “one step” optimization passing to a “Two-Step” approach, that work in a first moment only on the most important descend directions of the goal function, using an optimization method more suitable, and in a second step, starting from the correct seed matrix, obtain the final matrix using a different and more appropriate optimization method.
- Dividing the problem in two problems leads to a better result, reducing the noises in the goal function and so the errors in the measurements not directly included in the goal functions. So, in this approach, two fundamental indicators, the speed and the seed matrix, not directly considered in the goal function, are improved thanks to the use of the “Two-Step” approach. This is the demonstration that it is possible to work in a first step on the correction of the seed matrix, and only in a second time on the estimation problem
- There is a strong reduction of the computational time for the SBODE method, related to the new starting matrix.

Approach 2:

- It is possible to correct the starting matrix working only on the OD pairs that have a greater influence on the bottleneck, obtaining a very good matrix, close to the real one, as shown from the final value of the goal function
- The improvement with respect to the “one step” approach is low and not satisfactory taking into account the total computational time.
- Taking into account the results of the second experiment, and especially for big-size networks with a large number of OD pairs, it is possible to select a subset of them where to perform OD estimation.
- It is necessary to highlight as the high computational time is related not to the SPSA AD or the SBODE, but to the FDSA itself. FDSA has previously been rejected for calibration for this reason. In this case study it is used for simplicity and because it is a well-known algorithm. It is not advisable to use this method in other applications. One way to solve this problem, if an exact gradient is required in the first step, is to use the SBODE method itself in the first step.

Both the approaches show as dividing the space of the solutions, and solving with different methods the Two-Step could lead to better results. It is possible to see the first step as a way to obtain a new correct seed matrix, that is simpler to study and less influenced from the noise of the goal function.

3.5 Conclusions and Future Research

The main goal of the present Chapter is to propose a method for determining a starting demand that, when utilized in the dynamic demand estimation problem, improves the accuracy of the estimated matrix in reproducing the correct traffic regime on the network.

In this Chapter, different deterministic and stochastic solution procedures commonly adopted in literature are firstly presented and tested for the off-line dynamic demand estimation on the real case study of the inner ring of Antwerp in Belgium.

Both the deterministic and the stochastic procedures underline the same problem at the end of the estimation: an offset in the representation of the congestion pattern, with high differences in the congestion recovery part. This result leads to think that the final error is not related to the model adopted, but to the specific case study and in particular of the specific seed matrix adopted so highlighting the importance of a proper starting point.

Following a Two-Step procedure, the wrong seed matrix was modified to obtain a new correct matrix for the estimation problem. Specifically, the first step of the procedure focused on the optimization of a subset of OD variables, adopting two different approaches: the first approach considered as variables only those relative to ODs that generate the higher flows, while in the second approach only ODs generating bottlenecks on the network. In the second step, the optimization works on all the OD pairs, using as starting matrix the matrix derived from the first step. Using the new starting matrix from step 1 implies results that differ according to the adopted approach: specifically, using the deterministic method in the Two-Step procedure it was possible to obtain better solutions with the same of the goal function and with a reduction of the computational time. It was also possible to obtain a better result, with a higher improvement of the goal function, but with an increment in the computational time.

Working on the subset *N* is possible to arrive closer to the real demand matrix. If the computational time to obtain the final matrix from the correct seed matrix depend only from the adopted optimization method, is possible to choose in the first step how to obtain this new matrix. It is possible to arrive to the correct seed matrix following a “faster” way, as proposed by the approach 1, or it is possible to arrive to the correct seed matrix working only on the most important OD pairs, as in the Approach 2. It is possible, carefully choosing which elements insert into the *N*, to work on the effectiveness and efficiently of the results.

The most important result, however, is that it is possible to improve the quality of the estimated matrix without introducing new measurements or developing new models, but only working in different ways on the different OD pairs. In conclusion, it is necessary to highlight that using a two-step method it is possible to combine different kind of models, using not only path-search methods, but combining also random search and pattern search methods, based on the specific configuration of the network and of the problem.

Future developments will deal with more complex networks, because the case study focuses on highways, so the problem results quite simple with respect to an urban network. Moreover the goal function takes into account only the link flows, so it is necessary to understand if the method confirms the same features also if other measurements, more representative of the congestion state, are considered inside the goal function. Finally it is important not only to understand on which OD pairs is preferable to work, but also to develop a proper goal function that could take into account other information on the real matrix as a good OD trip distribution.

Acknowledgements

We wish to acknowledge the COST Action TU1004 ‘TransITS’, which has partly sponsored the collaboration between the authors. We wish also to acknowledge Ruben Corthout, Rodric Frederix and Willem Himpe for the assistance in the preparation of the tests and in the implementation of the algorithms.

3.6 References

Antoniou, C., R. Balakrishna, H.N. Koutsopoulos, and M. Ben-Akiva. 2009. "Off-Line and on-Line Calibration of Dynamic Traffic Assignment Systems." In , 104–11. doi:10.3182/20090902-3-US-2007.0056.

Ashok, K., and M.E. Ben-Akiva. 2000. "Alternative Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin-Destination Flows." *Transportation Science* 34 (1): 21–36.

Ashok, Kalidas. 1996. "Estimation and Prediction of Time-Dependent Origin-Destination Flows." Massachusetts Institute of Technology. https://its.mit.edu/sites/default/files/documents/AshokPhd_0.pdf.

Balakrishna, Ramachandran. 2006. "Off-Line Calibration of Dynamic Traffic Assignment Models." PhD Thesis, Massachusetts Institute of Technology. http://dspace.mit.edu/handle/1721.1/35120.

Balakrishna, Ramachandran, Haris N. Koutsopoulos, and Moshe E. Ben-Akiva. 2005. "Calibration and Validation of Dynamic Traffic Assignment Systems." In . http://trid.trb.org/view.aspx?id=758898.

Barceló, J., L. Montero, M. Bullejos, O. Serch, and C. Carmona. 2012. "Dynamic OD Matrix Estimation Exploiting Bluetooth Data in Urban Networks." In *Proceedings of the 14th International Conference on Automatic Control, Modelling & Simulation, and Proceedings of the 11th International Conference on Microelectronics, Nanoelectronics, Optoelectronics*, 116–121. ACMIN'12. Stevens Point, Wisconsin, USA: World Scientific and Engineering Academy and Society (WSEAS). http://dl.acm.org/citation.cfm?id=2230549.2230572.

Bierlaire, M., and F. Crittin. 2004. "An Efficient Algorithm for Real-Time Estimation and Prediction of Dynamic OD Tables." *Operations Research* 52 (1): 116–27. doi:10.1287/opre.1030.0071.

Caceres, N., J.P. Wideberg, and F.G. Benitez. 2007. "Deriving Origin-Destination Data from a Mobile Phone Network." *IET Intelligent Transport Systems* 1 (1): 15–26. doi:10.1049/iet-its:20060020.

Cascetta, Ennio, Domenico Inaudi, and Gerald Marquis. 1993. "Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts." *Transportation Science* 27 (4): 363–73.

Cascetta, Ennio, and Sang Nguyen. 1988. "A Unified Framework for Estimating or Updating Origin/Destination Matrices from Traffic Counts." *Transportation Research Part B: Methodological* 22 (6): 437–55. doi:10.1016/0191-2615(88)90024-0.

Choi, Keechoo, R. Jayakrishnan, Hyunmyung Kim, Inchul Yang, and Jungwoo Lee. 2009. "Dynamic Origin-Destination Estimation Using Dynamic Traffic Simulation Model in an Urban Arterial Corridor." *Transportation Research Record: Journal of the Transportation Research Board* 2133 (December): 133–41. doi:10.3141/2133-14.

Cipriani, E., M. Florian, M. Mahut, and M. Nigro. 2011. "A Gradient Approximation Approach for Adjusting Temporal Origin-Destination Matrices." *Transportation Research Part C: Emerging Technologies* 19 (2): 270–82. doi:10.1016/j.trc.2010.05.013.

Cipriani, E., A. Gemma, and M. Nigro. 2013. "A Bi-Level Gradient Approximation Method for Dynamic Traffic Demand Estimation: Sensitivity Analysis and Adaptive Approach." In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, 2100–2105. doi:10.1109/ITSC.2013.6728539.

Cipriani, Ernesto, Marialisa Nigro, Gaetano Fusco, and Chiara Colombaroni. 2013. "Effectiveness of Link and Path Information on Simultaneous Adjustment of Dynamic O-D Demand Matrix." *European Transport Research Review* 6 (2): 139–48. doi:10.1007/s12544-013-0115-z.

Cremer, M., and H. Keller. 1984. "A SYSTEMS DYNAMICS APPROACH TO THE ESTIMATION OF ENTRY AND EXIT O-D FLOWS." In *Proceedings of 9th ISTTT.*

Dixon, M.P., and L.R. Rilett. 2002. "Real-Time OD Estimation Using Automatic Vehicle Identification and Traffic Count Data." *Computer-Aided Civil and Infrastructure Engineering* 17 (1): 7–21.

Djukic, T., G. Flötteröd, Lint Van, and S. Hoogendoorn. 2012. "Efficient Real Time OD Matrix Estimation Based on Principal Component Analysis." In IEEE Conference on Intelligent Transportation Systems, 115–21. doi:10.1109/ITSC.2012.6338720.

Eisenman, S.M., and G.F. List. 2004. "Using Probe Data to Estimate OD Matrices." In , 291–96.

Frederix, R., F. Viti, R. Corthout, and C.M.J. Tampère. 2011. "New Gradient Approximation Method for Dynamic Origin-Destination Matrix Estimation on Congested Networks." *Transportation Research Record*, Transportation Research Record, , no. 2263: 19–25.

Frederix, R., F. Viti, and C. M. J. Tampère. 2010. "A Density-Based Dynamic OD Estimation Method That Reproduces within-Day Congestion Dynamics." In *13th International IEEE Conference on Intelligent Transportation Systems,* 694–99. doi:10.1109/ITSC.2010.5625220.

Frederix, Rodric. 2012. "Dynamic Origin-Destination Matrix Estimation in Large-Scale Congested Networks." PhD Thesis. https://lirias.kuleuven.be/handle/123456789/344248.

Frederix, Rodric, Francesco Viti, and M.J. Tampère. 2013. "Dynamic Origin–destination Estimation in Congested Networks: Theoretical Findings and Implications in Practice." *Transportmetrica A: Transport Science* 9 (6): 494–513. doi:10.1080/18128602.2011.619587.

Iannò, and M.N. Postorino. 2002. "A Generation Constrained Approach for the Estimation of O/D Trip Matrices from Traffic Counts." In *European Transport Conference,* Strasburg France.

Kiefer, J., and J. Wolfowitz. 1952. "Stochastic Estimation of the Maximum of a Regression Function." *The Annals of Mathematical Statistics* 23 (3): 462–66. doi:10.1214/aoms/1177729392.

Kim, H, and R Jayakrishnan. 2010. "The Estimation of a Time-Dependent OD Trip Table with Vehicle Trajectory Samples, Transportation Planning and Technology." *Transportation Planning and Technology* 33 (8): 747–68. doi:10.1080/03081060.2010.536629.

Lindveld, K. 2003. "Dynamic O-D Matrix Estimation: A Behavioural Approach." PhD Thesis, TU Delft.

Mitsakis, Evangelos, Josep-Maria Salanova Grau, Evangelia Chrysohoou, and Georgia Aifadopoulou. 2013. "A Robust Method for Real Time Estimation of Travel Times for Dense Urban Road Networks Using Point-to-Point Detectors." In . http://trid.trb.org/view.aspx?id=1242148.

Spall, James C. 2012. "Stochastic Optimization." In *Handbook of Computational Statistics,* edited by James E. Gentle, Wolfgang Karl Härdle, and Yuichi Mori, 173–201. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-642-21551-3_7.

Toledo, T., and T. Kolechkina. 2013. "Estimation of Dynamic Origin-Destination Matrices Using Linear Assignment Matrix Approximations." *IEEE Transactions on Intelligent Transportation Systems* 14 (2): 618–26. doi:10.1109/TITS.2012.2226211.

Yang, H. 1995. "Heuristic Algorithms for the Bilevel Origin-Destination Matrix Estimation Problem." *Transportation Research Part B* 29 (4): 231–42. doi:10.1016/0191-2615(95)00003-V.

Zhou, Xuesong. 2004. "DYNAMIC ORIGIN-DESTINATION DEMAND ESTIMATION AND PREDICTION FOR OFF-LINE AND ON-LINE DYNAMIC TRAFFIC ASSIGNMENT OPERATION," August. http://drum.lib.umd.edu/handle/1903/1819.

Zhou, Xuesong, Chung-Cheng Lu, and Kuilin Zhang. 2012. "Dynamic Origin–destination Demand Flow Estimation under Congested Traffic Conditions." In *Proceedings of the 91st Transportation Research Board Annual Meeting, Washington, DC*. http://www.civil.utah.edu/~zhou/Single-level_path-flow-OD-estimation_Zhou-Lu-Zhang.pdf.

Zhou, Xuesong, and Hani S. Mahmassani. 2007. "A Structural State Space Model for Real-Time Traffic Origin–destination Demand Estimation and Prediction in a Day-to-Day Learning Framework." *Transportation Research Part B: Methodological* 41 (8): 823–40. doi:10.1016/j.trb.2007.02.004.

4

Enhanced Two-Step Approach

This chapter focuses on extending the Two-Step approach proposed in the previous Chapter in order to further improve the reliability of the demand matrix and boost the robustness of the solution with respect to typical "Single-Step" formulation. This enhanced version of the Two-Step approach corrects sequentially generations and distributions in the demand matrix, reducing solution space size and the variance in the solutions of the calibration process. The proposed approach is again applied to the real network of Antwerp.

Content of this chapter has been presented in the following works:

Cantelmo, G., F. Viti, E. Cipriani, and N. Marialisa. 2015 "A Two-Step Dynamic Demand Estimation Approach Sequentially Adjusting Generations and Distributions." *In 2015 IEEE 18th International Conference on Intelligent Transportation Systems, 1477–82. doi:10.1109/ITSC.2015.241*

Cantelmo, G., F. Viti, E. Cipriani, and N. Marialisa. 2015. "A Two-Step Dynamic Demand Estimation Approach Sequentially Adjusting Generations and Distributions." *Transportation Research Board 2015*.

4.1 I Introduction

Simulation of traffic conditions requires, as main input, the knowledge of the travel demand. When dealing with transportation networks, traffic conditions are usually not stationary, hence it is recommended to adopt time-dependent profiles of the travel demand to best represent congestion and its propagation. If this information is not available or incorrect, the simulation output performances are compromised. In this chapter a Two-Step procedure that sequentially adjusts demand generations and distributions for improving its reliability is presented. In the first part of this introduction, a short overview of the demand estimation is provided, while in the second part the concept of the Two-Step procedure is introduced.

4.1.1 Literature Review

The problem of estimating travel demand in case of non-stationary conditions is well known in literature as the Dynamic Origin-Destination demand Estimation Problem. DODE searches for time-dependent Origin-Destination (OD) matrices able to best fit measured data. It can be applied for both within-day (intra-period) and day-to-day (inter-period) dynamic frameworks, as well as for offline (medium-long term planning and design) and on-line (real-time management) contexts. DODE is commonly classified in sequential, i.e. OD flows are estimated as sequence of separated intervals within which the estimated demand of one period is used to find the demand for the next period, or simultaneous approaches, where demand flows are estimated all at once (Cascetta, Inaudi, and Marquis 1993). Usually, the first is suited for online applications as it adopts a rolling-horizon approach, while the second for offline applications. Online solution algorithms, based on different state-space representations of traffic flow propagation, and tuned with advanced regression methods such as Kalman filtering (Ashok and Ben-Akiva 2002), are very popular for capturing within-day dynamics and calibrating traffic models (Zhou and Mahmassani 2007) using real-time data (Ashok and Ben-Akiva 2000); however, studies on Kalman filtering are also proposed for the offline context (Zhou and Mahmassani 2007; Ashok and Ben-Akiva 2002). In offline applications DODE is generally formulated as a bi-level optimization problem, where in the upper level demand matrices are corrected using measured data while in the lower level Dynamic Traffic Assignment (DTA) simulation is performed to obtain the synthetic data (Yang 1995; Tavana 2001). Generally, the upper level problem is solved using stochastic or deterministic path search approaches (Frederix 2012). Recently, stochastic solution approaches were proposed along this direction, as in (Antoniou et al. 2009; Cipriani et al. 2011).

Another classification of the DODE can be done according to the type of observed data adopted for the estimation: usually traffic counts are adopted, but recently also other measures such as speeds and occupancies are introduced to take into account the congestion state of the network (Balakrishna 2006). These data, as the traffic counts, are commonly link-based, while also other path-based data can be added as probe data from vehicle equipped by AVI tags (Balakrishna 2006; Dixon and Rilett 2002; Caceres, Wideberg, and Benitez 2007; Barceló et al. 2013). When only traffic counts are adopted for the estimation, the link between dynamic travel demand and measurements is usually captured by the assignment matrices (explicitly, as in (Cascetta et al. 2013), or by a linear approximation (Frederix et al. 2011; Toledo and Kolechkina 2013). Given the non-linear structure of the (bi-level) problem formulation, the complex mapping between OD flows, path flows and link flows represented by the assignment matrices, and the variety of solution algorithms available, one of the drawbacks of current dynamic demand estimation techniques is the reliability of its outcomes, i.e. often the solutions found by varying slightly the input, as well as the estimation approach, results in significant variations of the results, i.e. the resulting estimated demand flows can be significantly different even if observed link flows are not changing dramatically, and in case of stochastic solution approaches may even not change at all.

The reliability of demand estimation is well known to depend on the location of sensors as well as on the quality of the data (Cascetta et al. 2013). Focusing on link-based sensors, such as traffic counts, the number and position has a clear relation with the size and exploration possibilities of the demand estimation solution space (Yang, Iida, and Sasaki 1991; Djukic, Van Lint, and Hoogendoorn 2012). Approaches that work on the analysis of the solution space dimension are for instance the one of (Djukic, Van Lint, and Hoogendoorn 2012), which applies Principal Component Analysis (PCA) to study the high-dimensional data structure, and (Flötteröd and Bierlaire 2009), who propose to improve DODE using a new linearization of the network loading map in order to overcome the inadequacy of a proportional assignment in congested conditions. We have also recently investigated the opportunity to include information related to link flow observability, i.e. how a certain sensor can provide information about unobserved flows, to further reduce the solution space and therefore increase the solution reliability (Viti et al., 2015.). Due to the high number of approaches and algorithms for dealing with the DODE, recently a common evaluation and benchmarking platform has been proposed (Constantinos Antoniou et al. 2016) to allow for their comparison.

4.1.2 The Two-Step concept

To partially reduce the impact of the solution space and the consequent variety of possible solutions, classical methods, called Single-level in this chapter, often include information about a reference OD demand matrix (usually known as seed matrix) within the mathematical formulation of the problem. This demand works as constraint for the model, which provides solutions with similar demand levels with respect to the starting one. The Single-level DODE formulated in this way leads generally to a local calibration of the starting demand matrix. Therefore, if the seed matrix demand level is different from the real one, this localism can lead to significant errors (Rodric Frederix 2012). The need for methods dealing with the correction of the seed matrix in such applications was pointed out recently by (Guido Cantelmo et al. 2014), who proposed a Two-Step approach where the first step was focused on correcting the seed matrix by focusing on the OD flows having largest impact on the measured link counts, so the largest error according to the model. This Two-Step procedure demonstrated its ability in correcting the starting demand value without introducing new traffic measures, apart from traffic counts, or developing new models, and effectively improved the results on congested networks by correcting the seed matrix in the first step and directing it towards the real demand values. Though effective, the developed method can hardly be applied on large-sized networks, where the number of OD pairs to be selected in the first step may become significantly large. Moreover, identifying a subset of origin destination flows to modify it is not always easy. In (Guido Cantelmo et al. 2014), when different flows are used in the first step, the outputs of the model presented differences both in terms of quality of the results, and in terms of computational time.

To overtake this issue, in this chapter, the authors propose an enhancement of the previously developed Two-Step approach by exploiting temporal information on aggregated demand data such as generation data by zones. Specifically, the first step searches, in any time interval, for generation values that best represent the measurements (traffic counts); hence, in the first step, the variables are not the dynamic OD trips, but the total production values for any time interval, thus reducing the dimension of the problem considerably. In the second step, the classical DODE procedure is performed improving temporal and spatial matrix distributions. Framing the problem as such, one benefits of the right demand level identified in the first phase, avoiding the single-step localism problems previously mentioned.

The proposed Two-Step approach has been later applied on a real network case, resulting more robust in terms of goal function trends, link flows and traffic state representations. Conclusions and future research directions conclude this Chapter.

4.2 Methodology

The DODE is generally solved as an optimization problem. Its formulation requires the specification of the objective function (O.F.), its variables, the elements of the OD demand matrix to be estimated, and its constraints, related to feasibility and routing conditions. Considering different types of measures and by adopting a simultaneous approach the problem can be formulated as:

$$(\boldsymbol{d}_1^*, \dots, \boldsymbol{d}_n^*) = argmin \begin{bmatrix} z_1(\boldsymbol{l}_1, \dots, \boldsymbol{l}_n, \widehat{\boldsymbol{l}_1}, \dots, \widehat{\boldsymbol{l}_n}) + \\ +z_2(\boldsymbol{q}_1, \dots, \boldsymbol{q}_n, \widehat{\boldsymbol{q}_1}, \dots, \widehat{\boldsymbol{q}_n}) + \\ +z_3(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n, \widehat{\boldsymbol{d}_1}, \dots, \widehat{\boldsymbol{d}_n}) + \\ +z_4(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n, \widehat{\boldsymbol{r}_1}, \dots, \widehat{\boldsymbol{r}_n}) + \end{bmatrix} \quad (4.1)$$

Where

- $\boldsymbol{l}/\hat{\boldsymbol{l}}$ are respectively the simulated values and the corresponding measurements on the links;
- $\boldsymbol{q}/\hat{\boldsymbol{q}}$ are respectively the simulated values and the corresponding measurements on the nodes;
- $\boldsymbol{x}/\hat{\boldsymbol{d}}$ are respectively the estimated value and a-priori information on the dynamic demand (seed matrix);
- $\boldsymbol{r}/\hat{\boldsymbol{r}}$ are respectively the simulated values the and corresponding measurements on routes;
- $\boldsymbol{d}_n^*$ the estimated demand matrix for time interval n;

z :{z_1, z_2, z_3, z_4} is the estimator represented by the deviations between the simulated/estimated and the corresponding measured/a-priori values. The dependence between simulated information in Equation (4.1) and the estimated demand is obtained directly by simulation performing a dynamic traffic assignment (DTA).

4.2.1 Generation-Distribution adjustment process

In the proposed Two-Step procedure, the first step aims at optimizing the generation values of each zone in each time interval, while maintaining constant the dynamic trip distributions derived by the seed matrix. The objective function in Equation (4.1) can be generally rewritten for the first step as:

$$(\boldsymbol{E}_1^*, \dots, \boldsymbol{E}_n^*) = argmin \begin{bmatrix} z_1(\boldsymbol{l}_1, \dots, \boldsymbol{l}_n, \widehat{\boldsymbol{l}_1}, \dots, \widehat{\boldsymbol{l}_n}) + \\ +z_2(\boldsymbol{q}_1, \dots, \boldsymbol{q}_n, \widehat{\boldsymbol{q}_1}, \dots, \widehat{\boldsymbol{q}_n}) + \\ +z_3(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n, \widehat{\boldsymbol{d}_1}, \dots, \widehat{\boldsymbol{d}_n}) + \\ +z_4(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n, \widehat{\boldsymbol{r}_1}, \dots, \widehat{\boldsymbol{r}_n}) + \end{bmatrix} \quad (4.2)$$

Where $x_n^{OD} = E_n^O d_{D|O}^{Seed,n}$

and with:

- E_n^O = generation of origin zone O and time interval n;
- $\boldsymbol{E}_n^*$ = generation vector containing generation from all origins in time interval n.
- x_n^* = trips flow from origin zone O to destination zone D in time interval n.
- $d_{D|O}^{Seed,n}$ = seed matrix probability distribution between traffic zone D and traffic zone O in time interval n.

The idea of working on generation values in the first step, rather than on dynamic OD trips directly, derives by the increasing attention received by this type of aggregated information in the literature. Already (Iannò and Postorino 2002) proposed a generation-constrained approach for the static demand estimation problem where the objective function contains a specific term in order to prevent the emission from each origin zone to be greater than the actual one. Then, (Cipriani et al. 2011) proposed to introduce a generation constraint in the dynamic demand estimation. (Cascetta et al. 2013) proposed a quasi-dynamic approach where the main assumption is that the demand generation changes much faster than the distributions. Finally, in (G. Cantelmo et al. 2014) some remarks are reported about the possible adoption of the generation values as a constraint in the DODE.

The high significance given in literature to this aggregated information derives mainly by the following considerations:

- Total generated trips can act by limiting a demand overestimation during the DODE; the overestimation can usually occur when dealing with traffic measurements collected on congested networks;
- Total generated trips are more easily available than OD trips, and generation models, from which these data are obtained, are considered the most reliable models in transport engineering applications;
- Adopting the generation values inside the DODE, as in Equation (4.2), reduces the number of variables (from O×D×n to O×n): The expected result of this phase is the correct level of generated demand for each time interval.

The goal of the first step is to act on the seed matrix in order to obtain a reasonable generation value before moving to the second step, in which the dynamic distributions are corrected according to Equation (4.1). The present approach has analogies with the quasi-dynamic approach reported in (Cascetta et al. 2013). In the latter, distributions are explicitly considered in terms of probabilities and approximated as an average over a time period greater of the time slice itself; in this approach, in the first step we assume them constant and equal to the ones of the seed matrix, while they are considered as unconstrained variables in the second, removing any assumption on them. The goal of this formulation is to correct the generations to move on the right demand level, using constant seed distributions as an indirect constraint to the original demand matrix.

4.2.2 Solution Algorithm

For the solution of the First Step (4.2) a Finite Difference Stochastic Approach (FDSA (Spall 2012)) has been adopted to find the optimal descent direction. FDSA is a method usually adopted when stochastic measurements are adopted. At the first step we are mostly interested in investigating the effectiveness of our assumption about the ability of generation values to move the optimization towards the "right level of demand". Hence, the choice of using FDSA is done as it permits to obtain, at each iteration i, a deterministic gradient $\boldsymbol{G}^i$ from a finite-difference computation. Specifically each variable θ is perturbed as follows:

$$\boldsymbol{G}^i(\boldsymbol{\theta}^i) = \begin{bmatrix} \dfrac{z(\boldsymbol{\theta}^i + c^i\boldsymbol{\xi}^1) - z(\boldsymbol{\theta}^i)}{c^i} \\ \vdots \\ \dfrac{z(\boldsymbol{\theta}^i + c^i\boldsymbol{\xi}^r) - z(\boldsymbol{\theta}^i)}{c^i} \end{bmatrix} \tag{4.3}$$

Where ξ is a vector with all zeros, except for the variable to be perturbed, c^i is the step size and z the adopted objective function. In this method each variable is perturbed independently, so the number of simulations required for computing the gradient in any iteration is equal to the number of variables (in the first step, variables are equal to the generated trips from each origin zone O and time interval n) plus the value of z in the starting point.

Once computed $\boldsymbol{G}^i$, the solution is then updated at each iteration by:

$$\boldsymbol{\theta}^{i+1} = \boldsymbol{\theta}^i - \alpha^i\boldsymbol{G}^i \tag{4.4}$$

With α the step length for the update.

The FDSA can imply high computational times and it could become infeasible for large-scale networks: here, it is adopted only to validate the concept behind the Two-Step approach, since it is able to provide a good approximation of the exact gradient, not available in the DODE. When dealing with large-scale networks, it is anyway possible to substitute FDSA with other existing and less computational expensive gradient approximation methods as proposed in literature (Cipriani et al. 2011; R. Frederix et al. 2011; Constantinos Antoniou et al. 2016; G. Cantelmo et al. 2014; Lu et al. 2015).

At the second step, given the estimated total generated demand of the first step, the optimization works on dynamic OD trips in a more traditional manner. For the second step, the Simultaneous Perturbation Stochastic Approximation (SPSA (Spall 2012)) has been adopted. SPSA is a path search optimization method, where an approximation of the gradient is computed based on a simultaneous perturbation of all the variables. This approach has been successfully applied for different large-scale traffic optimization problems as for the off-line and on-line calibration of dynamic traffic assignment models (Balakrishna 2006), as also for the DODE problem (Balakrishna 2006; Barceló et al. 2013; G. Cantelmo et al. 2014; Lu et al. 2015). In the SPSA, the equation to update the solution is the standard formulation reported in Equation (4.4), while the approximated gradient at each iteration *i* is obtained as follows:

$$\widehat{\boldsymbol{g}}_k(\boldsymbol{\theta}^i) = \frac{z(\boldsymbol{\theta}^i + c^i\Delta^k) - z(\boldsymbol{\theta}^i)}{c^i}\begin{bmatrix}(\Delta_1^k)\\ \vdots \\ (\Delta_r^k)\end{bmatrix} \tag{4.5}$$

$$\boldsymbol{G^i} = \overline{\boldsymbol{g}}(\boldsymbol{\theta^i}) = \frac{\sum_{\boldsymbol{k=1}}^{\boldsymbol{Grad\_rep}} \widehat{\boldsymbol{g}}_k(\boldsymbol{\theta}^i)}{Grad\_rep} \tag{4.6}$$

with c^i the perturbation step, *Grad_rep* is the number of replications to compute the average gradient, and $\boldsymbol{\Delta}$ is a vector with elements {-1,1}.

With respect to the FDSA, the gradient has a stochastic component, but the computational time to obtain the descent direction is smaller being the variables perturbed simultaneously. It is possible, and recommended, to repeat the perturbation to obtain a good gradient approximation. In the equation (4.5), the formulation of the SPSA model is presented with the asymmetric design (SPSA-AD,(Cipriani et al. 2010)). The advantage of using this formulation is that the number of simulations needed to compute the gradient is halved with respect to the basic SPSA with symmetric design (SD).

4.2.2.1 P-SPSA

Since the approach aims to be applicable to real-sized networks, SPSA is appropriate for solving the second step problem. However, although the gradient computation is not dependent on the number of variables, approximation increases with the number of variables N:

$$\sum_{\boldsymbol{e=1}}^{\boldsymbol{M}} \frac{\partial z_e(\theta^k)}{\partial \theta^k}\widehat{\boldsymbol{\theta}} = \boldsymbol{G^{SPSA}}(\theta^k) + \boldsymbol{\varepsilon_1}(N) \tag{4.7}$$

Where *M* is the number of terms in the goal function and $\boldsymbol{\varepsilon_1}(N)$ is the error related to perturbing all the *N* variables simultaneously. A new variant, here called P-SPSA (Partial SPSA) is proposed in the second approach to reduce the approximation of the SPSA with respect to this problem. In every iteration only a percentage *P* of the matrix is perturbed and updated. Elements of the $\boldsymbol{\Delta}$ vector are now {-1,0,1}. Therefore by fixing the value of *P*, we regulate the share of non-zero in the $\boldsymbol{\Delta}$ vector. The variables to be perturbed are randomly selected in every iteration, so any of them is selected throughout the whole optimization process:

$$\sum_{e=1}^{M} \frac{\partial z_e(\theta^k)}{\partial \theta^k} \widehat{\boldsymbol{\theta}} = \boldsymbol{G}^{\boldsymbol{P-SPSA}}(\theta^k) + \boldsymbol{\varepsilon_1}^{\boldsymbol{k}}(N_p) \tag{4.8}$$

While it is easy to observe that the error $\boldsymbol{\varepsilon_1}^{\boldsymbol{k}}(N_p) < \boldsymbol{\varepsilon_1}(N)$ where $N_p < N$, we have to consider that the procedure could converge more slowly with respect to the SPSA since only a part of the variables are updated in an iteration. On the other hand, we know that $0 \leq P \leq 1$, and specifically the computational time is going to increase more and more the closer P gets to 0, while with P=1 is going to become the same of the SPSA. We can consider this problem inserting a second error in Equation (4.8) i.e.:

$$\begin{cases} \boldsymbol{\varepsilon}(\theta^k) = \boldsymbol{\varepsilon_1}^{\boldsymbol{k}}(N_p) & for\ \theta^k \in N_p \\ \boldsymbol{\varepsilon}(\theta^k) = \boldsymbol{f}(\theta^i - \theta_{spsa}^{i+i}) = \boldsymbol{\varepsilon}^{\boldsymbol{k}}(N - N_p) & for\ \theta^k \notin N_p \end{cases} \tag{4.9}$$

Where N_p is the ensemble of perturbed variables, $\boldsymbol{\varepsilon}^{\boldsymbol{k}}(N - N_p)$ is the error related to not updated variables, θ_{spsa}^{i+i} is the value that the variable θ^k, not updated in the current iteration, assumed in the next iteration when a full SPSA is performed. If *i* is the number of iterations, we can now assume that if:

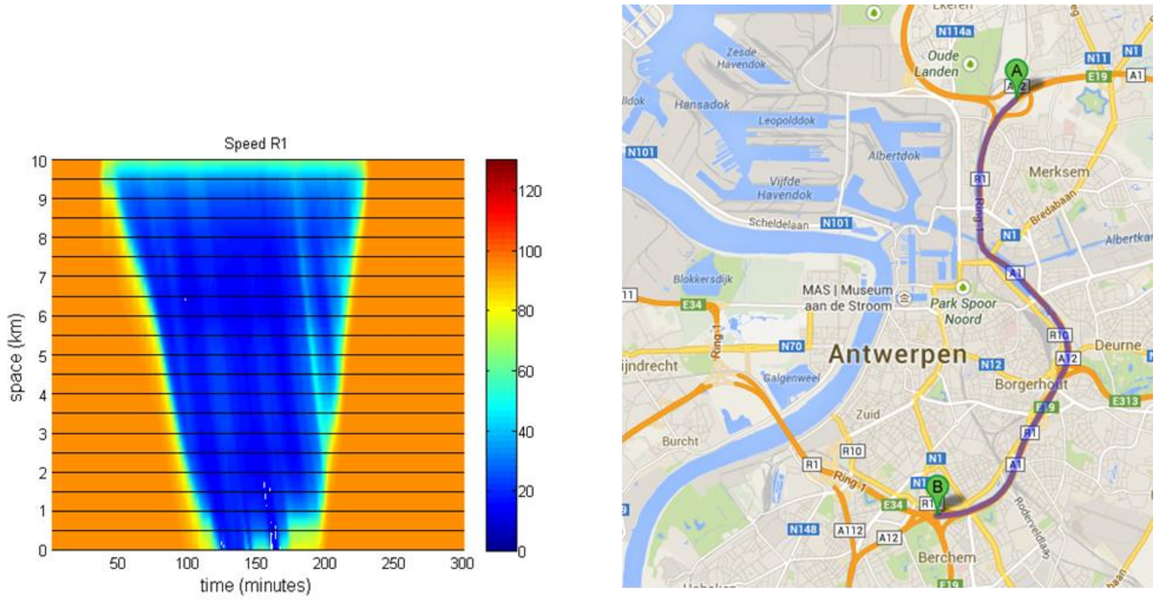
$$\sum_{i}\sum_{N_p} \boldsymbol{\varepsilon_1}(N_p) + \sum_{i}\sum_{\boldsymbol{N-N_p}} \boldsymbol{\varepsilon}(N - N_p) \leq \sum_{i}\sum_{\boldsymbol{N}} \boldsymbol{\varepsilon_1}(N) \tag{4.10}$$

Then the computational time of the P-SPSA is smaller or equal to the time of the SPSA. Since equality in Equation (4.10) is satisfied for P=1, our preliminary assumptions are that very low values of P (i.e. 0.25) the term $\boldsymbol{\varepsilon}(N - N_p)$ increases much more than the reduction in $\boldsymbol{\varepsilon_1}(N_p)$. It is reasonable to assume that opposite holds for high P values (i.e. 0.75).

Preliminary results in this chapter confirm these hypothesis on P-SPSA, which are relevant on big size networks where solving the problem presented in Equation (4.7) is well known to be cumbersome. This is shown in a test network in the next section.

4.3 Case Study

The setup of this case study is the same presented in (Rodric Frederix 2012), and used in (Guido Cantelmo et al. 2014). This scenario refers to the inner ring-way around Antwerp, Belgium. The network includes 56 links, 39 nodes, with 46 OD pairs, all mainly connecting the different entry and exit points of this motorway segment, making rerouting options not likely. The considered morning peak period occurs between 05:30 and 10:30. The field data – speeds and flows – were available every 5 minutes. The detectors are located at the on- and off-ramps and on some intermediate sections. The OD flows have been estimated for 15-minutes departure intervals, so the dynamic matrix contains 966 OD pairs; the seed matrix that amounts to 202,200 trips is derived from an existing static OD matrix by superimposing a time profile. Flows of a selection of OD pairs have been increased obtaining a congestion pattern similar to the actual one. As a consequence, the seed matrix captures the correct traffic regimes.



(a) (b)

Fig.4.1: (a) x,t plot of the measured speeds on the network, (b) Ring of Antwerp;

In order to start with the application of the two-step procedure, with respect to Equation (4.1) and (4.2), the objective function to be minimized contains only the z1 term, where the link measurements are the traffic counts. Specifically:

$$\min f\left(\boldsymbol{\theta}^{i}\right) = \left[h(\boldsymbol{l}-\hat{\boldsymbol{l}})\right] = \sum\nolimits_{\mathrm{l}\in\mathrm{D}} (\boldsymbol{l}-\hat{\boldsymbol{l}})^{2} \qquad (4.11)$$

With $\boldsymbol{l}$ and $\hat{\boldsymbol{l}}$ respectively being the simulated and measured flows on each link and *D* the subset of network links with sensors. The speed measurements have been used only for validation, since it is expected that if the initial traffic regime is accurately represented on any link, then the new estimated matrix reduces the link errors related to flows while preserving the correct traffic regimes. The simulations required to compute simulated flows on each link have been conducted adopting the Link Transmission Model described in (Corthout 2012).

4.3.1 First-Step Application

In the first step, the generation values for each zone and each time interval have been optimized using FDSA. In this first step, generations are corrected using increasing values for α: [3,6,12]. Since no line search is used, smaller step-sizes are expected to provide better solutions with higher computational times. In Table 4.1, it is possible to observe the results for different values of α. These have been compared with those obtained using the Two-Step Formulation proposed in (Guido Cantelmo et al. 2014), named “Data Based” in the table. (Details on the experiment set-up can be found in the Previous Chapter, Approach I, Table 3.2).

Table 4.1: Experiment Results

	α=3	*α=6*	*α=12*	***Data Based***
Final O.F. value	1.14E+09	1.33E+09	1.43E+09	5.8E+08
O.F. improvement [%]	45.02	36.48	30.04	72.26
Link flows RMSE	1031.5	1108.9	1163.7	730.3
Link flows RMSN [%]	31.91	34.31	36.55	22.6
Link speeds RMSE	19.00	15.75	17.48	18.47
Link speeds RMSN [%]	29.49	24.44	27.13	28.67
# iterations	91	71	58	53

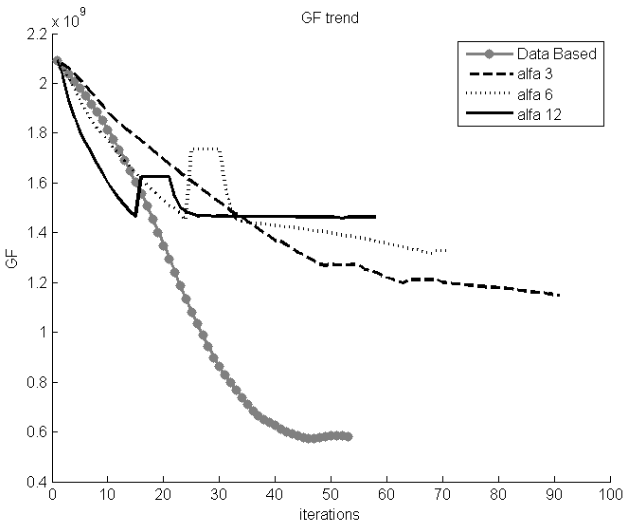


Fig.4.2: Goal Function Trend for α= [3,6,12] and Data Based

The reference *Data Based* model implies the best improvement, but as reported in the previous chapter (G. Cantelmo et al. 2014) it works on a specific subset of variables (the OD pairs that generated the highest error with respect to traffic counts). This sub-set of variables is not easy to capture in all the networks and changing the subset of variables, the quality of results and the computational time can present high variance. In the first step here proposed, there is not this type of problem since the subset of variables to correct is uniquely defined. To evaluate the reliability of the solutions with respect to the inputs, the error has been evaluated also in terms of r-square between simulated and observed flows. For α=3 after 91 iterations, the r-square is quite similar to the one obtained for α=12, after 31 iterations, changing from 0.838 to 0.854. This is also confirmed by RMSN values in Table 4.1. Moreover, in this case, we are working indirectly on all the variables, using distributions derived from the seed matrix as a constraint. Thus, it is reasonable to obtain a higher value for the goal function, while the greatest reduction is expected in the second step.

4.3.2 Second-Step Application

In this second step, the correction is mainly focused on distributions. The experiments are performed adopting as starting matrix the solution obtained using a step size of α=12, tested in the previous stage. Such solution is considered the most interesting case for two main reasons. First of all, it was the configuration for which the convergence has been reached earlier, finding the solution after 30 iterations (Fig. 4.3a). Since the matrix presents the highest value of the goal function, and since results are robust with respect to both link flows and speeds data, if matrices obtained with α value 3 or 6 are used, then the result should not be worse. Before performing the second step optimization, results from the single step are shown. In Table 4.2-A it is possible to observe results obtained applying conventional SPSA and P-SPSA in a single-step classical DODE. While for the SPSA the stop criterion is the convergence, P-SPSA is stopped after approximately 190 iterations. Since several single-step SPSA optimizations were performed, in table 4.2-A "best" represents the best value for each parameter, "worst" the worse value while "avg" is the average solution.

Results suggest that the hypothesis done in Equation (4.10) about computational time is reasonable: when the perturbation $P \geq 0.5$ computational time is not going to increase. Furthermore it is recommend to never use $P<0.5$: when P is small, the probability to work on all the variables during the optimization largely decreases. However, these results are experimental. P-SPSA allows to reduce the number of variables of the problem up to 50%, which is a fundamental property for big-sized networks, without affecting the quality of the results. Since the interest is to apply the Two-Step approach to all the networks, both SPSA and P-SPSA are tested. The most interesting goal for the P-SPSA is to reach the same result of SPSA without increasing the computational time, so the case with P=0.5 is considered to perform the second step.

Table 4.2: Experiment Results

4.2A – Single Step Results	P-SPSA P=0.25	P-SPSA P=0.50	P-SPSA P=0.75	SPSA BEST	SPSA AVG	SPSA WORST
Final O.F. value	6.67E+08	3.86E+08	3.93E+08	3.28E+08	3.96 E+08	5.01 E+08
O.F. improvement [%]	68.07	81.51	81.18	84.29	81.40	76.04
Link flows RMSE	786.29	601.35	602.75	552.03	598.44	681.79
Link flows RMSN [%]	24.32	18.60	18.60	16.58	18.39	21
Link speeds RMSE	18.94	18.42	18.63	17.59	19.30	21.01
Link speeds RMSN [%]	29.29	28.59	28.91	27.47	30.52	34.47
# of iterations	187	195	195	90	160	273

4.2B – Two Steps Results	SPSA	SPSA demand	P-SPSA P=0.5	Data based SPSA	Statistics results BEST	Statistics results AVG	Statistics results WORST
Final O.F. value	3.29E+08	3.18E+08	3.08E+08	4.49E+08	3.08E+08	3.23E+08	3.5E+08
O.F. improvement [%]	84.25	84.77	85.24	78.52	85.24	84.54	84.51
Improvement in 2th [%]	77.25	78.18	78.85	20.77	78.85	77.85	75.72
Link flows RMSE	547.42	538.20	534.75	644.82	534.34	545.76	571.92
Link flows RMSN [%]	16.93	16.65	16.53	19.95	16.53	16.88	17.69
Link speeds RMSE	19.98	18.44	17.29	13.67	16.22	18.69	20.7
Link speeds RMSN [%]	34.71	28.44	26.83	21.16	25.41	29.64	34.71
Regression coefficients Simulated Vs Real					*Regression coefficients*		
r2	0.936	0.937	0.939	0.920	0.939	0.936	0.936
Angular coefficient p1	0.99	1.00	0.99	0.97	1	1	0.997
Intercept coefficient p2	49.02	43.00	43.21	103.11	34.74	44.25	50.25

Concerning the Second-Step, all models use the same goal function presented in Equation (4.11). Furthermore, the enhanced algorithm is also tested using the demand matrix in the goal function. So equation becomes:

$$\min f\left(\boldsymbol{\theta}^{i}\right) = \sum\nolimits_{\mathrm{l} \in \mathrm{D}} (\boldsymbol{l} - \hat{\boldsymbol{l}})^2 + \sum\nolimits_{\mathrm{l} \in \mathrm{D}} (\boldsymbol{x} - \widehat{\boldsymbol{d}})^2 \tag{4.12}$$

Where N is the number of OD pairs and $\widehat{\boldsymbol{d}}$ is the target matrix, in this case the solution of the first step. This experiment is called "SPSA with Demand" in the rest of the Chapter. In Figures 4.3a and 4.3b goal functions trend are proposed for two independent optimizations. The trend shows again the robustness of the model. Results are compared with the old data-based Two-Step approach.

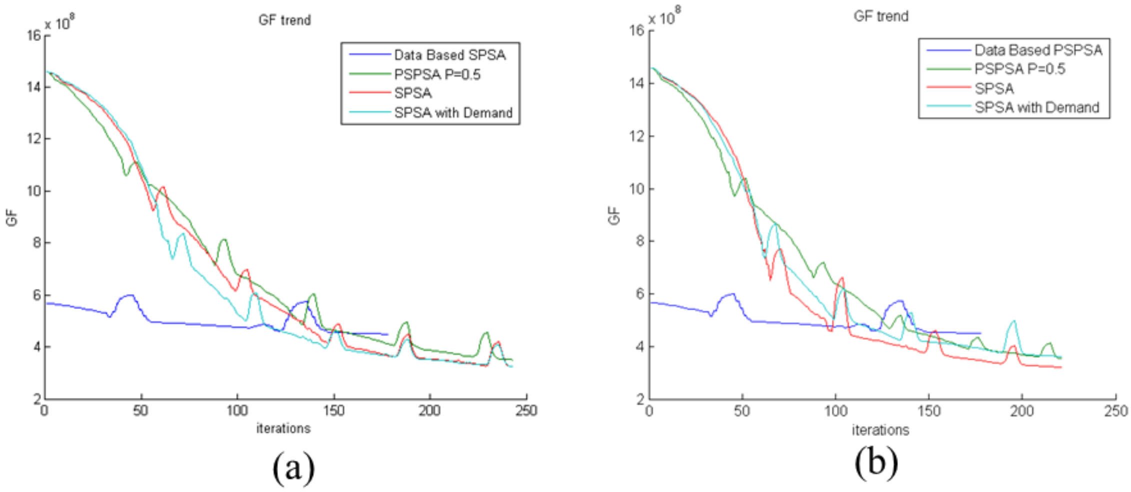


Fig.4.3: Goal Function Trend for α= [3,6,12] and Data Based

Stop criterion is the convergence or an RMSE on the speed lower than 20. Once more, the results highlight the robustness of the process with respect to the Data-Based approach. In fact in the Data Based the main contribution was in the first step, where only 126 OD pairs out of 966 were used. In the second step the model just added local adjustments on the matrix. The main positive results of the Data Based approach are the lower error in the speeds, with respect to the original seed matrix, and the lower number of iterations (Figure 4.3). Unfortunately these results are not easily generalizable since the subset is not uniquely defined and if another subset is chosen, results are completely different. In the previous chapter, another approach- called "Network Analysis Based Approach (Approach 2)- was proposed where a different subset of variables was used. The results were completely different from the Data Based one. If the goal function improvement was greater (89.9%), the error on the speed increased (RMSE=18) as well as the distance from the seed matrix (equal to 6.26 E+04 in Data Based and 1.15E+05 in the Network Analysis Based). The strong difference between results was the initial input to generate the current approach.

In this approach we can observe the advantages having a uniquely defined subset of variables in the first step. Results for each method are very close to each other. Moreover scatter plots of the results are very similar to each other: the parameters of the regression (r2, p1, p2) are very similar. About P-SPSA it is possible now to make some remarks. The main goal of P-SPSA it is to reach the same solution of SPSA without increasing computational time whilst reducing the number of variables. In Equation (4.10) we assume that if the number of variables perturbed in every iteration is at least the 50% computational time is not going to increase. Tests show that P=0.5 is, as expected, the limit case using P-SPSA. If the computational time is higher than the one of SPSA, such increase is limited. Setting as stop criterion the number of iterations, the goal function value at iteration 243 is 3.50E+08, while at iteration 269 is 3.32E+08. Results show that the approximation is not going to reduce the quality of the result. Although more tests are needed, this insight is important in real networks, where the number of variables is too high to use in an efficient way SPSA. P-SPSA is an appropriate alternative to manage problems two times bigger with respect to classic SPSA without compromising significantly the quality of the solution and the computational time.

Finally, some considerations have to be done with respect to the comparison with single step approach. Observing Table 4.2, differences in results are significant. In 4.2-B the procedure better fits measured data than the single step approach, as calibration parameters confirm. Furthermore, a strong reduction in variance results is observed. In the first case difference between the best and worst goal function value is almost 10%, while in the second case is approximately 1%. If variance in some parameters, like iterations number, seems to be good, these parameters are generally related to the worst cases. The number of iterations of the “best” case in Table 4.2-A is lower than those reported in figure 4.3. However, when convergence is reached too fast, model results in high goal functions values and not satisfactory solutions. Further, regression coefficients are worst with respect to the Two-Step approach (r2=0.934, p1=0.98, p2=71 for the best solution). About Euclidean distance from the seed matrix, the average value is similar in both cases while the distance between each solution matrix is different. The average distance between solutions matrices found using two step approach is 3.42E+04, while is 4.35E+04 in the single step. Further the variance of this value is higher in the single step with respect to the proposed approach, confirming robustness of our method with respect to the single step.

4.4 Conclusions and Future Research

In this Chapter, an enhanced Two-Step approach is proposed to improve performances of existing DODE algorithms. Since the reliability of the results in dynamics problem is one of the most critical aspects in using dynamics methods for real problems, the main contribution of this approach is finding robust results with respect to both the single-step approach and the previous version of the Two-Step approach. In this Chapter, a combination of deterministic and stochastic algorithms is used to perform offline estimation on the inner ring of Antwerp, Belgium. Speeds are used to validate quality of the solution and as stop criterion.

The main motivation in developing the proposed approach is obtaining accurate and reliable results by operating an adequate solution space reduction. Since the number of possible solutions generally increases with the size and the complexity of the network, it is relevant introducing general procedures to reduce step by step the solution space without increasing the problem complexity. The Two-Step approach is based on the correlation between the aggregate demand data – named generation data -, the disaggregate demand data – i.e. the OD flows – and supply data as link speeds and flows. Since, generally, aggregate data from statics models are more reliable with respect to the disaggregate one, it is natural to fix them in an aggregate level.

Following a Two-Step procedure, as initially proposed in previous studies by the authors, in the first step the total flow generated for each traffic zone is corrected. The demand at aggregate level can be used to catch the right demand level keeping constant the distributions. In this first phase, distributions are used as an indirect constraint for the demand, reducing the possible solutions for the problem without introducing new measurements or data. Vice versa, since aggregate data works as an indirect constraint, it is possible to eliminate the demand term from the goal function. In this way it is possible to strongly reduce the localism of the DODE. Results show the reliability of the approach with respect to the most important parameter, the step size. Is it so possible to increase the speed of the problem without having significant errors in the solution of the first step.

In the second step, correction of the demand is performed using SPSA algorithm obtaining good results. The used method is generally adopted to solve problem on big sized networks, since it is not dependent on the number of variables. On the other hand the stochastic nature of the model increases with the size of the problem. In the specific case study, SPSA obtains stable results. A variant of that model, called P-SPSA, is also presented. It should be pointed out that results are experimental and preliminary, since no test on other networks are still available; the model was tested together with the SPSA in the second step. P-SPSA reaches, in the case study, the same result of the SPSA, while working on no more than 50% of the OD pairs simultaneously. So in the current case study we are able to perform a full satisfactory OD estimation reducing the number of the variables to the only generation in the first step and to the 50% of the OD pairs in the second. The possibility to reduce number of variable is one of the most relevant aspects in DODE, since often in real practice is not possible to work on all of them. Results highlight the robustness of the proposed approach with respect to the classical single step.

Future research will still focus on small networks where however route choice is more significant than the network used in this network. If results are confirmed the last step is to apply it on medium/large sized networks.

Acknowledgements

The authors acknowledge for financing grant: AFR-PhD grant 6947587 IDEAS (Fonds National de Recherche FNR).

4.5 References

Antoniou, C., R. Balakrishna, H.N. Koutsopoulos, and M. Ben-Akiva. 2009. "Off-Line and on-Line Calibration of Dynamic Traffic Assignment Systems." In , 104–11. doi:10.3182/20090902-3-US-2007.0056.

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. doi:10.1016/j.trc.2015.08.009.

Ashok, K., and M.E. Ben-Akiva. 2000. "Alternative Approaches for Real-Time Estimation and Prediction of Time-Dependent Origin-Destination Flows." *Transportation Science* 34 (1): 21–36.

———. 2002. "Estimation and Prediction of Time-Dependent Origin-Destination Flows with a Stochastic Mapping to Path Flows and Link Flows." *Transportation Science* 36 (2): 184–98. doi:10.1287/trsc.36.2.184.563.

Balakrishna, Ramachandran. 2006. "Off-Line Calibration of Dynamic Traffic Assignment Models." PhD Thesis, Massachusetts Institute of Technology. http://dspace.mit.edu/handle/1721.1/35120.

Barceló, J., L. Montero, M. Bullejos, O. Serch, and C. Carmona. 2013. "A Kalman Filter Approach for Exploiting Bluetooth Traffic Data When Estimating Time-Dependent Od Matrices." *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 17 (2): 123–41. doi:10.1080/15472450.2013.764793.

Caceres, N., J.P. Wideberg, and F.G. Benitez. 2007. "Deriving Origin-Destination Data from a Mobile Phone Network." *IET Intelligent Transport Systems* 1 (1): 15–26. doi:10.1049/iet-its:20060020.

Cantelmo, G., E. Cipriani, A. Gemma, and M. Nigro. 2014. "An Adaptive Bi-Level Gradient Procedure for the Estimation of Dynamic Traffic Demand." *IEEE Transactions on Intelligent Transportation Systems* 15 (3): 1348–61. doi:10.1109/TITS.2014.2299734.

Cantelmo, Guido, Francesco Viti, Chris Tampère, Ernesto Cipriani, and Marialisa Nigro. 2014. "Two-Step Approach for Correction of Seed Matrix in Dynamic Demand Estimation." *Transportation Research Record: Journal of the Transportation Research Board* 2466 (December): 125–33. doi:10.3141/2466-14.

Cascetta, Ennio, Domenico Inaudi, and Gerald Marquis. 1993. "Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts." *Transportation Science* 27 (4): 363–73.

Cascetta, Ennio, Andrea Papola, Vittorio Marzano, Fulvio Simonelli, and Iolanda Vitiello. 2013. "Quasi-Dynamic Estimation of O–d Flows from Traffic Counts: Formulation, Statistical Validation and Performance Analysis on Real Data." *Transportation Research Part B: Methodological* 55 (September): 171–87. doi:10.1016/j.trb.2013.06.007.

Cipriani, E., M. Florian, M. Mahut, and M. Nigro. 2010. "Investigating the Efficiency of a Gradient Approximation Approach for the Solution of Dynamic Demand Estimation Problems." In *New Developments in Transport Planning: Advances in Dynamic Traffic Assignment*, 343–60.

———. 2011. "A Gradient Approximation Approach for Adjusting Temporal Origin-Destination Matrices." *Transportation Research Part C: Emerging Technologies* 19 (2): 270–82. doi:10.1016/j.trc.2010.05.013.

Corthout, Ruben. 2012. "Intersection Modelling and Marginal Simulation in Macroscopic Dynamic Network Loading." PhD Thesis. https://lirias.kuleuven.be/handle/123456789/344247.

Dixon, M.P., and L.R. Rilett. 2002. "Real-Time OD Estimation Using Automatic Vehicle Identification and Traffic Count Data." *Computer-Aided Civil and Infrastructure Engineering* 17 (1): 7–21.

Djukic, T., J. Van Lint, and S. Hoogendoorn. 2012. *Application of Principal Component Analysis to Predict Dynamic Origin-Destination Matrices*. Transportation Research Record.

Flötteröd, Gunnar, and Michel Bierlaire. 2009. "Improved Estimation of Travel Demand from Traffic Counts by a New Linearization of the Network Loading Map." In . http://infoscience.epfl.ch/record/152359.

Frederix, R., F. Viti, R. Corthout, and C.M.J. Tampère. 2011. "New Gradient Approximation Method for Dynamic Origin-Destination Matrix Estimation on Congested Networks." *Transportation Research Record*, Transportation Research Record, , no. 2263: 19–25.

Frederix, Rodric. 2012. "Dynamic Origin-Destination Matrix Estimation in Large-Scale Congested Networks." PhD Thesis. https://lirias.kuleuven.be/handle/123456789/344248.

Iannò, and M.N. Postorino. 2002. "A Generation Constrained Approach for the Estimation of O/D Trip Matrices from Traffic Counts." In *European Transport Conference,* Strasburg France..

Lu, Lu, Yan Xu, Constantinos Antoniou, and Moshe Ben-Akiva. 2015. "An Enhanced SPSA Algorithm for the Calibration of Dynamic Traffic Assignment Models." *Transportation Research Part C: Emerging Technologies* 51 (February): 149–66. doi:10.1016/j.trc.2014.11.006.

Spall, James C. 2012. "Stochastic Optimization." In *Handbook of Computational Statistics*, edited by James E. Gentle, Wolfgang Karl Härdle, and Yuichi Mori, 173–201. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-642-21551-3_7.

Tavana, Hossein. 2001. "Internally-Consistent Estimation of Dynamic Network Origin-Destination Flows from Intelligent Transportation Systems Data Using Bi-Level Optimization." PhD Thesis. http://repositories.lib.utexas.edu/handle/2152/1662.

Toledo, T., and T. Kolechkina. 2013. "Estimation of Dynamic Origin-Destination Matrices Using Linear Assignment Matrix Approximations." *IEEE Transactions on Intelligent Transportation Systems* 14 (2): 618–26. doi:10.1109/TITS.2012.2226211.

Viti, Francesco, Guido Cantelmo, Francesco Corman, and Marco Rinaldi. n.d. "Improving the Reliability of Demand Estimation Using Traffic Counts by Including Information on Link Flow Observability." *INSTR2015* Conference Proceedings: (In press).

Yang, H. 1995. "Heuristic Algorithms for the Bilevel Origin-Destination Matrix Estimation Problem." *Transportation Research Part B* 29 (4): 231–42. doi:10.1016/0191-2615(95)00003-V.

Yang, H., Y. Iida, and T. Sasaki. 1991. "An Analysis of the Reliability of an Origin-Destination Trip Matrix Estimated from Traffic Counts." *Transportation Research Part B* 25 (5): 351–63. doi:10.1016/0191-2615(91)90028-H.

Zhou, Xuesong, and Hani S. Mahmassani. 2007. "A Structural State Space Model for Real-Time Traffic Origin–destination Demand Estimation and Prediction in a Day-to-Day Learning Framework." *Transportation Research Part B: Methodological* 41 (8): 823–40. doi:10.1016/j.trb.2007.02.004

.

5

Effectiveness of the Two Step approach on large networks

The Two-Step approach provided excellent results for updating demand flows on the Ring of Antwerp. This chapter validates these findings on a general network: Luxembourg City. This network represents the typical mid-sized European city in terms of network dimension. Moreover, Luxembourg City has the typical structure of a metropolitan area, composed of a city centre, ring, and suburban areas.

An innovative element of the experiments proposed in this Chapter is to use mobile network data to create a time-dependent profile of the generated demand inside and outside the ring. To support the claim that the model is ready for practical implementation, it is interfaced with PTV Visum, one of the most widely adopted software tools for traffic analysis. Results of these experiments provide a solid empirical ground in order to further develop this model and to understand if its assumptions hold for urban scenarios.

Content of this chapter has been presented in the following work:

Cantelmo, G., Viti, F. & Derrmann, T. "Effectiveness of the two-step dynamic demand estimation model on large networks". *5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)* 356–361 (2017). doi:10.1109/MTITS.2017.8005697

5.1 Introduction

Dynamic Traffic Assignment (DTA) models represent the current state of the practice for managing transportation systems. To be able to make accurate predictions about the network condition or the effect of new traffic policies, these models require a good knowledge of the travel demand, which is usually represented in the form of an origin-destination (OD) matrix.(Caceres, Wideberg, and Benitez 2007).

In order to generate this matrix, while traditional demand generation models combine survey data and statistical tools (E Cascetta 2009; McNally 2007), more recent approaches have done a significant progress into including new data sources, such as Call Detail Records (CDR), GSM data, sensing data and geospatial data (Toole et al. 2015; Donna, Cantelmo, and Viti 2015). Although these works showed that big data can largely improve the overall quality of the result, the estimated demand matrix is at most a concise representation of the regular demand patterns. Unfortunately, since dynamics of traffic systems are complex and depend on partially predictable phenomena such as weather conditions, daily demand patterns can substantially differ from the regular ones, because of structural and random deviations (Zhou and Mahmassani 2007).

These deviations can be corrected by using traffic measurements, such as loop detectors, to update the existing (a-priori) OD matrix. This problem, which is known in literature as the Dynamic Demand Estimation Problem (DODE), searches for time-dependent OD demand matrices able to best fit measured data. It can be applied for both within-day (intra-period) and day-to-day (inter-period) dynamic frameworks (Ennio Cascetta, Inaudi, and Marquis 1993), as well as for offline (medium-long term planning and design) and on-line (real-time management) (C. Antoniou et al. 2009). While for a detailed overview, the interested reader can refer to (Constantinos Antoniou et al. 2016), we limit our discussion to recent works related to the off-line DODE.

Classical approaches solve two interconnected optimisation problems, according to a bi-level formulation: in the upper level, time-dependent OD matrices are corrected in order to replicate the observations, while the lower level relates OD with path and link flows (Constantinos Antoniou et al. 2016). However, the resulting optimisation problem is highly underdetermined (Marzano, Papola, and Simonelli 2009), and provides an accurate prediction only when the ratio between unknown and known variables (OD flows and traffic measurements, respectively) is close to one. From the modelling point of view, the easiest solution is to formulate the optimisation problem in a different way, in order to reduce the number of variables. This can be done, for instance, by introducing a parametric representation of the demand, as proposed in (Lindveld 2003), or performing a Principal Component Analysis (PCA) (Djukic et al. 2012). Recently, Ennio Cascetta et al. (2013) introduced the so-called "*quasi-dynamic assumption*", which assumes that the generated demand for a certain OD pair is time dependent, while its spatial distribution is constant. Under this assumption, as demonstrated in (Ennio Cascetta et al. 2013), the DODE problem becomes less underdetermined and more likely to find more robust results. Nevertheless, the authors point out that the resulting matrix will be "*intrinsically biased*", since this assumption introduces an "*intrinsic error*". To solve this problem Cantelmo et al. (2015) introduced a Two-Step procedure, which separates the problem in two sub-optimization problems. Through this procedure, authors correct sequentially generations and distributions in the demand matrix. In essence, the first step exploits the *quasi-dynamic* assumption in order to perform a broad evaluation of the solutions space, while in the second step the estimated OD flows are further updated in order to reduce the intrinsic error.

From the data-driven point of view, the most widely adopted procedure is to include new data sources, such as measured speeds (Yamamoto et al. 2009), link density (Frederix, Viti, and Tampère 2010) and route travel time (Nigro, Cipriani, and Del Giudice 2017), within the Objective Function (OF) to be minimised. As expected, by increasing the number of knowns in the optimisation problem, and by including information on the actual route choice, the solution reliability largely increases.

Driven by these considerations, in this Chapter we implement the Two-Step approach, already presented in (Cantelmo et al. 2015), to the network of Luxembourg, and we extend the goal function in order to include mobile network data within the DODE. The contribution is twofold. On one hand, we show that the Two-Step approach outperforms the standard formation on a real-life network. To support the claim that the model is ready for practical implementation, it is interfaced with PTV Visum, one of the most widely adopted software tools for traffic analysis. The second contribution regards the mobile network data. While these data have been widely adopted for generating dynamic OD matrix (Toole et al. 2015), their use within the OD/route flow estimation is still limited (Calabrese et al. 2011). The main reason is the low level of precision of this information, which makes the match between observations and road segments quite challenging. In the proposed work, mobile network data are used to directly estimate the time-dependent demand profile, thus no matching is required

5.2 Methodology

5.2.1 MAMBA-DEV Matlab Package

To be able to solve a DODE on a large real-sized network, we developed a Matlab package for solving the DODE using PTV Visum as DTA model (Demand Estimation for Visum - DEV).

The package allows performing assignment-free dynamic or static OD estimation, using a deterministic and/or stochastic approximation of the gradient (Cantelmo et al. 2015). The model also includes the Two-Step approach, which is presented in the next section. While the MAMBA-DEV package has been designed for Luxembourg City, it can work with any network.

5.2.2 The Two-Step Approach

While for a detailed overview of this model we refer to the previous chapter (Cantelmo et al. 2015), in this section we briefly present its main characteristics.

The standard DODE, called "S*ingle-Step*" in this thesis, is generally solved as an optimisation problem. Its formulation requires the specification of the OF, its variables and its constraints, which are related to feasibility and routeing conditions. Considering different types of measures and by adopting an offline approach, the OF can be formulated as:

$$(\boldsymbol{d}_1^*, \dots, \boldsymbol{d}_n^*) = argmin \begin{bmatrix} z_1(\boldsymbol{l}_1, \dots, \boldsymbol{l}_n, \widehat{\boldsymbol{l}_1}, \dots, \widehat{\boldsymbol{l}_n}) + \\ +z_2(\boldsymbol{q}_1, \dots, \boldsymbol{q}_n, \widehat{\boldsymbol{q}_1}, \dots, \widehat{\boldsymbol{q}_n}) + \\ +z_3(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n, \widehat{\boldsymbol{d}_1}, \dots, \widehat{\boldsymbol{d}_n}) + \\ +z_4(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n, \widehat{\boldsymbol{r}_1}, \dots, \widehat{\boldsymbol{r}_n}) + \end{bmatrix} \tag{5.1}$$

Where $\mathbf{l}/\hat{\mathbf{l}}$ are the simulated values and the corresponding measurements on the links, $\mathbf{n}/\hat{\mathbf{n}}$ are the simulated values and the corresponding measurements on the nodes, $\mathbf{x}/\hat{\mathbf{x}}$ are the estimated values and a-priori information on the dynamic demand, $\mathbf{r}/\hat{\mathbf{r}}$ are the simulated values the and the measurements on routes, $\mathbf{d}_n^*$ is the estimated demand matrix for time interval *n* and, finally, z :{z1, z2, z3, z4} is the estimator of the deviations between the simulated/estimated and the corresponding measured/a-priori values. The consistency between simulated traffic performances and the estimated demand is obtained directly by performing a Dynamic Traffic Assignment (DTA).

The applicability of Equation (5.1) is general, but has its shortcomings. Among others, when dealing with a large number of variables, Equation (5.1) collapses to a local adjustment of the a-priori OD flows, rather than a real estimation. As discussed in the introduction, this is one of the main reasons for which introducing the quasi-dynamic assumption sounds reasonable. On the one hand, this introduces an approximation, while on the other it allows the algorithm to avoid local minima.

In the proposed Two-Step procedure, the first step aims at optimising the generation values of each zone in each time interval, while maintaining constant the dynamic trip distributions derived by the seed matrix. To achieve this goal, the objective function in Equation (5.1) can be generally rewritten for the first step as:

$$(\boldsymbol{E}_1^*, \dots, \boldsymbol{E}_n^*) = argmin \begin{bmatrix} z_1(\boldsymbol{l}_1, \dots, \boldsymbol{l}_n, \widehat{\boldsymbol{l}_1}, \dots, \widehat{\boldsymbol{l}_n}) + \\ +z_2(\boldsymbol{q}_1, \dots, \boldsymbol{q}_n, \widehat{\boldsymbol{q}_1}, \dots, \widehat{\boldsymbol{q}_n}) + \\ +z_3(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n, \widehat{\boldsymbol{d}_1}, \dots, \widehat{\boldsymbol{d}_n}) + \\ +z_4(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n, \widehat{\boldsymbol{r}_1}, \dots, \widehat{\boldsymbol{r}_n}) + \end{bmatrix} \quad (5.2)$$

Where $x_n^{OD} = E_n^O d_{D|O}^{Seed,n} \; \forall O, \forall D, \forall n$

Where E_n^O is the generation of origin zone O and time interval *n*, $\boldsymbol{E}_n^*$ is the generation vector containing generation from all origins in time interval *n*, X_n^* is the number of trips originated in O with destination D in time interval *n* and $d_{D|O}^{Seed,n}$ is the matrix probability distribution between traffic zone D and traffic zone O in time interval *n*.

The goal of the first step is to act on the seed matrix in order to obtain a reasonable generation value before moving to the second step, in which the dynamic distributions are corrected according to Equation (5.1) in order to reduce the intrinsic error.

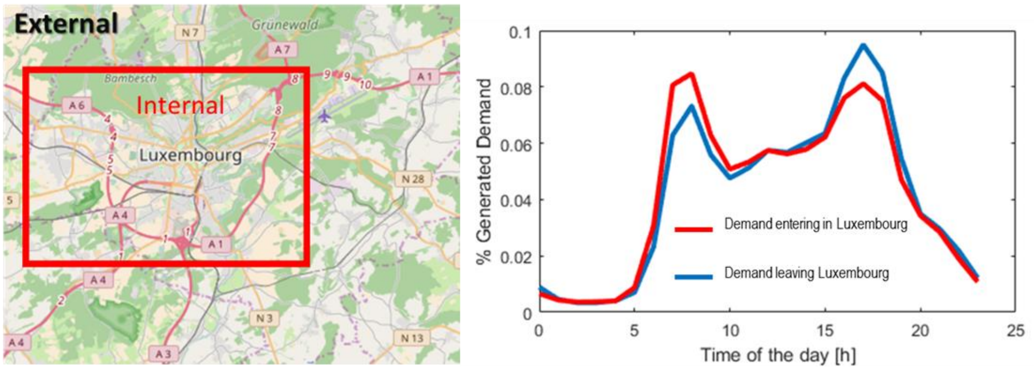
5.2.3 Including Mobile Network Data in the Objective Function

As pointed out in the introduction, it is commonly accepted that including more information within the goal function leads to a more robust result for the DODE. Clearly, this cannot be considered a general rule since, when different data sources are combined, the solution space of the OF can become more irregular. In this sense, mobile network technology, because of its spatial/temporal coverage and because of the great volume of information, seems a promising data source for the DODE. While the correlation between traffic demand and mobile data is well known (Toole et al. 2015), this source of information is hard to implement within the DODE. When dealing with GPS information, one of the most critical elements is to match the GPS coordinates and the road network. Mobile network data provide at most the geographic position at connected antenna level, so no direct road network match is possible. However, by clustering antennas located on the border of each traffic zone, it is possible to count active connections that are entering or exiting the zones (i.e. the number of *handovers*). Unfortunately, mobile network data cannot be considered as the sole source of information for the DODE, as they are subject to intrinsic errors such as the split of the user base between multiple network operators and the degree of activity on the network as well as the general mobile penetration rates. In this work, we use aggregated handover counts between antennas of 2G, 3G and 4G radio technologies of Luxembourg mobile network operator POST Luxembourg. The data consists of the hourly counts of connections being handed off between pairs of antennas, thus respecting users' privacy.

We propose the following two criteria to exploit demand emission flows estimated through the mobile network data:

- Antenna clusters need to be large enough to minimise the "ping-pong" effect, i.e. repeatedly counting the same users 'bouncing' back and forth between two antennas;
- Cluster edges shall be positioned so as to maximise the difference between number of people entering and leaving the study area;

Since we are focusing on Luxembourg City, we created two different clusters. One cluster captures the trips generated from the city to the external zones, while the other one captures those entering Luxembourg City, as shown in Figure (5.1).



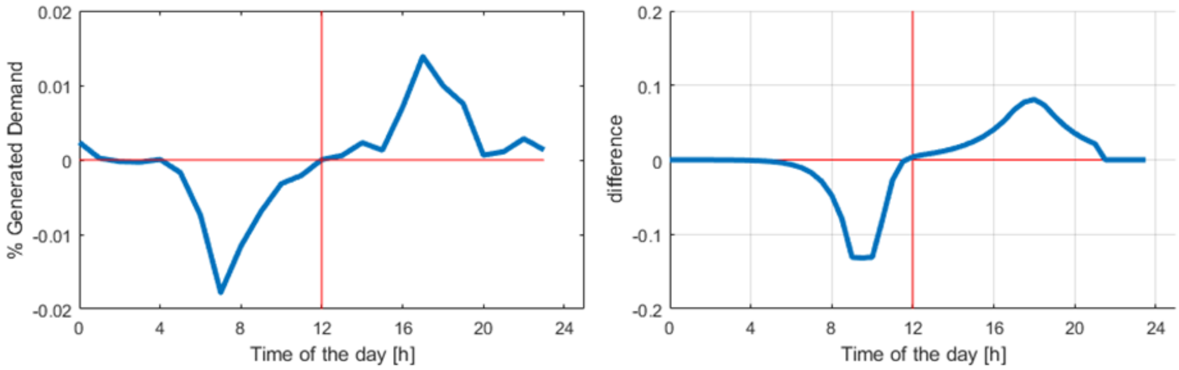
(a) (b)

Fig.5.1: (a) Internal and External antenna clusters for Luxembourg City; (b) Emission flow from and to Luxembourg City;

This procedure can be easily extended to any urban area, in which mobile connection handovers can be used to calculate the flows exchanged between the study area and the external centroids. Although the profile showed in Figure 1b looks realistic, we do believe that to simply include the emission flows within the goal function may still lead to a biased estimation, since it is equivalent to over-imposing a certain time-dependent profile to the demand. As pointed out in Chapter 2, this leads to substantial errors in the estimated mobility demand. Instead, we propose to use the difference between entering and exiting flow, as in Equation (5.3):

$$\Delta E_n^{GSM} = \frac{E_n^{GSM-IntZones}}{\sum_n E_n^{GSM-IntZones}} - \frac{E_n^{GSM-ExtZones}}{\sum_n E_n^{GSM-ExtZones}} \quad \forall n \tag{5.3}$$

Where $E_n^{GSM-IntZones}$ and $E_n^{GSM-ExtZones}$ are the mobile connection handovers to the internal and external zones, respectively. Figure 5.2 show the profile of ΔE_n^{GSM} for the real data (5.2a) and the a-priori OD matrix (5.2b).



(a) (b)

Fig.5.2: (a) Profile obtained through the real-data; (b) Profile obtained through the 4-step approach;

As showed in Figure (5.2), the profile obtained by combining the classical Four-Step approach with a departure time choice model (5.2b) is comparable to the one obtained with the (real) mobile network data (5.2a). We can also identify quite easily the two errors within the a-priori OD matrix. First, the average departure time for the morning peak is wrongly shifted in time. Second, there is a difference in the scale, on the y-axis. The reason is that, in this application, we calculate the OD flows for the morning and evening commute, thus the demand in the afternoon is highly underestimated. This suggests that, by including Equation (5.3) within the OF of the DODE, we can use mobile network data as a soft constraint to correct the demand obtained through classical demand generation models.

5.3 Case Study

Synthetic experiments have been conducted on the urban network of Luxembourg City (Figure 5.3). While real traffic measures are available in Luxembourg, authors believe that assessing the quality of the proposed algorithm in a controlled experiment is a fundamental step before moving to the practical implementation.

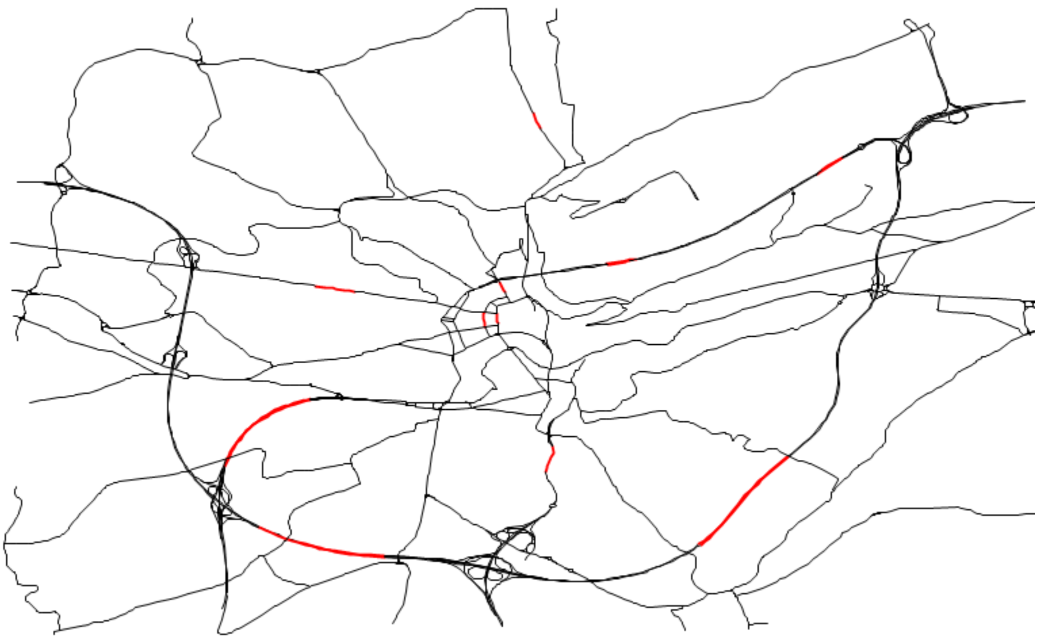


Fig.5.3: Network of Luxembourg City, Luxembourg ;

The network, which consists of 2744 active links, 1480 nodes and 17 traffic zones, represents the typical middle-sized European city in terms of network dimension. Moreover, Luxembourg City has the typical structure of a metropolitan area, composed of the city centre, ring, and suburb areas. OD flows are estimated over 24 hours assuming a 30-minutes departure interval. Under this assumption, the dynamic matrix contains 13872 variables to be estimated. The real matrix amounts to 239.966 trips, and with such an amount no congestion is expected on the network. Simulated measures for this network are available on a total of 32 counting sections – the links containing these sections are shown in red. Finally, the a-priori OD matrix, hereafter simply called *Seed* matrix, amounts to 171.060 trips, thus it significantly underestimates the number of trips in the network.

The DDEP is solved using both the Single-Step (SS) and Two-Step (TS) approaches. In both cases, the well-established Simultaneous Perturbation Stochastic Approximation (SPSA) is the numerical solution method adopted for the optimisation. In order to reduce the computational time, we adopted the one-sided version of this model. The interested reader can refer to (Cantelmo et al. 2015) for more details on the solution algorithm. Similarly, we performed two different sets of experiments:

Scenario 1. Only traffic counts are included within the OF.
Scenario 2. Traffic counts and mobile data are included within the OF.

Finally, the Root Mean Square Error (RMSE) metric is the estimator adopted to quantify the error.

5.3.1 Scenario I: Only Traffic Counts

We opted for an uncongested scenario to primarily assess the capability of the model in handling a large number of variables, while at the same time considering a smooth goal function. The gradient is calculated as the average of 300 stochastic perturbations of the current matrix for the SS and 100 stochastic perturbations for the TS model. As shown in Figure (5.4), results confirm that, when the number of variables is large, SS model performs a quite local adjustment of the OD demand. Specifically, to obtain a reliable estimation of the gradient, the number of stochastic perturbations should be approximately 10% of the number of variables (Cipriani et al. 2011). This entails 1382 DTA simulations for each iteration (~46 hours).

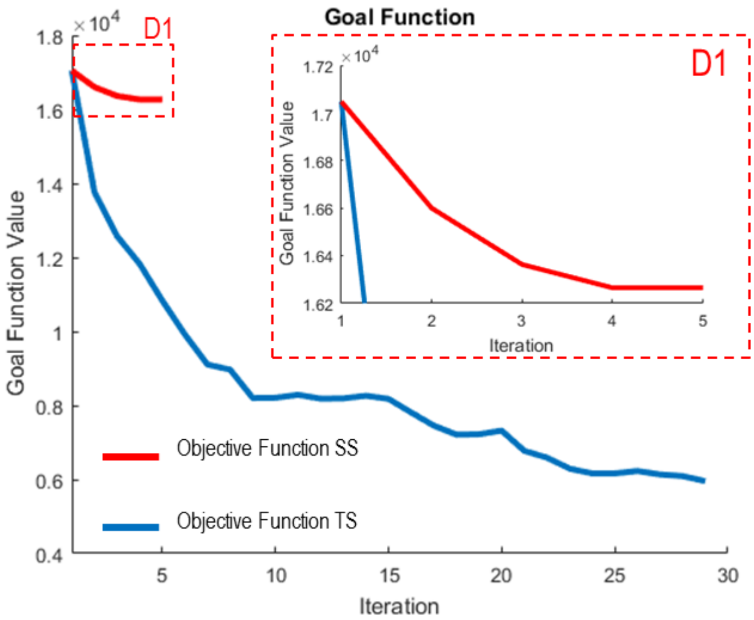


Fig.5.4: Goal Function trend;

By contrast, introducing the strict quasi-dynamic assumption, there are only 816 variables to update. As a consequence, even with fewer replications of the gradient, the estimation is more robust, and the improvement is network wide. This is shown in Figure (5.5), where the scatter of the link flows is presented, and in Table 5.1. Specifically, in Table 5.1 we reported the error with respect to real link speed and real OD flows, which are used for validation purposes.

Table 5.1: Experiment Results

	Seed	***Two-Step***	***Single Step***
RMSE Speed (Km/h)	3.73	2.47	3.66
RMSE OD (Veh/h)	42.25	37	43.00

It should be pointed out that, while the RMSE of the *Seed* may be considered low, we are considering in this experiment only morning and evening commute, thus a large number of OD/measures during the off-peak hours present a low error.

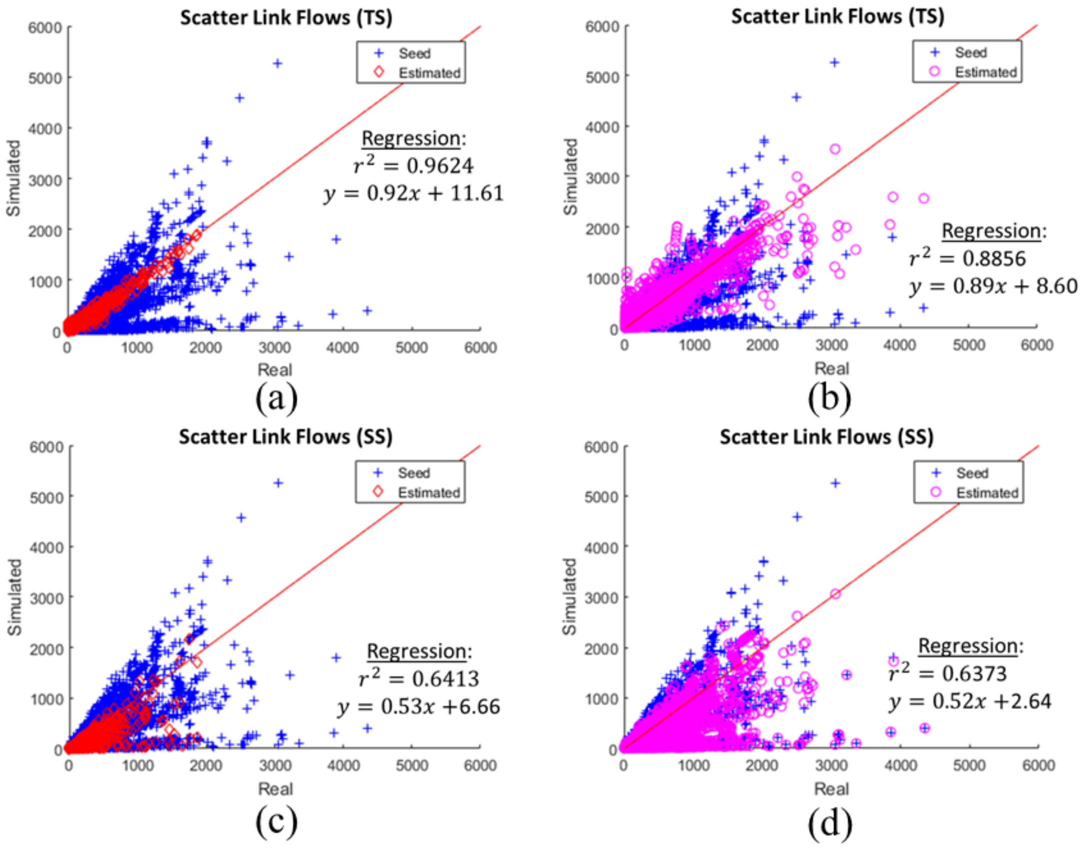


Fig.5.5: (a) Simulated vs Real link flows on the detector for the Two-Step; (b) Simulated vs Real link flows on all links for the Two-Step; (c) Simulated vs Real link flows on the detector for the Single Step; (d) Simulated vs Real link flows on all links for Single Step;

However, the resulting traffic pattern during the rush hour is substantially wrong, as shown in Figure 5.5, where we can clearly see that the *Seed* demand matrix is both overestimating and underestimating link flows.

Finally, results in Table 5.1 provide another important insight on the quality of the results. The SS model not only performed a local adjustment of the link flows but also increased the error with respect to the real matrix. By contrast, the proposed model is reducing the error according to all the performance measures.

5.3.2 Scenario II: Including Mobile Network Data

In this subsection, we show the improvement related to using the mobile network data within the goal function. In this case, the synthetic profile illustrated in Figure (5.2b) has been used to simulate the mobile network data for the synthetic experiment. Results of this experiment are quite unexpected.

Table 5.2: Experiment Results

	Seed	*TS*	*SS*
RMSE Speed (Km/h)	3.73	2.98	3.66
RMSE OD (Veh/h)	42.25	40.01	66.02

As showed in Table 5.2, the error in terms of OD flows is, at the end of the estimation, larger than in the previous case, showing that, for this uncongested network, the TS approach manages to find a better solution without the mobile data. However, as reported in Figure 5.6, when mobile data are included within the OF, the number of iterations required for solving the DDEP strongly decreases.

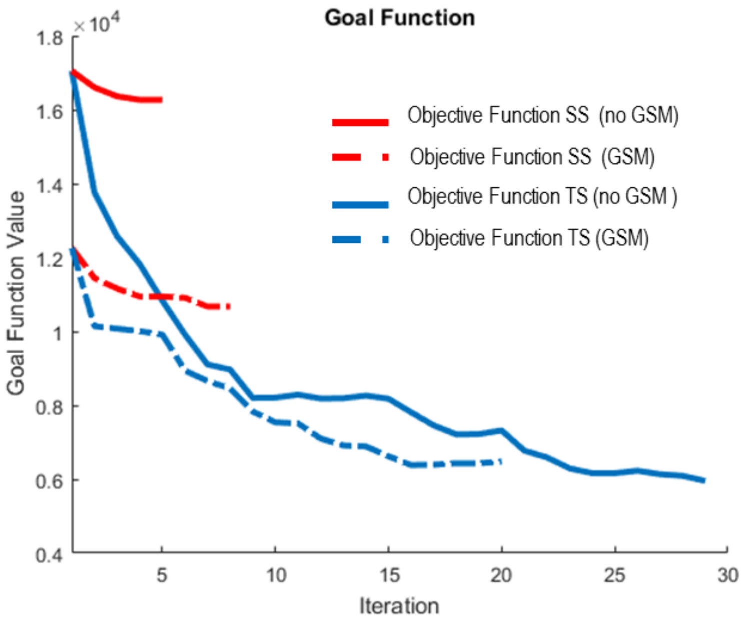


Fig.5.6: Goal Function trend, with and without GSM data;

As predictable, the same property is not observed for the SS model, which simply collapses on the closest local minima. However, when this model is combined with the mobile network data, the error on the link flows decreases with respect to the base case presented in Scenario I (the RMSE is 3% lower).

A second and fundamental result concerns the stability of the estimated matrices. The RMSE of the OD flows between the solution of Scenario I and II are 27 and 48 veh/h for the TS and SS model, respectively. Although the Single-Step model has a small OF improvement, the distance between the two estimated matrices is twice the distance of those estimated through the Two-Step approach. This means that the Two-Step approach not only manages to have a larger OF improvement but also to provide more reliable results. These findings are in line with the conclusions already presented in (Cantelmo et al. 2015). In general, we can claim that, since the Two-Step approach sequentially reduces the dimension of the solution space while keeping a lower number of variables with respect to the conventional Single-Step approach, it will provide a more reliable estimation (Marzano, Papola, and Simonelli 2009).

5.4 Conclusions and Future Research

The motivation for the research conducted in this chapter is twofold. First, we aimed to generalise the effectiveness of a two-step approach for the Dynamic Demand Estimation Problem already introduced in Chapter 4 (Cantelmo et al. 2015) for a general urban network. Second, we performed a systematic assessment for the network of Luxembourg City, a fundamental step in order to use the proposed methodology for real applications. More specifically, the proposed Two-Step approach is a simple procedure to iteratively reduce the solution space without increasing the problem complexity. Results presented in this Chapter suggest that this methodology is suited for improving the reliability of the estimated travel demand and for performing a broader analysis of the solution space with respect to the conventional approach. While this model has some similarity with the quasi-dynamic approach proposed by Cascetta et al. (Ennio Cascetta et al. 2013), by performing a double-optimization, it also manages to overcome limitations related to the so-called "*intrinsic error*" of the quasi-dynamic assumption.

From a practical point of view, the proposed model has been implemented within the MAMBA-DEV Matlab package for the OD estimation, which exploits PTV Visum as traffic assignment module. Thus, the proposed model can be easily implemented with other networks, and we can conclude that the model is ready for practice. On this point, authors incorporated GSM data as a soft constraint within the objective function, showing that this information largely increases the convergence speed.

Straightforward steps to future work are (i) validating the proposed results for a congested network and (ii) using the real data for performing the Dynamic Demand Estimation on Luxembourg City. More long-term objectives are to further extend MAMBA-DEV, in order to account for a larger set of models, including algorithms suited for solving on-line estimation and prediction problems.

Acknowledgements

We wish to acknowledge the COST Action TU1004 'TransITS', which has partly sponsored the collaboration between the authors. We wish also to acknowledge Ruben Corthout, Rodric Frederix and Willem Himpe for the assistance in the preparation of the tests and in the implementation of the algorithms.

5.5 References

Antoniou, C., R. Balakrishna, H.N. Koutsopoulos, and M. Ben-Akiva. 2009. "Off-Line and on-Line Calibration of Dynamic Traffic Assignment Systems." In , 104–11. doi:10.3182/20090902-3-US-2007.0056.

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. doi:10.1016/j.trc.2015.08.009.

Caceres, N., J.P. Wideberg, and F.G. Benitez. 2007. "Deriving Origin-Destination Data from a Mobile Phone Network." *IET Intelligent Transport Systems* 1 (1): 15–26. doi:10.1049/iet-its:20060020.

Calabrese, F., M. Colonna, P. Lovisolo, D. Parata, and C. Ratti. 2011. "Real-Time Urban Monitoring Using Cell Phones: A Case Study in Rome." *IEEE Transactions on Intelligent Transportation Systems* 12 (1): 141–51. doi:10.1109/TITS.2010.2074196.

Cantelmo, G., F. Viti, E. Cipriani, and N. Marialisa. 2015. "A Two-Step Dynamic Demand Estimation Approach Sequentially Adjusting Generations and Distributions." In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 1477–82. doi:10.1109/ITSC.2015.241.

Cascetta, E. 2009. *Transportation Systems Analysis*. 2nd ed. Vol. 29. Springer. http://www.springer.com/us/book/9780387758565.

Cascetta, Ennio, Domenico Inaudi, and Gerald Marquis. 1993. "Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts." *Transportation Science* 27 (4): 363–73.

Cascetta, Ennio, Andrea Papola, Vittorio Marzano, Fulvio Simonelli, and Iolanda Vitiello. 2013. "Quasi-Dynamic Estimation of O–d Flows from Traffic Counts: Formulation, Statistical Validation and Performance Analysis on Real Data." *Transportation Research Part B: Methodological* 55 (September): 171–87. doi:10.1016/j.trb.2013.06.007.

Cipriani, E., M. Florian, M. Mahut, and M. Nigro. 2011. "A Gradient Approximation Approach for Adjusting Temporal Origin-Destination Matrices." *Transportation Research Part C: Emerging Technologies* 19 (2): 270–82. doi:10.1016/j.trc.2010.05.013.

Djukic, T., G. Flötteröd, Lint Van, and S. Hoogendoorn. 2012. "Efficient Real Time OD Matrix Estimation Based on Principal Component Analysis." In , 115–21. doi:10.1109/ITSC.2012.6338720.

Donna, S. A. Di, G. Cantelmo, and F. Viti. 2015. "A Markov Chain Dynamic Model for Trip Generation and Distribution Based on CDR." In *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS),* 243–50. doi:10.1109/MTITS.2015.7223263.

Frederix, R., F. Viti, and C. M. J. Tampère. 2010. "A Density-Based Dynamic OD Estimation Method That Reproduces within-Day Congestion Dynamics." In *13th International IEEE Conference on Intelligent Transportation Systems,* 694–99. doi:10.1109/ITSC.2010.5625220.

Lindveld, K. 2003. "Dynamic O-D Matrix Estimation: A Behavioural Approach." PhD Thesis, TU Delft.

Marzano, V., A. Papola, and F. Simonelli. 2009. "Limits and Perspectives of Effective O-D Matrix Correction Using Traffic Counts." *Transportation Research Part C: Emerging Technologies* 17 (2): 120–32. doi:10.1016/j.trc.2008.09.001.

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

Nigro, M., E. Cipriani, and A. Del Giudice. 2017. “Exploiting Floating Car Data for Time-Dependent O-D Matrices Estimation.” *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* under review.

Spall, James C. 2012. “Stochastic Optimization.” In *Handbook of Computational Statistics*, edited by James E. Gentle, Wolfgang Karl Härdle, and Yuichi Mori, 173–201. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-642-21551-3_7.

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. doi:10.1016/j.trc.2015.04.022.

Yamamoto, Toshiyuki, Tomio Miwa, Tomonori Takeshita, and Takayuki Morikawa. 2009. “Updating Dynamic Origin-Destination Matrices Using Observed Link Travel Speed by Probe Vehicles.” In *Transportation and Traffic Theory 2009: Golden Jubilee*, 723–738. Springer. http://link.springer.com/chapter/10.1007/978-1-4419-0820-9_35.

Zhou, Xuesong, and Hani S. Mahmassani. 2007. “A Structural State Space Model for Real-Time Traffic Origin–destination Demand Estimation and Prediction in a Day-to-Day Learning Framework.” *Transportation Research Part B: Methodological* 41 (8): 823–40. doi:10.1016/j.trb.2007.02.004.

6

Application to real large congested networks

In the previous chapter, synthetic and real-life experiments demonstrated that the proposed Two-Step framework is suited for performing the DODE problem on uncongested urban networks and congested motorways.

In this chapter, we move one-step further, by testing this methodology with real data on the network of the Grand Duchy of Luxembourg, which includes the capital, Luxembourg City, the motorway network and the most important road arteries within the national borders.

Traffic counts and average speeds have been used to compare results obtained through the proposed methodology with the one obtained by using a standard bi-level formulation. Results show how the proposed model outperforms the standard one, as breaking the optimisation process in two parts strongly reduces the localism of the problem.

Content of this chapter has been presented in the following work:

Cantelmo, Guido, and Francesco Viti. 2018. "Assessing the Performances of a Two-Step Prediction Model on a Large Scale Congested Network Using Real Traffic Data." *Transportation Research Board 2018*.

6.1 Introduction

Dynamic Traffic Assignment (DTA) models represent essential tools for managing transportation systems. DTA models take as input the demand from each origin and destination and at each time period, and in turn estimate and/or predict route and link flows on transportation networks.

In order to generate the mobility demand, usually represented in the form of Origin-Destination (OD) matrices, traditional approaches combine survey data and mathematical tools (McNally 2007). Additionally, more recent works have done a significant progress into including new data sources, such as Call Detail Records (CDR), GSM data, sensing data and geospatial data (Toole et al. 2015). Unfortunately, the estimated demand matrix is at most a concise representation of the systematic component of the demand – such as the typical behaviour during a working day. However, daily demand patterns can substantially differ from the systematic ones because of several elements, including weather conditions or road works, as well as because of the inherent stochasticity of the travel demand. Deviations between estimated and actual demand patterns can be mitigated by using traffic measurements, which can be used to update an existing (a-priori) OD matrix. This problem, which is known in the literature as the Dynamic Origin-Destination Estimation (DODE) problem, exploits a properly specified objective function for estimating the time-dependent OD flows.

While the DODE problem has been initially treated as an extension of its static counterpart, the last decades have witnessed to a considerable effort by researchers in order to develop methodologies able to deal with the dynamic case (K. Ashok and Ben-Akiva 2002). As DTA models are applied in both *offline* (medium-long term planning and design) and *online* (real-time management) contexts, DODE is commonly classified between sequential or simultaneous approaches, where usually the first is adopted for *online* while the second for *offline* applications (K. Ashok and Ben-Akiva 2002). By limiting our focus to the *offline* case, DODE is usually formulated as a bi-level optimisation problem. In the upper level, OD flows are updated by minimising the error between simulated and observed traffic data, while in the lower level the DTA solves the combined Route Choice (RC) and Dynamic Network Loading (DNL) problems (Tavana 2001).

Earlier DODE models explicitly accounted for the assignment matrix – i.e. the set of rules linking OD and link flows - for updating the demand vector. However, this matrix assumes a linear relation between demand and supply parameters, assumption that does not hold for congested networks (Rodric Frederix, Viti, and Tampère 2013). In order to overcome this issue, Balakrishna, Ben-Akiva, and Koutsopoulos (2007) proposed a bi-level formulation that does not rely on this information. Instead, the authors suggested using a simulation-based DTA model to generate traffic measures and to include additional information, such as link speed, within the objective function, in order to represent the congested/uncongested network conditions. Following this seminal work, many researchers developed new and more robust assignment-free algorithms able to properly capture the non-linearity between link-flow propagation and time-varying OD demand (R. Frederix et al. 2011; Cipriani et al. 2011; Antoniou et al. 2015; Tympakianaki, Koutsopoulos, and Jenelius 2015). Despite this intense effort, the resulting optimisation problem remains highly non-linear and non-convex. To reduce the number of possible solutions, classical methods often include information about a reference OD demand matrix (usually known as historical or "seed" matrix) within the objective function. Therefore, if the structure of this seed matrix is different from the real one, this localism can lead to substantial errors (Rodric Frederix, Viti, and Tampère 2013).

Recently, Marzano, Papola, and Simonelli (2009) pointed out that DODE is generally unable to provide an effective estimation when the ratio between unknown and known variables (OD flows and traffic measurements, respectively) is greater than one. Hence, the easiest solution is to reformulate the objective function in order to reduce the number of variables. This can be done, for instance, by using Principal Component Analysis (PCA) (Djukic, Van Lint, and Hoogendoorn 2012). Alternatively, Cascetta et al. (2013) introduced the so-called "quasi-dynamic assumption", which assumes that the generated demand for a certain OD pair is time dependent, while its spatial distribution is constant. Under this assumption, as demonstrated in (Cascetta et al. 2013), the DODE problem becomes less underdetermined and more likely to find more robust results. Nevertheless, the authors point out that the resulting matrix will be "intrinsically biased" since this assumption introduces an "intrinsic error". Similarly, Cantelmo et al. (2015) proposed a Two-Step procedure, which separates the DODE in two sub-optimization problems. The first step searches for generation values that best fit the traffic data while keeping spatial and temporal distributions constant. In the second step, the standard bi-level procedure searches for a more reliable demand matrix.

Although the model has been tested on a simple or synthetic networks, the Two-Step approach has three characteristics that make it an ideal candidate for applications on large-scale networks. First, as pointed out by Antoniou et al. (2016), the starting matrix is still a key input for all state-of-the-art DODE models. The first step of this formulation improves the historical demand matrix by performing a broad evaluation of the solution space and estimating a "good" updated seed matrix to be used in the second step. Secondly, the proposed model reduces the number of variables in the first step, increasing the overall reliability of the results (Marzano, Papola, and Simonelli 2009; Cantelmo et al. 2015). On this point, the idea of performing successive iterations and linearization has been already introduced and validated in (Kalidas Ashok 1996) for the online DODE, showing that the reliability of the results generally increases.

Driven by these considerations, the contribution of this chapter is twofold. First, we apply the Two-Step approach to the real metropolitan network of Luxembourg. While the previous studies (Cantelmo et al. 2015, in press) tested the algorithm on a simple motorway, this chapter shows that the Two-Step approach outperforms the standard formulation on a real-life network. We numerically demonstrate that properties of robustness and reliability hold for a general network, and that the localism of the model strongly decreases. The test-network represents most of the country of Luxembourg, including urban roads, motorways and primary roads. Real traffic counts extracted from loop detectors are used within the calibration process to update the demand.

Second, as speed profiles on the counting stations were not available, we extend the objective function by including the average speeds over the analysis period, which have been calculated through Floating Car Data (FCD). We show that, when combined with a standard DODE procedure, this information leads to a poor calibration of the demand, as the DODE overfits the data within the objective function. However, as the Two-Step approach over-imposes a linear relation between distribution and generation for a certain traffic zone, it is more likely to capture congestion dynamics at network level, such as the systematic overestimation or underestimating of the demand, thus to avoid this issue.

Finally, to support the claim that the model is ready for practical implementation, it is interfaced with PTV-Visum, one of the most widely adopted software tools for traffic analysis (PTV-GROUP 2014).

The chapter is structured as follows. The next paragraph defines the methodology, including the "conventional" model (called Single-Step OD estimation in the rest of this chapter) and the proposed Two-Step approach. The chapter then describes the case study, including the network, the dataset used for the experiments and the results. Finally, in the last section conclusions are drawn.

6.2 Methodology

The DODE is usually formulated as a constrained optimisation problem, which requires the formulation of:

i. An objective function, which is composed of variables and constraints related to routeing conditions and behavioural assumptions;
ii. An optimisation method, which can be classified in Path Search, Pattern Search or Random Search approaches (Balakrishna, Ben-Akiva, and Koutsopoulos 2007);
iii. A parameter updating rule;

The remaining part of this section describes the set of functions and algorithms used for performing Single-Step and Two-Step demand estimation.

6.2.1 The Objective Function

6.2.1.1 Standard Simultaneous Generalized Least Squared

The most widely adopted goal function for solving the *offline* DODE is the Generalized Least Squared (GLS) proposed in (Cascetta, Inaudi, and Marquis 1993). Considering different types of measures and a *simultaneous* approach, the problem can be formulated as:

$$(\boldsymbol{d}_1^*, \dots, \boldsymbol{d}_n^*) = argmin \begin{bmatrix} z_1\left(\boldsymbol{l}_1, \dots, \boldsymbol{l}_n, \widehat{\boldsymbol{l}_1}, \dots, \widehat{\boldsymbol{l}_n}\right) + \\ +z_2\left(\boldsymbol{q}_1, \dots, \boldsymbol{q}_n, \widehat{\boldsymbol{q}_1}, \dots, \widehat{\boldsymbol{q}_n}\right) + \\ +z_3\left(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n, \widehat{\boldsymbol{d}_1}, \dots, \widehat{\boldsymbol{d}_n}\right) + \\ +z_4\left(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n, \widehat{\boldsymbol{r}_1}, \dots, \widehat{\boldsymbol{r}_n}\right) + \end{bmatrix} \tag{6.1a}$$

Where $\mathbf{l}/\hat{\mathbf{l}}$ represent, respectively, simulated and measured link performances, $\mathbf{q}/\hat{\mathbf{q}}$ calibrated and observed values on the node, $\mathbf{x}/\hat{\mathbf{d}}$ indicate the estimated and historical value for the OD flows (seed matrix) and $\mathbf{r}/\hat{\mathbf{r}}$ the simulated and observed route performances. Finally, $\boldsymbol{d}_{\boldsymbol{n}}^*$ designates the estimated demand matrix for time interval *n*, while $z\colon\{z_1, z_2, z_3, z_4\}$ is the estimator of the error between simulated/estimated and measured/a priori values.

The dependence between supply and demand in Equation (6.1a) is obtained directly by simulation performing a dynamic traffic assignment (DTA), so that:

$$\begin{aligned} \boldsymbol{l}_1, \dots, \boldsymbol{l}_n &= \mathbf{F}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \\ \boldsymbol{q}_1, \dots, \boldsymbol{q}_n &= \mathbf{F}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \\ \boldsymbol{r}_1, \dots, \boldsymbol{r}_n &= \mathbf{F}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n) \end{aligned} \tag{6.1b}$$

with function **F** representing the Dynamic Traffic Assignment (DTA) function. The objective function presented in Equation (6.1a) presents a series of agreeable properties that make it an ideal candidate for assignment-matrix free algorithms. First, apart from the traffic counts, the function may account for different sources of information, such as link speeds and densities – which have been proved to capture the non-linear relation between demand and supply parameters (Balakrishna, Ben-Akiva, and Koutsopoulos 2007; R. Frederix, Viti, and Tampère 2010). Moreover, recent works showed how more elaborate information, such as point-to-point data, can also be included in this function, largely improving the overall estimation accuracy (Barceló and Montero 2015; Mitsakis et al. 2013; Antoniou et al. 2016). An additional advantage of the simultaneous GLS presented in Equation (6.1a) with respect to the sequential case is that all variables are jointly estimated, which is formally more correct as OD flows over different time intervals are likely to be correlated (Rodric Frederix, Viti, and Tampère 2013). However, for large networks, this approach becomes less reliable and, if not enough traffic data is available (Marzano, Papola, and Simonelli 2009), the sequential approach is preferred.

6.2.1.2 Strict Quasi-Dynamic Simultaneous Generalized Least Squared

As suggested in (Cascetta et al. 2013), the objective function described in Equation (6.1) can be enhanced by exploiting information on aggregated socio-demographic data such as generation data by traffic zones. The objective function (6.1a) can be then reformulated as:

$$(\boldsymbol{E}_1^*, \dots, \boldsymbol{E}_n^*) = argmin \begin{bmatrix} z_1\left(\boldsymbol{l}_1, \dots, \boldsymbol{l}_n, \widehat{\boldsymbol{l}_1}, \dots, \widehat{\boldsymbol{l}_n}\right) + \\ +z_2(\boldsymbol{q}_1, \dots, \boldsymbol{q}_n, \widehat{\boldsymbol{q}_1}, \dots, \widehat{\boldsymbol{q}_n}) + \\ +z_3\left(\boldsymbol{x}_1, \dots, \boldsymbol{x}_n, \widehat{\boldsymbol{d}_1}, \dots, \widehat{\boldsymbol{d}_n}\right) + \\ +z_4(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n, \widehat{\boldsymbol{r}_1}, \dots, \widehat{\boldsymbol{r}_n}) + \end{bmatrix} \tag{6.2a}$$

S.t.

$$x_n^{OD} = E_n^O d_{D|O}^{Seed,n} \tag{6.2b}$$

Where:

- $E_n{}^O$ = generated flow from traffic zone *O* and time interval *n*;
- $\boldsymbol{E}_n^*$ = generation vector containing the generated flow from all zones in time interval n.
- x_n^{OD} = demand flow from origin zone O to destination zone D in time interval *n*.
- $d_{D|O}^{Seed,n}$ = seed matrix spatial/temporal distribution to move in traffic zone D from traffic zone O in time interval *n*.

Constraint (6.2b) is the main difference between the general quasi-dynamic formulation proposed in (Cascetta et al. 2013) and the one proposed in Equation (6.2). The former explicitly considers a probability function that captures the correlation between generation and distribution over a certain sub-period of time. As a consequence, $d_{D|O}^{Seed,n}$ is updated during the optimization process. Instead, constraint (6.2b) assumes a constant value of the distribution, resulting in a smoother objective function. Equation (6.2b) presents two major advantages with respect to the simultaneous GLS. First, as the number of unknown variables strongly decreases, the simultaneous approach can be applied to larger networks. Second, this approach does not necessarily require to explicitly account for historical OD flows within the objective function. As pointed out in the introduction, historical OD flows are usually included within equation (6.1a) in order to reduce the number of possible solutions. However, this information is already considered within constraint (6.2b), that over-impose, to the estimated matrix, the spatial/temporal structure of the historical demand. However, a main drawback of this formulation is that it is likely to provide a poor fit of the traffic data with respect to equation (6.1) or to the general quasi-dynamic formulation, as pointed out in (Cantelmo et al. 2015). Thus, it is an ideal candidate for being used in the first phase of the Two-Step approach, where the main purpose is to have a broad evaluation of the solution space, rather than to best fit the observations.

6.2.2 Optimization method: The SPSA

The optimisation method adopted in this work is the Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm proposed in (Spall 2012). While we adopt the original model in this Chapter, as has been proven to be very effective for tackling the DODE problem, many authors proposed enhanced versions (Antoniou et al. 2015; Tympakianaki, Koutsopoulos, and Jenelius 2015; Cipriani et al. 2011), which can also be combined with the proposed framework. The SPSA is a stochastic approximation of the deterministic finite difference gradient method, which has been proved to be very effective for tackling the DODE, but becomes computationally too expensive for large networks (R. Frederix et al. 2011). By assuming a one-sided perturbation (Cipriani et al. 2011), the SPSA computes the approximated gradient $\boldsymbol{G^i}$ at each iteration *–i* as:

$$\widehat{\boldsymbol{g}}_k(\boldsymbol{\theta}^i) = \frac{z(\boldsymbol{\theta}^i + c^i\Delta^k) - z(\boldsymbol{\theta}^i)}{c^i}\begin{bmatrix}(\Delta_1^k)\\ \vdots \\ (\Delta_r^k)\end{bmatrix} \tag{6.3a}$$

$$\boldsymbol{G^i} = \overline{\boldsymbol{g}}(\boldsymbol{\theta^i}) = \frac{\sum_{\boldsymbol{k=1}}^{\boldsymbol{Grad\_rep}} \widehat{\boldsymbol{g}}_k(\boldsymbol{\theta}^i)}{Grad\_rep} \tag{6.3b}$$

With $\boldsymbol{\theta}^i$ the vector with the estimated variables, $z(\boldsymbol{\theta}^i)$ the objective function value in $\boldsymbol{\theta}^i$, c^i the perturbation step, *Grad_rep* the number of replications to compute the average gradient and Δ is a vector with elements {-1,1}. Given the stochastic nature of the model, it is recommended to repeat the perturbation multiple times in order to obtain a good approximation. If only one replication is used, then $\boldsymbol{G^i} = \widehat{\boldsymbol{g}}_k$. In Equation (6.3a), the asymmetric design (SPSA-AD) model is showed. The main advantage of using this formulation is that it allows to reduce the number of simulations needed while still providing a proper approximation of the gradient (Cipriani et al. 2011).

6.2.3 Parameter updating rule

Given a properly specified objective function and a descent direction – the gradient $\boldsymbol{G}^i$– the parameters are updated at each iteration according to:

$$\boldsymbol{\theta}^{i+1} = \boldsymbol{\theta}^i - \alpha^i \boldsymbol{G}^i \tag{6.4}$$

Where α^i is the stepsize and $\boldsymbol{\theta}^i$ is again the vector of parameters to be updated, the OD or the Generation flows if we are minimizing, respectively, objective function (6.1) or (6.2). Concerning the value of α^i, we proposed to use a line search to find the optimal value in order to reduce the overall computational time.

6.2.4 Single-Step and Two-Step approaches

The Single-Step OD estimation is formulated in this work as a single constrained optimisation problem, which minimises Equation (6.1) according to a certain optimisation method, the SPSA, and the parameter updating rule showed in Equation (6.4). Results of this, quite general, optimisation framework depend on the overall quality of the initial seed matrix (Antoniou et al. 2016). While a more elaborate algorithm may improve the performances of the standard SPSA when applied to large networks, this critical element still remains (Antoniou et al. 2016, 2015). The main contribution of breaking the optimisation problem in two phases is to relax this strong limitation.

The proposed Two-Step approach combines the set of rules, functions and algorithms described in the previous sub-sections. Specifically, the first step minimises Equation (6.2) in order to optimise the generated demand flows for each zone in each time interval. Hence, in the first phase, the variables are the total generated demand flows, which reduces the dimension of the problem considerably. In the second step, the classical DODE procedure is performed by minimising Equation (6.1), improving temporal and spatial matrix distributions. Breaking the problem as such, one benefits of the right demand level identified in the first phase. As the objective function presented in (6.2) reduces the number of variables used, it becomes less sensitive to the network size. Thus, the estimated matrix can be used in the second level of the Two-Step approach, where the optimisation can exploit a better initial point in order to achieve overall better results.

The idea of updating the generation in the first step derives from the increasing attention received by this type of aggregated information in the literature (Cipriani et al. 2011; Cascetta et al. 2013). This high significance derives mainly by the following considerations:

- Total generated trips can limit a demand overestimation during the DODE, which is otherwise likely to occur when dealing with congested networks;
- As generation models are considered the most reliable models in transport engineering applications, total generated trips are more easily observable than OD trips;
- Adopting the generation values inside the DODE, as in (2), reduces the number of variables.

The goal of the first step is to act on the seed matrix in order to obtain a "right level of demand", then moving to the second step in order to optimise the dynamic distributions OD trips as in (6.1). However, it should be stressed out that the main advantage in using the generation is not related to the nature of the data – i.e. we have a better knowledge of the generated demand with respect to the number of trips. The main advantage of the proposed methodology is instead exploiting the mathematical relation between OD demand and generated trips in order to reduce the number of variables and create a smoother objective function.

6.3 Case Study: Luxembourg

We now test both approaches to the real large-sized network of Luxembourg, showed in Figure (6.1). The Grand Duchy of Luxembourg is a small country placed in the heart of Europe, bordered by Belgium to the west, Germany to the east and France to the south. As most of the activities are located in the capital, Luxembourg City, the country is facing mobility challenges, which are being made worse by the 170.000 workers - about 43% of the commuting demand (Sprumont, Astegiano, and Viti 2017) - coming every day to Luxembourg City from the neighbouring countries.

The main goal of the case study proposed in this section is to model the complex interaction between the cross-borders – commuters coming from France, Germany and Belgium – and the road users living within the Grand Duchy's borders. This latter demand segment can be further divided into people living in the capital and people living in the countryside, where the second one is the predominant component of the commuting demand. The ring of Luxembourg City represents the bottleneck for this system, as its capacity is not sufficient to properly serve the high volume of demand moving to the capital during the rush hour, hence major congestion patterns are reported every day.

The network, showed in Figure (6.1), includes all national motorways, which go from the city of Ettelbruck to Luxembourg City in the north, and from the capital to the east, west and south borders. Additionally, the network includes also primary and secondary roads, as they are commonly used by commuters.

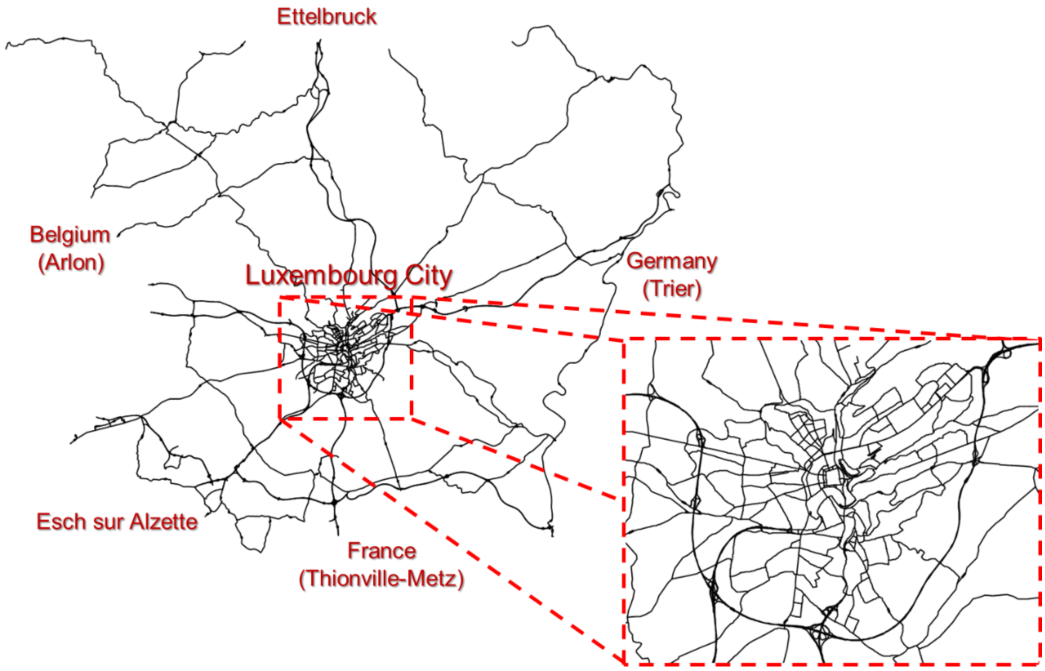


Fig.6.1: The Grand Duchy of Luxembourg network, with detail of Luxembourg City;

6.3.1 Description of the Data

As part of the initiative "Digital Luxembourg", the Grand Duchy is developing a new open-data portal (https://data.public.lu), which gathers different sources of information including socio-demographic data. These data, collected by the National Institute of Statistics (STATEC), include the growth of the population for each year, the population for each canton and number of cross-borders. Based on these statistics, a static matrix for the morning commute has been estimated through the classical Four-Step demand generation model. A departure time choice model based on the Vickrey/Small (Small 2015) formulation has been then used to derive a dynamic OD matrix from the static one. This dynamic matrix accounts for 46 traffic zones and represents the historical demand (Seed Matrix) for the experiments presented in the next sub-sections.

Concerning the supply side, the Luxembourgish Road Administration agency collects and provides traffic counts on most of the motorways and primary roads of the Grand Duchy. Unfortunately, these data present two major limitations. The first main limitation is that, based on the publicly available data, only three detectors are located inside the ring of Luxembourg. This means that we can expect to have a realistic representation of the demand on the regional network and on the ring, but it is not possible to validate the estimated solution inside the city. The second problem concerns the time interval aggregation for these data, as traffic counts are aggregated on an hourly basis. This time interval is clearly too large for a network with an average free-flow travel time of 20 minutes since basic congestion dynamics could not be properly captured. Finally, neither the open-data portal nor the Luxembourgish Road Administration provides information on the speed, which is an essential input when dealing with large congested networks such as the one proposed in Figure (6.1).

To deal with this lack of information, additional data have been provided by Motion-S Luxembourg. The company provided average speeds on the ring of Luxembourg for each time interval, which have been calculated by using Floating Car Data (FCD). The obtained information, depicted in Figure (6.2), is based on the average of all available information and does not contain specifications about time and location.

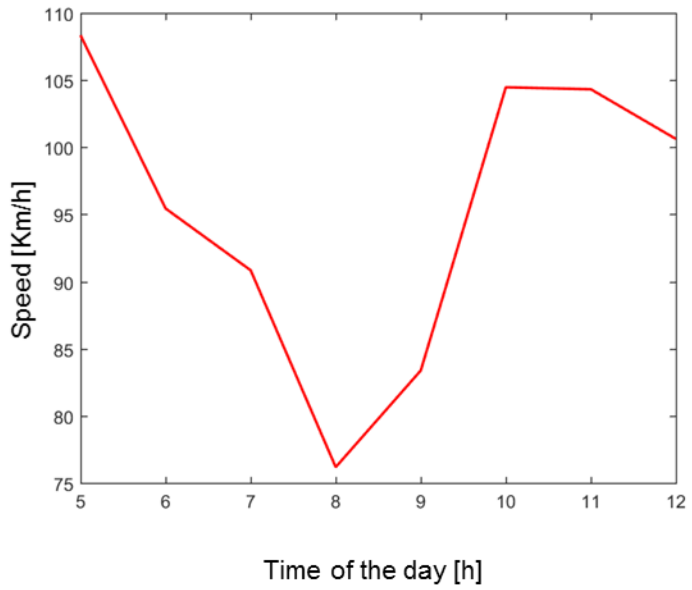


Fig.6.2: Measured Average Speed on the ring way around Luxembourg City;

FCD data carry definitely more information than just the average speed, as demonstrated in (Nigro, Cipriani, and Del Giudice 2017), but privacy laws do not allow sharing sensible data in Europe. Nevertheless, Figure (6.2) shows that the average speed properly capture the expected behaviour on the ring way at a network level, as we see a clear drop in the speed that is slowly recovered at the end of the peak hour.. The downside is that many possible solutions exist, which can create congestion on the ring. As a consequence, the most logical solution for the DODE should be to keep the demand as close as possible to the historical demand, while at the same time reproducing the speed profile. However, as this information is strongly aggregate, the Single-Step approach has the tendency to over-fit the average speed, while the Two-Step approach manage to provide more reliable results by exploiting the link flows as a constraint within the objective function. This claim is numerically illustrated in the next section.

6.3.2 Experiment Setup

The network introduced in the previous section consists of 3700 links and 1469 nodes. Luxembourg City, located in the heart of the system, represents the typical middle-sized European city in terms of network dimension and has the typical structure of a metropolitan area, composed of a city centre, the ring, and suburb areas. Considering the speed profile and that the infrastructure is composed of highways, primary roads and urban roads, we can classify this system as a large-sized heavily congested network. In this study, we consider the morning peak between 5 AM and noon (8 hours). After some data cleaning, 54 counting stations have been retained, all located on the main arterial roads going to Luxembourg City and on the ring. The seed-matrix accounts for 307.544 trips and 16928 time dependent OD pairs. Both traffic counts and the average speed are included in the objective function, where the Root Mean Squared Error (RMSE) is the chosen estimator z: $\{z_1, z_2\}$**:**

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(\hat{l}_i - l_i)^2}{N}} \tag{6.5}$$

Where N is the number of observations, $\hat{q}_i$ is the observed value for the measured data and q_i is the simulated one. In order to have a reliable estimation of the gradient, we performed 40 gradient replication for the first phase of the Two-Step approach, and 300 replications for both the second-phase of the Two-Step framework and the Single-Step approach. Although these settings lead to a high computational time, the SPSA is expected to produce an accurate approximation of the gradient. In terms of computational time, each model took about 3-4 days of simulation. Considering the size of the network, we consider this estimation time acceptable. However, it is possible to reduce the computational time by decreasing the number of gradient replication. Finally, the second phase of the Two-Step approach has the same parameters as the Single-Step, while the first phase has a larger perturbation step size.

Finally, to be able to solve the DODE on the network of Luxembourg, we developed a Matlab package for solving the dynamic O-D estimation using PTV Visum as DTA model. The package, named MAMBA-DEV, allows performing assignment-free dynamic or static OD estimation, using a deterministic and/or stochastic approximation of the gradient. The package also includes the Two-Step approach discussed in this thesis. While the MAMBA-DEV package has been designed for Luxembourg, it can work with any network in Visum, supporting the idea that the model is ready for practical implementation. Each simulation takes about 2 minutes.

6.3.3 Results

6.3.3.1 Comparison between Single-Step and Two-Step approaches

The first experiment proposed in this section aims to numerically validate two properties of the Two-Step approach formulated in the methodology section:

- The Two-Step approach outperforms the standard one on big sized networks;
- The first step is likely to find a good initial point to be updated through the objective function presented in Equation (6.1);

The starting point of this experiment is not a "good initial point", as it derives from a static matrix and has not previously been calibrated. The initial matrix provides in fact a rather poor fit with the traffic counts ($r^2 = 0.2686$ and $RMSE_{link-flows} = 452.61\, Veh/h$). In order to reduce this error, weights have been considered so that the traffic counts are responsible for 70% of the overall error within the objective function, while the average speed is responsible for the remaining 30%, thus a relatively poor representation of the average speed is expected.

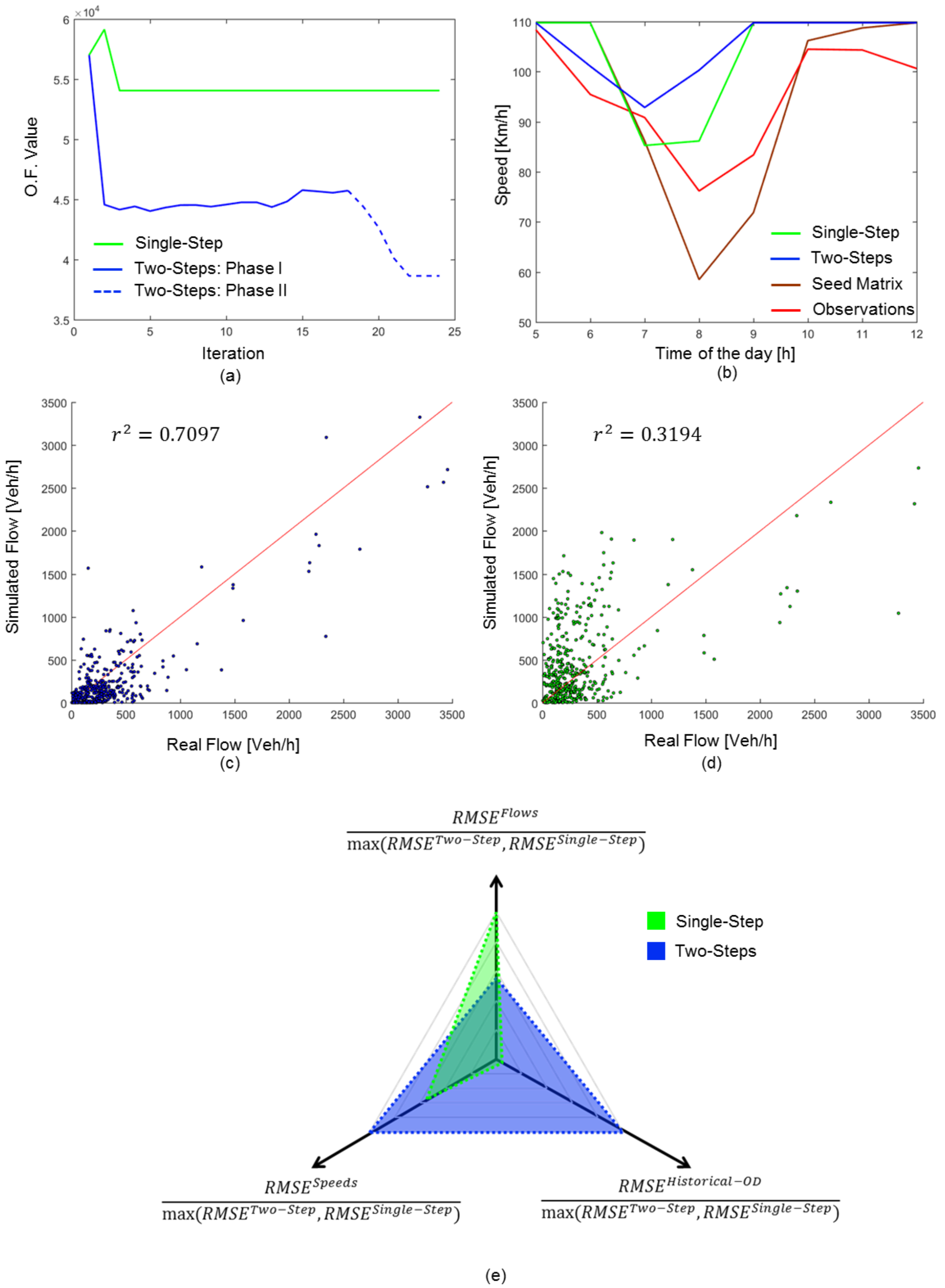


Fig.6.3: (a) Objective Function trend; (b) Estimated and Observed Average Speed; (c) Scatter Estimated and Observed Link Flows for the Two-Step; (d) Scatter Estimated and Observed Link Flows for the Single-Steps; (e) Spider Chart of the relative error for the estimated matrix in terms of Link-Flows, Distance from the Historical OD flows (Seed Matrix) and Average Speed;

As showed in Figure (6.3), the Two-Step approach clearly outperforms the Single-Step in terms of estimation results, as the latter just collapses on the closest local minima. While the model reduces error on the traffic counts ($r^2 = 0.3194$), these results are far from being acceptable for any practical application ($RMSE_{Link-Flow}^{Single-Step} = 438.89\ Veh/h$).

By contrast, results from the Two-Step approach seem more reasonable and similar to the expectations ($r^2 = 0.7097,\ RMSE_{Link-Flow}^{Two-Step} = 241.31\ Veh/h$). During the first phase, the model exploits Equation (6.2) to explore the solution space by updating only the generations. After finding a local minimum, the model switches to Equation (6.1) in order to find the best fit with the observations.

It should be pointed out that the second step of the model is basically adopting the same algorithm as the Single-Step approach. The only difference is the starting point, which has been updated during the first step of the algorithm. While this framework collapsed in a few iterations when coupled with the historical seed-matrix, exploiting the more reliable demand matrix estimated through Equation (6.2) gives a relevant contribution to the overall optimisation, stressing how both phases of the Two-Step approach are complementary and, thus, necessary.

Figure (6.3e) depicts the Spider Chart plot of the estimation error for speeds, flows and seed-matrix – i.e. the initial point. For each measure, this relative error has been calculated as:

$$Rel_{error} = \frac{RMSE^{Model}}{\max(RMSE^{Two-Step}, RMSE^{Single-Step})} \tag{6.6}$$

Figure (6.3e) intuitively shows the dynamics behind the optimization. The Single-Step approach does not manage to move from the initial point, thus to reduce the error on the Link Flows. As the Two-Step approach moves to a new solution during the first phase of the optimization, the distance with respect to the initial matrix is larger, while the error on the link flows is two times smaller than the one for the Two-Step. In terms of estimated demand, the demand matrix estimated through the Two-Step approach accounts for 220.360 trips, while the one estimated through the conventional Single-Step approach accounts for 343.000 trips, supporting the claim that the Two-Step approach is capable do perform a broader exploration of the solution space avoiding local optima. The initial demand was 307.544 trips.

Although the Two-Step approach outperforms the Single-Step one, it also increases the error on the speeds, which was expected as this information has a low weight in the goal function. Thus, in the section we introduce a second experiment, which aims at finding a consistent solution for both counts and speeds.

6.3.3.2 Improving the results obtained: Good Starting Matrix

The second experiment presented in this section aims at demonstrating that, even when a "good" a prior demand matrix is available, the Single-Step approach is more likely to over-fit the data with respect to the proposed methodology. Results illustrated in Figure (6.3) show how using the Two-Step approach reduces the localism of the standard single-step DODE, relaxing the dependency on the starting matrix. However, although the model outperformed the Single-Step formulation, the overall estimation is still unsatisfactory. While the model largely reduced the error on the link flows, increasing the r^2 from 0.2686 to 0.7097, the estimated OD matrix significantly underestimates the congestion on the ring. Thus, we performed a second experiment to correct this error. The OD matrix obtained through the Two-Step approach in the previous estimation is now used as initial point for this second experiment, simulating the situation for which a "good" a priori OD matrix is available. The objective function still accounts for both traffic counts and average speed, but this time the latter accounts for 70% of the error, while former are mostly used as a constraint to reduce the search space, avoiding the model to move too far from the current solution.

Results, shown in Figure (6.4), prove that both Two-Step and Single-Step methods estimate a reasonable approximation of the congestion pattern. While congestion between 8 AM and 9 AM is still slightly underrepresented, the average speed on the ring seems more realistic, as the congestion period begins and terminates approximately at the same time for both models.

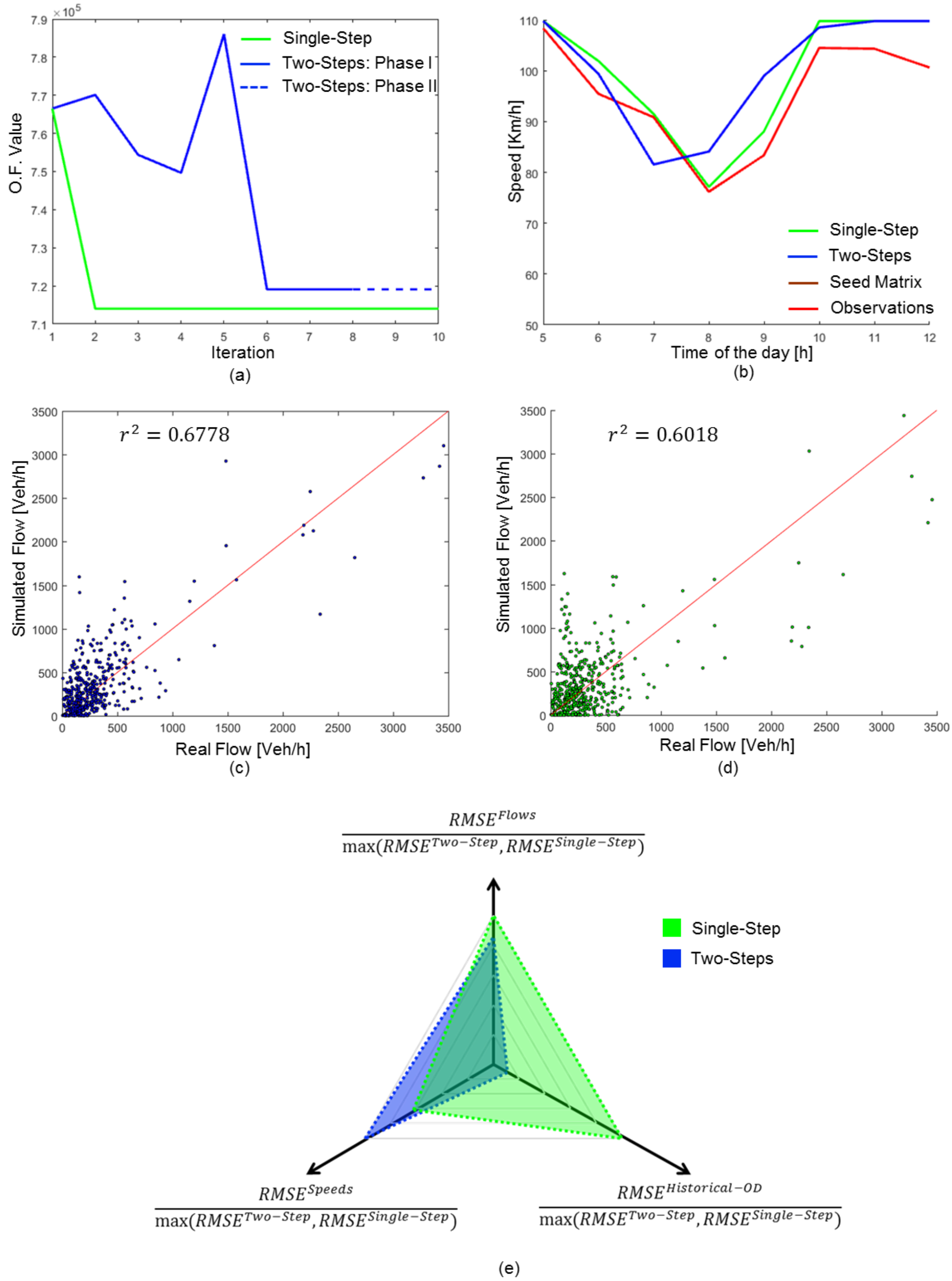


Fig.6.4: (a) Objective Function trend; (b) Estimated and Observed Average Speed; (c) Scatter Estimated and Observed Link Flows for the Two-Step; (d) Scatter Estimated and Observed Link Flows for the Single-Steps; (e) Spider Chart of the relative error for the estimated matrix in terms of Link-Flows, Distance from the Historical OD flows (Seed Matrix) and Average Speed;

However, the Single-Step clearly approximates the average speed on the ring better than the Two-Step approach. By contrast, the error on the link flows clearly shows that the Single-Step is overfitting the speed data, which was expected given the aggregate nature of this information, while strongly increasing the error with respect to the link flows ($RMSE_{Link-Flow}^{Single-Step} = 338.61\ Veh/h$).

Instead, the Two-Step approach manages to provide a realistic fitting for both traffic counts and speed. Although the error on the Link Flows increases with respect to the starting point ($RMSE_{Link-Flow}^{Two-Step} = 291.10\ Veh/h$), the difference is not as big as for the Single-Step approach, as the r^2 shows in Figure (6.4). This brings to a second important consideration. In this second experiment, no improvement is observed in the second step of the Two-Step approach.

Constraint (6.2b) imposes a linear relation between temporal and spatial distribution, meaning that the spatial and temporal structure of the demand is constant during the first step of the optimisation. The direct consequence of that is that the matrix estimated through Experiment II keeps the same structure as the one obtained through Experiment I, while the total demand is different. Although the real OD matrix is not available, as we are dealing with real traffic information, we can easily calculate the error in terms of Euclidean distance with respect the initial matrix, as we would like to keep the distance with respect to the "good" historical OD flows as small as possible. While the Euclidean distance between the estimated matrix and the initial one is only 718 trips for the Two-Step approach, this error increases up to 6449 trips when using the Single-Step approach as optimization framework.

In essence, we may argue that the Two-Step approach kept the structure of the demand from the Seed-Matrix, but sensed and increased the demand in order to move the traffic state on the ring from the uncongested to the congested branch of the fundamental diagram. This suggests that the Two-Step approach is more likely to exploit aggregate data, without altering the structure of the demand in order to overfit the available data.

This is further illustrated in the Spider Chart (Figure (6.4e)). While the Single-Step provides a substantial improvement with respect to the Two-Step in terms of speeds, Figure (6.4e) shows that it clearly alters the structure of the demand, moving to a new local minimum and increasing the error on the link flow. Instead, the Two-Step estimation seems more robust. Although it does not provide an extremely accurate fit of the speed, it keeps the original structure of the demand and provides a reasonable approximation for both speeds and traffic counts, which is in line with the expectations. This is demonstrated by checking the difference between the "good" initial matrix and the estimated one. The final demand matrix estimated through the proposed Two-Step approach amounts at 232.500 trips, against 244.490 estimated with the conventional Single-Step approach. While these numbers seem similar, the root mean square error between estimated and historical seed matrix is ten time higher when using the Single-Step approach (49.57 trips/hour against 5.53 trips/hour). This supports the claim that the model is more likely to provide more robust results

6.4 References

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. doi: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. doi:10.1016/j.trc.2015.04.030.

Ashok, K., and M.E. Ben-Akiva. 2002. "Estimation and Prediction of Time-Dependent Origin-Destination Flows with a Stochastic Mapping to Path Flows and Link Flows." *Transportation Science* 36 (2): 184–98. doi:10.1287/trsc.36.2.184.563.

Ashok, Kalidas. 1996. "Estimation and Prediction of Time-Dependent Origin-Destination Flows." Massachusetts Institute of Technology. https://its.mit.edu/sites/default/files/documents/AshokPhd_0.pdf.

Balakrishna, Ramachandran, Moshe Ben-Akiva, and Haris Koutsopoulos. 2007. *Offline Calibration of Dynamic Traffic Assignment: Simultaneous Demand-and-Supply Estimation*. Vol. 2003. Transportation Research Record. http://trb.metapress.com/openurl.asp?genre=article&id=doi:10.3141/2003-07.

Barceló, Jaume, and Lídia Montero. 2015. "A Robust Framework for the Estimation of Dynamic OD Trip Matrices for Reliable Traffic Management." *Transportation Research Procedia*, 18th Euro Working Group on Transportation, EWGT 2015, 14-16 July 2015, Delft, The Netherlands, 10: 134–44. doi:10.1016/j.trpro.2015.09.063.

Cantelmo, Guido, Francesco Viti, Ernesto Cipriani, and Marialisa Nigro. 2015. "Improving the Reliability of a Two-Step Dynamic Demand Estimation Approach By Sequentially Adjusting Generations and Distributions." Proceedings of theTransportation Research Board conference 2018

Cantelmo, Guido, Francesco Viti, Chris M. J. Tampère, Ernesto Cipriani, and Marialisa Nigro. in press. "A TWO-STEP APPROACH FOR THE CORRECTION OF THE SEED MATRIX IN THE DYNAMIC DEMAND ESTIMATION." *Transportation Research Record*.

Cascetta, Ennio, Domenico Inaudi, and Gerald Marquis. 1993. "Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts." *Transportation Science* 27 (4): 363–73.

Cascetta, Ennio, and Sang Nguyen. 1988. "A Unified Framework for Estimating or Updating Origin/Destination Matrices from Traffic Counts." *Transportation Research Part B: Methodological* 22 (6): 437–55. doi:10.1016/0191-2615(88)90024-0.

Cascetta, Ennio, Andrea Papola, Vittorio Marzano, Fulvio Simonelli, and Iolanda Vitiello. 2013. "Quasi-Dynamic Estimation of O–d Flows from Traffic Counts: Formulation, Statistical Validation and Performance Analysis on Real Data." *Transportation Research Part B: Methodological* 55 (September): 171–87. doi:10.1016/j.trb.2013.06.007.

Cipriani, E., M. Florian, M. Mahut, and M. Nigro. 2011. "A Gradient Approximation Approach for Adjusting Temporal Origin-Destination Matrices." *Transportation Research Part C: Emerging Technologies* 19 (2): 270–82. doi:10.1016/j.trc.2010.05.013.

Djukic, T., J. Van Lint, and S. Hoogendoorn. 2012. *Application of Principal Component Analysis to Predict Dynamic Origin-Destination Matrices*. Transportation Research Record.

Frederix, R., F. Viti, R. Corthout, and C.M.J. Tampère. 2011. "New Gradient Approximation Method for Dynamic Origin-Destination Matrix Estimation on Congested Networks." *Transportation Research Record*, Transportation Research Record, , no. 2263: 19–25.

Frederix, R., F. Viti, and C. M. J. Tampère. 2010. "A Density-Based Dynamic OD Estimation Method That Reproduces within-Day Congestion Dynamics." In *13th International IEEE Conference on Intelligent Transportation Systems*, 694–99. doi:10.1109/ITSC.2010.5625220.

Frederix, Rodric, Francesco Viti, and M.J. Tampère. 2013. "Dynamic Origin–destination Estimation in Congested Networks: Theoretical Findings and Implications in Practice." *Transportmetrica A: Transport Science* 9 (6): 494–513. doi:10.1080/18128602.2011.619587.

GROUP, PTV. 2014. *VISUM 14 User Manual*. epubli.

Marzano, V., A. Papola, and F. Simonelli. 2009. "Limits and Perspectives of Effective O-D Matrix Correction Using Traffic Counts." *Transportation Research Part C: Emerging Technologies* 17 (2): 120–32. doi:10.1016/j.trc.2008.09.001.

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

Mitsakis, Evangelos, Josep-Maria Salanova Grau, Evangelia Chrysohoou, and Georgia Aifadopoulou. 2013. "A Robust Method for Real Time Estimation of Travel Times for Dense Urban Road Networks Using Point-to-Point Detectors." In . http://trid.trb.org/view.aspx?id=1242148.

Nigro, M., E. Cipriani, and A. Del Giudice. 2017. "Exploiting Floating Car Data for Time-Dependent O-D Matrices Estimation." *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* under review.

Small, Kenneth A. 2015. "The Bottleneck Model: An Assessment and Interpretation." *Economics of Transportation*, Special Issue on Collective Contributions in the Honor of Richard Arnott, 4 (1–2): 110–17. doi:10.1016/j.ecotra.2015.01.001.

Spall, James C. 2012. "Stochastic Optimization." In *Handbook of Computational Statistics*, edited by James E. Gentle, Wolfgang Karl Härdle, and Yuichi Mori, 173–201. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-642-21551-3_7.

Sprumont, François, Paola Astegiano, and Francesco Viti. 2017. "On the Consistency between Commuting Satisfaction and Traveling Utility: The Case of the University of Luxembourg." *European Journal of Transport and Infrastructure Research* 17 (2). http://orbilu.uni.lu/handle/10993/30973.

Tavana, Hossein. 2001. "Internally-Consistent Estimation of Dynamic Network Origin-Destination Flows from Intelligent Transportation Systems Data Using Bi-Level Optimization." PhD Thesis. http://repositories.lib.utexas.edu/handle/2152/1662.

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. doi:10.1016/j.trc.2015.04.022.

Tympakianaki, Athina, Haris N. Koutsopoulos, and Erik Jenelius. 2015. "C-SPSA: Cluster-Wise Simultaneous Perturbation Stochastic Approximation Algorithm and Its Application to Dynamic Origin–destination Matrix Estimation." *Transportation Research Part C: Emerging Technologies*, Engineering and Applied Sciences Optimization (OPT-i) - Professor Matthew G. Karlaftis Memorial Issue, 55 (June): 231–45. doi:10.1016/j.trc.2015.01.016.

PART II

Regular Mobility Patterns

7

Combining Utility and Dynamic Traffic Assignment Theories

The second part of this thesis deals with the estimation of the regular component of the demand. As pointed out in the introduction, the model presented in this thesis is based on two components. First, the DTA model is paired with a Utility-Based Departure Time Choice model in order to account for heterogeneous user behaviour when different activity patterns are considered. Then, the parameters of this departure time choice model are updated together with the demand flows within a DODE framework in order to estimate the mobility demand.

This chapter introduces the problem of jointly modelling activity scheduling/duration within a DTA problem framework and the effect of the underlying model assumptions.

Content of this chapter has been presented in the following works:

Cantelmo, Guido, and Francesco Viti. 2016. “Effects of Incorporating Activity Duration and Scheduling Utility on the Equilibrium-Based Dynamic Traffic Assignment*,” Proceeding of the 6th Symposium on Dynamic Traffic Assignment*

Cantelmo, Guido, and Francesco Viti. 2016. “Effects of Incorporating Activity Duration and Scheduling Utility on the Equilibrium-Based Dynamic Traffic Assignment,” *Submitted to Transportation Research Part B: Methodological journal (Second round of review)*

7.1 Introduction

Dynamic Traffic Assignment (DTA) models are important tools in the analysis of congested networks since, differently from their static counterpart, they properly consider network capacity constraints and the propagation of demand flows and traffic congestion over time, providing relevant insights into optimising the network usage. Seminal works (Merchant and Nemhauser 1978; Carey 1986) proposed to exploit analytical approaches to solve this problem and to demonstrate their essential properties (see e.g. Peeta and Ziliaksopoulos, 2001, or Viti and Tampere, 2010 for an overview). Properties of existence and uniqueness of a solution could be derived, hence guaranteeing algorithmic convergence and stability. However, in order to ensure such properties, these formulations usually lack realism and are generally not applicable for analysing big-sized networks. Hence, researchers developed new simulation-based DTA models (e.g. Mahmassani and Herman 1984; Ben-akiva et al. 1998), which focus on having a more realistic on-route behaviour.

Simply stated, DTA models provide an approximation of the network conditions for a given demand according to two main sub-components: the travel choice model and the Dynamic Network Loading model (DNL). The former deals with pre-trip/en-route user decisions, while the latter with the physical propagation of the vehicles on the network. While in early approaches the choice model was mainly limited to the routeing strategy, while considering other decision-making levels as “exogenous”, researchers stressed that both departure time (Mahmassani and Herman 1984), and mode of transport (Fu and Lam 2014) should be explicitly considered within the assignment process. In general, authors agree that these decisions are strongly interconnected, and the DNL model could provide inaccurate predictions if these three elements are not explicitly represented within the choice model (Peeta and Ziliaskopoulos 2001).

Additionally, many researchers stressed the relevance of shifting from a single trip-based approach to a tour/schedule-based representation, able of capturing more complex activity patterns (Adnan 2010; Abdelghany and Mahmassani 2003; Lam and Yin 2001; Lin et al. 2008; Zockaie et al. 2015). Empirical studies (Bowman and Ben-Akiva 2001; Zhang et al. 2005) show that the advantage is twofold. First, trips are intrinsically correlated: according to a trip-based representation, the choice to perform or not a specific trip is independent with respect to the other trips in the analysis period. However, users have the tendency to change their activity pattern as a response to the growing congestion levels and/or policies, meaning that a dependency upon different trips might exist. For example, decisions such as the departure time for the morning and evening commute are intrinsically interconnected. Secondly, “travel decisions are activity based” (Bowman et al. 1999), meaning that each user travels in order to reach his/her destination and perform a certain activity. For this reason, the correlation between different trips depends on the duration of such activities, as well as their spatial and temporal distributions (e.g. opening hours of shops), meaning that trip-based DTA models, which rely only on the disutility of travelling, are likely to provide inaccurate predictions of the traffic state. In contrast, by taking advantage of a more reliable choice model, DTA models increase their responsiveness when dealing with, for instance, large-scale events, activity relocation problems and, in general, for short/long term planning applications.

One of the most classical and widely adopted procedures to account for a more complex choice model is adopting Utility Maximisation Theory (UMT), which assumes that users (re-)schedule their activities in order to maximize their perceived utility (Yamamoto et al. 2000; Ettema and Timmermans 2003). Although many works exploit UMT within the DTA framework (Balmer et al. 2008; Fu, Lam, and Xiong 2016), not enough attention has been given to understanding how the different assumptions on UMT will affect the simulation outputs. Since departure time, activity duration and location are intrinsically correlated, we can expect that the assumptions we make on each of these elements have direct implications on the predicted traffic states.

In this chapter, the authors show that, by relaxing the assumptions of the UMT model, the Activity-Based DTA framework can turn into a trip-based, tour-based or schedule-based approach. While many authors already assess the results of the UMT-based choice model under specific conditions, this chapter generalizes these findings, showing the conditions in which UMT is able of capturing this dependency. To reach this goal, the authors contribute to the state of the art by:

(i) Formulating a set of properties, which allow predicting the effect of a certain activity function on the departure time choice module;
(ii) In order to evaluate the effect of the utility functions on the departure time choice model, we extend the bottleneck model proposed in (Li, Lam, and Wong 2014), initially formulated only for commuting trips, in order to account for all type of activities;
(iii) We introduce a new utility function which properly captures activity duration and can be adapted to model different activities, including special events;
(iv) We introduce a new metric to calculate the Degree of Correlation ¬– the DoC – between different activities.

We test the proposed set of rules with an analytical formulation for simple – i.e. constant – values of the utility, while we use a simulation-based approach to show that the proposed general rules hold for non-linear utility functions.

It is important to stress that the main goal of this work is not just to compare different utility functions in order to test their performances. The more general purpose of the proposed study is instead to assess the implications of using a certain UMT framework. Different assumptions within the choice model lead to a different congestion pattern. While simple functions might be sufficient when dealing with simple dynamics, it is still useful to investigate the error behind this assumption.

7.2 Literature review

7.2.1 Dynamic Traffic Assignment

When it comes to choose in the broad range of existing models, DTA approaches can be categorised as analytical or simulation-based models. The analytical models (Merchant and Nemhauser 1978; Ziliaskopoulos 2000) provide strong theoretically sound mathematical structures in terms of analytical tractability, focusing on finding conditions for solution uniqueness and convergence. While formulating the DTA as a rigorous mathematical problem, these approaches usually yield unrealistic congestion patterns when applied to real sized networks. Some authors (Mahmassani and Herman 1984; Hendrickson and Kocur 1981) pointed out how the simultaneous choice of route and departure time is responsible for the length of the congestion period. Taking inspiration from transport economic models (Vickrey 1969; Small 1982), many authors developed (analytical) DTA frameworks able to properly capture this phenomenon (Arnott, de Palma, and Lindsey 1990; Noland and Small 1995; Ben-Akiva, de Palma, and Kanaroglou 1986; Ben-Akiva, Cyna, and de Palma 1984; de Palma et al. 1983; De Palma and Arnott 1986; Lindsey and Verhoef 2000; Li, Lam, and Wong 2014a; Jiang et al. 2016). However, while these models provide useful insights to evaluate transport policies (e.g. tolling) on simple networks, they still lack of realism when dealing with real networks under the user equilibrium assumption. To overcome the abovementioned shortcomings, new simulation-based algorithms have been developed over the last decades in order to reproduce realistic traffic patterns (Ben-akiva et al. 1998, 1998; S. Peeta and Mahmassani 1995; Gentile 2015; Lu et al. 2015). While these models efficiently reproduce congestion phenomena, most of these approaches assume that departure time is exogenous and fixed, hence including only the route choice dimension in the choice model. If departure time is considered exogenous, we might assume that the choice model is mostly based on the route cost, which, from a theoretical point of view, is an incorrect assumption (Zhang et al. 2005). Two alternatives have been explored to overcome this critical issue: 1) simultaneously calibrating the DTA parameters and the demand model, hence changing the departure time and the route choice parameters simultaneously (Balakrishna 2006), or 2) explicitly considering the departure time choice within the DTA (Mahmassani and Herman 1984; Abdelghany and Mahmassani 2003).

A second key distinction in the DTA approaches is related to the choice model type. In the last decades, many authors stressed how the complex congestion dynamics are derived from complex activity patterns (Adnan 2010): since users change activity schedule as a consequence of congestion, if equilibrium does not include this interdependence the DTA might lead to biased results. Lam and Yin (2001) developed an activity-based time-dependent traffic assignment model, where time-dependent utility at the destination is used in a discrete choice formulation to understand the sequence of activities. Lam and Huang (2002) extended this model to consider the elasticity of the demand with respect to changing the activity pattern by having a new *non-work* location. While this framework properly captures activity sequences, it does not take into account activity duration, which has been considered of paramount relevance to model activity interdependence (Zhang et al. 2005). Abdelghany and Mahmassani (2003) proposed a stochastic microsimulator including activity scheduling. The input demand is the travel plan at the user level, including origin, preferred arrival time, activity duration, intermediate and final destination. Polak and Heydecker (2006) proposed a traffic assignment formulation where users choose a tour constrained to a base point, home, where the tour starts and ends. At the equilibrium, the utility of each user with the same activity pattern has to be identical. Ramadurai and Ukkusuri (2010) proposed a model to estimate simultaneously activity scheduling, route choice, activity duration and departure time using a *supernetwork* representation of the problem subject to a dynamic user equilibrium condition. Adnan (2010) presented a macroscopic framework, showing that UMT is capable of capturing activity correlation for the daily commuting. Li et al. (2014) modelled the utility at the user level, using macroscopic equilibrium conditions to constrain the single user decision level. Liu et al. (2015) have recently developed a new model where the link representing the activity is modelled through a cost function. User (dis-)utility is then calculated through a waiting time term, which measures the gap between preferred and actual activity starting time, and a duration time term, measuring the gap between preferred and actual activity duration. This concept has been also considered in a more general framework, called ATS-SAM network (Fu and Lam 2014), which takes into account two streams of research in one single framework. The supernetwork platform integrates activity-time-space (ATS) and state-augmented-multimodal (SAM) networks. The first one is an expanded network, where activity links are considered, as in Ramadurai and Ukkusuri (2010), to represent activity and travel choices together. The second one takes into account the multi-modal aspect in a realistic way. While all the aforementioned studies are based on the UMT, it should be pointed out that other models have been developed to generate realistic dynamic activity patterns through different approaches originating from activity-based modelling theory. One of the most famous examples is Albatross (Arentze and Timmermans 2004), which is based on decision trees and time constraints. Similarly, Pendyala et al. (2012) propose an integrated model that takes into account location choices, activity-travel choices and individual vehicles on networks. Differently from other works, where these choices are modelled through a sequential approach, the proposed approach integrates the activity-travel module, dynamic traffic assignment module, and land use model within a unique framework. Specifically, it integrates minute-by-minute the ABM – which provide the list of trips to be routed during the current time interval - and the DTA model – which provide the list of users who finished their trip during that specific time interval. However, the focus in these cases is more on generating more comprehensive activity patterns for planning purposes rather than to analyse the network congestion

While we considered only the distinction between analytical and simulation-based DTA together with the choice model type, it should be pointed out that DTA models can be further classified according to the type of assignment (*user equilibrium* vs. *system optimum*), aggregation level (*macroscopic*, *mesoscopic* and *microscopic* models), and user equilibrium principle (*stochastic* vs. *deterministic*). While for a detailed overview we refer to other works (Peeta and Ziliaskopoulos 2001; Viti and Tampere 2010; Barceló 2010), we introduce the following categorization of DTA models, which will be used throughout this thesis:

- *Trip-Based Dynamic Traffic Assignment* (TB-DTA): if the decision of travelling for each trip is independent from all other trips and is based only on the travel costs related to that trip;
- *Activity-Based Dynamic Traffic Assignment* (AB-DTA): if the decision of travelling for each trip is independent from all other trips and depends on the benefit of performing an activity at the destination and the travel costs;
- *Trip-Chain based Dynamic Traffic Assignment* (TC-DTA): if decisions of performing certain trips during the analysis period are interdependent;
- *Schedule-Based Dynamic Traffic Assignment* (SB-DTA): if decisions of travelling for different trips during the analysis period are interdependent and time constraints are explicitly considered.

7.2.2 Utility Maximization Theory

Many studies show how activity scheduling is strongly related to the satisfaction of performing a specific activity at a certain time. While for an extensive literature review we refer to other authors (Dick Ettema and Timmermans 2003; D. F. Ettema et al. 2007; Small 2015), we focus here on a specific key issue: the final daily activity pattern is a function of travel time, activity duration and the preferred arrival time at the destination (Zhang et al. 2005). Hence, there are (at least) three main motivations which drive our activity scheduling approach:

1) The departure/arrival time influences the time each user allocates to a specific activity;
2) If a certain time constraint exists for a certain activity (i.e. fixed schedule), the user could skip a specific activity if he/she has scheduling constraints;
3) The level of congestion of the system influences the departure time. This is an additional constraint for the user, who might not be able to perform an activity when congestion levels rise.

Different authors (D. Ettema and Timmermans 2003; D. F. Ettema et al. 2007, among others), assume that the utility varies with the time of the day and that it can be modelled through a continuous function, meaning that the marginal utility (MU) would change with the time of the day. Hereby, and in the following of this chapter, we will refer to this time-of-day-dependent marginal utility as *clock-based* MU. The advantage of *clock-based* MU is to properly represent the connection between time of the day and utility. A well-known limitation of these functions is however that they assume no correlation existing between marginal utility value and activity duration, meaning that they do not consider the satisfaction (or, inversely, fatigue) effect of performing the same activity for several hours.

A second utility function can be defined as *duration-based* MU. In this case, the utility is assumed to be proportional to the duration of the activity (Yamamoto et al. 2000). *Duration-based* MU models consider the fatigue effect, but they are not able to fully represent the evolution of the utility over time. For this reason, researchers agree that both these ingredients are relevant for properly modelling the choice process in a dynamic setting and hence they should be properly combined (D. F. Ettema et al. 2007; Adnan 2010).

According to the classical formulation proposed by Yamamoto et.al (2000), each user chooses a departure time that maximizes his/her utility U:

$$U^n = (U^t + U^a); \tag{7.1}$$

where U^n is the utility for a certain user *n*, U^t represents the disutility of travelling, while U^a the utility of performing one or more activities. This quite general formulation has been broadly used in the literature (Dick Ettema and Timmermans 2003; Zhang et al. 2005; Kim, Oh, and Jayakrishnan 2006; Polak and Heydecker 2006; D. F. Ettema et al. 2007; Adnan 2010; Li, Lam, and Wong 2014a; Fu and Lam 2014). The first component of Equation (7.1) has been analysed in both traffic engineering and transport economics research. Specifically, in order to calculate U^t, the classical schedule delay-based formulations (Vickrey 1969; Small 1982) are useful tools in dealing with activity scheduling and disutility of travelling. While on one hand they consider the disutility of travelling, on the other they allow a flexible behaviour, since users can change their departure time, hence losing some utility. Furthermore, when MU is considered a continuous function, the schedule based formulation provides time constraints which lead to more realistic conditions (Ettema et al. 2007).

Focusing on the second term, U^a, Ettema and Timmermans (2003) proposed a *clock-dependent* utility function. In this model, each user has a different utility by performing a specific activity at a different time of the day. However, an additional parameter, which considers the flexibility of the user to change the starting time of the activity, is introduced. This parameter considers if the activity is clock dependent or time-dependent. If the value is equal to 0, the *maximum* utility of the activity is clock dependent, while if equal to 1 it is considered duration dependent, i.e. independently from the activity starting time, the user can reach the maximum utility.

One of the main problems of this formulation and, in general, of the *clock-based* MU function, is that it does not consider the component of satisfaction related to performing an activity for a certain period of time. This leads to the fact that the utility of performing an additional hour of activity is assumed to be not related to how many hours the user already performed that activity. Furthermore, the satisfaction related to one more hour of activity decreases when the duration increases due to a fatigue effect. This issue has been considered for the first time by Becker (1965) as a time allocation problem and formulated as an optimization problem where the user tries to allocate the resource time in the best possible way. Other authors (Yamamoto et al. 2000; Bhat and Misra 1999) assume that, in order to consider the fatigue effect, the utility decreases according to a logarithmic scale. A model including all these elements is proposed by Ettema et al (Ettema et al. 2007) as an extension of their previous work (Ettema, Ashiru, and Polak 2004). To consider heterogeneity in the users' classes, authors take into account two elements. An "observed" heterogeneity is modelled through socio-demographic data, in terms of wages and habits. The authors consider also a "non-observed" heterogeneity, related to users, which are in the same social class but exhibit different preferences. This element is considered by directly including an error component in the utility function.

7.2.3 Remarks and conclusions

The literature review presented so far provides useful insights into identifying advantages and opportunities in pairing UMT and DTA models, in terms of route choice, mode choice, departure time choice modelling and activity scheduling.

With respect to the DTA model, two main considerations should be pointed out. Firstly, if a UMT-based choice model is considered, both flow based (Adnan 2010) and microscopic traffic simulators (Balmer et al. 2008) can account for activity patterns. Secondly, since our goal is of evaluating the effect of a specific DTA sub-component - the (departure time) choice model - properties of convergence and uniqueness of the adopted DTA model are of paramount relevance. Thus an analytical approach is more suited than a simulation-based one for the current study.

Concerning the critical issue of the utility functions effect on the DTA, previous works highlighted the following shortcomings:

1) Adnan (2010) demonstrates mathematically and numerically that, if only *clock-based* MU functions are considered, the DTA models become AB-DTA, according to the above-mentioned classification. As a consequence, the author concludes that *duration-based* utility functions should be considered for modelling the work activity;
2) Li, Lam, and Wong (2014) demonstrate that, for the home-work commuting, if flexibility is considered in the *clock-based* MU, departure ratios for the morning and evening commute are correlated. In this case, the models can be considered as a fully-fledged SB-DTA;

3) Feil, Balmer, and Axhausen (2009) show that, if the standard *duration-based* MU formulation proposed by Yamamoto et al. (2000) is implemented for modelling longer activities (i.e. work), then the model is very likely to provide biased estimations.

By following these shortcomings, we can easily observe that there is a contradiction in the literature on when a model can be categorized as a *Scheduled-based* or an *Activity-Based DTA*. The reason is that, in these works, the authors analyse a very specific framework, thus their findings cannot be directly applied to other frameworks. Thus, the challenge is to propose a new set of conditions that hold for all Utility-Based DTA frameworks and, even more important, when a different activity type or duration is included in the system.

7.3 Methodology

In this section, we introduce our methodology for evaluating the effect of the UMT-based choice model on the DTA. The first section introduces a general DUE formulation, which takes into account departure time, route choice, mode of transport and activity location. Then, in Section 7.3.2, we introduce the utility functions for representing activity participation and travel cost. Table 7.1 reports the most relevant notations.

7.3.1 Utility based Dynamic User Equilibrium

Following traditional UMT (Yamamoto et al. 2000) already introduced in Equation (7.1), we assume that each user maximises his/her own utility. The decision variables are 1) the departure time, 2) the mode of transport and 3) the route.

Thus, knowing the set of departure times, modes of transport and routes, for each user, the total utility can be calculated. We assume that, at equilibrium, the overall utility is maximal, and can be calculated through the following set of equations:

$$U^n(\boldsymbol{t},\boldsymbol{m},\boldsymbol{r}) = \max_{\boldsymbol{t},\boldsymbol{m},\boldsymbol{r}}(U_n^T(\boldsymbol{t},\boldsymbol{m},\boldsymbol{r}) + U_n^A(\boldsymbol{t})) \quad \forall\, n \in N \tag{7.2}$$

Where

$$U_n^A(\boldsymbol{t}) = \sum_{p=1}^{P} U_{n,p}^a(\boldsymbol{t}(p,n)); \tag{7.2a}$$

$$U_n^T(\boldsymbol{t},\boldsymbol{m}.\boldsymbol{r}) = \sum_{s=1}^{S} U_{n,s}^t\big(\boldsymbol{t}(p,n),\boldsymbol{m}(p,n),\boldsymbol{p}(p,n)\big); \tag{7.2b}$$

With P and S the number of activities and number of trips, respectively, N the number of users, $U_n^T(\boldsymbol{t},\boldsymbol{r},\boldsymbol{m})$ the overall dis-utility of traveling and $U_n^A(\boldsymbol{t})$ the overall utility of performing activities for user n. Lastly, $\boldsymbol{t}$ is the vector of the *selected* departure times, $\boldsymbol{m}$ the vector of the *selected* transport modes and $\boldsymbol{r}$ the vector of the *selected* routes.

Equation (7.2) might lead or not to a pure Nash equilibrium, based on the assumptions we make on U^t, U^a and, in general, on the mobility demand in our system. For a general transport system, we have different users with different goals. Let us consider, for the sake of illustration, two classes of users: workers and shoppers.

Table 7.1: Notations

Notations
U = *Total net utility*
U^t = *Disutility of travelling for a specific trip in the system*
U^a = *Utility related to performing a certain activity*
U^T = *Overall disutility of travelling*
U^A = *Overall utility of performing activities*
n = Subscript for the user
N = Number of users
p = Subscript for the number of activities/purposes
P = Number of activities/purposes
s = Subscript for the trip
S = Number of trips
m = Subscript for the mode of transport
M = Set of feasible modes of transports
r = Subscript for the route
R = Set of feasible routes
t = Analysis time interval
t^0 = *Preferred arrival time*
t^{d-0} = *Departure time for which user n arrives on time at the destination*
t_p^s = *Actual starting time for a specific activity p*
t_p^e = *Actual ending time for a specific activity p*
EA =Scheduling delay - Early Arrival
LA =Scheduling delay – Late Arrival
VoT = *Value of Time of unit duration for the travel time*
VoE = *Value of Time of unit duration for arriving early*
VoL = *Value of Time of unit duration for arriving late*
U^{MAX} *is the maximum accumulated utility for an activity p*
$\beta_p, \alpha_p, \gamma_p$ = *parameters for the clock-based MU for an activity p*
τ_p = *Parameter considering the flexibility of an activity p*
η_p = *Parameter for the duration-based MU for an activity p*
T = Travel Time
T^f = *Free flow travel time*
T^b = *Time spent at the bottleneck*
D(t) = length of the queue during time interval t
S = Capacity of the bottleneck
ε *time interval in which congestion is observed on the network*
G= parameters accounting for the fatigue effect

The maximum net utility for these two different classes will be different since, even when they join the same activity, their gain in terms of utility will be different. However, these users will share the same common resource, which is the capacity of the transport network. We define here an *ideal-activity-pattern* as the activity pattern each user chooses regardless of the travel time spent to reach the activities. In other words, the *ideal-activity-pattern* corresponds to maximise (7.2) when $U_n^T = 0$, hence each user can allocate his/her time to different activities without considering any constraint related to transport costs. The latter costs will be instead controlled by the disutility of travelling as explained later.

We assume that users with the same *ideal-activity-pattern* belong to the same class. Hence, for each class of users, we assume that U^t and U^a are evaluated through the same functions, while different functions might be considered for different classes. However, equilibrium may not exist for users of the same class. In fact, utility U^a captures the utility of performing an activity, but it does not consider other elements, among others, the travel time to reach the zone where the activity takes place, i.e. two commuters might live/work in different zones of the city, hence the disutility of travelling U^t will be different even in free flow conditions. For TB-DTA, the most straightforward way to consider this issue is to assume that equilibrium exists for each OD pair. However, if we consider multiple destinations and complex tours this solution becomes unfeasible. One possible solution could be to assume that equilibrium exists for users having the same tour – thus same *ideal-activity-patterns* and *activity location*. However, we need in this case to explicitly map all the possible alternative tours for each user. Instead, we exploit the concept of "*base-point*" introduced in (Polak and Heydecker 2006). In this case, rather than having one single base-point (home), we consider that multiple base-points may exist. For commuters, this is the OD pair related to the home-work commuting. Given the *base-points*, we assume that these users might change the other locations in order to increase their utility. In order to consider the above-mentioned elements, we can now formulate the following general definition:

Definition: For a given class of users with the same *base-points* and *ideal-activity-pattern,* at equilibrium, each combination of $\boldsymbol{t}, \boldsymbol{m}$ and $\boldsymbol{r}$ leads to the same net utility.

Following and generalizing (Li, Lam, and Wong 2014a), equilibrium can be now formulated as a complementarity problem:

$$\begin{cases} N(\boldsymbol{t},\boldsymbol{r},\boldsymbol{m})[U^n(\boldsymbol{t},\boldsymbol{r},\boldsymbol{m}) - U^*] = 0 \\ N(\boldsymbol{t},\boldsymbol{r},\boldsymbol{m}) \geq 0, \qquad U^n(\boldsymbol{t},\boldsymbol{r},\boldsymbol{m}) - U^* \leq 0 \quad \forall\, \boldsymbol{t},\boldsymbol{r},\boldsymbol{m} \end{cases} \tag{7.3}$$

Where N$(\boldsymbol{t},\boldsymbol{r},\boldsymbol{m})$ is the number of user for a certain class, $\boldsymbol{t},\boldsymbol{r}$ and $\boldsymbol{m}$ is their set of chosen departure times, routes and transport modes, $U^n(\boldsymbol{t},\boldsymbol{r},\boldsymbol{m})$ is the utility related to this set of choices, and U^* is the maximum net utility.

7.3.2 Modelling the utility for performing an activity

In this section, we introduce the positive component of the utility for Equation (7.2). In order to evaluate the *clock-based* utility $U^a_{n,p}$ related to performing an activity *p* for a generic user *n*, the following function is considered (Ettema and Timmermans (2003)):

$$U^a_{n,p}(t) = \frac{\gamma_p \beta_p U^{MAX}}{\exp[\beta_p(t-(\alpha_p + t^s_p \tau_p))](1+\exp[-\beta_p(t-(\alpha_p + t^s_p \tau_p))])^{\gamma_p+1}} \tag{7.4}$$

Where:

- U^{MAX} is the (calibrated) maximum accumulated utility for activity *p;*
- $\beta_p, \alpha_p, \gamma_p$ and τ_p are parameters to be calibrated;
- t^s_p is the starting time for activity *p;*

The term U^{MAX} in Equation (7.4) scales the utility distribution. Specifically, the parameters β_p, α_p and γ_p are responsible of the (bell) shape of the function, while τ_p allows to identify the flexibility with respect to the arrival time. If τ_p is equal to 1, the maximum value of the utility $U^a_{n,p}(t)$ can be obtained for any value of t^s_p. If τ_p= 0, the formulation becomes a standard *clock-based* MU formulation.

To capture the fatigue effect, we use the *duration-based* formulation originally proposed by Yamamoto et al. (2000):

$$U_{n,p}^{a}(t, t_p^s) = \eta_p \ln(t - t_p^s) \tag{7.5}$$

Where η_p is a scale parameter to be calibrated. Equations (7.4) and (7.5) will be used to evaluate the effect of using a pure *clock-based* or *duration-based* MU function on the DTA. The main reason is that Equations (7.4) and (7.5) (or similar versions of the same utility functions) have been used in many utility-based frameworks (Zhang et al. 2005; Adnan 2010; Fu and Lam 2014; Li, Lam, and Wong 2014a; Feil, Balmer, and Axhausen 2009).

Additionally, Ettema et al. (2007) proposed a utility function which is *clock-based* and, in the same time, *duration-based*:

$$U_{n,p}^{a}(t, t_p^e, t_p^s) = \frac{wU^{MAX}}{\pi}\left(\left(\arctan\left(\frac{t_p^e - \alpha_p}{\beta_p}\right)\right) - \left(\arctan\left(\frac{t_p^s - \alpha_p}{\beta_p}\right)\right)\right) - \eta_p(1 - w)\ln(t - t_p^s) \tag{7.6}$$

Where t_p^e is the ending time for activity *p*. Equation (7.6) presents a minor difference with respect to the original version proposed in Ettema et al. (2007). The authors did not consider the weight w in the original formulation, since they include this term in η and U^{MAX}. Here we prefer to explicitly include this weight for having an intuitive comparison with the other methods by explicitly representing the trade-off between *clock-based* and *duration-based* components. The main problem of Equation (7.6) is that the *duration-based* and *clock-based* components are considered as additive functions. As we show in the next sections, this may lead to an overestimation of the utility for those time intervals in which the clock-based component is negligible. Thus, we suggest here a new equation to jointly model the time dependent and duration dependent utility.

Let us consider Equation (7.5). By computing the partial derivative over time, we get the marginal (time-dependent) utility for each value of *t*:

$$U'^{a}_{n,p}(t, t_p^s) = \frac{\partial}{\partial t}\left(U_{n,p}^{a}(t, t_p^s)\right) = \eta_p \frac{1}{(t - t_p^s)} \tag{7.7}$$

Specifically, after the first hour of work, when $t = t_p^s + 1$, the utility will be η_p, then it will start decreasing according to a logarithmic scale for any additional hour of work. Equation (7.7) assumes that the utility of the modelled activity, for instance shopping, is constant during the day, i.e. $U_{n,p}^{a}(t) = \eta_p \; \forall \; t$, but decreases with time according to a logarithmic trend. After one hour of shopping, the user might decide to perform a different activity. If the same concept is applied to a time dependent utility function, in which the utility in certain time interval *t* is calculated through a function $U_{n,p}^{a}(t)$, we get:

$$U'^{a}_{n,p}(t, t_p^s) = U_{n,p}^{a}(t)\frac{1}{(t - t_p^s)} \tag{7.8}$$

According to this formulation, we might simply consider that $U_{n,p}^{a}(t)$ is a general *clock-dependent* utility function – such as the one presented in Equation (7.4), while $U'^{a}_{n,p}(t, t_p^s)$ is the same function, when the fatigue effect is considered. If the utility is constant and equal to η_p, we obtain Equation (7.7). While Equation (7.8) sounds correct from a mathematical point of view, behaviourally speaking, it is biased. As pointed out in (D. F. Ettema et al. 2007), the effect of fatigue should be weighted according to the specific activity. While, in Equations (7.5) and (7.7), η_p includes this weight, it has to be explicitly modelled in Equation (7.8). If not, activities with a longer duration will be penalized (Feil, Balmer, and Axhausen 2009).

The overall utility for an activity *p*, for a given value of t_p^s and t_p^e can be estimated as:

$$U_{n,p}^a(t_p^s, t_p^e) = \int_{t_p^s+1}^{t_p^e} U_{n,p}^a(t) \left(\frac{1}{(t - t_p^s)}\right)^G dt \qquad (7.9)$$

Where G is a parameter to be calibrated. Specifically, if *G* is equal to 0, then Equation (7.9) is the integral of the original *clock-based* formulation if G=1 the fatigue effect will be dominant, as shown in Figure 7.1.

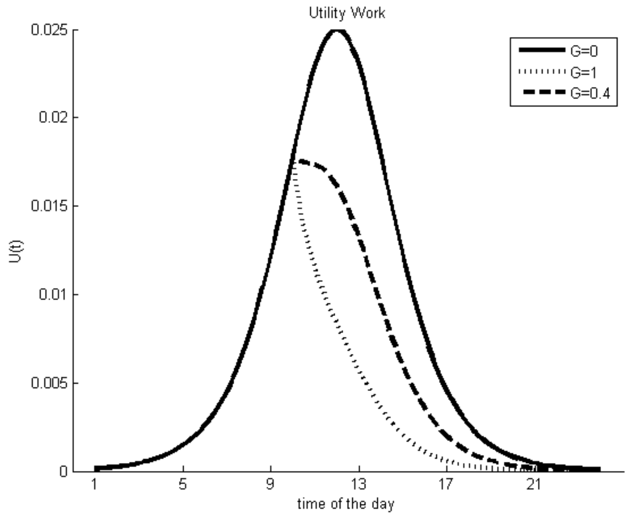


Fig.7.1: Marginal utility obtained using equation 7.8 (t_p^s= 9);

The main advantage of Equation (7.9) is to be able to consider the effect of the satisfaction within a *clock-based* approach. While Adnan (2010) points out that a *duration-based* function should be used when dealing with such activities like work, Feil et. al. (2009) show that the *duration-based* formulated in Equation (7.5) can hardly model activities with a relevant duration. Hence, equation (7.9) is a good compromise for modelling such activities. Note that, for G=1 and $U_{n,p}^a(t) = \eta_p \; \forall \; t$, Equation (7.9) and Equation (7.5) are the same.

It should also be pointed out that Equations (7.4-7.9) are here assumed error-free. This is possible when they are used to model a population of synthetic identical users travelling on a certain network. However, these functions are usually calibrated through survey data, where clearly different users have different preferences. To account for this problem, in the calibration phase, the modeller should consider a perception error for each variable, which accounts for the idiosyncrasy of the population. Clearly, different assumptions of the distribution of the error term will lead to a different calibration result. While we do not dive further in this direction, should be pointed out that, in the literature, Equations (7.4-7.9) are usually calibrated using Logit-type models, and assuming a Weibull/Gamble distribution of residuals (Bowman and Ben-Akiva 2001). An example of the different outputs we can obtain when calibrating Equation (7.4-7.9) is proposed in Appendix C.1.

7.3.3 Modelling the disutility of travelling: the Bottleneck Model

In order to evaluate the disutility of travelling, the cost function U^t in Equation (7.2) for the user *n* during a generic trip *s* with purpose *p* can be defined as:

$$U_{n,p,s}^{t} = VoT_{s,p} \cdot (T(t_n, m_n, r_n)) + VoE_{s,p} \cdot (EA(t_n, m_n, r_n)) + VoL_{s,p} \cdot (LA(t_n, m_n, r_n)) \quad (7.10)$$

Where *T* is the travel time, EA and LA are scheduling costs for the early and late arrival, respectively. $VoT_{s,p}$, $VoE_{s,p}$ and $VoL_{s,p}$ are parameters to calibrate, which are purpose/trip dependent. These parameters take into account the value of time related to travel and arrive late/early. For simplicity of notation we refer to the parameters $\{VoT_{s,p},\ VoE_{s,p}, VoL_{s,p}\}$ as $\{VoT,\ VoE, VoL\}$ for the rest of this chapter. This function is the same proposed by Vickrey (1969) and Small (1982), and presents two main advantages: it deals with the problem of departure time rescheduling and it tackles problems related to having continuous *clock-based* MU functions, which do not capture the rigid scheduling of specific activities.

To calculate the disutility function presented in Equation (7.10), under the assumption of user equilibrium, we apply a recent extension of the classical bottleneck model introduced by Li et al. (2014). While using a bottleneck model might look simplistic, the following considerations endorse this decision:

1) Zhang et al. (2005) show that the bottleneck model is able to capture the correlation between morning and evening congestion;
2) Li et al. (2014) analyse the analytical properties of the bottleneck model, when the utility of performing an activity is considered in the equilibrium problem. The properties of convergence and uniqueness make this model an ideal candidate for a preliminary analysis;
3) By definition, the bottleneck model is able to consider the interaction between users related to the capacity constraint;
4) Utility at the destination can be considered as an additional component in the trip costs. The bottleneck model is the most suited model for evaluating the sensitivity with respect to different utility functions;
5) We might imagine that a simulation-based DTA will incur *at least* in the same shortcomings of a bottleneck model, hence the bottleneck model can provide useful insights into the conditions under which a utility-based equilibrium model provides realistic outputs.

Lastly, the bottleneck model matches the requirements of the DUE formulated in Equation (7.2). The bottleneck model represents the connections between the *base-points* through a single link-route. Figure 7.2 shows the model for the home-work commuting.

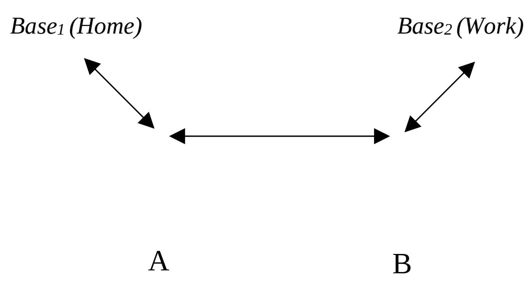


Fig.7.2: Bottleneck model for the Home-Work commuting;

By focusing on only car users, we obtain that the only decision variable is the vector of departure times ***t***. While for a detailed overview of this model, authors refer to Li et al. (2014), here we present the fundamental equations, based on the seminal work of Arnott, de Palma, and Lindsey (1990). We define *S* as the capacity of the link AB, which is expressed in vehicles per hour (veh/h). Assume to have a single class of *N* users, with *N*>*S*, travelling between the two *base-points*. Since the demand exceeds the capacity, a queue will occur on the link A-B, which is the bottleneck of our *system*. The travel time from A to B can be calculated as:

$$T(t) = T^f + T^b(t) \tag{7.11}$$

where *t* is the departing time, T^f is the free flow travel time and T^b is the time spent at the bottleneck, given a certain departing time *t*. The free flow travel time is here considered equal to 0 for illustration's sake, meaning that the travel time between A and B is equal to the time spent at the bottleneck. The time spent at the bottleneck can be calculated given the length of the queue, according to Little's law (Little, 1961), i.e. Equation (7.12):

$$T^b(t) = \frac{D(t)}{SB} \tag{7.12}$$

Where *D(t)* is the length of the queue and *SB* is the capacity at the bottleneck. The maximum length of the queue is calculated as the integral of all vehicles queuing after a certain time interval t^*, which represents the last time interval in which no queue was observed at the bottleneck. Defined as *r(t)* the departing rate for a certain time interval *t*, we can obtain the length of the queue as follows:

$$D(t) = \int_{t^*}^{t} r(t)dt - SB(t - t^*) \tag{7.13}$$

The derivative with respect to the departure time provides the number of vehicles queuing in the time interval *t*. $sb(t - t^*)$ represents the capacity during the time interval $(t - t^*)$.

$$\frac{\partial\, D(t)}{\partial t} = r(t) - sb \quad for\, D(t) > 0 \tag{7.14}$$

For each class of users, we assume the same preferred arrival time t^o. We can define the t^{d-o} the departure time for which the user arrives at work on time:

$$t^{d-0} = t^0 - T^b(t^{d-0}) \tag{7.15}$$

Therefore, we can now quantify the early and late departure times as:

$$\begin{cases} EA = t^0 - T^b(t) - t & for\ t < t^{d-0} \\ LA = t + T^b(t) - t^0 & for\ t > t^{d-0} \end{cases} \tag{7.16}$$

The cost function of one specific trip *s* is a linear combination of the following three elements, as showed in Equation (7.10), and can be now formulated as:

$$U_{n,s}^{t} = VoT \cdot (T) + VoE \cdot (EA) + VoL \cdot (LA) = \\ = VoT \cdot \left(T^{b}(t)\right) + VoE \cdot max\left(0; t^{0} - t - T^{b}(t)\right) + VoL \cdot \left(t + T^{b}(t) - t^{0}\right) \quad (7.17)$$

The parameters VoT, VoE and VoL usually change and should be calibrated according to the activity located in the origin or destination of the trip *s*.

While the equilibrium point can be analytically derived for the standard bottleneck model (Arnott, de Palma, and Lindsey 1990), Li et al. (2014) show that, when utility is explicitly considered, an analytical solution is available only for constant values of the utility. Thus, in order to evaluate complex functions, they suggest to find the equilibrium through the well-known method of successive averages (MSA) algorithm. We also apply here the same numerical scheme:

Step 0: Define an initial departure flow value*:* $N^{k}(\boldsymbol{t})$ by setting $t = t^{0}\ \forall\ n \in N$ and the iteration number *k*=1;

Step 1: Load the demand on the network, obtaining the travel time *T*;

Step 2: Calculate the Utility $U^{n} = (U^{t}(t) + U^{a}(t))\ \forall\ n$;

Step 3: Perform an *all-or-nothing assignment* in order to obtain the auxiliary flows $N_{k}^{*}(\boldsymbol{t})$;

Step 4: Update the solution through the MSA algorithm as $N^{k+1}(\boldsymbol{t}) = N^{k}(\boldsymbol{t}) + \frac{1}{k+1}\left(N_{k}^{*}(\boldsymbol{t}) - N^{k}(\boldsymbol{t})\right)$

Step 5: Check the convergence. If $\max(abs(q^{k+1}(\boldsymbol{t}) - q^{k}(\boldsymbol{t})) > \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ is the vector containing precision tolerance, then *k=k+1*, and go to Step 1.

This numerical procedure will be used for evaluating the different utility functions in the numerical experiments. As pointed out in section (7.2.3), an analytical model is more suited for this study than a simulation-based one. Since properties of convergence and uniqueness of the model have been already discussed in the original paper (Li, Lam, and Wong 2014b), this framework seems ideal for analysing the effect of using different utility functions to model the activity purpose. However, the model proposed in the original paper focuses on the commuting trip only, while in this thesis we are interested in modelling all types of activity, thus we expect our model to be able to capture more reliably congestion dynamics. To achieve this goal, in the next chapter we generalize the conditions introduced in (Li, Lam, and Wong 2014b) in order to hold for all activity types.

7.4 Properties of the Utility-Based DTA

In this section, the bottleneck model is combined with the proposed utility functions in order to evaluate the effect of dynamic travel time and congestion on the departure time. To do so, we study the behaviour of a homogeneous population of drivers, travelling on the network shown in Figure 7.2. Users are assumed to maximise Equation (7.2) using as decision variable the vector $\boldsymbol{t}$. First, we generalize the Bottleneck model in order to account for different activities. Then, we analyse the conditions under which a TB-DTA turns to be an AB-DTA and, lastly, we focus on the scheduled based case.

7.4.1 Extending the Activity-Based DTA

To capture the effect that utility at the destination has on the congestion pattern, we analyse the critical moment in which users switch from one activity to another. Let Activity 1 be performed at the origin of a trip, while Activity 2 is to be performed at the destination.

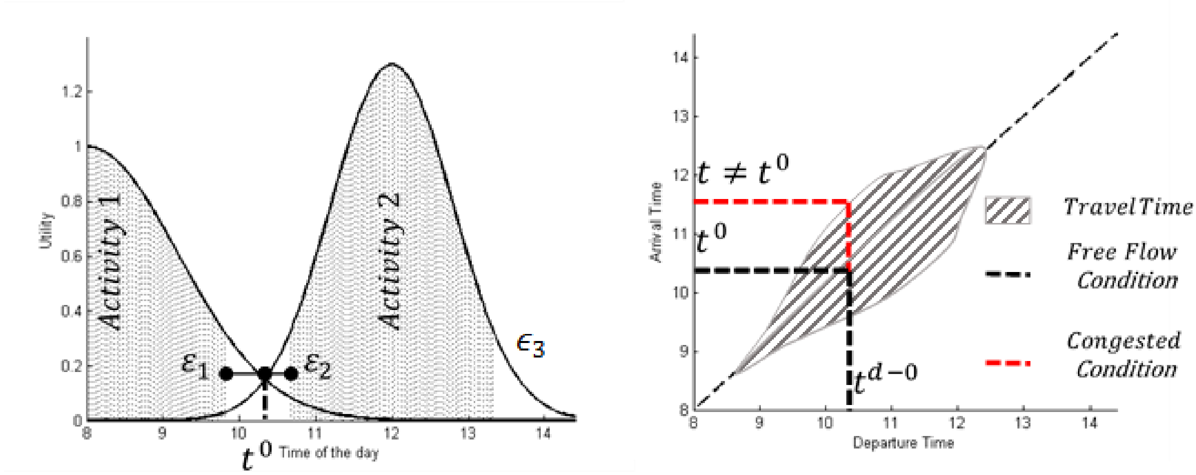


Fig.7.3: (a) clock-based MU for consequent activities; (b) Departure and arrival time;

In an ideal situation, when *N<S* and *T=0,* the arrival time at destination and the departure time from the origin are equal to the preferred arrival time $t^0 = t^{d-0}$, which is purely determined by maximising (the positive components of) the utility.

In a congested case, this option is not feasible anymore, and each user will choose between arriving at t^0 while incurring in some congestion, or changing his/her scheduling. Considering that the length of the congestion period will be equal to $\frac{N}{S}$, let $\frac{N}{S} = \varepsilon_2 - \varepsilon_1$, where ε_1 and ε_2 are the first and last moment in which congestion occurs according to the bottleneck model presented in section 7.3.3.

As demonstrated in (Li, Lam, and Wong 2014), a queue occurs at the bottleneck, for the home-work morning commute, when:

$$\begin{cases} U_1^a(t) - U_2^a(t) > -VoE \quad for \;\; t \in (\varepsilon_1, t^0] & (7.18a) \\ U_1^a(t) - U_2^a(t) < VoL \quad for \;\; t \in [t^0, \varepsilon_2) & (7.18b) \end{cases}$$

Where $U_1^a(t)$ and $U_2^a(t)$ are the utilities lost with respect to the activity Activities 1 and 2, respectively. Equation (7.18) has been demonstrated to hold for the home-work commute, but does not hold for general activities since it relies on the assumption that the first and the last user will not face congestion. In general, we cannot accept this assumption in this study. While there will always be a first user not facing congestion, for many activities (such as special events or concerts) it can happen that different users will arrive together and face some congestion. Specifically, Equation (7.18a) has been demonstrated under the assumption that the first user does not face congestion, while Equation (7.18b) under the assumption that the last user in the system does not face congestion. Thus, while it is intuitive to accept that Equation (7.18b) does not hold for the type of analysis carried on in this thesis, we might argue that Equation (7.18a) still holds (i.e. the assumption that the first user does not face queue is accepted). Thus, through a simple example, we show that Equation (7.18a) alone it is not sufficient when the last user faces congestion.

Consider that the first commuter entering in the system at $t = \varepsilon_1$ will face no queue, while for any departure time $t = \hat{t}$, users will face congestion and/or some schedule delay. Considering the case of early penalty, we calculate the utility until $t = \varepsilon_2$ as:

$$U(\varepsilon_1) = \int_{\varepsilon_1}^{\varepsilon_2} U_2^a(t)dt - U^t\ (\varepsilon_1) \tag{7.19a}$$

$$U(\hat{t}) = \int_{\varepsilon_1}^{\hat{t}} U_1^a(t)dt + \int_{\hat{t}+T^b(\hat{t})}^{\varepsilon_2} U_2^a(t)dt - U^t\ (\hat{t}) \tag{7.19b}$$

Where $U^t\ (t)$ represents the disutility of travelling. In the condition in which $U_1^a(t) = 0\ for\ t \in (\varepsilon_1, \varepsilon_2)$, $U_2^a(t^0) = \infty$ and $U_2^a = 0 \quad \forall\, t > t^0$, we have that at equilibrium the total net utility is infinite for both users, thus we have:

$$U(\varepsilon_1) - U(\hat{t}) = \left[\int_{\varepsilon_1}^{\varepsilon_2} U_2^a(t)dt - U^t\ (\varepsilon_1)\right] - \left[\int_{\varepsilon_1}^{\hat{t}} U_1^a(t)dt + \int_{\hat{t}+T^b(\hat{t})}^{\varepsilon_2} U_2^a(t)dt - U^t\ (\hat{t})\right] =$$

$$-U^t\ (\varepsilon_1) + U^t\ (\hat{t}) = \tag{7.22}$$

$$= \left[-VoE \cdot (t^0 - \varepsilon_1)\right] + \left[VoT \cdot \left(T^b(\hat{t})\right) + VoE \cdot (t^0 - \hat{t} - T^b(\hat{t}))\right] = 0$$

As for the standard bottleneck model, the equilibrium conditions depend upon the schedule delay and the travel time. However, at equilibrium we expect $\int_{\hat{t}+T^b(\hat{t})}^{\varepsilon_2} U_2(t)dt = \infty\ \forall\, \hat{t}$. In order to achieve this goal, the last user entering in the network chooses the departing time in order to be at destination just in time: $\hat{t} + T^b(\hat{t}) = t^0$. Since the user does not face late penalty, Equation (7.20) becomes:

$$U(\hat{t}) - U(\varepsilon_1) = -\,VoE \cdot (t^0 - \varepsilon_1) + VoT \cdot \left(T^b(\hat{t})\right) = 0 \tag{7.21a}$$

$$\frac{VoE}{Vot} = \frac{\left(T^b(\hat{t})\right)}{(t^0 - \varepsilon_1)} \tag{7.21b}$$

As we can see, the case in which the utility is infinite in t^0, the last user will face congestion while no user will arrive late. For the special case $VoE = VoT$, the equilibrium solution is $(\hat{t} = \ \varepsilon_1\ \forall\, n \in N)$. This is shown in Figure 7.4, where the y-axis shows the cumulative of the departure ratio, and the x-axis the time. In this case all the demand (in the example N=5000 users) chooses the same departure time interval $t = \varepsilon_1$. The length of the congestion period is, as previously indicated, equal to N/S, where S is the saturation flow, 2000 veh/h in this case.

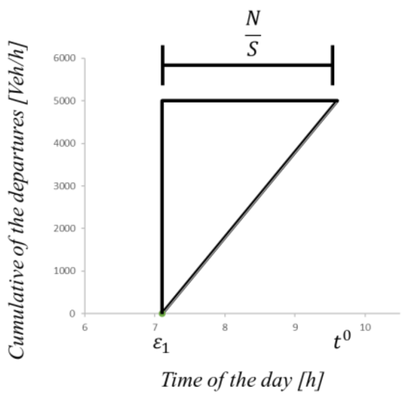


Fig.7.4: Cumulative of the departures for the case $\alpha = \beta$, *when* $t = \varepsilon_1 \forall\, u \in N;$

Equations (7.19-7.21) show that the Equation (7.18a) alone does not hold when we accept that the last user might face some queue. Thus, we introduce a new and more general condition to explicitly take into account this phenomenon:

Proposition 1: *Given the preferred departure time t^{d-0}, the departure time of the first user ε_1 and of the last user ε_2, a queue must exists at the bottleneck if and only if at least one of the two following conditions is satisfied:*

$$\begin{cases} \int_{\varepsilon_1}^{t^0} \left(U_2^a(t) - U_1^a(t)\right)dt < VoE \cdot (t^0 - \varepsilon_1) & \text{(7.22a)} \\ \int_{\varepsilon_1}^{\varepsilon_2} \left(U_2^a(t) - U_1^a(t)\right)dt < VoE \cdot (t^0 - \varepsilon_1) - VoL \cdot (\varepsilon_2 - t^0) & \text{(7.22b)} \end{cases}$$

Equation (7.22a) takes into account the modelled behaviour already discussed in (Li, Lam, and Wong 2014b), while Equation (7.22b) considers that the last user might face congestion. It should be pointed out that the system of equations (7.22) is a generalization of the system of equations (7.18). While Equation (7.22a) is basically the same as Equation (7.18a), Equation (7.22b) it seems to point in the opposite direction with respect to Equation (7.18b). There are three main reasons to explain this difference. First, as we said at the beginning of this section, Equation (7.18b) holds under the assumption that there is no queue at time interval ε_2. By removing this assumption, the equation does not hold. Second, Equation (7.22) specifies if a queue exists or not, while Equations (7.18) divide the system in queue before and after the preferred arrival time t^0. This assumption is unreasonable for the current study since we want to model the case for which all users decide to leave before the preferred departure time, as we showed in Figure 7.4. Lastly, in Equations (7.22) we accept that the condition $T^b(\hat{t}) < T^b(t^0)\ \forall\ \hat{t}\ \in (\varepsilon_1, \varepsilon_2)$ does not hold. Thus, the system of equations (7.22) is more general to identify if a queue exists in the period $(\varepsilon_1, \varepsilon_2)$. The proof of Proposition 1 is discussed in Appendix C.2.

7.4.2 Activity-Based DTA

After identifying the conditions under which a queue is expected at the bottleneck, we investigate in this section the effect of the utility on the dimension of the rush hour in terms of the length of the queue and beginning of the congestion period. With respect to Figure 7.3 we can calculate the utility lost as:

$$U^{lost} = U_1^a(\varepsilon_1, t^0) + U_2^a(\varepsilon_2, t^0) = \int_{\varepsilon_1}^{t^0} U_1^a(t)dt + \int_{t^0}^{\varepsilon_2} U_2^a(t)dt \tag{7.23}$$

Where $U_1^a(\varepsilon_1, t^0)$ and $U_2^a(\varepsilon_2, t^0)$ are the utilities lost with respect to the ideal activity pattern for Activities 1 and 2, respectively. U^{lost} represents the maximum possible amount of lost utility in the system. In general, we can argue that the size of U^{lost} will affect the departure ratio. When U^{lost} tends to zero, the model calculates the departure ratio by taking into account scheduled delay and travel time, i.e. it coincides with the standard bottleneck model. Considering instead $U^{lost} > 0$, we can calculate ΔU as:

$$\Delta U(\varepsilon_1, \varepsilon_2, t^0) = U_1^a(\varepsilon_1, t^0) - U_2^a(\varepsilon_2, t^0) \tag{7.24}$$

This consideration suggests one reflection, i.e. the bigger $\Delta U(\varepsilon_1, \varepsilon_2, t^0)$ is, the more the utility-based equilibrium will diverge with respect to the trip based one.

Equations (7.23-7.24) capture only the positive effect of the utility on the rush hour, neglecting the effect that the disutility has on the duration of the rush hour. However, the two terms ε_1 and ε_2 depend on the early and late penalty coefficients *VoE* and *VoL*. Specifically, if *VoE=VoL*, then we have a symmetric distribution of the demand around the preferred departure time, and Equations (7.16-7.17) properly capture the trade-off between departure time and utility. Otherwise, we need to explicitly consider this trade-off when calculating $\Delta U(\varepsilon_1, \varepsilon_2, t^0)$. Given this motivation, we can re-formulate the first property of the Activity Based DTA models as:

$$\Delta U({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB}) = U_1^a({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB}) - U_2^a({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB}) \tag{7.25}$$

$$U_1^a({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB}) = \int_{{\varepsilon_1}^{TB}}^{{\varepsilon_2}^{TB}} U_1^a(t)dt \tag{7.25}$$

$$U_2^a({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB}) = \int_{{\varepsilon_1}^{TB}}^{{\varepsilon_2}^{TB}} U_2^a(t)dt \tag{7.25}$$

Where coefficients ${\varepsilon_1}^{TB}$ and ${\varepsilon_2}^{TB}$ in Equation (7.25) are calculated with the TB-Bottleneck model, thus without considering the positive effect of the utility. The main advantage is that they can be calculated easily through the equations of the standard bottleneck model (Arnott, de Palma, and Lindsey 1990).

Given Equation (7.25), we can now formulate three properties for the Activity-Based DTA:

Property 1: *If $U^{lost} \neq 0$, the congestion period may shift. Specifically, TB-DTA anticipates the congestion if $\Delta U > 0$, while for $\Delta U < 0$ the congestion period is postponed;*

Property 2: *If $U^{lost} > 0$ and $\Delta U = 0$, then TB models underestimate/overestimate the queue with respect to AB models, while congestion shift is not observable;*

Property 3: *Under condition (22), when $U_1^a(t) > U_2^a(t)\ \forall\ t \in (\varepsilon_1, \varepsilon_2)$, the queue will reach its maximum length sooner or in the same time with respect to the case in which $U_1^a(t) < U_2^a(t)\ \forall\ t \in (\varepsilon_1, \varepsilon_2)$.*

Since an analytical solution exists for constant utilities, in order to demonstrate this concept, Figure 7.5 shows the results for constant values of $U_1^a(t)$ and $U_2^a(t)$. The numerical experiments will then generalize our findings for non-constant utility functions.

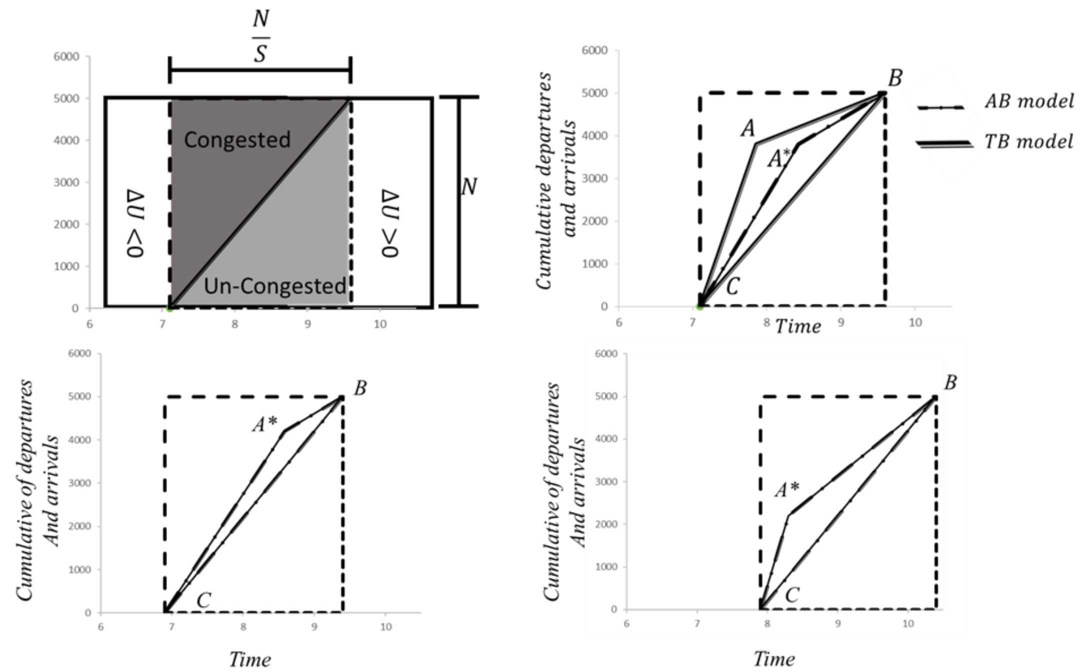


Fig.7.5: (a) Congestion as a function of $\Delta U(\varepsilon_1, \varepsilon_2, t^0)$; Difference between TB and AB model when (b) $U^{lost} \neq 0$ and $\Delta U = 0$, (c) $\Delta U < 0$ and (d) $\Delta U > 0$;

Figure 7.5 shows the typical cumulative representation for the standard bottleneck model. In Figure 7.5a, we can see the general effect of U^{lost} on the congestion. If **Proposition 1** holds, congestion occurs. In Figure 7.5b, the curve CAB represents the cumulative departure ratio for the standard TB bottleneck model, while $CA^{*}B$ is the cumulative departure ratio for the AB version. The curve CB represents the cumulative arrival ratio, which is the same for the AB and TB model, and depends upon the capacity of the bottleneck. Before moving to the next step, which is considering that activities are interconnected, we provide here some numerical analysis to prove Properties (1-3) to hold.

Remarks on property 1: Considering the TB-bottleneck model, we can easily calculate that, at equilibrium, the following equation has to be satisfied:

$$-VoE \cdot (t^0 - {\varepsilon_1}^{TB}) = -VoL \cdot ({\varepsilon_2}^{TB} - t^0) \tag{7.26}$$

At the equilibrium, both the first and last user does not face congestion and the overall utility is related to the late/early arrival penalty. Introducing a positive cost component in the system, Equation (7.26) becomes:

$$\int_{{\varepsilon_1}^{TB}}^{{\varepsilon_2}^{TB}} U_2^a(t)dt - VoE \cdot (t^0 - {\varepsilon_1}^{TB}) = \int_{{\varepsilon_1}^{TB}}^{{\varepsilon_2}^{TB}} U_1^a(t)dt - VoL \cdot ({\varepsilon_2}^{TB} - t^0) \tag{7.27}$$

Or, substituting Equations (7.26,7.25b, 7.25c) within Equation (7.27):

$$U_2^a({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB}) = U_1^a({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB}) \tag{7.28}$$

Equation (27) holds if and only if the overall utility at the origin and at the destination during the time interval $({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB})$ is the same. In all other conditions, equilibrium is not satisfied. As a consequence, users will have to reschedule their activities in order to reach an equilibrium point. There are two options. The last user is anticipating his trip, thus anticipating the beginning of the rush-hour, or the last user will postpone his/her own departure time. Equation (7.27) can be written as a difference, obtaining again Equation (7.25). If $U_2^a({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB}) > U_1^a({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB})$, then $\Delta U({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB}) < 0$, meaning that the utility at the destination is larger than at the origin. Thus, demand in time interval ${\varepsilon_1}^{TB}$ will increase, while decreasing in time interval ${\varepsilon_2}^{TB}$. Since the assumption that the first user will never face congestion, the only solution is that at equilibrium $\varepsilon_1 < {\varepsilon_1}^{TB}$. The opposite phenomena can be observed if $\Delta U({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB}, t^0) > 0$.

Remarks on property 2: Equation (7.27) shows that, when $U_2^a({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB}) = U_1^a({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB})$, then Equation (7.26) holds, meaning that there is no shift in the congestion period, i.e. $\varepsilon_1 = {\varepsilon_1}^{TB}$ and $\varepsilon_2 = {\varepsilon_2}^{TB}$. However, travel costs during the time interval $({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB})$ will still change. Intuitively, we thus expect that the following assumption to be true:

If

$$U_2^a({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB}) = U_1^a({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB}) \tag{7.29}$$

Then the output of the TB and AB Bottleneck model is the same.

However, this is not necessarily the case. In fact, users will still perceive a benefit at the origin/destination, which will still influence the departure time choice within the interval $({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB})$. This can be easily observed for the situation in which we have constant utility functions, under the assumption that equation (7.29) holds.

The analytical solution for the problem is described through the following system of equations (7.30). Given the preferred arrival time t^{a-0} for a certain activity, $\varepsilon_1, \varepsilon_2$ and t^{d-0} can be calculated as follows (Li, Lam, and Wong 2014b)

$$\varepsilon_1 = t^{a-0} - \frac{U_2 - U_1 + VoL}{VoE + VoL} \cdot \frac{N}{S} \tag{7.30}$$

$$\varepsilon_2 = t^{a-0} + \frac{U_1 - U_2 + VoE}{VoE + VoL} \cdot \frac{N}{S} \tag{7.30}$$

$$t^{d-0} = t^{a-0} - \frac{(U_2 - U_1 + VoL)(U_1 - U_2 + VoE)}{(VoE + VoL)(VoT + U_1)} \cdot \frac{N}{S} \tag{7.30}$$

We can clearly see that, if $U_2 = U_1$, the utility disappears from Equations (7.29a-7.29b). Specifically, the equation to identify ε_1 and ε_2 become the same as the one of the TB-Bottleneck model, confirming what already showed in Equations (7.26-7.28). However, even if Equation (7.29) holds, the utility does not disappear in equation (7.30c). We can calculate then the preferred departure time for both the TB and AB Bottleneck model:

$$t_{TB}^{d-0} = t^{a-0} - \frac{(VoL)(VoE)}{(VoE + VoL)(VoT)} \cdot \frac{N}{S} \tag{7.31}$$

$$t_{AB}^{d-0} = t^{a-0} - \frac{(VoL)(VoE)}{(VoE + VoL)(VoT + U_1)} \cdot \frac{N}{S} \tag{7.31}$$

Since the utility is only present in the denominator, this leads to:

$$t_{TB}^{d-0} < t_{TB}^{d-0} \qquad \forall\, U_1 > 0 \tag{7.32}$$

Equation (7.32) also implies that $T(t_{TB}^0) > T(t_{AB}^0)$, thus that the TB-DODE is overestimating the length of the queue. This proofs property 2.

Remarks on property 3:

Again, we adopt the assumption to have constant utility functions, since this assumption has an analytical solution. The departure time between time interval (ε_1, t^0) can be calculated as:

$$DeP(\varepsilon_1, t^0) = \frac{VoT + VoE + U_1}{VoT + U_2} \tag{7.33}$$

We can easily observe that $DeP(\varepsilon_1, t^0)$ increases for increasing values of U_1, while decreases for increasing values of U_2. This show that Property (3) holds.

7.4.3 Schedule-Based DTA

The discussion so far focused on the critical difference between AB-DTA and TB-DTA models. As pointed out in the introduction, considering the utility at the origin and at the destination is not a sufficient condition for linking different purpose-specific trips. According to the literature, *clock-based* MU functions are not able of capturing the correlation between morning and evening commute since arrival time and departure time for the same location are independent (Adnan 2010). However, if this correlation is explicitly included, for example by assuming $\tau \neq 0$ in equation (7.3), *clock-based* functions can capture this dependence (Li, Lam, and Wong 2014). We can thus formulate the following property:

Property 4: *Trips belonging to the same tour of activities are independent if and only if:*

1) U^{lost} *is constant.*
2) *If an equilibrium solution for users travelling from Activity 1 to Activity 2 exists, as shown in equation (7.22); If does not, equilibrium might exist over multiple trips.*

Then the AB-DTA model is trip based, i.e. is not able to capture the dependency between different trips. All utility functions proposed in Section 7.3.2 violate condition (1). It should be pointed out that, according to this condition, the utility function presented in Equation (7.5) is the least appropriate with respect to modelling the working activity. In general, the working activity takes about 8 hours. As observed in (Feil, Balmer, and Axhausen 2009), according to Equation (7.5), the best option is to perform more activities with a shorter duration (2-3 hours). If the duration is higher, then the utility of working one more hour is almost zero. Thus, we introduce a new metric to measure the Degree of Correlation (DoC) between activities. First, we define $dU(\hat{t})$ as the utility related to last hour at the origin:

$$dU(\hat{t}) = \begin{cases} \int_{\hat{t}-1}^{\hat{t}} U^{a}(t)dt & for\ constant\ values\ of\ U^{lost} \\ 0 & otherwise \end{cases} \tag{7.34}$$

In Equation (7.34), $\hat{t}$ represents the departure time from the origin and $U^{a}(t)$ the utility at the origin. Thus, $dU(\hat{t}-1,\hat{t})$ measures the utility during the last time interval before the departure. In general, the smaller $dU(\hat{t})$ is, the weaker is the correlation between activities. Then, the proposed metric measures the impact of the last time interval of activity with respect to the overall activity duration:

$$DoC = \frac{dU(\hat{t})}{\int_{t^{a-0}}^{\hat{t}} U^{a}(t)dt} \tag{7.35}$$

To demonstrate why the DoC can properly capture the degree of correlation, let us consider the utility of the activity pattern Home – Work – Home for a generic user, departing from home in $\hat{t}_1$, and from the working place at $\hat{t}_2$:

$$\begin{aligned} U(\hat{t}_1,\hat{t}_2) &= \int_0^{\hat{t}_1} U^{MH}(t)dt + \int_{\hat{t}_1+T_1^b(\hat{t}_1)}^{\hat{t}_2} U^{W}(t)dt + \int_{\hat{t}_2+T_2^b(\hat{t}_2)}^{24} U^{EH}(t)dt - U^t(\hat{t}_1) - U^t(\hat{t}_2) = \\ &= \int_0^{\hat{t}_1} U^{MH}(t)dt + \int_{\hat{t}_1+T_1^b(\hat{t}_1)}^{\hat{t}_2-1} U^{W}(t)dt + \int_{\hat{t}_2-1}^{\hat{t}_2} U^{W}(t)dt + \int_{\hat{t}_2+T_2^b(\hat{t}_2)}^{24} U^{EH}(t)dt - U^t(\hat{t}_1) - U^t(\hat{t}_2) = \\ &= \int_0^{\hat{t}_1} U^{MH}(t)dt + \int_{\hat{t}_1+T_1^b(\hat{t}_1)}^{\hat{t}_2-1} U^{W}(t)dt + dU(\hat{t}_2-1,\hat{t}_2) + \int_{\hat{t}_2+T_2^b(\hat{t}_2)}^{24} U^{EH}(\hat{t}_1)dt - U^t(\hat{t}) - U^t(\hat{t}_2) \end{aligned} \tag{7.36}$$

Where U^{MH}, U^{W} and U^{EH} are respectively the utility of staying home in the morning, working in the afternoon and staying home in the evening. $\int_{\hat{t}_2-1}^{\hat{t}_2} U^{W}dt = dU(\hat{t}_2)$ represents the utility related to the last hour for Activity "Work". If the utility related to the last hour of the activity "work" is extremely low, then $dU(\hat{t}_2) \approx 0$:

$$\begin{aligned} U(\hat{t}_1,\hat{t}_2) &= \int_0^{\hat{t}_1} U^{MH}(t)dt + \int_{\hat{t}+T^b(\hat{t})}^{\hat{t}_2-1} U^{W}(t)dt + dU(\varepsilon_3,\hat{t}_2) + \int_{\hat{t}_2+T_2^b(\hat{t}_2)}^{24} U^{EH}(t)dt - U^t(\hat{t}) - U^t(\hat{t}_2) \\ &\approx \left[\int_0^{\hat{t}_1} U^{MH}(t)dt + \int_{\hat{t}_1+T_1^b(\hat{t}_1)}^{\hat{t}_2-1} U^{W}(t)dt - U^t(\hat{t})\right] + \left[\int_{\hat{t}_2+T_2^b(\hat{t}_2)}^{24} U^{EH}(t)dt - U^t(\hat{t}_2)\right] = \\ &= U_{Morning\ Commute} + U_{Evening\ Commute} \end{aligned} \tag{7.37}$$

Equation (7.37) shows that the assumption over dU allows to separate the two trip costs. This has been demonstrated to be a sufficient condition for modelling the two trips as separate trips (Adnan 2010). Hence, *duration-based* MU might not be the best function to model this correlation, since, for $\eta = 18£/\mathrm{min}$ (Yamamoto et al. 2000; Adnan 2010), the utility of working 7 hours is about 2% less than working 8 hours. As a consequence, we can conclude that Equation (7.5) captures this interdependence, but the resulting degree of correlation is fairly weak, which suggests that the conclusions reported in (Adnan 2010) and (Feil, Balmer, and Axhausen 2009) are complementary.

7.5 Numerical analysis

In this section, by adopting the numerical method presented in Section 7.3.3, we perform an in-depth sensitivity analysis of the effects of using the different utility functions presented in Section 7.3.1 through a set of numerical experiments. Specifically, we show in this section that:

1. Properties (1, 2, 3) hold for general functions and under general conditions. While we used an analytical model to demonstrate that these properties hold when an analytical solution exists, in this section we use the MSA algorithm to support the claim that these properties hold when the uniqueness of the (DTA) solution is not guaranteed;
2. The situation identified in Proposition 1 – i.e. the last user faces congestion – is a possible solution when the utility is properly considered within the DTA model;
3. The Utility Function (7.22) is more suited for modelling different type of activities;
4. The *DoC* metric presented in Equation (7.35) properly captures the degree of correlation;
5. When the degree of correlation DoC is low, AB-DTA can properly approximate SB-DTA models.

We replicate the experiment proposed in Li, Lam, and Wong (2014), for sake of comparability. In the proposed example, we assume to have a single class of users (N=5000 vehicles), performing the home-work-home commuting trip chain on the simple network shown in Figure 7.2. Demand and supply parameters are introduced in Table 7.2. The decision variables for each user are the arrival time at work (preferred arrival time $t^0 = 9$) and the departure time in the afternoon (preferred departure time $t_d^0 =$ *17*). A detailed overview of the parameters for the utility function is presented in Appendix C.3, while hereafter we introduce the most important characteristics for each case we analysed.

Table 7.2: Experiment Setup

Table 7.2: Supply and Demand Parameters
N = 5000 veh
t^0 *= 9*
t_d^0 *= 17*
S = 2000 veh/h

Case 1: We model the standard bottleneck model, i.e. without considering the utility at the destination.
Case 2: We consider constant utilities at the destination ($U_{Work} > U_{Home-Evening} > U_{Home-Morning}$).
Case 3: Home and work activities are modelled through Equation (7.4), as *clock-based* ($\tau_p = 0$).
Case 4: As in the previous case, we model all the activities through Equation (7.4), but considering the flexibility term ($\tau_p = 0.5$).
Case 5: In order to observe E1 with respect to Case 3, the utility of the activity work is increased, in order to observe a shift of the congestion period.
Case 6: Home and work activities are modelled through Equation (7.9), assuming that the fatigue effect is greater for the working activity ($G_{work} > G_{home}$).
Case 7: As in the previous case we calculate the utility through Equation (7.9). We increased significantly the utility for the second activity, simulating a special event, in order to observe the phenomena described in Equations (7.22-7.23).

Case 8: As proposed in Adnan (2010), we represent the home activitiy as *clock-based* and the work activity through Equation (7.5).

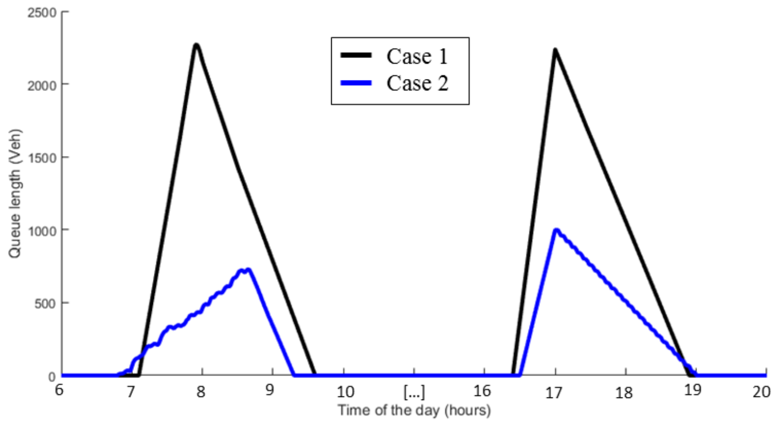
Case 9: Home and work activities are modelled through Equation (7.6), for the case in which w=0.5.

7.5.1 Activity-Based case

At this stage, our interest is to test the Properties (1, 2, 3) presented in Section 7.4.2. These conditions can be verified only for *clock-based* functions since it is not possible to uniquely define $\Delta U({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB})$ for *schedule-based* utility functions. In fact, when the duration is considered, different users perceive different utilities. For this reason, we first focus on analysing the results for Cases 1 until 5. In Figure 7.6, the most important parameters are shown. Specifically, Table 7.3, shows the value of $\Delta U({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB})$ for morning and evening commute, while Figures 7.6a-b the starting and ending time of the congestion period and the length the queue. Clearly when no utility, at the origin/destination is considered (Case 1), ΔU is equal to 0.

Again, this setup is the same presented in (Li, Lam, and Wong 2014b) in order to be replicable. In that paper, the authors also calculated the analytical solution for the model, which is properly approximated in Figure (7.6a), showing that the solution is a reliable approximation of the analytical result. In our study results have been divided into two sections. Figure (7.6a) shows the results when no utility or constant utility at the destination is used, while Figure (7.6b) shows the results when time-dependent (*clock-based*) utility functions are considered.

Table 7.3:	$\Delta U^{Morning}_{Commute}$	$\Delta U^{Evening}_{Commute}$
Case 1	0	0
Case 2	-7.5	2.5
Case 3	48*	-72.86
Case 4	48	-72.86*
Case 5	68.34	-93.07



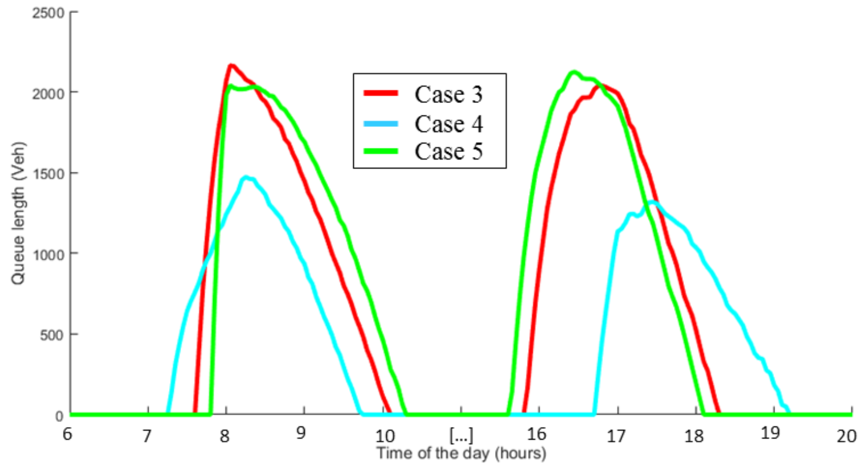


Fig.7.6: Queue length during the morning and evening commute for (a) Case 1-2 and (b) Case 3-4-5;

We first analyse **Property 1**. Through Figure 7.6, we can observe that in all cases in which a utility at the origin/destination is considered, there is a shift in the congestion period. If we focus on the morning commute, when ΔU is negative (Case 2), we can see that the rush hour begins 20 minutes before with respect to the TB-DTA bottleneck model ($\varepsilon_1^{Morning} = 7.1$). In all the other cases, in which ΔU is positive, the congestion period is shifted to a later time period. Similar considerations can be done for the evening commute. There is one main exception, where Property 1 cannot be observed, which is Case 4 in the evening commute. In this case, although $\Delta U({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB})$ is negative, the congestion period is postponed ($\varepsilon_1^{Evening} = 16{,}7$). However, we set the parameter $\tau_p = 0.5$. As a consequence, $U^{lost}(\varepsilon_1, \varepsilon_2, t^0)$ is not constant, since the utility for the morning and evening commute are chained, and the model turns to be an SD-DTA. This condition is also confirmed checking the $\Delta U({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB})$ values in Table 7.3, which is the same for Cases 3 and 4. This is because the flexibility with respect to the activity *work* neglects Properties (1-2-3), which are applicable if and only if $U^{lost}(\varepsilon_1, \varepsilon_2, t^0)$ is constant (as discussed in Property 4).

Concerning **Property 2**, when the utility at origin/destination is considered, the standard bottleneck model systematically overestimates the congestion. As we are going to show in the next section, TB-DTA models can also underestimate the length of the queue in certain conditions, such as special events.

Lastly, we focus on **Property 3**. Pointing out that the utility functions respect the condition listed in Equation (7.20), it is possible to observe that for all Cases 2, 3 and 5, when $\Delta U({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB})$ is positive, the queue reaches its maximum value sooner with respect to the case in which $\Delta U({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB})$ is negative.

This result supports the claim that the properties discussed in Section 7.4.2 hold for general functions, even when a uniqueness of the solution is not guaranteed.

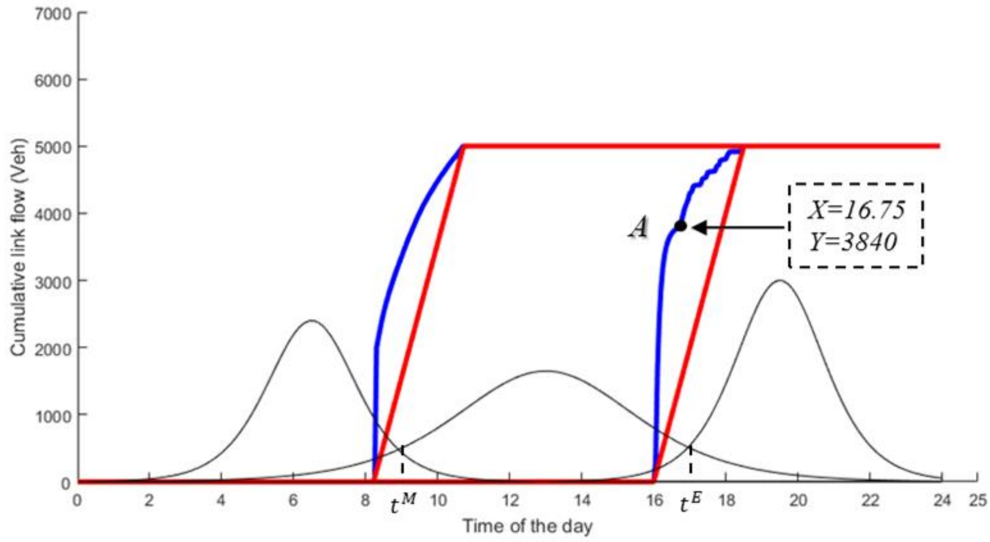
7.5.2 Schedule-Based case

The last part of this chapter aims at demonstrating the generalization proposed in Equation (7.22) and Equations (7.34-7.37). Hence, we focus on those utility functions that include the fatigue effect. In order to show the contribution, we applied Equation (7.9) to simulate activity *work* (Case 6) and to simulate a *special event*, e.g. a concert (Case 7). In the second case, the utility is extremely high in a single time interval. This leads to have a very high value of the utility in a specific time of the day (α=13 am). As shown in Figure 7.7b, the consequence is that users adapt their scheduling in order to arrive just in time at the concert. Moreover, since the value of the utility strongly decreases after a few hours because of the logarithmic effect, a similar behaviour is observed during the evening, where users spread around the preferred departure time, in order to still arrive at home when the utility is maximum. Simply stated, everybody wants to be there when the main star is starting the show, while most of the people wish to leave immediately after.

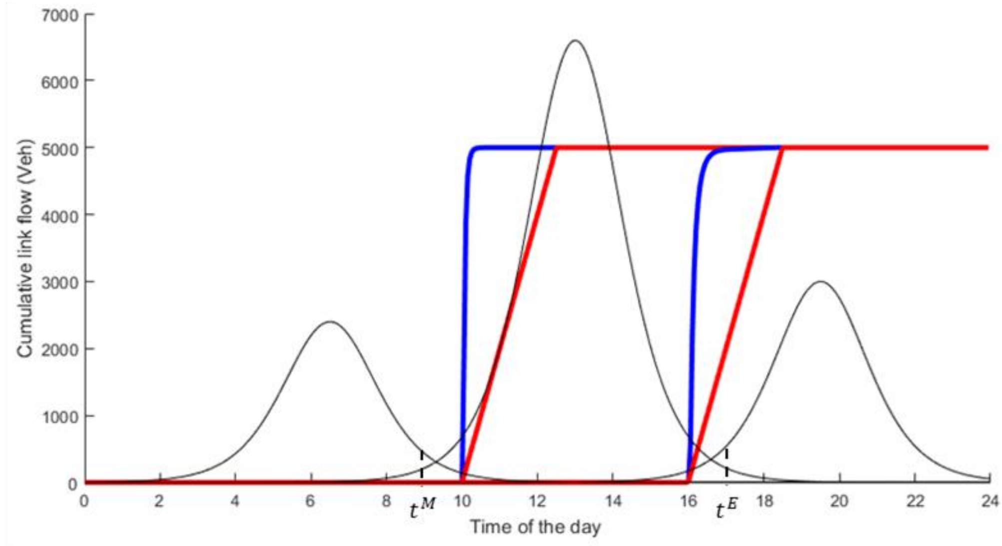
Figure 7.7a shows the interesting effect of considering activity duration within the utility *work*. As first consideration, it is important to compare Cases 6 and 7 with Cases 3 and 5 previously discussed. In fact, the parameters $\{\beta_p, \alpha_p, \gamma_p\}$ of the utility functions are the same, with the major difference that, in Cases 6 and 7, the *fatigue* (or *duration*) effect is considered. When comparing Cases 3 and 5, although we observed a shift in the congestion period, the behaviour of the users in the two experiments was similar, and both cases have been used to simulate the activity work. When duration is properly considered in the problem, Equation (7.22) allows modelling a completely different behaviour, which allowed us to model a different activity.

A second consideration regards the activity *work*. Although the model considers a single, homogeneous class of users, Equation (7.7) identifies two different behaviours in the evening commute. Before and after 16,75 pm, the departure ratio presents two different trends. The reason is that for those users that arrived too early, the fatigue effect is dominant, thus these users are highly motivated to leave the workplace and return home. Since the utility at 12-13 am is very high, those users that arrived late takes a higher utility in the beginning, thus they are not motivated to leave work before 17 – which is the preferred departure time. The different trend in the evening commute shows that this model properly represents this phenomenon.

Concerning Case 7, it should be stressed out how the last user is facing congestion, and, specifically, the last user is the one experimenting the largest delay. This supports the idea that, if non-working related activities are considered, the assumption that the last user does not face congestion is unrealistic. Moreover, we can also notice that all users in this experiment have a similar departure time both in the morning and in the evening. This means that, for this type of activity, all users would prefer to have the same activity duration, and utility before and after is near zero. This means that Case 7 can be probably approximated by a simplified model, like the one proposed in Equations (7.19-7.21) and showed in Figure (7.4).



(a)



(b)

Fig.7.7: Cumulative departure/arrive ratio according to the setting presented in Case 6 (a) and in Case 7 (b);

7.5.3 Comparison with previous works

While we discussed so far the general difference between AB-DTA and SB-DTA, we pointed out in Section 7.2 that other solutions have been proposed in the literature (Adnan 2010; Feil, Balmer, and Axhausen 2009; Li, Lam, and Wong 2014b). From a theoretical point of view, we can now point out that all these approaches satisfied the condition for which $U^{lost}(\varepsilon_1, \varepsilon_2, t^0)$ is not constant, showing that the general conditions proposed in this thesis hold. However, for the sake of comparability, we are now interested to (i) compare their result with the one proposed in Case 6 and (ii) investigate the DoC for all these cases, in order to understand the effectiveness of the proposed metric in estimating the degree of correlation between activities. In, Figure (7.8) the evolution of the queue is shown, while in Table 7.4 the DoC value for activity work is reported for each Case.

Table 7.4:	DoC_{work}
Case 6	4.84 %
Case 7	1.45 %
Case 8	1.87 %
Case 9	2.11 %

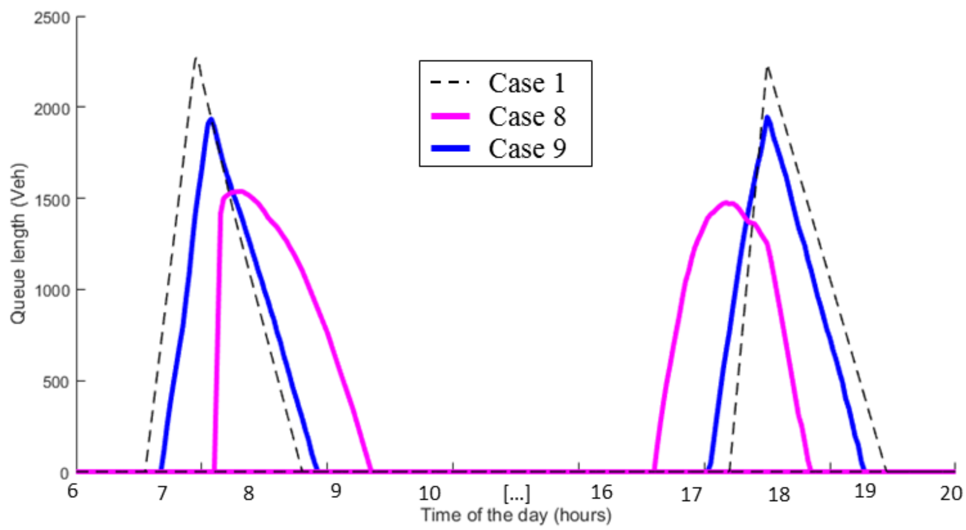


Fig.7.8: Cumulative departure/arrive ratio according to the setting presented in Case 6 (a) and in Case 7 (b);

Case 8 reproduces the UMT framework similar to the one proposed in (Adnan 2010), while the Case 9 exploits the utility function presented in (D. F. Ettema et al. 2007), which could replace Equation (7.9) for modelling the joint choice of activity time and duration. Results are shown in Figure 7.8.

First, we calculate the *DoC* coefficient for all Schedule-Based approaches tested so far. Results, showed in Table 7.4, show that the correlation is weak in all cases, with exception of Case 6. Specifically, when adopting Equation (7.8) to model the Activity Work, the correlation is almost 5 - i.e. the last hour contributes about 5% to the overall utility. For all the other cases, the *DoC* coefficient is less than half, meaning that the influence of the last hour is limited, and users are more likely to anticipate their trip. As we already discussed, this is expected for Case 7, which model a special event where all users have approximately the same activity duration and the utility in the last time interval is naturally low. However, Case 8 and 9 are supposed to model the Home-Work commute, so a higher value of the DoC is expected.

If we consider the results obtained applying Case 8, results look similar to the results for Case 3 (Figure 7.6d), with the main difference that the maximum number of vehicles queuing is lower, while we can observe that the shift of the congestion is more significant. The main reason is again in the dimension of $\Delta U({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB})$ and $U^{lost}(\varepsilon_1, \varepsilon_2, t^0)$. Although we cannot calculate analytically this number, because $U^{lost}(\varepsilon_1, \varepsilon_2, t^0)$ is not constant, we can easily observe that if we were to calculate $\Delta U({\varepsilon_1}^{TB}, {\varepsilon_2}^{TB})$ for a single user, this will always be positive for the morning commute and negative for the evening commute. In fact, when Equation (7.5) is used to model the work activity, the utility is not time dependent – i.e. is a constant utility combined with a *fatigue effect*. Thus, while users lose utility for the activity home, this is not the case for activity *work*. Moreover, by applying Equation (7.5) to model activity *work*, as discussed in section 7.4.2, if a certain user has an activity duration of 7 hours – instead of 8 – the reduction in the utility is relatively small compared to the utility lost at home in the evening. To support this claim, we in Figure 7.9 we show that we can approximate Case 8 by adopting an AB-DTA. Specifically, by exploiting Properties (1-2-3) of the AB-DTA, we can modify Case 3 in order to shift the congestion period and reduce the queue. Although the approximation is far from been perfect, these results do not derive from an optimization process but by simply implementing the properties showed in Section 7.4.2, and to support the claim that these properties can be used to adapt the available Utility-Functions to different applications.

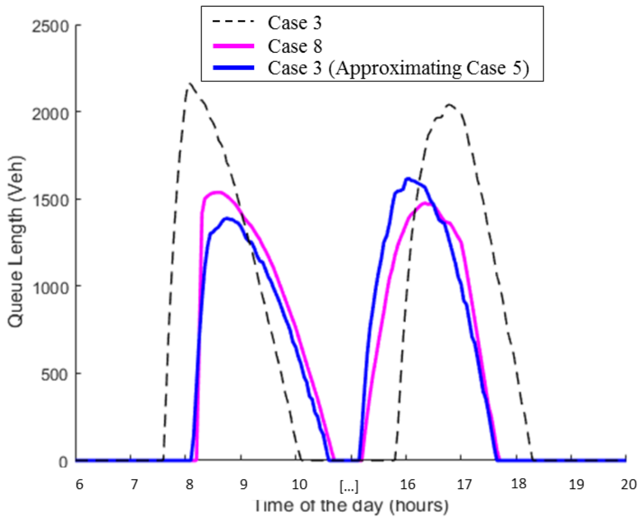


Fig.7.9: Cumulative departure/arrive ratio according to the setting presented in Case 6 (a) and in Case 7 (b);

It should be pointed out that results showed in Figure 7.8 for the evening commute are different from those presented in (Adnan 2010). The main reason is that, in the original paper, the framework was considering a schedule-based utility function for capturing the dis-utility of travelling only for modelling the morning commute, while we also applied the equation to the evening commute.

Lastly, let focus on Case 9. Also in this case, results look similar to the results for Case 3. The major difference is that the queue has the triangular shape typical of constant utility functions (Figure 7.6a). This result can be explained by analysing the structure of Equation (7.6), which is the linear sum of the *clock based* and *duration based* components.

If we were to consider only the *clock-based* component, then we would expect to observe again the same result as in Case 3. This is actually the case when we observe that the congestion period shifted and the maximum length of the queue is lower. However, when we sum the *duration-based* component, the function has the tendency to overestimate the utility for all those time intervals in which the *clock-based* component is low. This is shown also in the calibration results presented in Appendix C.1. Since the *duration-based* component is constant over time, as long as the duration is fixed, the final result is a congestion pattern that fallows the time-dependent utility – because of the *clock based* component – with a more linear shape – because of the *duration based* component. Also in this case, if we were to use a clock based function – although the model turns to be an AB-DTA – the predicted congestion pattern would not be too different from the real one.

7.6 Conclusions

This chapter studied the effect of combining Dynamic Traffic Assignment (DTA) models and Utility Maximization Theory (UMT) to model within-day traffic patterns. This research contributes to the state of the art by: 1) establishing that a correlation exists between the utility lost and the overall level of congestion; 2) extending the Bottleneck model in order to consider different type of activities; 3) establishing the condition under which a model can be considered as Trip-Based, Activity-Based, Trip-Chain Based or Schedule-Based; 4) Measuring the degree of correlation between multiple activities; 5) Defining a new Utility-Function, which allows to model different activities while considering activity duration in a realistic way. This has been proven analytically, by formulating a set of Properties and Propositions for the bottleneck model, and through numerical experiments.

We also showed that – under certain conditions - activity-based DTA can properly approximate scheduled-based models. This is extremely relevant since, from an algorithmic point of view, it is possible to significantly simplify the problem. To support this conclusion, we showed that, although some quite popular utility functions capture the correlation between different trips, the degree of correlation and/or the effect on the congestion pattern are fairly low. The proposed method shows how to calculate this degree of correlation so that DTA modeller can have a better understanding of the consequences of oversimplifying the choice model.

Following this concept, the novel utility function proposed in this chapter matches the above-stated conclusions. By jointly modelling *clock-based* and *duration-based* utility, the proposed function not only simulates a more realistic behaviour, but it is also able of modelling different activities such work or special events and capturing in a realistic way the correlation between arrival time and departure time.

Future research focuses on generalizing these findings to more complex scenarios. First, we showed that a simple Activity-Based DTA can approximate a Schedule-Based DTA. However, we need to also analyse if the approximation holds when new traffic policies -such as tolling – are applied. It can be that the approximated model will lead to a different result. Another key issue is also to further extend the choice model in order to take into account other decision dimensions which also influence the travel decision, such as weather conditions, accessibility of a certain area or the household composition. Then, we will test the properties analysed in this thesis with a more realistic Simulation-Based DTA, and generalize the problem for different routes, locations and transport modes.

Acknowledgements

We acknowledge for financing the following grant: – AFR-PhD grant 6947587 IDEAS.

7.7 References

Abdelghany, Ahmed, and Hani Mahmassani. 2003. “Temporal-Spatial Microassignment and Sequencing of Travel Demand with Activity-Trip Chains.” Transportation Research Record: Journal of the Transportation Research Board, no. 1831:89–97.

Adnan, Muhammad. 2010. “Linking Macro-Level Dynamic Network Loading Models with Scheduling of Individual’s Daily Activity–Travel Pattern.” In New Developments in Transport Planning: Advances in Dynamic Traffic Assignment, 251–74.

Arentze, Theo A, and Harry J. P Timmermans. 2004. “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.

Arnott, Richard, André de Palma, and Robin Lindsey. 1990. “Economics of a Bottleneck.” Journal of Urban Economics 27 (1):111–30. https://doi.org/10.1016/0094-1190(90)90028-L.

Balakrishna, Ramachandran. 2006. “Off-Line Calibration of Dynamic Traffic Assignment Models.” PhD Thesis, Massachusetts Institute of Technology. http://dspace.mit.edu/handle/1721.1/35120.

Balmer, M., K. Meister, M. Rieser, K. Nagel, and Kay W. Axhausen. 2008. Agent-Based Simulation of Travel Demand: Structure and Computational Performance of MATSim-T. ETH, Eidgenössische

Technische Hochschule Zürich, IVT Institut für Verkehrsplanung und Transportsysteme. http://svn.vsp.tu-berlin.de/repos/public-svn/publications/vspwp/2008/08-07/PortlandMATSim-T.pdf.

Barceló, Jaume, ed. 2010. Fundamentals of Traffic Simulation. Vol. 145. International Series in Operations Research & Management Science. New York, NY: Springer New York. http://link.springer.com/10.1007/978-1-4419-6142-6.

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

Ben-akiva, Moshe, Michel Bierlaire, Haris Koutsopoulos, and Rabi Mishalani. 1998. DynaMIT: A Simulation-Based System for Traffic Prediction.

Ben-Akiva, Moshe, Michèle Cyna, and André de Palma. 1984. “Dynamic Model of Peak Period Congestion.” Transportation Research Part B: Methodological 18 (4):339–55. https://doi.org/10.1016/0191-2615(84)90016-X.

Ben-Akiva, Moshe, Andre de Palma, and Pavlos Kanaroglou. 1986. “Dynamic Model of Peak Period Traffic Congestion with Elastic Arrival Rates.” Transportation Science 20 (3):164–81. https://doi.org/10.1287/trsc.20.3.164.

Bhat, Chandra R., and Rajul 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. 2001. “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., M. Bradley, Y. Shiftan, T. K. Lawton, and M. Ben-Akiva. 1999. “DEMONSTRATION OF AN ACTIVITY-BASED MODEL FOR PORTLAND.” In . http://trid.trb.org/view.aspx?id=639619.

Carey, Malachy. 1986. “A Constraint Qualification for a Dynamic Traffic Assignment Model.” Transportation Science 20 (1):55–58. https://doi.org/10.1287/trsc.20.1.55.

De Palma, André, and Richard Arnott. 1986. “Usage-Dependent Peak-Load Pricing.” Economics Letters 20 (2):101–5. https://doi.org/10.1016/0165-1765(86)90153-9.

Ettema, D., O. Ashiru, and J.W. Polak. 2004. Modeling Timing and Duration of Activities and Trips in Response to Road-Pricing Policies. Transportation Research Record.

Ettema, D. F., F. Bastin, J. Polak, and O. Ashiru. 2007. “Modelling the Joint Choice of Activity Timing and Duration.” Transportation Research A 41 (9):827–41.

Ettema, Dick, and Harry 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.

Feil, Matthias, Michael Balmer, and Kay W. Axhausen. 2009. New Approaches to Generating Comprehensive All-Day Activity-Travel Schedules. ETH, Eidgenössische Technische Hochschule Zürich, IVT, Institut für Verkehrsplanung und Transportsysteme. http://www.strc.ch/conferences/2009/Feil_2.pdf.

Fu, Xiao, and William Lam. 2014. “A Network Equilibrium Approach for Modelling Activity-Travel Pattern Scheduling Problems in Multi-Modal Transit Networks with Uncertainty.” Transportation 41 (1):37–55. https://doi.org/10.1007/s11116-013-9470-9.

Fu, Xiao, William H. K. Lam, and Yiliang Xiong. 2016. "Modelling Intra-Household Interactions in Household's Activity-Travel Scheduling Behaviour." Transportmetrica A: Transport Science 12 (7):612–28. https://doi.org/10.1080/23249935.2016.1189710.

Gentile, Guido. 2015. "Using the General Link Transmission Model in a Dynamic Traffic Assignment to Simulate Congestion on Urban Networks." Transportation Research Procedia, SIDT Scientific Seminar 2013, 5:66–81. https://doi.org/10.1016/j.trpro.2015.01.011.

Hendrickson, Chris, and George Kocur. 1981. "Schedule Delay and Departure Time Decisions in a Deterministic Model." Transportation Science 15 (1):62–77. https://doi.org/10.1287/trsc.15.1.62.

Jiang, Yu, W. Y. Szeto, Jiancheng Long, and Ke Han. 2016. "Multi-Class Dynamic Traffic Assignment with Physical Queues: Intersection-Movement-Based Formulation and Paradox." Transportmetrica A: Transport Science 12 (10):878–908. https://doi.org/10.1080/23249935.2016.1190421.

Kim, Hyunmyung, Jun-Seok Oh, and R. Jayakrishnan. 2006. "Application of Activity Chaining Model Incorporating a Time Use Problem to Network Demand Analysis." Transportation Research Record: Journal of the Transportation Research Board 1977 (January):214–24. https://doi.org/10.3141/1977-27.

Lam, W.H.K., and H.-J. Huang. 2002. "A Combined Activity/Travel Choice Model for Congested Road Networks with Queues." Transportation 29 (1):3–29.

Lam, W.H.K., and Y. 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.

Li, Zhi-Chun, William H.K. Lam, and S.C. Wong. 2014b. "Bottleneck Model Revisited: An Activity-Based Perspective." Transportation Research Part B: Methodological 68 (October):262–87. https://doi.org/10.1016/j.trb.2014.06.013.

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

Lindsey, C. Robin, and Erik T. Verhoef. 2000. "Traffic Congestion and Congestion Pricing." Tinbergen Institute Discussion Paper. http://www.econstor.eu/handle/10419/85524.

Liu, Peng, Feixiong Liao, Hai-Jun Huang, and Harry Timmermans. 2015. "Dynamic Activity-Travel Assignment in Multi-State Supernetworks." Transportation Research Procedia, 21st International Symposium on Transportation and Traffic Theory Kobe, Japan, 5-7 August, 2015, 7:24–43. https://doi.org/10.1016/j.trpro.2015.06.002.

Lu, Y., R. Seshadri, F. Pereira, A. Osullivan, C. Antoniou, and M. Ben-Akiva. 2015. "DynaMIT2.0: Architecture Design and Preliminary Results on Real-Time Data Fusion for Traffic Prediction and Crisis Management." In , 2015–October:2250–55. https://doi.org/10.1109/ITSC.2015.363.

Mahmassani, Hani, and Robert Herman. 1984. "Dynamic User Equilibrium Departure Time and Route Choice on Idealized Traffic Arterials." Transportation Science 18 (4):362–384. https://doi.org/10.1287/trsc.18.4.362.

Merchant, Deepak K., and George L. Nemhauser. 1978. "A Model and an Algorithm for the Dynamic Traffic Assignment Problems." Transportation Science 12 (3):183–99. https://doi.org/10.1287/trsc.12.3.183.

Noland, Robert B., and Kenneth A. Small. 1995. “Travel-Time Uncertainty, Departure Time Choice, and the Cost of Morning Commutes.” Transportation Research Record, no. 1493:150–58.

Palma, Andr&dacute; de, Moshe Ben-Akiva, Claude Lefèvre, and Nicolaos Litinas. 1983. “Stochastic Equilibrium Model of Peak Period Traffic Congestion.” Transportation Science 17 (4):430.

Peeta, S., and H.S. Mahmassani. 1995. “System Optimal and User Equilibrium Time-Dependent Traffic Assignment in Congested Networks.” Annals of Operations Research 60 (1):81–113. https://doi.org/10.1007/BF02031941.

Peeta, S., and Athanasios K. Ziliaskopoulos. 2001. “Foundations of Dynamic Traffic Assignment: The Past, the Present and the Future.” Networks and Spatial Economics 1 (3–4):233–65. https://doi.org/10.1023/A:1012827724856.

Peeta, Srinivas, and Athanasios K. Ziliaskopoulos. 2001. “Foundations of Dynamic Traffic Assignment: The Past, the Present and the Future.” Networks and Spatial Economics 1 (3–4):233–65. https://doi.org/10.1023/A:1012827724856.

Pendyala, Ram, Karthik Konduri, Yi-Chang Chiu, Mark Hickman, Hyunsoo Noh, Paul Waddell, Liming Wang, Daehyun You, and Brian Gardner. 2012. “Integrated Land Use-Transport Model System with Dynamic Time-Dependent Activity-Travel Microsimulation.” Transportation Research Record: Journal of the Transportation Research Board 2303 (December):19–27. https://doi.org/10.3141/2303-03.

Polak, J., and B. G. Heydecker. 2006. “The Dynamic Assignment of Tours in Congested Networks with Pricing.” In . First International Symposium on Dynamic Traffic Assignment: Conference Proceedings.

Ramadurai, Gitakrishnan, and Satish Ukkusuri. 2010. “Dynamic User Equilibrium Model for Combined Activity-Travel Choices Using Activity-Travel Supernetwork Representation.” Networks and Spatial Economics 10 (2):273–92. https://doi.org/10.1007/s11067-008-9078-3.

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

———. 2015. “The Bottleneck Model: An Assessment and Interpretation.” Economics of Transportation, Special Issue on Collective Contributions in the Honor of Richard Arnott, 4 (1–2):110–17. https://doi.org/10.1016/j.ecotra.2015.01.001.

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

Viti, F., Chris MJ Tampere, R. Frederix, Castaigne, Cornelis, and Walle. 2010. “Analyzing Weekly Activity-Travel Behaviour from Behavioural Survey and Traffic Data.” In .

Viti, F, and C.M.J. Tampere. n.d. “New Developments in Transport Planning.” Accessed March 4, 2010. http://www.abebooks.com/New-Developments-Transport-Planning-Tampere-Chris/12383047976/bd.

Yamamoto, Toshiyuki, Satoshi Fujii, Ryuichi Kitamura, and Hiroshi Yoshida. 2000. “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.

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.

Ziliaskopoulos, A.K. 2000. “Linear Programming Model for the Single Destination System Optimum Dynamic Traffic Assignment Problem.” Transportation Science 34 (1):37–49.

Zockaie, Ali, Meead Saberi, Hani S. Mahmassani, Lan Jiang, Andreas Frei, and Tian Hou. 2015. “Activity-Based Model with Dynamic Traffic Assignment and Consideration of Heterogeneous User Preferences and Reliability Valuation: Application to Toll Revenue Forecasting in Chicago, Illinois.” Transportation Research Record: Journal of the Transportation Research Board 2493 (January):78–87. https://doi.org/10.3141/2493-09.

8

Utility-Based OD Estimation

This chapter proposes a DODE framework that explicitly accounts for activity scheduling and duration. By assuming a Utility-Based departure time choice model, the time-dependent OD flow becomes a function, whose parameters are those of the utility function(s) within the departure time choice model. In this way, the DODE is solved using a parametric approach, which, on one hand, has less variables to calibrate with respect to the classical bi-level formulation while, on the other hand, it accounts for different trip purposes. Properties of the model are analytically and numerically discussed, showing that the model is more suited for estimating the systematic component of the demand with respect to the standard GLS formulation.

Content of this chapter has been presented in the following works:

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 Procedia, Papers Selected for the 22nd International Symposium on Transportation and Traffic Theory* Chicago, Illinois, USA, 24-26 July, 2017., 23 (Supplement C): 440–59. doi:10.1016/j.trpro.2017.05.025.

Cantelmo, Guido, Francesco Viti, Ernesto Cipriani, and Marialisa Nigro. 2017. "A Utility-Based Dynamic Demand Estimation Model That Explicitly Accounts for Activity Scheduling and Duration." *Accepted for publication to Transportation Research Part A*

8.1 Introduction

Simulation of traffic conditions requires accurate knowledge of the travel demand. In a dynamic context, this entails estimating time-dependent demand matrices, which are a discretized representation of the dynamic origin-destination (OD) flows. This problem, referred to as Dynamic OD Estimation (DODE) in literature, seeks for the best possible approximation of OD flows which minimize the error between simulated and available traffic data (Cascetta 1984; Cascetta, Inaudi, and Marquis 1993).

Traditional DODE models solve two interconnected optimization problems, according to a bi-level formulation: in the upper level, time-dependent OD matrices are corrected in order to replicate the observations, while the lower level relates OD with path and link flows. For an extensive overview of these models one can refer to (Antoniou et al. 2016; Cascetta, Inaudi, and Marquis 1993).

An important role in DODE problems is assigned to the Dynamic Traffic Assignment (DTA), which has the paramount role of determining the relation between link flows and OD flows. This is typically done by specifying a (dynamic) process for assigning OD flows to the routes connecting each OD pair, and by dynamically propagating the route flows onto the links. The combination of both processes determines the dynamic relation between link and OD flows, which is commonly known as the assignment matrix. Restricting our discussion on only the bi-level approach, the vast majority of the existing works uses a deterministic or a stochastic user equilibrium formulation to assign the time-dependent OD flows to the routes (Maher, Zhang, and Vliet 2001; Zhou, Lu, and Zhang 2012). The resulting dynamic OD matrices are therefore assumed to satisfy equilibrium principles at each time period. To guarantee consistency across time periods, sequential and simultaneous estimation approaches have been proposed and compared (Cascetta, Inaudi, and Marquis 1993; H. Yang, Iida, and Sasaki 1991). Consistency between the time-dependent route flows and link flows is instead guaranteed by choosing an opportune Dynamic Network Loading (DNL) model. Examples of DNL approaches used in DODE problems are those based on exit flow functions (Cremer and Keller 1984), on Kinematic Wave Theory (R. Frederix et al. 2011) or using simulation-based DTA models (Lu 2013; Tympakianaki, Koutsopoulos, and Jenelius 2015). In the last decades many researchers developed utility-based DTA models, which consider both the utility of performing an activity and the disutility of traveling. Microscopic agent-based DTA focus on generating comprehensive activity patterns (Flötteröd, Chen, and Nagel 2012), while flow-based models stress the correlation between morning and evening commute, hence their effect on congestion (Li, Lam, and Wong 2014). In this case, DODE research is relatively poor, and only a few works are available for estimating comprehensive demand patterns capable of reproducing realistic traffic conditions (Flötteröd 2009).

The DODE problem is usually underdetermined because of the high number of unknown variables and their mutual dependencies (Marzano, Papola, and Simonelli 2009). Spatial dependencies are related to the multiple mapping between OD, route and link variables and therefore depend on the complexity of the network topology, on the chosen route set and on the number and location of sensors (Simonelli et al. 2012; Viti, Verbeke, and Tampère 2008; Hai Yang and Zhou 1998). Under determinedness can also characterise DODE solutions because of the nonlinear relation between link and demand flows, which is related to congestion propagation phenomena such as spillback (Zhou, Lu, and Zhang 2012). Rodric Frederix, Viti, and Tampère (2013) analyse the effect of congestion, pointing out that, if link flows are assumed to be separable, biased solutions are likely to be found. To deal with these problems, many authors recommend that good starting values of the demand (seed matrices) should be available (Cipriani et al. 2013), and that a distance measure between these values and the estimated OD flows should also be included in the upper level. In existing DODE problems, seed matrices are often coarse adaptations of static matrices, which are derived from socio-demographic data. However, the more the network is complex and congested, the more this condition becomes tight, meaning that the DODE procedure collapses to local adjustments of the demand (Marzano, Papola, and Simonelli 2009). In addition, static matrices are often calibrated in order to match specific traffic patterns (e.g., the morning rush hour), while empirical evidence shows that spatial and temporal distribution of demand flows changes considerably along the day and in between days. Hence a good static matrix may be suited only for a relatively short evaluation time period.

We argue in this study that not enough attention is drawn on identifying and estimating reliable time-dependent seed matrices, which should take into account the underlying daily and weekly activity-travel patterns. Our research hypothesis is that incorporating information about daily activity scheduling and duration is of paramount importance to derive dynamic OD flows, which are consistent across time periods. We do so by first formulating the lower level of the traditional bi-level problem as a utility maximisation problem where the departure time choice to perform a certain set of activities is endogenously estimated. Then we extend the proposed model to consider also activity location and duration information.

We demonstrate that, if we extend the bi-level approach by taking into account such information, the number of free parameters in the DODE problem systematically decreases, reducing the under determinedness of the solution. In this Chapter we demonstrate that the proposed utility-based formulation brings the following scientific and practical contributions:

- By adopting a parametric approach, the number of decision variables is systematically reduced, and a smoother objective function is obtained by exploiting the relation between utility and dynamic user equilibrium (DUE);
- By extending the utility-based approach to account for activity scheduling, location and duration, richer information can be contained in the generated demand matrices;
- By estimating the demand as simultaneous route and departure time choice, the localism of the general DODE formulation is reduced, hence the reliability of the estimated dynamic OD matrices is improved.

We will show how the above contributions are achieved by means of a theoretical analysis and with numerical tests performed on toy networks.

8.2 Methodology

This work focuses on extending the classical DODE problem to account for the utility of performing different activities. For each considered activity, users are assumed, in this study, to maximize their own utility, which represents the perceived net benefit of performing the activity at some location. Hence, while most of the DODE models consider only the cost of travelling, the utility at the destination represents the main reason for travelling and significantly determines travellers' decisions such as where and when to perform the said activity.

We first formulate the joint route and departure time choice model as a utility maximisation problem within a bi-level formulation. We adopt the "classical" Generalised Least Squares (GLS) estimator presented in (Cascetta, Inaudi, and Marquis 1993), which is widely adopted in practice, as the benchmarking model in the next sections. In order to properly extend the bi-level formulation, both the upper level and lower level need to be modified for explicitly considering heterogeneous demand and user behaviour. Two sources of information are considered in the goal function: historical OD flows and traffic counts. The motivation is twofold. On one hand, a mathematical relation between these measures exists, which is relevant for demonstrating the properties of the proposed formulation. On the other hand, these data have been widely used in the literature as, even with the great advancements in telecommunications and sensor technologies, they still represent the most commonly used sources of information in practice. All the advantages related to using other sources of information data, such as probe vehicles or link densities, hold for both models.

8.2.1 Lower Level – the DTA model

DTA models aim to describe the mutual interaction between demand and supply systems. A key building block of DTA is the assignment process, which is often characterised by a choice model. Many authors stressed that travellers jointly choose route and departure time with respect to an (expected) experienced travel time (Arnott, de Palma, and Lindsey 1990; Mahmassani and Herman 1984; Zockaie et al. 2015) and a preferred arrival time at the destination (Vickrey 1969). Thus, we consider, as an explicit requirement for the DTA, which is used in the lower level of our DODE formulation, to include a Departure Time choice Model (DTM) as simultaneous decision process to the route choice. In this section we assume that the destinations of where to perform the activities are known, hence the location of the activities is assumed fixed and assigned a-priori. We will relax this assumption in Section 3.

The advantage of formulating a DTM within the DODE is twofold: first, given a number of users, N_{od}, travelling along an origin-destination pair, *od*, the DTM will estimate the temporal distribution of the N_{od} users within the DTA model. Second, we argue that utility-based DTM models are suited for estimating activity scheduling and duration, since the activity pattern is a function of activity type, duration, travel time, and the preferred arrival time at the destination (Zhang et al. 2005). Hence, including a DTM is seen in this study as a natural step to include the utility of performing an activity in the DTA model.

The utility-based DTM is assumed to depend on the joint choice of departure time and route, and it is formulated as follows:

$$U_n^{s,p}(t,r) = U_n^{s} + U_n^p = [V_n^s(t,r) + \varepsilon_n^s(t,r)] + [V_n^p(t,r) + \varepsilon_n^p(t,r)] \qquad n \in N_{od}, t \in T, r \in R^{od}, s \in S, p \in P \qquad (8.1)$$

Where $U_n^{s,p}$ is the utility for a certain user *n*, U_n^s represents the (dis-)utility of performing a travel *s*, U_n^p the (positive) component of the utility of performing an activity *p*, *t* the departure time at the origin and *r* the route chosen to reach the destination, which for the time being we consider pre-determined. *P* and *S* are, respectively, the set of all activities and trips for a certain user *n*, whose locations are all known. The utility is a random variable, which is generally expressed as the sum of an observable (or systematic) component, and an unobservable component (or error). In Equation (8.1), the two terms V_n^s and V_n^p represent the systematic component, while ε_n^s and ε_n^p the random components of the utility, i.e. the perception error in the costs associated to the trip and the benefits for performing the activity, respectively. Following the general theories of individual choice behaviour (Ben-Akiva and Lerman 1986), for a certain trip *s* with purpose *p*, the probability $\Pr_n[(t^s,r^s)|(\tau,p,s)]$ of jointly choosing the departure time-route couple (t^s,r^s) within the set of time intervals Θ and the set of routes R^{od}, can be obtained as (Ben-Akiva and Lerman 1986):

$$\Pr_n[(t^s,r^s)|\,(\tau^p,p,s)] = \text{Prob}\left[V_n^{s,p}(t^s,r^s) + \varepsilon_n^{s,p}(t^s,r^s) \geq \max_{\substack{\theta\in\Theta \\ r\in R^{od} \\ (\theta,r)\neq(t^s,r^s)}} \left(V_n^{s,p}(\theta,r) + \varepsilon_n^{s,p}(\theta,r)\right)\right] \qquad (8.2)$$

Where R^{od} is the set of feasible routes for a certain *od* and Θ the set of simulation time intervals, t^s is the actual departure time and r^s the route. Usually Equation (8.2) is solved using a Discrete Choice modelling approach. In fact, the alternatives can hardly be represented within a continuous space and are usually mutually exclusive. The error term is included in the utility function to account for the fact that the analyst is not able to completely and correctly model or identify all attributes that determine travellers' behaviour. Many distributions could be used to represent the distribution of error terms over individuals and alternatives, such as the multivariate-normal or the Weibull-Gumble distribution, which lead to a different DTM.

Many functional forms have been proposed for calculating the V_n^s and V_n^p. To calculate the disutility of travelling, the bottleneck model formulation proposed by Vickrey and extended to a more general scheduling problem by Small (Vickrey 1969; Small 1982), looks suitable for the purpose of estimating the temporal distribution of the demand. This model is reformulated in Equation (8.3):

$$V_n^s(t^p, r^p) = \alpha_{od}^p \cdot \left(TT(\tau_{od}^p, t_{od}^p, r_{od}^p)\right) + \beta_{od}^p \cdot \left(EA(\tau_{od}^p, t_{od}^p, r_{od}^p)\right) + \gamma_{od}^p \cdot \left(LA(\tau_{od}^p, t_{od}^p, r_{od}^p)\right) \qquad (8.3)$$

Where *TT* is the travel time, *EA* and *LA* are scheduling delay for the early and late arrival, respectively, while α_{od}^p, β_{od}^p and γ_{od}^p are purpose and OD-dependent parameters to be calibrated, representing respectively the cost of travelling, of early arrival and late arrival. Finally, τ_{od}^p represents the preferred arrival (or departure) time for activity *p*. Equation (8.3) differs from the original version formulated in (Vickrey 1969), since we include here different values of the parameters for different purposes and different OD pairs. The original version was assuming a homogeneous population travelling on a simple network with one single OD pair and purpose. While it is intuitive to observe that parameters $\{\alpha_{od}^p$, β_{od}^p, γ_{od}^p, $\tau_{od}^p\}$ are purpose-dependent, we might expect them not to depend on different OD pairs, e.g. the penalty for arriving late at work is assumed to be the same for all the users. However, this penalty is different for different users, belonging to different classes. In the literature, more elaborated utility functions exist, which assume a heterogeneous user behaviour for users with the same trip purpose (Small 2015). While these utility functions can be adopted with the proposed model, at the current stage we assume having homogeneity between users travelling to/from a certain traffic zone. For instance, the preferred arrival time for all the users working in the business district is the same, but might be different from the preferred arrival time for users working in the city centre. Thus, the current model considers both a geographical and purpose-dependent heterogeneity of the demand. For simplicity of notation we refer to the parameters $\{\alpha_{od}^p, \beta_{od}^p, \gamma_{od}^p, \tau_{od}^p\}$ as $\boldsymbol{\varpi} = \{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\tau}\}$ in the rest of this Chapter.

For the utility at the destination, we do not suggest any specific function in this thesis. In the literature, many different formulations have been proposed for calculating U_n^p (D. Ettema and Timmermans 2003; Yamamoto et al. 2000). The most widely adopted functions are time-dependent and/or duration-dependent, meaning that they are not linear-in-parameters. Moreover, they assume different levels of correlation between time of the day, activity duration and departure time across different trips. If these choices are modelled as independent, a simple function with few parameters can be used for modelling the utility, while more parameters need to be considered for properly capturing their correlation. Regardless of their number, the utility function parameters are usually calibrated through survey data. Specifically, the most established way is to formulate Equation (7.2) as a Logit-type model, and then to calibrate the parameters through the maximum-log-likelihood estimation approach (Ben-Akiva and Lerman 1986; D. Ettema and Timmermans 2003). Since utility functions are not linear, nonlinear programming approaches should be implemented to solve the log-likelihood estimation. For further details on the correlation between duration/departure time U_n^p, the interested reader can refer to (D. F. Ettema et al. 2007).

The model presented in Equations (7.1-7.2) can be extended to represent the daily activity pattern. By assuming that U_n^s and U_n^p include their idiosyncratic term, for a generic user *n*, given a departure time *t* and a route *r*, the overall utility over the entire study period (e.g., a day), can be calculated as:

$$U_n(\boldsymbol{t}, \boldsymbol{r}) = \max_{\boldsymbol{t},\boldsymbol{r}}(U_n^S(\boldsymbol{t}, \boldsymbol{r}) + U_n^P(\boldsymbol{t}, \boldsymbol{r})); \qquad n \in N_{od}, t \in T, r \in R^{od} \qquad (8.4)$$

Where

$$U_n^S(\boldsymbol{t}, \boldsymbol{r}) = \sum_{s \in S} U_n^s(t^s, r^s); \qquad (8.4a)$$

$$U_n^P(\boldsymbol{t}, \boldsymbol{r}) = \sum_{p \in P} U_n^p(t^s, r^s); \qquad (8.4b)$$

Where t^s and r^s are the chosen departure time and route for performing a certain travel *s*. We also assume that the user *n* has a preferred departure time τ^p for each activity *p*. Alternatively, one can assume a preferred arrival time without affecting the model generality. To explain the relation between the preferred departure time τ, the actual departure time t, and the simulation time interval θ, we schematically depicted an example in Figure 8.1.

Simply stated, the simulation time interval θ is an input of the Dynamic Network Loading (DNL) model, which requires the duration of the analysis period (e.g. 24 hours), and how many time intervals we want to simulate; relatively smaller simulation time periods (such as 5 minutes) lead to more accurate simulation results. The DNL provides outputs for each simulation time interval θ, where if no vehicle is loaded on the link, the link flow will be equal to zero. Non-zero OD demand flows will propagate to reach the link represented in Figure 7.1 after some time. Since often the time intervals chosen for the OD matrices are different than those chosen for simulating the link flows, and since travel times to propagate flows are continuous variables, then fractions of OD flows stemming from two or more OD matrices can result in the same link flow at a time interval θ. By contrast, in our research, τ is a parameter of the departure time choice model. In this example, τ represents the most likely preferred departure time from the origin chosen by user *n* to reach the location where to perform the activity *p*. At equilibrium, and for a set of users, if travel times are flow-dependent, then the departure time choice model will spread the demand around the preferred value, providing as output the actual departure time *t*. Potentially, each simulation time interval θ could be an actual departure time, and the relation between t and θ is the one formulated in a generic way in Equation (8.2).

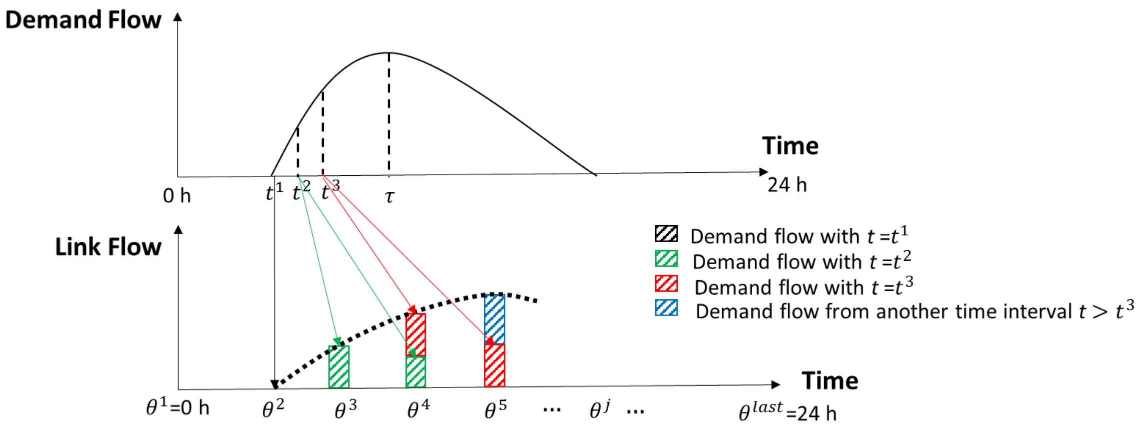


Fig.8.1: Relation between the preferred departure time τ, the actual departure time t, and the simulation time interval θ, their relation is showed;

Li et al. (Li, Lam, and Wong 2014) formulated the problem presented in Equation (8.4) as complementarity problem:

$$\begin{cases} N^{S,P}(\boldsymbol{t})[U^d(\boldsymbol{t},\boldsymbol{r}) - U_d^*] = 0 \\ N^{S,P}(\boldsymbol{t}) \geq 0, \qquad U^d(\boldsymbol{t},\boldsymbol{r}) - U_d^* \leq 0 \quad \forall\, \boldsymbol{t},\boldsymbol{r} \end{cases} \tag{8.5}$$

Where $N^{S,P}$ is the number of users with set of activities P and trips S, for a certain set of departure times t. System of Equations (8.5) leads to the user equilibrium condition, where any used set of departure times *t* leads to the same utility.

If we consider a single trip *s* with a single purpose *p*, then $N^{S,P}$ is equal to the time dependent OD flow $N^{S,P}(t) = x_{od}^t$. U_d^* is the equilibrium maximum utility for those sets of trips/activities, and $U^d(\boldsymbol{t},\boldsymbol{r})$ the utility related to a certain set of departure times and routes.

Before moving to the next section, it should be pointed out that many functional forms exist for calculating U^s and U^p, which lead to different DTM output. For a detailed overview on the properties of the utility functions we refer to (Adnan 2010; Cantelmo and Viti 2016; Li, Lam, and Wong 2014).

However, without loss of generality, we can observe that when the marginal utility of an activity U_n^p is considered as a function of the time-of-the-day only, system of Equations (8.4) detaches different trips (i.e. they are not correlated). When duration is explicitly considered, system of Equations (8.4) jointly estimates departure time and activity duration. As a consequence, when the proposed Utility-Based DTM is considered within the DTA in the lower level, DODE estimates demand flows that are consistent with the scheduling estimated in the lower level.

8.2.2 Upper Level – the DODE Formulation

At the upper level, the decision variables used to determine the objective function values represent the main difference between the proposed formulation and the classical DODE formulation. According to the original formulation proposed in (Cascetta, 1984), the OD flows for each time interval are the unknown variables to be estimated. However, when a parametric relation between OD pairs across a reference time period is assumed, the number of unknown variables decreases, leading to a more effective estimation of the time-dependent OD flows (Cascetta et al., 2013; Marzano et al., 2009). In the proposed formulation, the DTA model, formulated in Section 8.2.1, explicitly accounts for a parametric relation through the departure time choice model. As a consequence, we might expect that the only variable to be estimated is the number of users $N_{od} \forall \; od \in OD$, where OD is the set of all the (physical) OD pairs in the network. By extending this intuition to multiple activities, the overall number of variables becomes $n_p \cdot n_{od}$, where n_p is the number of activities and n_{od} the number of (physical) OD pairs. However, this condition holds if the departure time choice parameters are assumed to be known and constant. Balakrishna (2006) demonstrated that, if the route choice parameters are not estimated together with the demand flows, the DODE leads to unrealistic results and biased estimations. Since we expect this condition to hold for the departure time choice model, also the vector $\boldsymbol{\varpi}$, including all the most relevant parameters $\{\alpha, \beta, \gamma, \tau\}$ of the departure time choice model, needs to be calibrated. As for the parameters of the utility function U_n^p, a reasonable solution is to calibrate off-line the utility function parameters, i.e. through surveys data or using values from the literature (Adnan, 2010; Ettema and Timmermans, 2003). In this work, we consider the parameters of the dis-utility function presented in Equation (8.3), which properly models the mean and the variance of the departure time distribution when congestion occurs on the network. When no congestion occurs, we can use a different approach as proposed in (Ettema and Timmermans, 2003). Simple statistical tests, such as the likelihood-ratio test, can be used for evaluating the improvement due to considering more parameters in the model. Defined n_θ the number of simulation time intervals, the DODE problem can be formulated as:

$$\left(\boldsymbol{x}^{1,1}(\boldsymbol{\varpi},\boldsymbol{n}) \dots \boldsymbol{x}^{n_p,n_\theta}(\boldsymbol{\varpi},\boldsymbol{n})\right) =$$

$$= \min_{\boldsymbol{\varpi},\boldsymbol{n}}\left[z_1(\boldsymbol{d}^{1,1} \dots \boldsymbol{d}^{n_p,n_\theta}, \boldsymbol{x}^{1,1}(\boldsymbol{\varpi},\boldsymbol{n}) \dots \boldsymbol{x}^{n_p,n_\theta}(\boldsymbol{\varpi},\boldsymbol{n})) \cdot w_1 + z_2\left(\boldsymbol{f}^{e,1} \dots \boldsymbol{f}^{e,n_\theta}, \hat{\boldsymbol{f}}^1 \dots \hat{\boldsymbol{f}}^{n_\theta}\right) \cdot w_2\right] \tag{8.6a}$$

Subject to:

$$(\boldsymbol{f}^{e,1} \dots \boldsymbol{f}^{e,n_\theta}) = \max_{\boldsymbol{t},\boldsymbol{r}}(U^S(\boldsymbol{t}(\boldsymbol{\varpi},\boldsymbol{r},\boldsymbol{n})) + U^P(\boldsymbol{t}(\boldsymbol{\varpi},\boldsymbol{r},\boldsymbol{n}))); \tag{8.6b}$$

Where:

- $\boldsymbol{d}^{n_p,n_\theta} / \boldsymbol{x}^{n_p,n_\theta}$ is the vector including the starting/estimated demand values for each activity p and time interval θ;
- $\boldsymbol{f}^{e,\theta}/\hat{\boldsymbol{f}}^{\theta}$ represents the simulated/observed link flows during time interval θ;
- z_1, z_2 are the functions measuring the error between current and target OD flows and link flows, respectively;
- w_1, w_2 are the weights assigned to each error component depending on the trust one has on either the seed matrix or on the traffic data;
- $\boldsymbol{n}$ is the vector including the total number of users over the entire analysis period for each od/activity;

- $\boldsymbol{\varpi}$ is the vector with the DTM parameters $\{\alpha, \beta, \gamma, \tau\}$ to be calibrated, for each od/activity;
- $\boldsymbol{t}$ is the vector with the actual departure time, which is function of the departure time choice model;
- θ is the simulation time interval;

The vector $\boldsymbol{n}$ is a column vector including the demand for each OD and each purpose for the entire analysis period. Similarly, $\boldsymbol{\varpi}$ is the vector of the parameters of the DTM, for each OD and each trip purpose.

8.3 Properties of the model

In this section we show the advantages of the new Utility-Based model for solving the demand estimation (UB-DODE) with respect to the classical GLS, and the opportunity it offers to include additional information on activity location and duration. To do so, we divided this section in two parts. First, we consider the simple case of only one purpose and no utility at destination. This represents the classical situation in which only historical OD flows and traffic counts data are available, without any information on the activity done at destination. The model becomes therefore a standard trip-based approach. Thus we consider at this stage only a geographical heterogeneity for the user travelling on the network. Advantages of including scheduling information and/or parameters with respect to the classical approach are highlighted in this section. Then, we assume that activity location and duration are known or can be estimated within the DTA, thus we extend the UB-DODE model to include also the positive component in the utility function for performing an activity at destination.

8.3.1 Single-purpose: Including schedule function

In this section, we consider that the demand travelling on the network has one single purpose p, while the DTM model considers only the dis-utility of traveling U^S. As a consequence, the UB-DODE can be formulated as:

$$\left(\boldsymbol{x}^{1,1}(\boldsymbol{\varpi},\boldsymbol{n}) \dots \boldsymbol{x}^{n_p,n_\theta}(\boldsymbol{\varpi},\boldsymbol{n})\right) =$$

$$= \min_{\boldsymbol{\varpi},\boldsymbol{n}}\left[z_1(\boldsymbol{d}^{1,1} \dots \boldsymbol{d}^{n_p,n_\theta}, \boldsymbol{x}^{1,1}(\boldsymbol{\varpi},\boldsymbol{n}) \dots \boldsymbol{x}^{n_p,n_\theta}(\boldsymbol{\varpi},\boldsymbol{n})) \cdot w_1 + z_2\left(\boldsymbol{f}^{e,1} \dots \boldsymbol{f}^{e,n_\theta}, \hat{\boldsymbol{f}}^1 \dots \hat{\boldsymbol{f}}^{n_\theta}\right) \cdot w_2\right] \tag{8.7a}$$

Subject to:

$$(\boldsymbol{f}^{e,1} \dots \boldsymbol{f}^{e,n_\theta}) = \max_{\boldsymbol{t}}\left(U^s(\boldsymbol{t}(\boldsymbol{\varpi},\boldsymbol{r},\boldsymbol{n}))\right); \tag{8.7b}$$

The standard trip-based approach adopted in this thesis, which we will refer to as Assignment-Based DODE (AB-DODE), is detailed in Equation (8.8):

$$(x^1 \dots x^{n_\theta \cdot n_{od}}) = \min_{\mathbf{x}}\left[z_1(d^1 \dots d^{n_\theta \cdot n_{od}}, x^1 \dots x^{n_\theta \cdot n_{od}}) \cdot w_1 + z_2\left(\boldsymbol{f}^{e,1} \dots \boldsymbol{f}^{e,n_\theta}, \hat{\boldsymbol{f}}^1 \dots \hat{\boldsymbol{f}}^{n_\theta}\right) \cdot w_2\right] \tag{8.8a}$$

Subject to:

$$(\boldsymbol{f}^{e,1} \dots \boldsymbol{f}^{e,n_\theta}) = \boldsymbol{M}(\boldsymbol{x})\boldsymbol{x} \tag{8.8b}$$

Where $\boldsymbol{M}(\boldsymbol{x})$ is the assignment matrix and $\boldsymbol{x}$ the column vector containing all the time-dependent OD flows. While the utility-based approach still relies on the assignment matrix, since has been developed to work with macroscopic DTA, the difference is that an AB-DODE relies *only* on the assignment matrix in the lower level for ensuring consistency between the demand and link flows. By contrast, the Utility-Based formulation is schedule based, i.e. each trip has a specific time schedule.

Assuming that U_n^s is calculated according to Equation (8.3), then the decision variables to be estimated for each OD pair are $\alpha, \beta, \gamma, \tau$ and the total demand N_{od} for the whole analysis period, thus the UB-DODE approach has a number of free parameters equal to $\left[n_{var\_UB-DODE} = n_{od} \cdot 5\right]$, while for the AB-DODE is $\left[n_{var\_AB-DODE} = n_{od} \cdot n_\theta\right]$, where n_θ is the number of time intervals and n_{od} the number of (physical) OD pairs. In general, by defining n_ω the number of variables to estimate for each OD according to the schedule based approach, when $n_\theta > n_\omega$, then $n_{var\_UB\_DODE} < n_{var\_AB\_DODE}$. This observation leads to the following property:

<u>*Property 1-*</u> *If the number of time intervals in which the demand is divided is larger than the number of parameters of the Utility-Based formulation, then the number of variables of the Utility-Based Approach is lower than the one of the Assignment-Based.*

The most important consequence of this property is that the number of parameters n_ω is not dependent on the number of time intervals n_θ. As a consequence, the Utility-Based formulation can be considered a parametric approach, in which the DTM is the parametric relation between different time intervals for the same OD pair. This condition leads to an important advantage. On one hand, as observed in (Marzano et al., 2009), an effective estimation is achievable when the unknown/equation ratio is close to one. On the other hand, as pointed out in (Cascetta et al., 2013), for the UB-DODE, when a longer analysis period is adopted, e.g. the entire day, and shorter simulation time intervals are considered, then the number of equations increases considerably, while the number of parameters to be estimated in our approach stays constant. Consequently, the value of the unknowns/equations ratio increases, and a better estimation is obtained.

Apart from the systematic reduction of the number of free parameters for all cases in which $n_\theta > n_\omega$, two more properties should be highlighted, which are related to the spatial and temporal propagation of the OD flows. More specifically

i) The first property is related to the observation by Frederix et al. (2010), who point out that, if a linear correlation between OD-flows and link flows is assumed, there will be a biased estimation of the gradient of the goal function. We show that this bias is reduced using the UB-DODE approach;
ii) The second property is related to the fact that, if there exists a relationship between time-dependent flows for a specific OD pair, then the solution space of the OD estimation problem is reduced. To demonstrate the latter, we will make use of the Maximum Possible Relative Error (MPRE) metric, originally introduced by Yang et al. (1991).

The consequence of the combined effect of (i) and (ii) is that the AB-DODE formulation properly approximates the total demand, while has a hard time in estimating its temporal and/or spatial distribution.

Equations (8.7) and (8.8) are constrained optimization problems, which can be solved through well-known methods such as the gradient projection method. For a detailed overview we refer to other works (Balakrishna, 2006; Lindveld, 2003), while hereafter we stress that one of the key problems with these approaches is the calculation of the gradient ***Gr***, which can be formulated as the explicit gradient of the objective function (Cascetta, 2009; Cascetta et al., 1993) or as a numerical approximation (Cantelmo et al., 2014; Lu, 2013).

Equation (8.8b) assumes that a linear relation between OD and link flows exists. As a consequence, by assuming an explicit approximation of the gradient, we assume that the link flow on a certain section and during a certain time interval can change always and only by changing the OD flows passing on that link and in that time interval. However, in a dynamic environment this assumption cannot be accepted, since it does not consider basic dynamics such as congestion spillback and flow interactions at the intersections. Numerical approaches (Antoniou et al., 2015; Cantelmo et al., 2014; Frederix et al., 2011) might overcome this issue (Frederix et al., 2013), however there are two drawbacks: firstly, these methods are usually computationally more expensive because of the high number of variables, secondly, for an increasing number of variables, the effect of the numerical perturbation on the goal function decreases.

The gradient numerical approximation for Equation (8.8) can be generally assumed to be, for instance:

$$\boldsymbol{Gr} = \begin{bmatrix} \frac{\partial z(x)}{\partial x^1} \\ \vdots \\ \frac{\partial z(x)}{\partial x^{n_\theta \cdot n_{od}}} \end{bmatrix} \cong \begin{bmatrix} \frac{z(x^1 + c) - z(x^1 - c)}{2c} \\ \vdots \\ \frac{z(x^{n_\theta \cdot n_{od}} + c) - z(x^{n_\theta \cdot n_{od}} - c)}{2c} \end{bmatrix} \quad (8.9)$$

Equation (8.9) properly approximates the real gradient if the numerical perturbation c is small. This perturbation in the demand flow leads to a new route flow, thus a new generalized cost for each link l belonging to that route. This new cost triggers a rerouting effect for all the OD pairs $od \in OD_l$, where OD_l is the set of OD pairs passing for link l. By assuming that the numerical perturbation is a small proportion "ρ" of the overall demand flow, thus $c = \rho \cdot x_{od}^{\theta}$, the magnitude of the effect of the perturbation on network can be expressed as:

$$MAG_{AB\_DODE,(od,\theta)} = \frac{\rho \cdot x_{od}^{\theta}}{\sum_{l \in r} \sum_{\theta \in \Theta} \sum_{od \in OD_l} x_{od}^{\theta}} \quad (8.10)$$

Where ρ is a proportional factor and $l \in r$ is the set of links in the network belonging to route r. Equation (8.10) shows, for one OD and one time interval θ, that the effect of perturbation is inversely proportional to the number of links and OD pairs. In other words, the larger the network, the lower will be the effect of the numerical perturbation on the link flows, which is usually the case for urban networks. As a consequence, the only option to increase the impact of the perturbation is to increase the parameter ρ, which however leads to an incorrect approximation of the gradient in Equation (8.9). On the other hand, by assuming a UB-DODE formulation, the perturbation becomes $c = \rho \cdot \sum_{\theta \in \Theta} \cdot x_{od}^{\theta}$, leading to:

$$MAG_{UT\_DODE(od)} = \frac{\rho \cdot \sum_{\theta \in \Theta} \cdot x_{od}^{\theta}}{\sum_{l \in r} \sum_{\theta \in \Theta} \sum_{od \in OD_l} x_{od}^{\theta}} \quad (8.11)$$

Equation (8.11) shows that the classical bi-level formulation, even under the conditions in which a numerical derivative is assumed, is not able to properly calibrate the *spatial-distribution* – i.e. the distribution of trips over multiple OD pairs – for a given time period when multiple OD flows cross the same link. The main reason is that in Equation (8.10) the time interval is fixed, while in (8.11) we perturb the demand across the entire analysis period. As a consequence, not only the perturbation is larger, but the DTM will provide a different temporal distribution, which affects more time intervals and the overall duration of the rush hour. This property is analysed in more detail in Appendix D.1.

While Equations (8.10)-(8.11) measure the effect of the perturbation c on the entire network, the goal function is sensible only to a subset of observed links, for which traffic counts are available. If the perturbation affects only links that are not within this set, then the change in the link flow does not influence the goal function value. For this reason, link number *and* location have a significant impact on the DODE problem.

Concerning this issue, to measure the error of using the normal formulation, we exploit the MPRE proposed by Yang et al. (1991). This method allows estimating the maximum possible relative deviation between a given estimated matrix and the real one – therefore its *Maximum Possible Relative Error* (MPRE). First, we introduce hereafter the concept of *observability* of a variable, which is used in this thesis:

Definition of observability: We refer to observability of a variable as the probability of that variable of being observed at at least one counting station.

The MPRE can be estimated by solving a quadratic programming problem (See Yang et al. (1991) for more details). One of the main problems with the AB-DODE is that, since time-dependent OD flows are treated as free parameters, the correlation between different OD flows derives only by congestion phenomena. As a consequence, the problem becomes highly non-linear, thus it is very easy to have extremely high values of the MPRE. Specifically, if one variable is not captured at at least one counting station, then the variable is not *observable*. In this case, the MPRE is unbounded and evolves to infinity. In the UB-DODE, the DTM constrains the solution space of the MPRE, since the only feasible solutions are those for which the OD flows are consistent with the departure time choice model. Since this constraint reduces the solution space size, the following property can be formulated:

Property 2- Under the user equilibrium assumption, if a departure time choice model based on the only dis-utility of traveling is considered in the lower level of the demand estimation, and this single function is a concave continuous function, then the MPRE is less than or equal to the case where the departure time is exogenous.

Demonstration of property 2 is provided in Appendix D.2. A second property derives from the OD-coverage rule. Regardless of the estimation model, if one variable is not observed at any counting station, then the MPRE is infinite. If, for instance, an OD pair is observable for n_θ-1 time intervals, then the AB-DODE MPRE goes to infinite. However, if a relation exists between OD flow belonging to different time intervals –i.e. the DTM – then the variable is still observable. According to this observation, the missing time interval can be estimated.

Property 3- By assuming a parametric relation between different time intervals, the demand flows observability increases.

As a consequence, if an OD flow has been observed in *at least* one time interval, then the MPRE is not infinite. This property is analysed in Appendix D.3.

8.3.2 Multiple-purpose: Including activity location and duration

In this section we study the model when including multiple activity locations and duration. Many studies show how the trip/activity scheduling is strongly related to the utility of performing a specific activity at a certain time. As a result, the final activity pattern is a function of travel time, activity duration and the preferred arrival time at destination (Zhang et al., 2005).

Before analysing the consequences of considering the activity duration within the DODE, we relax the assumption imposed in Section 8.2.1, for which the location of all the activities is known. First, we categorize the activities in two classes: *rigid* and *flexible*. Activities such as work and home, where the location is fixed, belong to the first class (*rigid*), while those activities in which each user can choose multiple locations, such as daily shopping, belong to the second one (*flexible*). According to this categorization, the following information is needed for mapping the demand:

a) Activity location for the *rigid* activities, for each user;
b) Candidate locations for the *flexible* activities, assigning to each location the appropriate utility function;

If a sufficient amount of data is available, activity location can be derived through collected data such as travel surveys and/or probe vehicles (Cipriani et al., 2015; Eisenman and List, 2004). Alternatively, they can be estimated within the DTA itself (Fu and Lam, 2014; Polak and Heydecker, 2006; Ramadurai and Ukkusuri, 2011). These approaches can be used to estimate the sequence of *rigid* and *flexible* activities for each user. Then, users can be grouped in *macroscopic* activity patterns ν, based on the activity scheduling and the location of the *rigid* activities:

$$\nu = \{(g^1, \tau^1, \vartheta^1), \dots, (g^P, \tau^P \vartheta^P)\} \tag{8.12}$$

For a group of users, the macroscopic activity pattern ν is the list of all activities P, their locations (or node) g, and the preferred departure times τ. The variable ϑ is a Boolean variable indicating whether the activity is *rigid* or not. Thus, all the users belonging to a macroscopic activity pattern ν, will have the same activity pattern in terms of activity scheduling, constrained to the same location only for *rigid* activities ($\vartheta = 1$). For the flexible activities, the activity location can be modelled as a route choice problem, where users choose jointly route and activity location for which they maximize their own utility according to Equation (1). An equilibrium model that can take into account this type of activity pattern is proposed in (Cantelmo and Viti, 2016). The main consequence is that, instead of estimating the number of users for each OD pair, we can estimate the number of users *for each activity pattern.* The advantage is twofold. On one hand, the number of decision variables decreases with respect to estimating the demand flow for each activity, on the other hand, by considering daily patterns, estimating the demand for a certain activity pattern brings consistency in the system. Given the last location visited during the day g_n^{final}, we can write:

$$\begin{cases} t_n^p \leq t_n^{p+1} & \forall\, n \in N \\ g_n^{origin-1} = g_n^{final} & \forall\, n \in N \end{cases} \tag{8.13}$$

Where t_n^p is the departure time for the trip with purpose *p*, for a certain user *n*, and $g_n^{origin-1}$ is the origin zone for the first trip in the activity pattern. Equation (8.13) assumes that there is conservation of users in the system, and imposes a constraint that each user will return to his/her original location (home) at some point in time. If travel time TT^s and the activity duration for activity p $\Delta t^p = (t_n^{p+1} - t_n^p - TT^s)$ are considered, then Equation (8.13) becomes:

$$\begin{cases} \Delta t^p + t_n^p + TT^s \leq t_n^{p+1} & \forall\, n \in N \\ g_n^{origin-1} = g_n^{final} & \forall\, n \in N \end{cases} \tag{8.14}$$

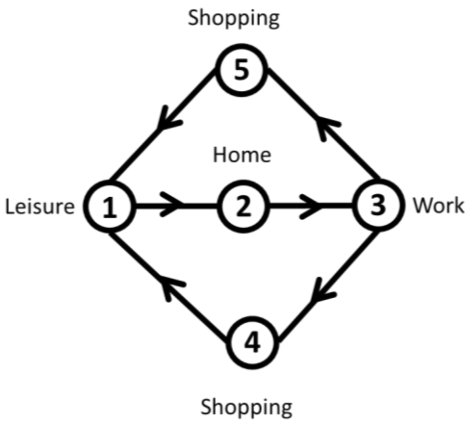
Equation (8.14) can be formulated for the macroscopic activity pattern detailed in Equation (8.12) as:

$$N_{od}^{p,\nu} \geq N_{od}^{p+1,\nu} \cdot \vartheta^{p+1} \qquad p, p+1 \in \nu, \;\; 1 < p < P-1 \tag{8.15}$$

Where, for a given activity pattern ν, $N_{od}^{p,\nu}$ represents the demand for a certain OD, N_{od}^{p+1} the OD demand for the activity *p+1.* Equation (8.15) shows that if both activities at origin and destination for a certain OD pair are rigid, for a certain activity pattern ν, then the OD flow from to the next activity location(s) is less than or equal to $N_{od}^{p,\nu}$. As a consequence, the number of unknowns further decreases, resulting in a more consistent estimation of the demand flows. Based on this discussion, we can now extend Property 3 into the following property 4:

<u>Property 4-</u> *By assuming a parametric relation between different time intervals, if activity duration is estimated within the DTA, the demand flow observability increases with respect to a trip based approach.*

This property depends on Equation (8.15), which creates a relation between demand flows belonging to the same activity pattern. If we can observe this demand in at least one trip, then the MPRE for all the OD pairs belonging to that activity pattern is not infinite. This consideration is discussed in Appendix C, while Figure 8.2 shows intuitively an example of the macroscopic activity pattern {Work-Shopping-Leisure-Home} on a simple network, for a demand of $N^\nu = 100$ users.



$$\nu = \{\boldsymbol{g}, \boldsymbol{\tau}, \boldsymbol{\vartheta}\}$$

$$\boldsymbol{\tau} = \begin{Bmatrix} 8 \\ 17 \\ 18 \\ 19 \end{Bmatrix} \quad \boldsymbol{g} = \begin{Bmatrix} 3 \\ 0 \\ 1 \\ 2 \end{Bmatrix} \quad \boldsymbol{\vartheta} = \begin{Bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{Bmatrix}$$

$\underline{1)\ Calculate\ N_{od}^{p}}$

$$N_{2-3}^{Work} = \sum_t x_{2-3}^t = 100$$
$$N_{3-4}^{Shop} = \sum_t x_{3-4}^t = 50$$
$$N_{3-5}^{Shop} = \sum_t x_{3-5}^t = 50$$
$$N_{5-1}^{Leis} = \sum_t x_{5-1}^t = 50$$
$$N_{4-1}^{Leis} = \sum_t x_{5-1}^t = 50$$
$$N_{1-2}^{Home} = \sum_t x_{1-2}^t = 100$$

$\underline{2)\ Apply\ Equation\ 18}$

$$N_{2-3}^{Work} \geq N_{3-4}^{Shop}(0) \;\rightarrow\; 100 \geq 0$$
$$N_{2-3}^{Work} \geq N_{3-5}^{Shop}(0) \;\rightarrow\; 100 \geq 0$$
$$N_{3-4}^{Shop} \geq N_{4-1}^{Leis}(1) \;\rightarrow\; 50 \geq 50$$
$$N_{3-5}^{Shop} \geq N_{5-1}^{Leis}(1) \;\rightarrow\; 50 \geq 50$$
$$N_{5-1}^{Leis} \geq N_{1-2}^{Home}(1) \;\rightarrow\; 100 \geq 100$$

Fig.8.2: Example of consistency for the activity Demand

In Figure (8.2), we first define the macroscopic activity pattern v, which is composed of the set of preferred departure times $\boldsymbol{\tau}$ for each activity, the geographical information $\boldsymbol{g}$ for the *rigid* activities and the variable $\boldsymbol{\vartheta}$**.** Then, by taking all the OD flows for each time interval together, we calculate the value of N_{od}^{P} for each "spatial" OD and each activity. Clearly all the 100 users will visit those locations for which $\vartheta = 1$, while they will choose one location among the possible candidates when $\vartheta = 0$. In this symmetric network, the demand for the "shopping" activity is equally distributed among nodes 4 and 5. Finally, we can verify that, when the macroscopic activity pattern is formulated as in Equation (8.12), Equation (8.15) is satisfied for all OD pairs.

8.4 Numerical analysis

8.4.1 Solution algorithm

In this section, two test networks are used to illustrate the properties of the proposed UB-DODE model. The DTA model adopted is the Iterative Link Transmission Model (I-LTM) introduced in (Himpe et al., 2016; Tampère et al., 2011), which allows a realistic network loading which properly reproduces complex traffic dynamics such as congestion spillback. While the AB-DODE assumes exogenous departure times, i.e. each OD pair is a free parameter to be calibrated, the Utility-Based formulation implements the DTM presented in Section 8.2.1 within the DTA, which is solved through the well-known Method of Successive Averages (MSA), as described in (Cantelmo and Viti, 2016; Li et al., 2014).

Step 0: Define an initial departure flow value $\boldsymbol{x}^i$ by setting $x_{od}^{\tau} = N_{od} \;\; \forall\, od$, and set the iteration number i=1;

Step 1: Load the demand on the network, obtaining the travel times TT_{od};

Step 2: Calculate the Utility $U^n = (U^s + U^p)\ \forall\, n$;

Step 3: identify the time interval θ^*in which U^n is maximum, and estimate the auxiliary flows $\ddot{\boldsymbol{x}}_{od}^{\theta^*}$ by setting $\ddot{x}_{od}^{\theta^*} = N_{od} \;\; \forall\, od$

Step 4: Update the solution through the MSA algorithm as $\boldsymbol{x}^{ite+1} = \boldsymbol{x}^{ite} + \frac{1}{ite+1}\left(\ddot{\boldsymbol{x}}_{od}^{\theta^*} - \boldsymbol{x}^{ite}\right)$

Step 5: Return to Step 1;

Different solution algorithms have been proposed to tackle the different task of estimating the demand from the traffic counts. In this work we apply an iterative path-search method for solving the DODE.

Specifically, we apply a Gradient Descent method, which assumes that, if the multi-variable objective function(s) (8.6a)-(8.7a)-(8.8a) are defined and differentiable, then the goal function decreases in the direction of the gradient:

$$\boldsymbol{Z}_{i+1} = \boldsymbol{Z}_i - \Lambda_i \boldsymbol{Gr}_i \tag{8.15}$$

Where $\boldsymbol{Z}_i$ is the set of variables to be updated at iteration i, Λ_i is the step size and $\boldsymbol{Gr}_i$ is the gradient. As pointed out in Section 8.3.1, the gradient can be obtained through an explicit calculation of the gradient of the objective function or as a numerical approximation. While the first one is very popular when only link-flows/demand-flows are used in the goal function, and when the dynamic assignment matrix $\boldsymbol{M}$ is known, it may lead to biased estimations, as discussed in section 8.3.1 of this work and in (Frederix et al., 2013). Thus we get its numerical approximation by applying two different approaches, the FDSA and the SPSA, which have been widely adopted in the literature.

The FDSA (Finite Difference Stochastic Approximation, Kiefer and Wolfowitz (1952)) is a method usually adopted when there is stochasticity in the measurements. It obtains the descent direction by perturbing every variable in $\boldsymbol{Z}_i$: The gradient is obtained as follows

$$\boldsymbol{G}^i(\boldsymbol{Z}_i) = \begin{bmatrix} \dfrac{z(\boldsymbol{Z}_i + c^i \boldsymbol{\xi}^1) - z(\boldsymbol{Z}_i - c^i \boldsymbol{\xi}^1)}{c^i} \\ \vdots \\ \dfrac{z(\boldsymbol{Z}_i + c^i \boldsymbol{\xi}^{\boldsymbol{n_v}}) - z(\boldsymbol{Z}_i - c^i \boldsymbol{\xi}^{\boldsymbol{n_v}})}{c^i} \end{bmatrix} \tag{8.16}$$

where ξ is the vector with all zeros, except for the variable to be perturbed, c^i is the perturbation and n_v the number of variables to be perturbed. In this method each variable is perturbed independently, so the number of simulations required for computing the gradient in any iteration is equal to the number variables times two. For big-sized networks, if each OD pair is assumed to be an independent variable to be calibrated, this approach becomes unfeasible. As a consequence, for many real-life applications, the SPSA (Simultaneous Perturbation Stochastic Approximation, Spall, 2012) is usually adopted. The SPSA is a stochastic approximation of the FDSA, formulated as follows:

$$\widehat{\boldsymbol{g}}_k(\boldsymbol{Z}_i) = \frac{z(\boldsymbol{Z}_i + c^i \Delta^{\boldsymbol{j}}) - z(\boldsymbol{Z}_i - c^i \Delta^j)}{c^i} \begin{bmatrix} (\Delta_1^j) \\ \vdots \\ (\Delta_{n_v}^k) \end{bmatrix} \tag{8.17a}$$

$$\boldsymbol{G}^{\boldsymbol{i}} = \overline{\boldsymbol{g}}(\boldsymbol{Z}_i) = \frac{\sum_{\boldsymbol{j=1}}^{\boldsymbol{J}} \widehat{\boldsymbol{g}}_j(\boldsymbol{Z}_i)}{Grad\_rep} \tag{8.17b}$$

With J the number of the replications to compute the average gradient and Δ is a vector with elements {-1,1}. The SPSA perturbs all the variables at the same time, thus decreasing the number of simulations required. Since one single perturbation leads, usually, to a wrong approximation of the real gradient, J stochastic approximations are performed, and then the real gradient is calculated of their average, as detailed in Equation (8.17b).

The step Λ in equation (8.15) is assumed to decrease at each iteration according to the rules detailed in (Spall, 2012). For the Utility-Based approach, we use three different values, specifically one for updating the number of trips $\boldsymbol{N}$, a second value for updating the departure time $\boldsymbol{\tau}$ and a third one for the parameters of the utility functions. Lastly, the demand weight term w_1 is settled equal to 0, so that only the link flow error is considered in the optimization.

The function used to assess the estimation results is the Root Mean Square Normalized Error (RMSN) metric:

$$RMSN = \frac{\sqrt{\chi \sum_{i=1}^{\chi} (f_i^e - \hat{f}_i)^2}}{\sum_{i=1}^{\chi} \hat{f}_i} \quad (8.18)$$

Where χ is the number of measured variables.

8.4.2 Numerical results

8.4.2.1 Utility-Based Demand Estimation – Single destination case (Trip-Based)

The first experiment is performed on the network shown in Figure 8.3, which is composed of 5 nodes and 4 links. Traffic counting is done only on link 3. The network, whose details are reported in Figure 3, consists of a two-lane motorway with two origins, i.e. node 1 and node 2, and one common destination located in node 5. This network is similar to the one analysed in (Frederix et al., 2013). However, in this case the network is not symmetric, i.e. the free flow travel time from 1 to 5 is higher than the one from 2 to 5, since link 1 is two times longer than link 2. By applying the FDSA method, the UB-DODE formulated in 4.1 is tested on this network through two experiments:

Scenario I. Right temporal distribution, wrong spatial distribution: The seed matrix has the correct temporal distribution with respect to the real matrix to be estimated, while the spatial distribution is biased. Specifically, the demand from node 2 is 1.5 times the real one, while from node 1 is 0.5 the real demand flow.

Scenario II. Wrong temporal distribution and spatial distribution: The spatial distribution keeps the same error as in Scenario (I), while considering a second perturbation to the temporal distribution.

Link ID	*Length (Km)*	*free Speed (Km/h)*	*Capacity (Veh/h)*	*Jam Density (Veh/Km)*
1	*120*	*120*	*5000*	*300*
2	*60*	*120*	*5000*	*300*
3	*60*	*120*	*7500*	*450*
4	*60*	*120*	*4000*	*240*

Network characteristics

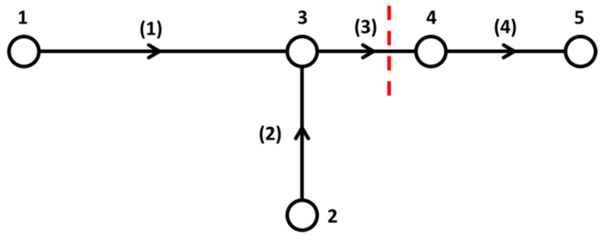


Figure 8.3: Test network and network characteristics for the trip-based case. Nodes 1 and 2 are origins, Node 5 is destination. The counting station is located on link 3;

Scenarios I and II are solved using both the UB-DODE and the AB-DODE. However, the starting matrix for the two scenarios is different. In the case of the UB-DODE, to have the correct temporal distribution we assume that the DTM parameters are correct. That leads to have a realistic departure time choice model, thus a correct temporal distribution. Even if DTM parameters are correct, they are still variables of the optimization model –i.e. they are updated in order to match the observations. If we assume an exogenous departure time, we assume that the percentage of users leaving in a certain time interval is correct, but not their number. As a consequence, the generated seed matrix is not the same. Similarly, in Scenario II we strongly perturb the real solution, thus we assume wrong demand and wrong DTM parameters for the UB-DODE. For the AB-DODE, we assume that the demand is uniformly distributed over different time intervals.

Figures 8.4-8.5 introduce the results. They show clearly that, both for Scenarios I and II, according to the goal function, the UB-DODE finds a solution that fits well the observed link flows (although not as well as the AB-DODE) (Figure 8.4-8.5e, 8.4-8.5f), while at the same time moving closer to the real solution with respect to the AB-DODE case. Furthermore, as expected, when a more reliable seed matrix is available, i.e. in the case of Scenario I, then a better estimation is achieved both in terms of OD flows and of link flows. Scenario I shows how the concept of temporal distribution is completely different when applied to AB-DODE or UB-DODE. In the first case, the two OD pairs overestimate or underestimate the demand in each time interval. In the second case, when we increase/decrease the demand, we have a longer/shorter period in which the demand flow for that OD pair has high values (Figure 8.4b-8.4d). This depends on the properties of the dis-utility function presented in Equation (3). According to the DTM, the demand will spread around the preferred departure time. However, according to Equation (8.3), and in general to the bottleneck model, for a given capacity i.e. 5000 veh/h, the DTM suggests the users to anticipate or postpone the trip, rather than starting the trip when the network is already congested. For this reason, in the UB-DODE we have more time intervals in which the demand flow is close to the capacity, but it is not possible to observe time intervals in which this value is too high with respect to the road capacity, differently from the AB-DODE, in which the maximum demand is around 8000 veh/h. In the standard bottleneck model, by increasing the number of users, we increase the length of the congestion period; since the length of link 1 is two times the length of link 2, while the departure time is the same, there is a first time interval at 8.5 h in which only the demand from 2 is loaded on the network, and a last time interval at 9.25 h in which only the demand from 1 is observable. By identifying this error, the UB-DODE identifies the right demand. By contrast, the AB-DODE is not able to identify this problem, estimating a solution, which is completely biased for one OD (Figure 8.4a), while keeping constant the other one (Figure 8.4c).

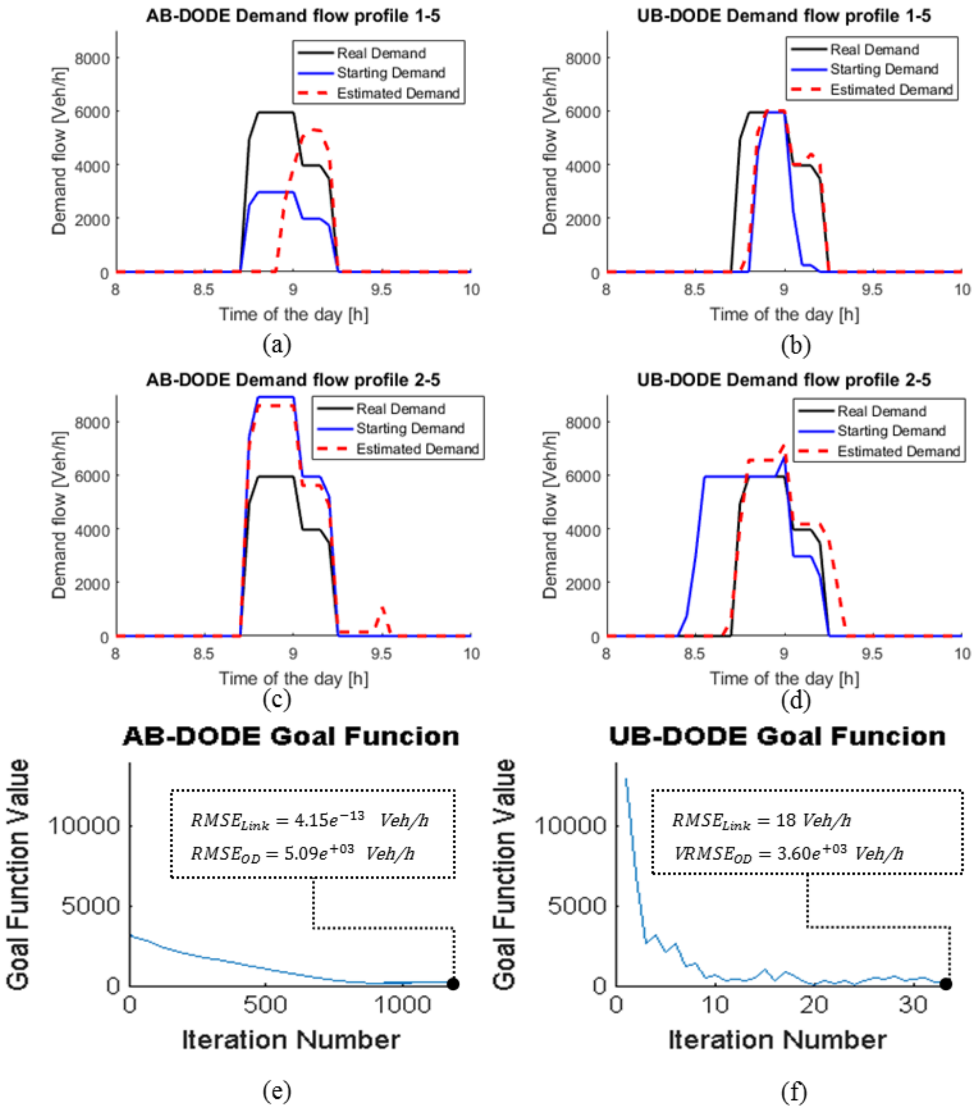


Fig.8.4: Results for Scenario I– (a-b) Demand Profile from 1 to 5; (c-d) Demand Profile from 2 to 5; (e-f) Goal Function trend with respect to the iteration number;

In Scenario II (Figure 8.5), although the demand profile is far from the real one, AB-DODE performs better than in the previous case, offering a reasonable estimation of the demand in the first time intervals for one OD (Figure 8.5b) and in the last time intervals for the other one (Figure 8.5c). The reason is, again, that in these time intervals the correlation between the two OD flows is limited because of the difference in the travel time. However, in the other time intervals the demand estimation is not able to improve the starting situation. On the other hand, the UB-DODE is able to estimate a reasonable solution, although not as precise as in the previous case. It should also be pointed out that the error on the link flows is still higher for the UB-DODE, indicating that the goal function is more representative of the error between real and estimated OD flows. A last comment considers the computational time: it looks clear that the number of iterations needed for the UB-DODE is smaller with respect to the AB-DODE. This is related to the fact that we have less variables and a relatively smoother problem. However, when the DTM is combined with I-LTM, and solved through the MSA algorithm, the single simulation becomes up to 100 times more expensive. By considering this issue, together with the reduction in the number of variables and number of iterations, the computational time for the AB-DODE and UB-DODE is respectively of 28 and 18 hours. The reason is that, again, FDSA is an effective but not efficient algorithm for this problem. Scenario I and II numerically show Properties 1, 2 and 3 to hold. Specifically, we show that UB-DODE has a lower number of variables (10 for the UB-DODE and

42 for the AB-DODE), and that the proposed formulation is more sensitive with respect to errors in the spatial distribution. While Frederix et al. (2013) identified that, already on symmetric toy networks, conditions in which AB-DODE provides a proper estimation are tight and usually unrealistic for applications on real-sized networks, the proposed Utility-based formulation, instead, by taking into account the travel time in the utility function, is able to move closer to the real demand matrix in more relaxed conditions.

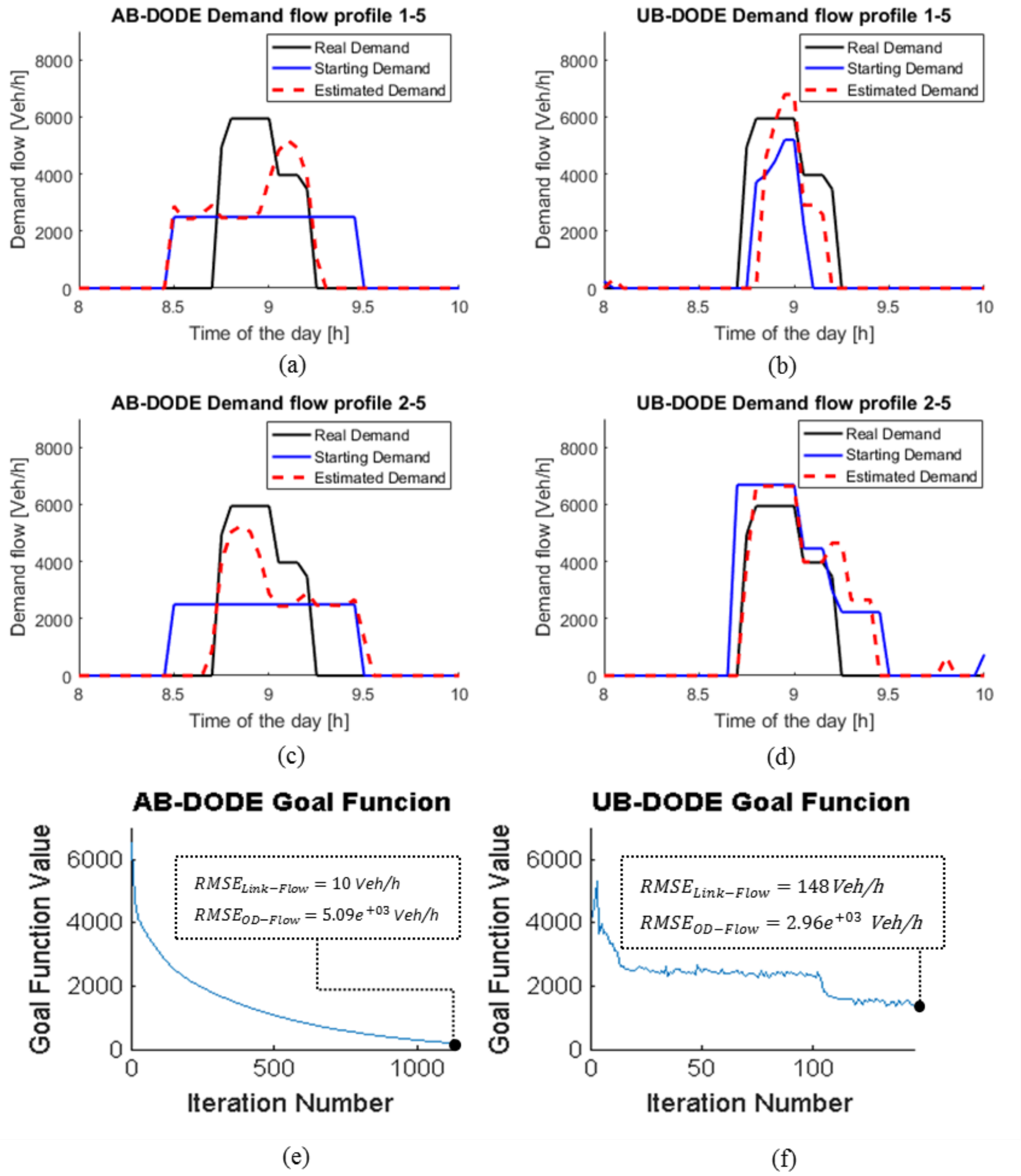


Fig.8.5: Results for Scenario I– (a-b) Demand Profile from 1 to 5; (c-d) Demand Profile from 2 to 5; (e-f) Goal Function trend with respect to the iteration number;

8.4.2.2 Utility-Based Demand Estimation – Including activity location and duration

The second network, showed in Figure 8.6, presents a more comprehensive case, and it is used to show the improvement related to considering the activity duration as formulated in Section 3.2. The network has five OD pairs {1-4, 4-5, 4-6, 5-1, 6-1}. Links 1 and 5 are connectors, thus link jam density and capacity are infinite. Traffic counting is done only on link 3. Nodes 1 and 4 represent, respectively, home and work locations, while secondary activities, such as fitness activity, are located on nodes 5 and 6. We have one macroscopic activity pattern {work-leisure-home} on this network, where home and work are rigid activities. For this network the secondary activity has constant duration, equal to one hour, and the same (constant) utility. Under this assumption, for both the AB-DODE and UB-DODE the activity location problem turns out to be equivalent to a route choice problem, where two alternative routes, connecting nodes 4 and 1, have the same cost. In this case, the two models have the same starting value of the demand. Utility of performing the activity work is, instead, calculated through equation 8.19:

$$U_{n,p}^{a}\left(t_n^{start}, t_n^{end}\right) = \int_{t_n^{start}+1}^{t_n^{end}} U_n^p(t) \left(\frac{1}{(t - t_n^{start})}\right)^{0.5} dt \tag{8.19}$$

Where t_n^{start} and t_n^{end} are the starting and ending time for the activity "work", $U_n^p(t)$ is the *bell-shaped* time-dependent utility function proposed in (Ettema and Timmermans, 2003), while the second term include the duration – or *fatigue* – effect. For more details on this function, we refer to (Cantelmo and Viti, 2016). Results of this experiment are presented in Figure 7.

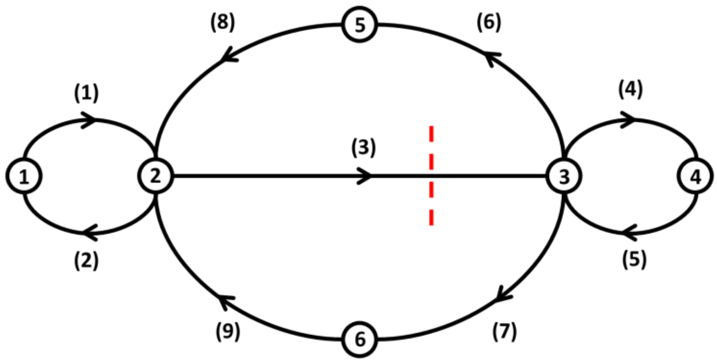


Fig.8.6: Test Network for the activity-based approach. The counting station is located on link 3;

The analysis period covers 24 hours with a simulation time interval of 15 minutes. The AB-DODE uses the SPSA since, already for this simple toy-network, the number of variables is too high for using FDSA. The Utility-based approach has been tested instead with both FDSA and SPSA. As expected, the two models provide almost the same result, since SPSA is assumed to properly approximate the FDSA for small networks (Spall, 2012). However, for the sake of a better readability, in this Chapter we show the results for the FDSA. The main reason is to avoid issues related to the stochastic component of the SPSA when analysing UB-DODE results.

As shown in Figure 8.7, the Utility-based estimation clearly outperforms the standard AB-DODE in this example. The main reason is a better consistency in the demand pattern. The extremely accurate estimation is fully explained by Equations (8.14)-(8.15), which impose that the same users observed in the morning will also return home, thus $\mathrm{N}^{\mathrm{Home}} = \sum_{\mathrm{od}} \mathrm{N}_{\mathrm{od}}^{\mathrm{Leisure}} = \mathrm{N}^{\mathrm{Work}}$. However, the link flow on the links connecting nodes 5 and 6 are unobserved. The Utility-based formulation observes this flow in the morning, when estimating the demand from Node 1 to 4. Figure 8.7a intuitively shows this issue. The AB-DODE partially corrects the demand morning commute, as long as the correction leads to an improvement in the goal function. However, it is not able to correct the evening commute, since no detector intercepts that demand. As consequence, technically, the maximum error between simulated and estimated demand in the evening is infinite. On the other hand, the UB-DODE properly estimated both morning and evening commute because of the relation detailed in Equation (8.15). However, a significant error between estimated and observed traffic counts is observable (Figure 8.7b-8.7c).

Lastly, while in Section 8.3.1 we stressed that there is a consistency in the demand N, the estimation updated also the values of the parameters {α, β, γ} and the departure time in the evening commute. Since the utility of performing the activities home and work are modelled through the realistic utility function presented in Equation (8.19), which takes into account activity scheduling and duration, by changing the disutility of travelling in the afternoon we observe a different behaviour in the morning, leading to a different link flow on link 3.

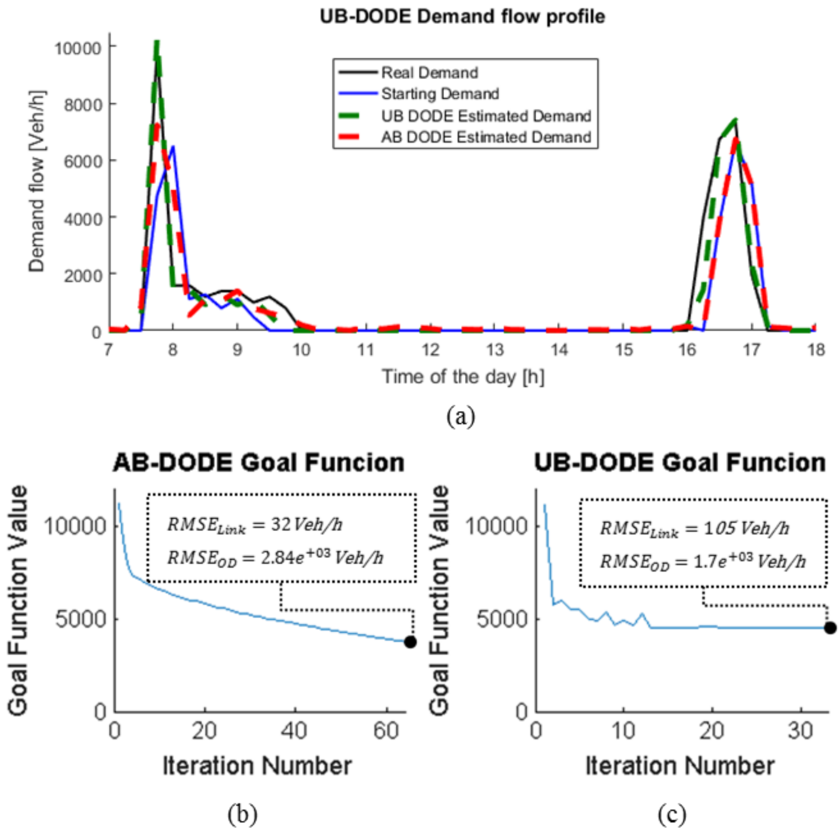


Fig.8.7: Results for Scenario II– (a) Demand Profile for morning and evening commute; (b-c) Scatter Simulated vs Real link flows; (e-f) Goal Function trend wrt the iteration number;

8.5 Conclusions

In this Chapter, the authors presented a Utility-Based formulation for the demand estimation, which is able of incorporating activity scheduling and duration. The building block of this methodology is to adopt a utility-based departure time choice model in the DTA, and to exploit this model to derive the temporal distribution of the demand. Since most of the state-of-the-art DTA models do not include activity location, properties of the model have been evaluated in two conditions: (i) when no knowledge about activity purpose and location is available – i.e. the standard trip-based representation, and (ii) when activity location and characteristic are considered. Mathematical properties of the model are presented and tested on simple networks with a realistic trip-based macroscopic DTA model, showing that the model can be implemented in practice. Hereafter, the main properties of the model are listed:

I. The utility-based formulation has, usually, a lower number of variables. Since the number of variables does not depend on the number of simulation time intervals. This is particularly true when the analysis period is large enough;
II. The utility-based formulation is more sensitive to the spatial distribution of the demand, since the derivative of one specific variable influences several time intervals, generating a meaningful perturbation on the network flows;

III. Since the demand over different time intervals is highly correlated, the proposed model increases the demand flow observability over time;
IV. If activity duration is considered in the problem, since demand over different trips is correlated, demand observability increases over both time and space.

Most of these properties derive from the adopted departure time choice model, which bridges the Goal Function at the upper level and the DTA at the lower level. While usually the DTA model is simply a constraint, in this case the temporal distribution is estimated within the DTA itself. Numerical examples provide additional insight into the model performances. First, when a gradient approach is used, this model is extremely sensitive with respect to the step size. While the standard approach estimates demand flows, the proposed formulation estimated demand flows *and* departure time choice model parameters at the same time, meaning that different step sizes need to be considered. Thus, for more realistic implementations, a line search should be included in the solution algorithm, improving the quality of the result while reducing the computational time. Second, the last experiment shows that, when duration is included in the problem, if one OD flow is well calibrated, then the overall demand for the entire day/activity pattern can be properly estimated. This is extremely relevant for implementations on real networks, where due to topological reasons few OD flows are easier to be estimated than the others.

Theoretically, the approach could be used for both online and offline demand estimation. However, as other approaches, it is extremely demanding in terms of computational times. Thus, is more suited for off-line calibration. However, this model has been developed in order to estimate the systematic component of the demand, thus a reasonable option is to use this methodology to create a good dynamic matrix, to be corrected with other approaches such as Kalman Filter or Sequential GLS. Lastly, the scalability of the model depends on the assumptions on the DTM. The trip-based version of the UB-DODE presented in this Chapter can be already implemented in practice. The only needed information for implementing this model is the OD-travel time, while Property 1 shows that, although the number of parameters increase with the number of ODs, it is constant with respect to the number of time intervals. The tour-based version can be implemented only with those DTA that can model macroscopic activity patterns, such as the one proposed in (Ramadurai and Ukkusuri, 2011).

Future research will focus on extending and to strengthening the findings presented in this work. On one hand, the first challenge is to test the Utility-based formulation for big sized networks, in order to evaluate the mathematical properties under more general conditions. A second difficult challenge is to properly map activity location for macroscopic activity patterns, in order to implement the multiple-purpose version on generic networks with first-order macroscopic DTA. Once these two points have to be solved, the effect of considering different information, such as probe vehicles data, which carry a lot of information regarding activity location, should be investigated. Lastly, since the proposed formulation is utility-based i.e. based on the perceived utility rather than traffic performances, an interesting direction is the multi-modal demand estimation.

Acknowledgements

We acknowledge for financing the following grant: – AFR-PhD grant 6947587 IDEAS.

8.6 References

Adnan, Muhammad. 2010. "Linking Macro-Level Dynamic Network Loading Models with Scheduling of Individual's Daily Activity–Travel Pattern." In *New Developments in Transport Planning: Advances in Dynamic Traffic Assignment*, 251–74.

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. doi:10.1016/j.trc.2015.08.009.

Arnott, Richard, André de Palma, and Robin Lindsey. 1990. "Economics of a Bottleneck." *Journal of Urban Economics* 27 (1): 111–30. doi:10.1016/0094-1190(90)90028-L.

Ben-Akiva, M.E., and S.R. Lerman. 1986. "Discrete Choice Analysis." *MIT Press*. https://mitpress.mit.edu/books/discrete-choice-analysis.

Cantelmo, Guido, and Francesco Viti. 2016. "Effects of Incorporating Activity Duration and Scheduling Utility on the Equilibrium-Based Dynamic Traffic Assignment," June.

Cascetta, Ennio. 1984. "Estimation of Trip Matrices from Traffic Counts and Survey Data: A Generalized Least Squares Estimator." *Transportation Research Part B: Methodological* 18 (4–5): 289–99. doi:10.1016/0191-2615(84)90012-2.

Cascetta, Ennio, Domenico Inaudi, and Gerald Marquis. 1993. "Dynamic Estimators of Origin-Destination Matrices Using Traffic Counts." *Transportation Science* 27 (4): 363–73.

Cipriani, Ernesto, Marialisa Nigro, Gaetano Fusco, and Chiara Colombaroni. 2013. "Effectiveness of Link and Path Information on Simultaneous Adjustment of Dynamic O-D Demand Matrix." *European Transport Research Review* 6 (2): 139–48. doi:10.1007/s12544-013-0115-z.

Cremer, M., and H. Keller. 1984. "A SYSTEMS DYNAMICS APPROACH TO THE ESTIMATION OF ENTRY AND EXIT O-D FLOWS." In *Proceedings of 9th ISTTT*.

Ettema, D. F., F. Bastin, J. Polak, and O. Ashiru. 2007. "Modelling the Joint Choice of Activity Timing and Duration." *Transportation Research A* 41 (9): 827–41.

Ettema, Dick, and Harry 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. doi:10.3141/1831-05.

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

Flötteröd, Gunnar, Yu Chen, and Kai Nagel. 2012. "Behavioral Calibration and Analysis of a Large-Scale Travel Microsimulation." *Networks and Spatial Economics* 12 (4): 481–502.

Frederix, R., F. Viti, R. Corthout, and C.M.J. Tampère. 2011. "New Gradient Approximation Method for Dynamic Origin-Destination Matrix Estimation on Congested Networks." *Transportation Research Record*, Transportation Research Record, , no. 2263: 19–25.

Frederix, Rodric, Francesco Viti, and M.J. Tampère. 2013. "Dynamic Origin–destination Estimation in Congested Networks: Theoretical Findings and Implications in Practice." *Transportmetrica A: Transport Science* 9 (6): 494–513. doi:10.1080/18128602.2011.619587.

Li, Zhi-Chun, William H. K. Lam, and S. C. Wong. 2014. "Bottleneck Model Revisited: An Activity-Based Perspective." *Transportation Research Part B: Methodological* 68 (October): 262–87. doi:10.1016/j.trb.2014.06.013.

Lu, Lu. 2013. "W-SPSA : An Efficient Stochastic Approximation Algorithm for the off-Line Calibration of Dynamic Traffic Assignment Models." Thesis, Massachusetts Institute of Technology. http://dspace.mit.edu/handle/1721.1/88395.

Maher, Michael J., Xiaoyan Zhang, and Dirck Van Vliet. 2001. "A Bi-Level Programming Approach for Trip Matrix Estimation and Traffic Control Problems with Stochastic User Equilibrium Link

Flows." *Transportation Research Part B: Methodological* 35 (1): 23–40. doi:10.1016/S0191-2615(00)00017-5.

Mahmassani, Hani, and Robert Herman. 1984. "Dynamic User Equilibrium Departure Time and Route Choice on Idealized Traffic Arterials." *Transportation Science* 18 (4): 362–384. doi:10.1287/trsc.18.4.362.

Marzano, V., A. Papola, and F. Simonelli. 2009. "Limits and Perspectives of Effective O-D Matrix Correction Using Traffic Counts." *Transportation Research Part C: Emerging Technologies* 17 (2): 120–32. doi:10.1016/j.trc.2008.09.001.

Simonelli, Fulvio, Vittorio Marzano, Andrea Papola, and Iolanda Vitiello. 2012. "A Network Sensor Location Procedure Accounting for O–d Matrix Estimate Variability." *Transportation Research Part B: Methodological* 46 (10): 1624–38. doi:10.1016/j.trb.2012.08.007.

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

———. 2015. "The Bottleneck Model: An Assessment and Interpretation." *Economics of Transportation*, Special Issue on Collective Contributions in the Honor of Richard Arnott, 4 (1–2): 110–17. doi:10.1016/j.ecotra.2015.01.001.

Tympakianaki, Athina, Haris N. Koutsopoulos, and Erik Jenelius. 2015. "C-SPSA: Cluster-Wise Simultaneous Perturbation Stochastic Approximation Algorithm and Its Application to Dynamic Origin–destination Matrix Estimation." *Transportation Research Part C: Emerging Technologies*, Engineering and Applied Sciences Optimization (OPT-i) - Professor Matthew G. Karlaftis Memorial Issue, 55 (June): 231–45. doi:10.1016/j.trc.2015.01.016.

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

Viti, Francesco, Wouter Verbeke, and Chris Tampère. 2008. "Sensor Locations for Reliable Travel Time Prediction and Dynamic Management of Traffic Networks." *Transportation Research Record: Journal of the Transportation Research Board* 2049 (September): 103–10. doi:10.3141/2049-12.

Yamamoto, Toshiyuki, Satoshi Fujii, Ryuichi Kitamura, and Hiroshi Yoshida. 2000. "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. doi:10.3141/1725-13.

Yang, H., Y. Iida, and T. Sasaki. 1991. "An Analysis of the Reliability of an Origin-Destination Trip Matrix Estimated from Traffic Counts." *Transportation Research Part B* 25 (5): 351–63. doi:10.1016/0191-2615(91)90028-H.

Yang, Hai, and Jing Zhou. 1998. "Optimal Traffic Counting Locations for Origin–destination Matrix Estimation." *Transportation Research Part B: Methodological* 32 (2): 109–26. doi:10.1016/S0191-2615(97)00016-7.

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. doi:10.1016/j.tra.2004.04.005.

Zhou, Xuesong, Chung-Cheng Lu, and Kuilin Zhang. 2012. "Dynamic Origin–destination Demand Flow Estimation under Congested Traffic Conditions." In *Proceedings of the 91st Transportation Research Board Annual Meeting, Washington, DC*. http://www.civil.utah.edu/~zhou/Single-level_path-flow-OD-estimation_Zhou-Lu-Zhang.pdf.

Zockaie, A., M. Saberi, H.S. Mahmassani, L. Jiang, A. Frei, and T. Hou. 2015. “Activity-Based Model with Dynamic Traffic Assignment and Consideration of Heterogeneous User Preferences and Reliability Valuation: Application to Toll Revenue Forecasting in Chicago, Illinois.” *Transportation Research Record* 2493: 78–87. doi:10.3141/2493-09.

9

Utility-Based OD Estimation for general networks

The previous Chapter introduced the concept of Utility-Based OD Estimation, proved its properties and validated them on toy networks. In this Chapter, we show that the model can, in principle, be applied to a general network and DTA model. To achieve this goal, we test model performances by mean of a synthetic experiment on the network of Luxembourg City. The adopted DTA model is PTV-Visum, which is a widely established commercial software for Traffic Analysis, showing that the model might be applied in practice,

The content of this chapter has been presented in the following works:

Cantelmo, Guido, and Francesco Viti. 2018. "Evaluating the reliability of the Utility-Based Dynamic OD Estimation on Large Networks" *In 7th International Conference on Transport Network Reliability (Sydney, 17-19 January, 2018)*

Cantelmo, Guido, and Francesco Viti. 2017. "Assessing the applicability of the Utility-based Dynamic Demand Estimation on large, real Networks" *In 6th Symposium of the European Association for Research in Transportation -September 12-14, Haifa (Israel)*

9.1 Introduction

Simulation of traffic conditions requires accurate knowledge of the travel demand. In a dynamic context, this entails estimating time-dependent demand matrices, which are a discretised representation of the dynamic origin-destination (OD) flows. This problem, referred to as Dynamic OD Estimation (DODE) in literature, seeks for the best possible approximation of OD flows, which minimises the error between simulated and available traffic data. Traditional DODE models solve two optimisation problems, according to a bi-level formulation: the upper level updates the time-dependent OD flows, while in the lower level a dynamic traffic assignment model ensures consistency between demand and supply models.

Since DODE problems are usually underdetermined because of the high number of unknown variables (Marzano, Papola, and Simonelli 2009), researchers have dealt with the critical issue of decreasing the number of decision variables in order to (i) obtain a smooth approximation of objective function (Djukic et al. 2012) and (ii) to reduce the overall computational time (Cipriani et al. 2011). Additionally, issues have been addressed, among others, to the nonlinear relation between link and demand flows (Zhou, Lu, and Zhang 2012), pointing out how having a reliable a-priori knowledge of the demand (a-priori seed matrix) is of paramount relevance in order to achieve a satisfactory outcome. On this point, Zhou and Mahmassani (Zhou and Mahmassani 2007) highlight that, in order to provide a robust and reliable estimation, the demand should be considered as a convolution of three functional components: the "regular pattern", the "structural deviation" and the "random fluctuation". The regular pattern can be considered as the systematic component of the demand, the structural deviation is the influence of the specific conditions for which we are estimating/updating the OD matrix (weather conditions, road works,…) and, finally, the random component takes into account the random fluctuations of the demand. Since having a reliable knowledge of the "a-priori seed matrix" is equivalent to know the systematic component of the demand – or regular pattern - we can observe that the overall reliability of the DODE depends on how accurate the knowledge of this component is.

The contribution of this chapter is threefold. First, we test the Utility Based on the real network of Luxembourg city, in order to assess its applicability to real life scenarios. Then, we assess the reliability of this methodology. Finally, in order to increase the reliability of the model, we propose a modification of the SPSA algorithm. By imposing a soft constraint to the research space of the model, we systematically increase results reliability in terms of how likely we are to estimate the "regular patter" of the OD matrix.

9.2 Methodology

While for a detailed overview of the model the interested reader can refer to the previous chapter, in this section, before introducing the proposed extension of the SPSA, we briefly recall the main features of the model. The main difference with respect to the standard DODE formulation is in the lower level. We include within lower level DTA procedure a Departure Time Choice (DTC) model that performs the equilibrium through the utility maximisation theory, as proposed in (Li, Lam, and Wong 2014). The advantage of using this approach is twofold.

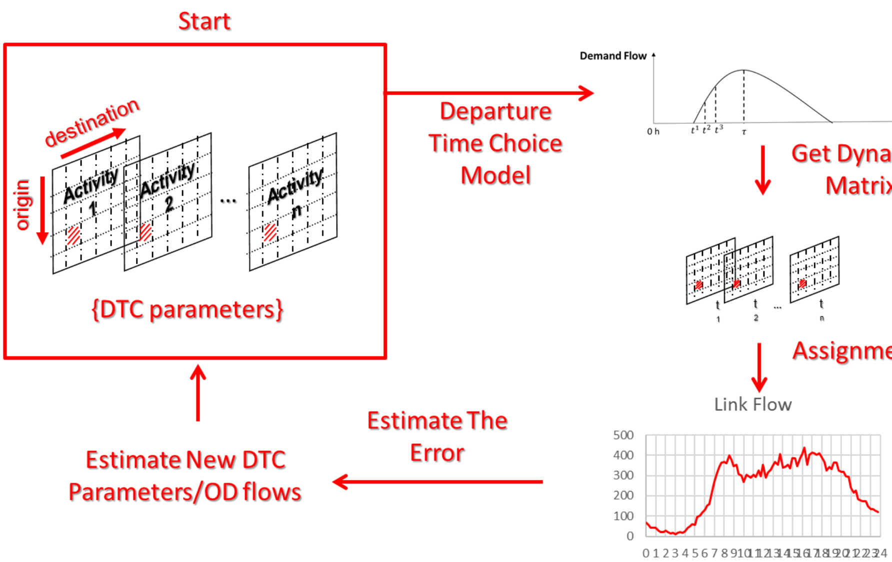


Fig.9.1: Illustrative representation of the UB-DODE model;

Firstly, this formulation automatically decreases the number of decision variables. The reason is that the parameters of the departure time choice model become the decision variables of the model. As a consequence, the UB-DODE becomes a parametric approach in which, for each OD pair, the model estimates the average departure time and its variance. The second advantage is that the DTC model can include different parameters for different activities, thus it explicitly accounts for the trip purpose within the DODE. Lastly, since each parameter directly affects a large number of time-dependent OD flows, the locality of the optimization problem strongly decreases. Figure 9.1 shows the main steps for the proposed UB-DODE model.

9.2.1 Enhanced SPSA for the UB-DODE

The proposed methodology can be implemented with most of the existing solution algorithms, including the well-established SPSA. In this chapter, we proposed to use the C-SPSA (Cluster-SPSA), as it is intuitive to create different clusters for a different type of variables (OD flows, preferred departure time,…) when calculating the gradient.

Unfortunately, many of the desirable properties of convergence of the SPSA derive from the assumptions that the variables are independent. Clearly, this assumption does not hold for the UB-DODE, as OD and DTC parameters are highly correlated. Therefore, the SPSA has the tendency of exploring unrealistic solutions during the optimization. To avoid this behaviour, we proposed the following equation to update the solution at each iteration:

$$\boldsymbol{X}_{i+1} = \boldsymbol{X}_i + \alpha \cdot \boldsymbol{G}_i \cdot P(\boldsymbol{X}_i + \alpha \cdot \boldsymbol{G}_i) \tag{9.1}$$

Where $\boldsymbol{X}_i$ is the vector of the variables to be updated at iteration i, α is the step size and $\boldsymbol{G}_i$ is the gradient. $P(\boldsymbol{X}_i + \alpha \cdot \boldsymbol{G}_i)$ represents the probability that a certain value is realistic for a certain variable. If for instance we are estimating the value for the preferred departure time for commuting to work in the morning, $P(\boldsymbol{X}_i + \alpha \cdot \boldsymbol{G}_i)$ will have a very high value between 6 and 9 am, while will be close to zero for unrealistic values, such as 1 am or 17 pm. This probability, whom parameters an input for the optimization, acts like a constraint during the optimization, reducing drastically the number of unrealistic solutions.

9.3 Case study and results

The aim of this section is to assess the applicability of the novel procedure, by applying the procedure to the network of Luxembourg City. This network, shown in Figure 9.2, consists of more than 3400 links and 1400 nodes and represents the typical middle-sized European city in terms of network size. Moreover, Luxembourg City has the typical structure of a metropolitan area, composed of a city centre, ring, and suburb areas.

Lastly, the simulation environment employed is PTV Visum, which is one of the most widely adopted software for traffic analysis.

Fig.9.2: Network of Luxembourg City;

After generating a realistic starting matrix through the well-known 4-Step model, we performed the DODE on the network showed in Figure (9.2), comparing the results obtained through the proposed UB-DODE and a conventional DODE procedure. Both methods use the well-established SPSA to estimate the gradient, and the traffic counts within the GLS-type goal function.

9.3.1 Utility-Based vs non-Utility-Based DODE

Results, reported in Figure 7.3, show that the proposed model outperforms the standard DODE approach in inferring the OD demand. Although both models fail in estimating the real demand, which is expected when only link flows are used in the objective function, the standard approach creates an unrealistic demand pattern, that can be hardly combined with any prediction model. On the other side, the demand profile for the proposed approach is smoother and, in general, more realistic. This suggests that the model is more suited for estimating the systematic component of the demand, or “*regular pattern*”.

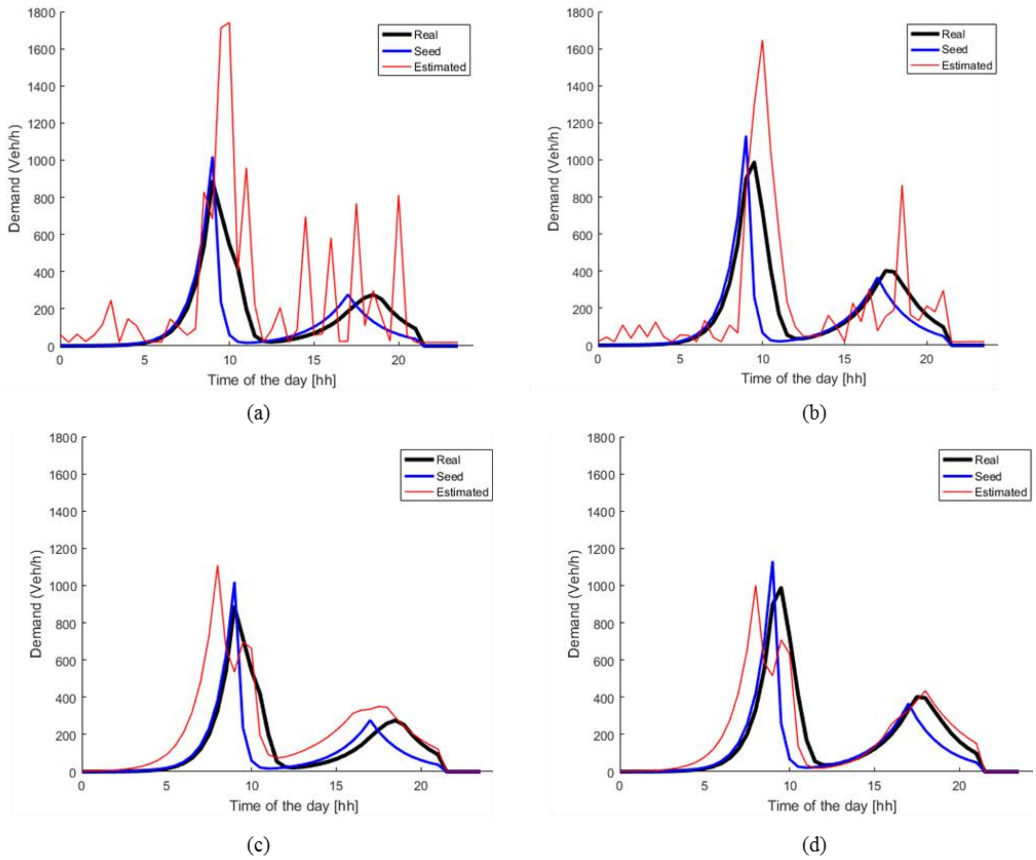


Fig.9.2: (a) Generated Demand For Traffic Zone 12 (Bridel) according to the standard DODE; (b) Generated Demand For Traffic Zone 17 (France-Longwy) according to the standard DODE; (c) Generated Demand For Traffic Zone 12 (Bridel) according to the UB-DODE; (d) Generated Demand For Traffic Zone 17 (France-Longwy) according to the UB-DODE;

9.3.2 C-SPSA vs Enhanced SPSA

In order to assess the reliability of the model, we performed three different experiments, using the UB-DODE to estimate the systematic component of the demand. Figure (9.3) reports some of the results. The first test – Experiment 1 - exploits the standard C-SPSA to estimate the purpose-dependent demand. As shown in Figure (7.3a), the estimated OD demand does not provide an adequate approximation of the real demand. In the second test – Experiment 2 - we investigated the possibility of using a different set of parameters, which were more likely to provide a realistic result. Figure (7.3b) shows how this new set of parameters (step size, perturbation, weights within the goal function) leads to an extremely accurate estimation of the demand.

However, in this case study, the real-demand is known, so it is relatively easy to properly calibrate the parameters of the model in order to improve the performances. Unfortunately, this is not always the case for real applications. As a third option, in Experiment 3 we applied the C-SPSA combined with the soft constraint reported in Equation (9.1). Results show that the constrained C-SPSA achieves a satisfactory estimation of the demand, even with sub optimal set of parameters.

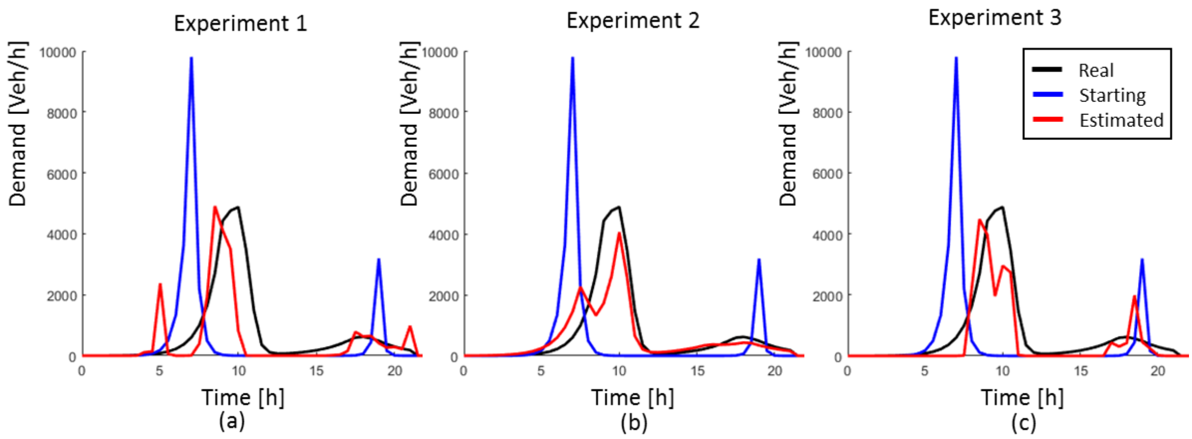


Fig.9.3: Demand generated from France to Luxembourg according to Experiment 1 (a),2 (b) and 3(c);

Figure (9.3) reports an intuitive representation of how accurate the model is in reproducing a realistic demand profile. However, we are interested in the behaviour of the model al network level. While Experiment 2 seems to outperform the other twp for that specific traffic zone, we need to analyse if this observation olds at a network level. Hence, we calculated the relative improvement in terms of RMSE (Root Mean Squared Error) for the Generated Demand Flows for each traffic zone as:

$$\Delta\_RMSE_{zone} = RMSE_{Zone}^{Starting} - RMSE_{Zone}^{Estimated} \quad (9.2)$$

The Δ_RMSE_{zone} term represents how close we are to reality in terms of temporal distribution. If Δ_RMSE is negative, it means that the error is higher for the estimated matrix than for the starting demand, if it is equal to zero there is no improvement, while if Δ_RMSE >0 we improved the situation with respect to the initial situation. The larger Δ_RMSE is, the bigger is the improvement in the Estimated matrix. Figure (9.4) report the probability (9.4a) and cumulative probability (9.4b) of having a certain Δ_RMSE value for a generic traffic zone according to the three Experiments.

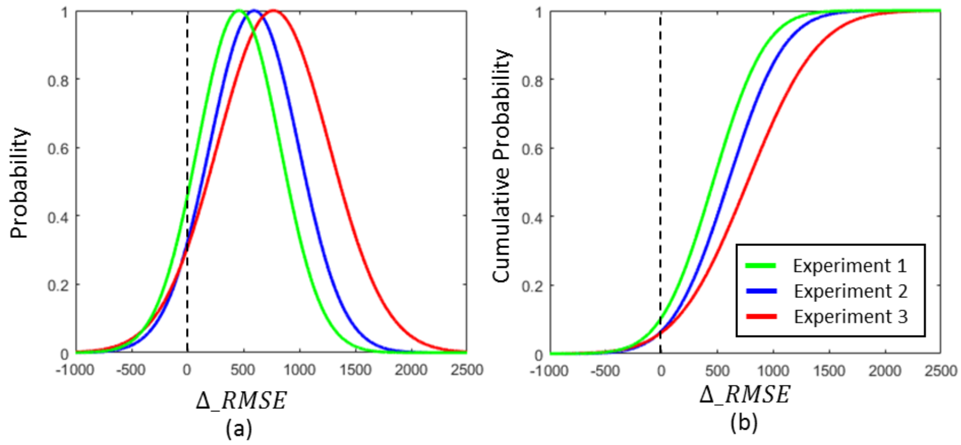


Fig.9.4: (a) Probability and (b) Cumulative Probability of improving the starting OD matrix for each Experiment;

Figure 9.4 clearly shows that, at a network level, if the constraint formulation is applied, we systematically improve the quality of the estimated OD matrix with respect to the unconstraint scenario. Even when we adopt a good set of parameters – Experiment 2 – at network level the model is not as good as the constrained one, meaning that for each traffic zone properly calibrated – such as the one in Figure 2a – there is another one with a larger error. By contrast, the constrained formulation provides a more reliable estimation at a network level.

9.4 Conclusions

This chapter provided some insights on the possibility of adopting the proposed Utility-Based DODE on a general network. Experiments suggest that the model can indeed provide more reliable results with respect to the standard approach. Nevertheless, we should point out some main limitations. First, this chapter shows that, from a theoretical point of view, the methodology can be implemented on a general network and provide reasonable results. However, as pointed out in Chapter 2, from a practical point of view different activity patterns have to be considered in order to properly recreate the mobility demand. Thus, to estimate the number of function to consider within the model and their shape is a fundamental step for applying this methodology in practice. A first attempt of answering this question is reported in (Sceffer, Cantelmo, and Viti 2017). Additionally, we can observe that additional work in integrating the DTC model and the adopted DTA and for creating more efficient optimization frameworks is needed in order to achieve reliable results in a real life application.

Acknowledgements

We acknowledge for financing the following grant: – AFR-PhD grant 6947587 IDEAS.

9.5 References

Cipriani, E., M. Florian, M. Mahut, and M. Nigro. 2011. "A Gradient Approximation Approach for Adjusting Temporal Origin-Destination Matrices." *Transportation Research Part C: Emerging Technologies* 19 (2): 270–82. doi:10.1016/j.trc.2010.05.013.

Djukic, T., G. Flötteröd, Lint Van, and S. Hoogendoorn. 2012. "Efficient Real Time OD Matrix Estimation Based on Principal Component Analysis." In , 115–21. doi:10.1109/ITSC.2012.6338720.

Li, Zhi-Chun, William H. K. Lam, and S. C. Wong. 2014. "Bottleneck Model Revisited: An Activity-Based Perspective." *Transportation Research Part B: Methodological* 68 (October): 262–87. doi:10.1016/j.trb.2014.06.013.

Marzano, V., A. Papola, and F. Simonelli. 2009. "Limits and Perspectives of Effective O-D Matrix Correction Using Traffic Counts." *Transportation Research Part C: Emerging Technologies* 17 (2): 120–32. doi:10.1016/j.trc.2008.09.001.

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

Sceffer, Ariane, Guido Cantelmo, and Francesco Viti. 2017. "Generating Macroscopic, Purpose Dependent Production Factors through Monte Carlo Sampling Techniques." *Transportation Research Procedia*, 19th Euro Working Group on Transportation, EWGT 2017, 17-19 September 2018, Budapest, Hungary, .

Zhou, Xuesong, Chung-Cheng Lu, and Kuilin Zhang. 2012. "Dynamic Origin–destination Demand Flow Estimation under Congested Traffic Conditions." In *Proceedings of the 91st Transportation Research Board Annual Meeting, Washington, DC*. http://www.civil.utah.edu/~zhou/Single-level_path-flow-OD-estimation_Zhou-Lu-Zhang.pdf.

Zhou, Xuesong, and Hani S. Mahmassani. 2007. “A Structural State Space Model for Real-Time Traffic Origin–destination Demand Estimation and Prediction in a Day-to-Day Learning Framework.” *Transportation Research Part B: Methodological* 41 (8): 823–40. doi:10.1016/j.trb.2007.02.004.

10

Conclusions

In this final Chapter, conclusions about this thesis are summarized, shortcomings are highlighted and future research is presented.

10.1 Research questions

The purpose of this thesis was to develop general frameworks for solving the dynamic OD estimation. The proposed solution should be feasible in practice, provide robust estimation and create a good dynamic matrix when only static information is available. To pursue this goal, we answered three research questions, which have been formulated in the first chapter:

- *R.Q.1: What is the impact of considering activity information within the DODE problem?*

  The accurate estimation of the activity-based demand is extremely important when the "systematic" component of the demand is unknown. Specifically, a wrong estimation of the activity-based demand flows, such as a wrong estimation of the morning commute, lead to a biased estimation of the demand. Thus there are two solutions. When a good "seed matrix" is available, thus the "systematic" component of the demand has been properly estimated, then we can achieve good results even without considering the activity dimension. Otherwise, this component has to be explicitly modelled within the model in order to achieve a reliable estimation. Given this observation, a model has been proposed in Chapter 8 for tackling this issue, while Chapter 9 showed that the model can be applied in practice.

- *R.Q.2: Can we estimate purpose dependent OD flows without increasing the number of parameters?*

  Chapter 8 introduced the Utility-Based OD Estimation framework (UB-DODE). The proposed framework can be used to estimate the activity-based demand flows without increasing the number of parameters. Chapter 8 introduces the mathematical properties of the model, showing under which conditions the model achieve reliable solutions. Then, Chapter 9 shows that this parametric approach can be applied for general networks.

- *R.Q.3: Can we increase the reliability of the OD estimation?*

  As reported several times in this manuscript, the DODE is a highly non-linear problem, which is usually solved though simple first-order optimization models. Although more sophisticated approaches have been proposed, no existing model can guarantee to reach the global optimum of the problem. Even worse, a small perturbation of the model parameters can strongly affect the quality of the solution, even when a reliable starting matrix is available. We proposed a model that separates the OD estimation in two sub-optimization problem – the Two-Step approach. The first problem aims at finding a reliable initial point for the second problem, which looks for the best fit with the available traffic data. Experiments show that this approach is more likely to provide a more robust estimation.

In the next sub-sections, we recapitulate these points, highlighting the complementarity of the two models presented in this manuscript and the relevance of considering multiple mobility patterns within the DODE problem.

10.2 Interacting mobility patterns

One of the critical point identified in this manuscript is the need of dealing with multiple mobility patterns. Specifically, the demand can be classified into regular and irregular patterns. Following the Classification proposed in (Zhou and Mahmassani 2007), the demand can be classified in:

- *Regular Pattern*: This mobility pattern represents the "*systematic*" component of the demand and accounts for its typical structure (length of the rush hour, activity scheduling, etc). This component is expected to follow a day-to-day dynamic behaviour, meaning that users might behave differently on different days of the week, while we can assume it constant in a week-to-week context (same behaviour for two consecutive Mondays).

- *Structural Deviation (Irregular Pattern):* This component accounts for those deviations from the regular pattern that we can explain, for instance, a longer rush hour due to roadworks and/or severe weather phenomena.
- *Random Deviation (Irregular Pattern):* This component takes into account the random deviations from the regular pattern, which cannot be explained. Although users usually go to work at a certain time, some unpredictable event might force them to anticipate or postpone their trip.

The interaction between these mobility patterns is a trivial problem during the DODE process, as errors in the regular patterns lead to even larger errors in the estimated Irregular component. Additionally, the DODE is the last step in order to estimate the mobility demand, meaning that results inherit all the errors from previous demand models, as well as those within the underlying DTA model. Thus it is extremely relevant to explicitly account for both mobility patterns within the estimation process in order to control this error.

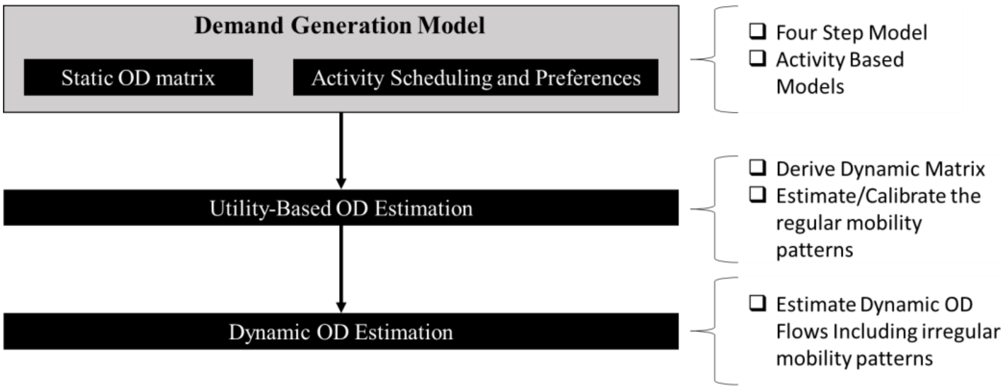


Fig. 10.1: From Static Models to Dynamic OD Estimation;

10.2.1 Regular Demand Patterns

As shown in Figure 10.1, this thesis aims at reducing the error by introducing the Utility-Based OD Estimation (UB-DODE) technique to correct the regular demand pattern. The idea is that traffic data can be directly used to reduce the error within the regular pattern inherited by the Demand Generation Model. To achieve this goal, we decide to include information at activity level within the DODE, and to use a probabilistic approach to estimate purpose dependent OD flows (Figure 10.2). The probabilistic approach assumes that, given some preferences about the activity scheduling, users will more likely perform activities in a certain time frame. As we pointed out in the previous section, this "preferences" might be affected by substantial errors that generally affect also the final estimation. To avoid this problem, observed preferences (such as "average departure time") are also updated during the UB-DODE process so that, given a set of observations, the most likely regular pattern can be estimated.

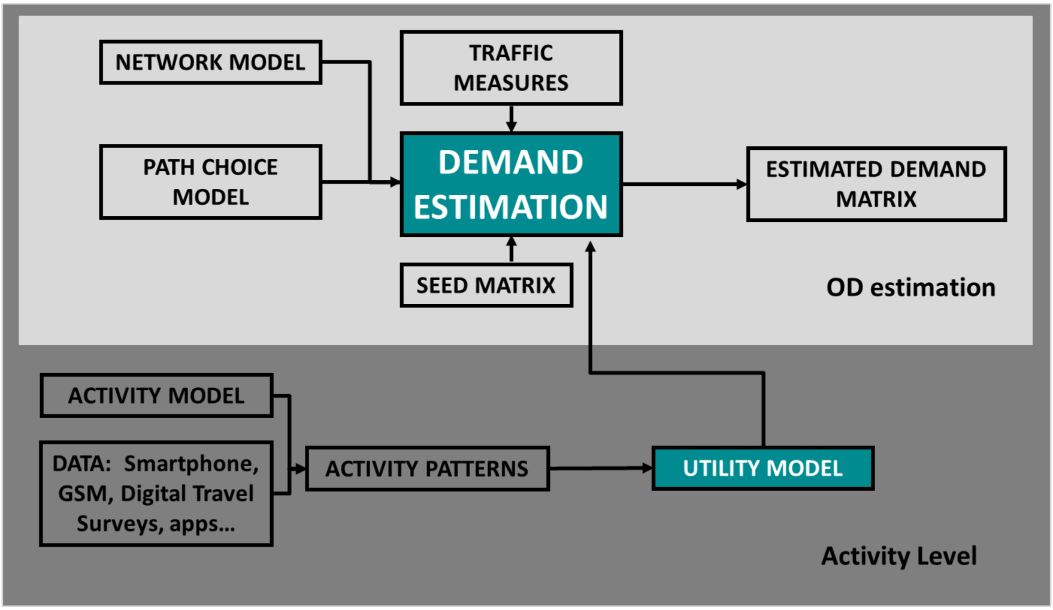


Fig. 10.2: Illustrative representation of the UB-DODE model;

10.2.2 Irregular Demand Patterns

Even if a good demand matrix can be obtained through the UB-DODE, this one will hardly provide an accurate fit with the traffic data. There are two main reasons:

(i) The UB-DODE approach denies "unusual" user behaviour, meaning that might be not sensitive enough to capture heterogeneous user behaviour. Different users will react differently to unexpected events such as severe weather conditions. Thus the parametric approach will hardly be able to replicate the observations;

(ii) From a mathematical point of view, the Assignment-Based DODE (AB-DODE) has more free parameters then the UB-DODE. In essence, although the AB-DODE is more likely to over-fit the data, for the same reason is also more likely to provide more realistic traffic conditions on the network.

Point (ii) is extremely important since, in some applications, it results more important to reproduce a realistic congestion pattern rather than an accurate structure for the mobility demand. For these reasons, the ideal solution is to combine the two models as presented in Figure 10.1, achieving the goal of having a correct structure for the demand and an accurate fit with the data.

10.3 Main findings

Here we summarize the main conclusions of this research, distinguishing between advances in the DODE framework, contribution in the DTA and practice oriented insights.

Concerning the DODE framework, this thesis brings the following innovations:

Irregular Mobility Patterns:

a) The proposed Two-Step approach allows reducing the localism issue of the problem. It is possible to use Principle Component Analysis or Quasi-Dynamic estimators in order to perform a broader exploration of the solution space, handling larger scenario than the typical "one-step formulation", while at the same time allowing the same level of fitting accuracy as the conventional models;

b) The proposed Two-Step approach allows combining different optimization algorithms – such as stochastic and deterministic models – improving both computational time and estimation performances.
c) The Two-Step procedure present more reliable solution with respect to the typical “single-step” procedure.
d) The Two-Step procedure capture demand dynamics at a network level, recreating the uncongested/congested dynamics on the network without significantly modifying the structure of the demand.
e) Mobile phone network data can be used as a proxy for the generated/attracted demand, improving model performances. However, this is an option when large traffic zones are available.

Regular Mobility Patterns:

f) Few parameters representing the activity schedule can be used to approximate the mobility demand through parametric models.
g) Activities can be explicitly included within the DODE increasing the overall reliability of the estimation.
h) The proposed parametric Utility-Based DODE model accounts for activities while reducing the overall number of the variable.
i) The UB-DODE can account for activity duration, creating a more consistent approximation of the demand over the 24 hours period.
j) When the model explicitly accounts for the simulated travel time –as for the UB-DODE – the DODE framework is more sensitive to the spatial distribution of the demand, since the derivative of one specific variable influences several time intervals, generating a meaningful perturbation on the network flows

Concerning the DTA model in the context of dynamic OD estimation, the main achievements are that:

a) By assuming a departure time choice model within the DTA, the number of the relation between observations and unknown variables increases.
b) Utility-functions can be used for modelling different activities in the network and their effect on the congestion.
c) A new utility function has been proposed, which can be used for modelling different activities, from work to special events.
d) The effect of the utility function on the DTA has been analysed, showing that including a positive cost not only changes the congestion in terms of timing and duration but also correlates it over different trips.

Finally, practice oriented insights are that:

a) Errors in both data and models need to be properly accounted when implementing DODE. A possible way to solve this problem is to identify regular and irregular demand components and fix these errors one after the other.
b) A methodology for creating a seed matrix when a good dynamic matrix is not available is presented.
c) For large networks, there is an unavoidable trade-off between accurate mobility demand and model capability of reproducing the available traffic data. When more heterogeneity is considered, the DODE is more likely to find an accurate fit with the traffic measures but the

problem becomes more local. Vice versa, a simple structure lead to a smooth objective function but will fail in reproducing the available data, as user behaviour is oversimplified.

10.4 Future research

Although this thesis answered the research questions addressed in Chapter 1, more effort is needed to address problems that have been listed in the previous chapters and to generalize the framework proposed in Figure 10.1. Existing issues can be divided into three main challenges.

First, in order to implement the Utility-Based DODE in practice, it is important to understand how many activity patterns should be considered and the proper function to represent them. Although some explorative analysis has been presented in Chapter 2 and some functions in Chapter 7, this only partially answers the question. From a mathematical point of view, we need to understand the number and type of functions that can properly reproduce the overall mobility demand, while at the same time are consistent with the requirements presented in this manuscript.

The second problem consists in the solution algorithm for the UB-DODE. This thesis focused on the formulation of a new objective function and its relation with the DTA model in the lower level. We just applied and extended existing algorithms to solve the optimization problem. Although these algorithms provided reasonable results, it is relevant to investigate alternative solutions. For example, Simulating Annealing has been proven to be less efficient than the SPSA for the normal DODE since the solution space is highly non-linear. However, it might provide a better estimation with UB-DODE, as the research space is smoother.

Lastly, the UB-DODE has been tested for car users only in this thesis. However, transportation is moving towards multimodality, meaning that in the future more and more interest will be on estimating the mobility demand in a multimodal environment. The UB-DODE seems ideal for this purpose, as it is based on utility functions. The demand adjustment passes through a departure time choice model based on the generalized cost of travelling, thus the extension from (departure time + route choice) to (departure time + mode + route choice) is straightforward. However, proper utility functions have to be included. Moreover, the interaction between different data coming from different transport modes has to be properly investigated, as it is likely to make the problem even more complex.

Future research on the Two-Step approach is more limited. This framework is the extension of well-established models and has been deeply analysed in this thesis, on both simple and real networks. The main research effort in this direction is to extend the MAMBA-DEV tool, in order to include more algorithms, models and data. From the methodological point of view, two main questions still need to be investigated. First, if online and offline models can be combined with the Two-Step of the model.

Second, to formulate the Two-Step approach in an iterative way. Experiments in Chapter 6 show that iterating the two step approach while updating the parameters of the goal function might lead to a significant improvement with respect to the current formulation. Thus update rules might be included in an iterative approach in order to adapt the objective function and escape local minimum.

10.5 References

Zhou, Xuesong, and Hani S. Mahmassani. 2007. “A Structural State Space Model for Real-Time Traffic Origin–destination Demand Estimation and Prediction in a Day-to-Day Learning Framework.” *Transportation Research Part B: Methodological* 41 (8): 823–40. doi:10.1016/j.trb.2007.02.004.

Appendix A

Cluster analysis

A.1 Activity Classification

In this section, we clarify the way we performed the cluster analysis in Chapter 2. Let us consider the definition of *Daily-Systematic-Activity* (DSA), *Within-week-Systematic-Activity* (WSA) and *Not-Systematic-Activity* (NSA) provided in Chapter 2. If we observe a single user for N_{obs} days, we can approximate the probability to do or nor a certain activity during a certain day with the frequency of observation. Assume that we observe a certain user *i* for four weeks. Assume that user has a strongly systematic behaviour. Every day he goes to work (DSA), on Wednesday goes to the swimming pool (WSA), on Saturday does the weekly shopping (WSA) and on Sunday visits the family (WSA). During this month, user *i* goes one time to the hospital and one time to the bank (NSA). Time profile of user's agenda is:

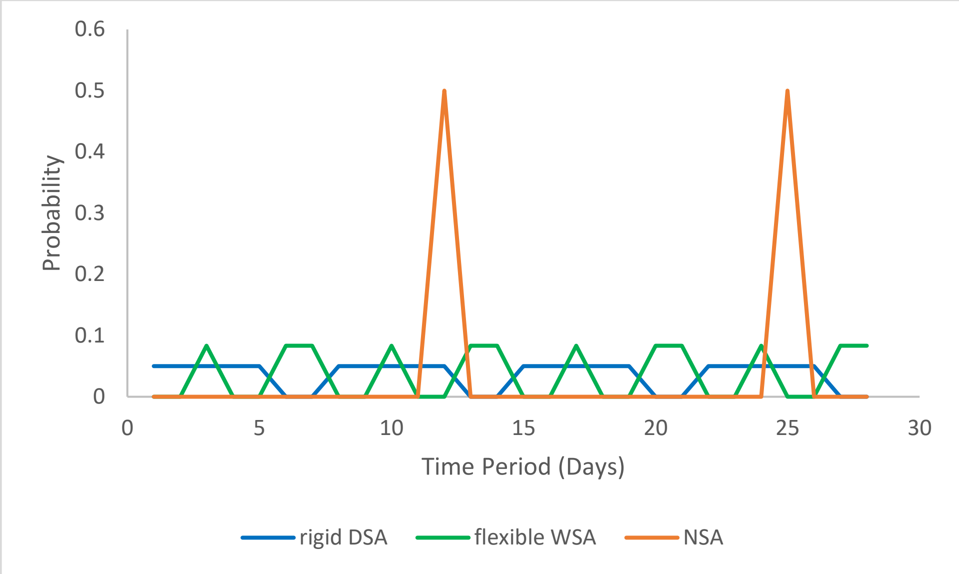


Fig. A.1: Probability to observe WSA/DSA/NSA activities for a single user;

In this case, the probability is assumed to be equal to the frequency of observation. This means that if we sum all these probabilities during the entire period, we obtain 1. Assume now that we have a population of users. Since it is a theoretical situation, we can assume that all users work only during the working days, from Monday to Friday. We can assume in this framework that all users will behave in a similar way to the single user, with one or more WSA activities. However, it is unrealistic to imagine that all the users will perform their WSA activities the same day. The aggregate version of graph plotted in Figure A.1 becomes:

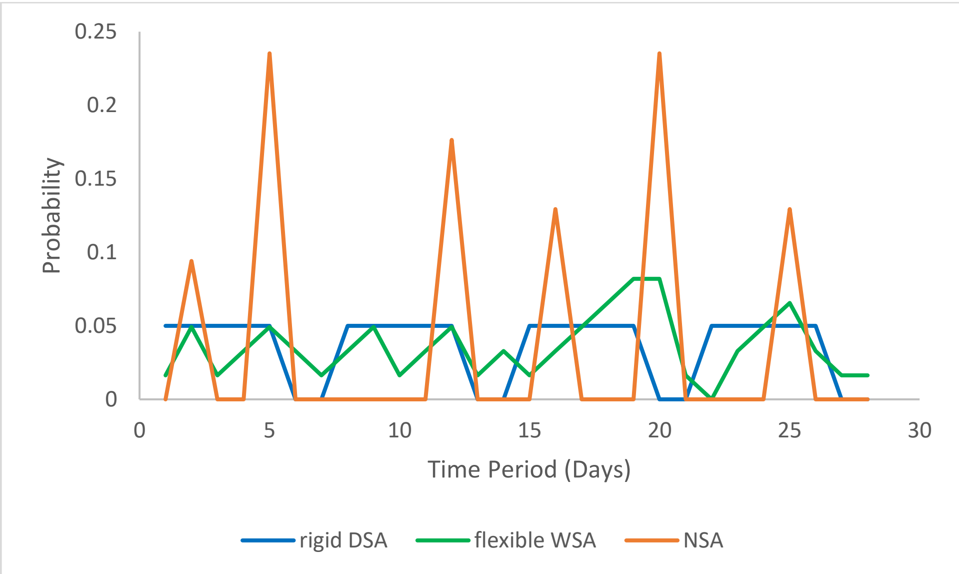


Fig. A.2: Probability to observe WSA/DSA/NSA activities for a theoretical population;

The assumptions used to build the synthetic population represented in Figure A.2 are unrealistic and very simples. However, interpreting the information behind this graph is not an easy task, like in the single user case. This suggests that a different approach should be considered to classify our data in case of multiple users. Let's sort our Probability in ascending order. The x-axis does not represent anymore the Time Period, but the day in which a certain number of observations occur. According to our representation, DSA will present a constant probability.

$$P_{DSA} = \frac{1}{N_{obs}} \quad \text{(A.1)}$$

This probability is correct if we evaluate the probability to come back home, which every user does every day. Activity Work could be considered an extreme case of WSA activity since it is systematic five days out of seven. However, *Within-Week-Systematic Activities* are *flexible* by definition, while Work is a *rigid* component of the demand. For the activity component *Home-Work*, we can observe two probabilities equal to 0 on Sunday and Saturday and a constant probability the other days:

$$\begin{cases} P_{Work}(x) = \dfrac{1}{N_{wd}}; & \forall\, x > P_b \cdot N_{obs} \\ P_{Work}(x) = 0; & \forall\, x \leq P_b \cdot N_{obs} \end{cases} \quad \text{(A.2)}$$

Where P_b is the percentage of not working days during the week, and N_{wd} is the number of working days.

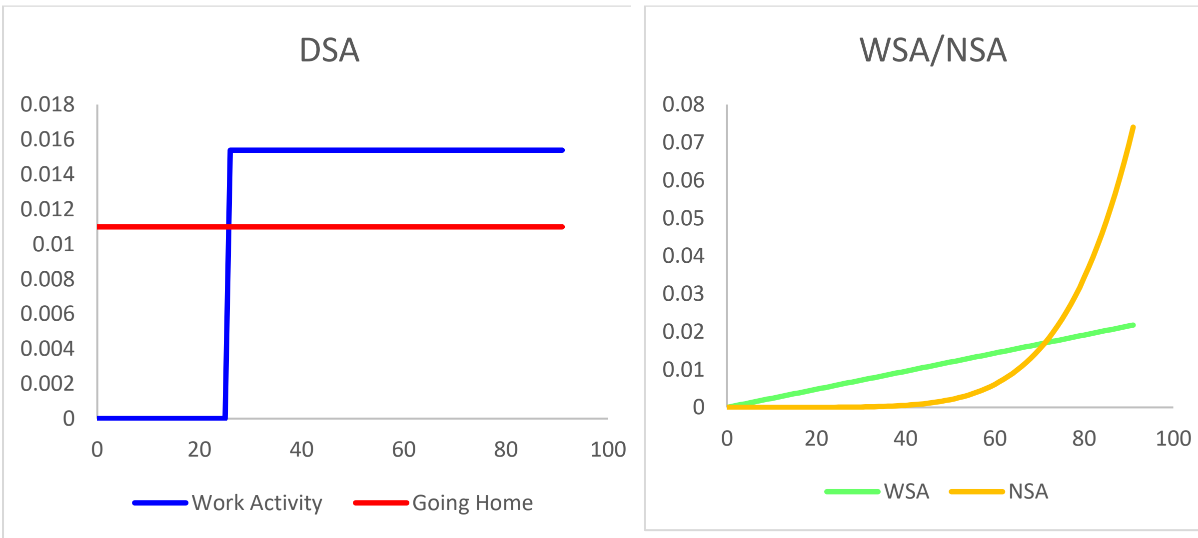
WSA activities are *flexible* because, generally, the user chooses to make or not that activity, based on different elements. Still, these activities are systematic within the week, which means that the individual utility is a function of performing or not the WSA activity. If the User does not perform that activity a certain day, the probability to perform that activity the day after increases. For this reason, the probability to observe or not WSA activity will increase with the number of observations. In this work, we assume that the probability increases linearly:

$$P_{WSA}(x) = x \cdot \gamma; \qquad \forall\, x \in N_{obs} \quad \text{(A.3)}$$

Where γ is the angular coefficient. In the NSA case, the probability will take the shape of an exponential, since normally users don't perform these activities:

$$P_{NSA}(x) = \frac{(1 + x^{\alpha})}{\beta}; \qquad \forall\, x \in N_{obs} \quad \text{(A.4)}$$

Where α and β are parameters to calibrate. Figure A.3 shows the theoretical trend of the system of equations 1-4, for ninety days of observations.



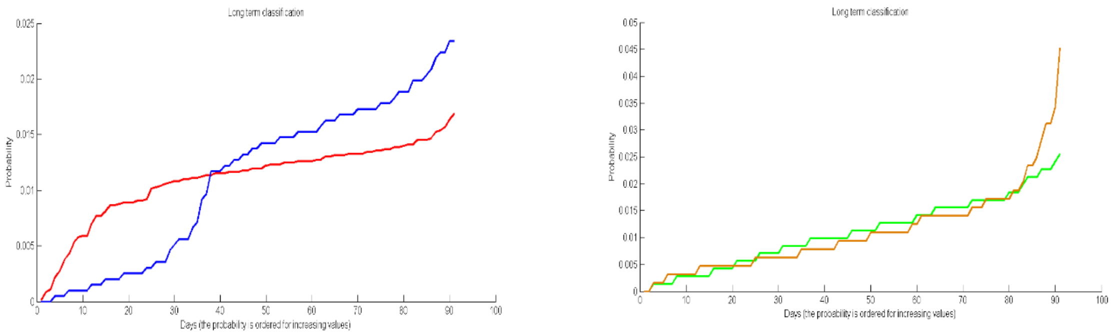
(a) *(b)*

Fig. A.3: (a) Theoretical Probability Trend for DSA activities; (b) Theoretical Probability Trend for WSA and NSA activities;

The trend showed in Figure A.3 is purely theoretical and has been obtained through strong assumptions on the user behaviour. Now the same analysis is performed on the real dataset. Specifically, the individuals in the BMW survey can classify their trip according to twelve different purposes:

1. Pick up/ Drop off s.o.
2. Home
3. Work
4. School
5. Eat
6. Daily Shopping
7. Long term shopping
8. Personal Business
9. Visit the family
10. Walking/Riding/…
11. Leisure
12. Other

By way of example, work, home, personal business and walking/riding purposes are plotted. The observed probability function, as expected, is different from the theoretical one. However, we can observe similarity with the theoretical one. Specifically, the work probability is characterized by two branches, a very low probability in the beginning and a very high probability at the end.



(a) *(b)*

Fig. A.4: (a) Observed Probability Trend for Home (red) and Work (blue) activities; (b) Observed Probability Trend for personal business (green) and walking/riding (orange) activities;

The activity “Home” presents almost a constant probability. Personal business (WSA) activity presents linear functions, while “walking/riding” is a convolution of WSA and NSA Activity components. In Figure 4, the cumulative probability, for each purpose observed in the survey, is showed. The observed probabilities are not clearly identifiable as DSA/WSA/NSA. The trend for different activities looks like a convolution of the observed probabilities reported in Figure A.5.

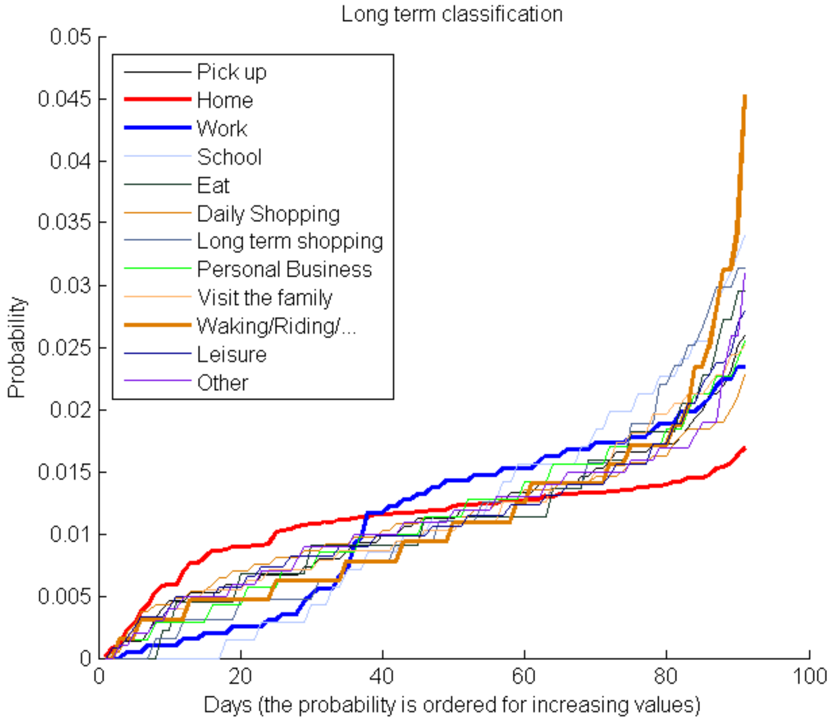


Fig. A.4: Observed probabilities for each observed purpose ;

Activities have been clustered according to the similarity with respect to the theoretical trend expected for DSA, WSA and NSA activities.

Appendix B

The Bottleneck Model

This section introduces the equations behind the bottleneck model. The equations and results reported derive mostly from the following woks:

R. Arnott, A. de Palma, and R. Lindsey, "Economics of a bottleneck," *Journal of Urban Economics*, vol. 27, no. 1, pp. 111–130, Jan. 1990.

X. Zhang, H. Yang, H.-J. Huang, and H. M. Zhang, "Integrated scheduling of daily work activities and morning–evening commutes with bottleneck congestion," *Transportation Research Part A: Policy and Practice*, vol. 39, no. 1, pp. 41–60, Jan. 2005.

Z.-C. Li, W. H. K. Lam, and S. C. Wong, "Bottleneck model revisited: An activity-based perspective," *Transportation Research Part B: Methodological*, vol. 68, pp. 262–287, Oct. 2014.

B.1 The Bottleneck Model

As the utility models presented in Chapters 7-9 are based on the Bottleneck model theory, we introduce in this Appendix the Bottleneck model and the activity based Bottleneck model. According to Vickrey, six different types of congestion exist, which can even be observed simultaneously: simple interaction, multiple interactions, bottleneck, trigger neck, network control and general density. The Vickrey model considers the bottleneck situation when a relatively small segment of a link or a route presents a smaller capacity with respect to the demand for that link/route. The model has been formulated in a more detailed way from Arnott, de Palma and Lindsey in the following system:

Assume the simple network in Figure B.1. We have one origin zone A (Home), one destination B (Work) connected through one link [c-d] with capacity S.

Fig. B.1: Network ;

Assume to have a population of *N* identical users, with *N*>S, traveling between A and B. Since the demand exceeds the capacity, a queue will occur on the link c-d, which is the *bottleneck* of our system. The travel time from A to B can be calculated as:

$$T(t) = T^f + T^b(t) \quad \text{(B.1}$$

Where *t* is the departing time, T^f is the free flow travel time and T^b is the time spent at the bottleneck, given a certain departing time *t*. The free flow travel time is generally considered equal to 0, which means that the travel time between A and B is equal to the time spent at the bottleneck. The time spent at the bottleneck can be calculated given the length of the queue, according to equation B.2:

$$T^b(t) = \frac{D(t)}{S} \quad \text{(B.2}$$

Where the *D(t)* is the length of the queue. The maximum length of the queue is calculated like the integral of all the vehicle queuing after a certain time interval t^* , which is the last time interval in which no queue was observed at the bottleneck. Defined as r(t) the departing rate for a certain time interval *t*, we can obtain the length of the queue as follows:

$$D(t) = \int_{t^*}^{t} r(t)dt - s(t - t^*) \quad \text{(B.3}$$

Where $s(t - t^*)$ is the total capacity for all the departing times between $t^*$and *t*. The derivative with respect to the time, provide the number of vehicle queueing in the time interval *t.*

$$\frac{\partial\, D(t)}{\partial t} = r(t) - s \quad for\, D(t) > 0 \quad \text{(B.4}$$

Since we assume that all the travellers are identical, we define them as *homogeneous population*, and we assume that all of them wants to arrive at the same *preferred* time t^o. We can define the t^{d-o} the departure time for which the user arrive at work on time:

$$t^{d-0} = t^0 - T^b(t^{d-0}) \quad \text{(B.5a}$$

So we can now quantify the early and late arrive as:

$$\begin{cases} \Delta t = t^0 - t - T^b(t) & for\ t < t^{d-0} \\ \Delta t = t + T^b(t) - t^0 & for\ t > t^{d-0} \end{cases} \quad \text{(B.5b}$$

The cost function of one specific trip U is a linear combination of three elements: The travel time *TT*, the early arrive *EA* and the late arrive *LA:*

$$U(t) = \alpha \cdot (TT) + \beta \cdot (EA) + \gamma \cdot (LA) = \\ = \alpha \cdot \left(T^b(t)\right) + \beta \cdot max\left(0; t^0 - t - T^b(t)\right) + \gamma \cdot \left(t + T^b(t) - t^0\right)$$ (B.5c

Where the parameters α, β and γ are the coefficients representing the penalty for late or early arrival. Since every individual wants to maximize his own utility, and every individual is identical to the others by definition of *homogeneous* population, at the *equilibrium* we can argue that utility will be the same for all the travelers. The solution of the model to maximize U for the users is a pure Nash equilibrium with respect to the control variable *t*.

Normally in literature, we assume α > β, both because is a relevant condition for the existence of an equilibria condition, both for empirical and modeling reasons. Specifically Small suggests using β/α = 0.5 and γ/α = 2. In this condition, all the vehicles, except for the first and the last traveler, face the congestion. The departure time rate is piecewise (i.e. has different segments) given by the following system:

$$r(t) = \begin{cases} s + \dfrac{\beta s}{\alpha - \beta} & for\ t \in [t_q, t^{d-0}) \\ s - \dfrac{\gamma s}{\alpha + \gamma} & for\ t \in (t^{d-0}, t_{q'}] \end{cases}$$ (B.6

Where t_q is the time at which the queue begins and $t_{q'}$ is the time at which the queue ends. Equation (A.6) take into account that the departure time is chosen in order to balance the disutility in the travel time. The system of equation 6 is obtained by finding the best possible solution for equation A.5, i.e. by finding the value of *t* for which the derivative is equal to 0. The demonstration for the case of early arrive is here provided:

$$U(t) = \alpha \cdot \left(T^b(t)\right) + \beta \cdot \left(t^0 - t - T^b(t)\right) + \gamma \cdot (0) = \\ (\alpha - \beta) \cdot T^b(t) + \beta \cdot t^0 - \beta \cdot t$$ (B.7a

By substituting equations (B.2) in (B.7a) we get:

$$U(t) = (\alpha - \beta) \cdot \frac{D(t)}{s} + \beta \cdot t^0 - \beta \cdot t$$ (B.7b

To find the best possible departure time *t* we need to maximize equation 7b, which means find *t* for which the derivative is zero. Substituting equation 4 in 7b we get:

$$\frac{\partial U(t)}{\partial t} = 0 = \frac{(\alpha - \beta)}{s} \cdot \frac{\partial\, D(t)}{\partial t} + 0 - \beta \\ \frac{(\alpha - \beta)}{s} \cdot r(t) - (\alpha - \beta) = \beta \\ r(t) = s + \frac{s\beta}{(\alpha - \beta)}$$ (B.7c

Applying the same procedure, it is possible to find also the equation in the case of late arrival time. The arrival time at the bottleneck will be equal to the saturation flow S over the entire simulation period. To estimate the departure times t_q, t^{d-0} and $t_{q'}$, Arnott, De Palma and Lindsey define the following system of three equations:

$$\left(t^{d-0} - t_q\right)\left(s + \frac{\beta s}{\alpha - \beta}\right) + \left(t_{q'} - t^{d-0}\right)\left(s - \frac{\gamma s}{\alpha + \gamma}\right) = N$$
$$\left(t^{d-0} - t_q\right) \cdot \frac{\beta s}{\alpha - \beta} = \left(t_{q'} - t^{d-0}\right) \cdot \frac{\gamma s}{\alpha + \gamma} \qquad \text{(B.8}$$
$$t^{d-0} + \left(t^{d-0} - t_q\right) \cdot \frac{\beta}{\alpha - \beta} = t^0$$

Where the first equations take into account that all the vehicles will be served during the entire period $\left[\left(t^{d-0} - t_q\right); \left(t_{q'} - t^{d-0}\right)\right]$, the second equation considers that congestion disappears at instant $t_{q'}$. In fact the meaning is that all the N vehicles passed through the bottleneck in $t_{q'}$. The last equation is a consequence of equation 5a-C (plus the consideration that at equilibrium the travel time has to be equal to the disutility of that guy leaving earlier). In fact by substituting equation 6 in equation 3 and in equation 2 we get $= \int_{t^*}^{t} \frac{r(t)}{s} dt - (t - t^*) = \left(t^{d-0} - t_q\right) \cdot \frac{\beta}{\alpha - \beta}$. Solving the system of equations:

$$t_q = t^0 - \left(\frac{N}{s}\right)\left(\frac{\gamma}{\beta + \gamma}\right)$$
$$t_{q'} = t^0 + \left(\frac{N}{s}\right)\left(\frac{\beta}{\beta + \gamma}\right) \qquad \text{(B.9}$$
$$t^{d-0} = t^0 - \left(\frac{N}{s}\right)\left(\frac{\beta\gamma}{\alpha(\beta + \gamma)}\right)$$

Constrained to $t_q < t^{d-0} < t_{q'}$. Knowing these three points it is, possible to plot the cumulative for arrival and departure times. Specifically, t_q will be the departure time for the first user in the system, the queue will reach the maximum in t^{d-0} and, since $\beta < \alpha$, it is decreasing after t^{d-0} until $t_{q'}$. For arrival time, since when congestion occurs the main constraint is the capacity of the bottleneck, and since the travel time after the bottleneck is zero, the cumulative is starting in t_q, and it presents a constant slope S until $t_{q'}$, when all the demand reaches the destination. Considering the trip cost for the users who are leaving at t_q. He will not experience any congestion, which means the cost will be equal to the *dis-utility* to arrive early:

$$C(t_q) = \beta \cdot \left(t^0 - t_q\right) = \left(\frac{N}{s}\right)\left(\frac{\beta\gamma}{\beta + \gamma}\right) \qquad \text{(B.10}$$

Since at equilibrium everyone has the same cost, the cost for all the users is obtained by multiplying equation B.10 times the number of users N:

$$TC^e = \left(\frac{N^2}{s}\right)\left(\frac{\beta\gamma}{\beta + \gamma}\right) \qquad \text{(B.11}$$

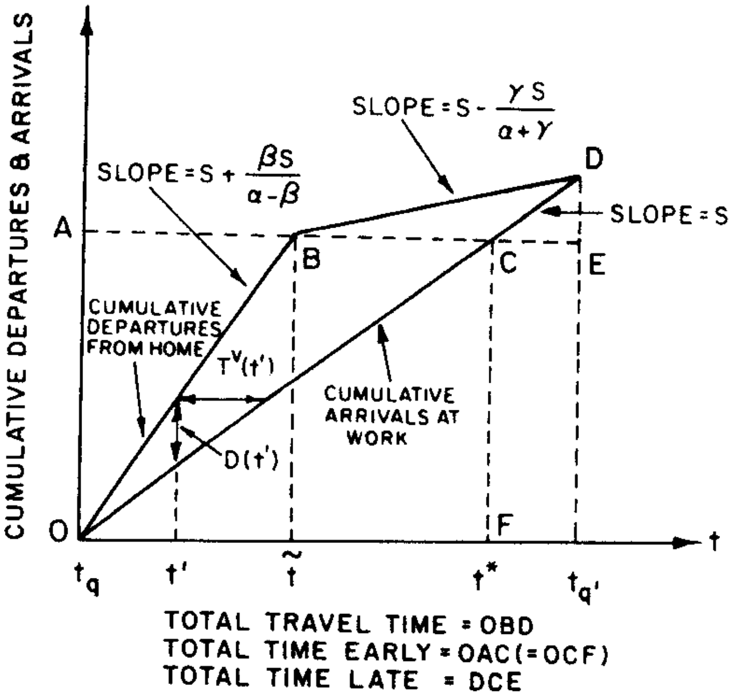


Figure B.2: Cumulative for arrival and departure times
(Figure From the Original Paper "Economics of a bottleneck"-Arnott de Palma Lindsey 1987)

The total travel time can be computed as the area of the triangle (OBD). By multiplying the area for the cost (Equation B.9), we can obtain the total travel cost. It is relevant to point out that, according to the authors (Arnott, De Palma and Lindsey) both the total cost presented in Equation B.11 and the Total Travel Cost are not dependent by α. This is explained by the authors with the motivation that length of the rush hour is not related to the travel time. In fact, the last and the first users don't experience travel time at all, but only scheduling costs and, at equilibrium, their scheduling delay has to be the same. Authors explain this exploiting Equation. B.9, where the departing time for the first and the last user can be calculated independently on α, which means is not dependent on the travel time cost.

I fill to add something about this last point. First of all, we can immediately argue that this is true only because we are considering an unrealistic situation, where the only travel time is the delay at the bottleneck. We are not considering any physic in the propagation of the vehicles. If we consider these elements, results could be different. On the other side, also considering the simple case of the Vickrey bottleneck model, this explanation is only a mathematical trick. If we consider that:

1) The travel time represents only the delay at the bottleneck
2) No physic represents the vehicles moving on the network

We can immediately understand that, given a certain departure time, the travel time will be a function of the number of users N and the capacity of the road S. The delay will be linearly related to these two terms, which is exactly what happens in Equation B.9. This is, in my opinion, demonstrated in equations B.3-B.4, where the delay is related to the incoming flow and the saturation flow. In fact, by changing the number of users N, travel time increase for equations B.3-B.4, departure time for the earliest user decreases according to Equation B.9, which means that travel time directly influences the sizes of the rush hour period.

It follows now the discussion on the case where a toll is applied in order to reduce the congestion during the rush hour. While in literature this topic is extremely popular, since Vickrey bottleneck model has been broadly used to evaluate the effect of pricing policies on the departure time, here we main focus on the user behaviour component. This means that we are more interested in evaluating the behaviour of the users in the model, rather than evaluating their reaction with respect to a specific toll policy. For this reason, the "toll equilibrium" is explained quickly in this section. The main idea is that each user pays a toll directly proportional to the queue he is joining. In this way, the users will try to avoid congestion in order to have a lower toll, resulting in smaller queues for a longer rush hour period. The idea proposed by the authors is that normally tolls are not flexibles, but only constants over a certain time period. In this case, the main problem is that the demand is flexible, and a "static" toll does not consider this flexibility. The simplest toll of this family is called "coarse" toll, and it is defined as a toll paid at the front of the queue for a certain time period. The goal of this problem is to find the best toll and the best time period. Figure B.3 shows the result by applying the optimum coarse toll (t^{+} and t^{-} represents the begin and the end of the toll):

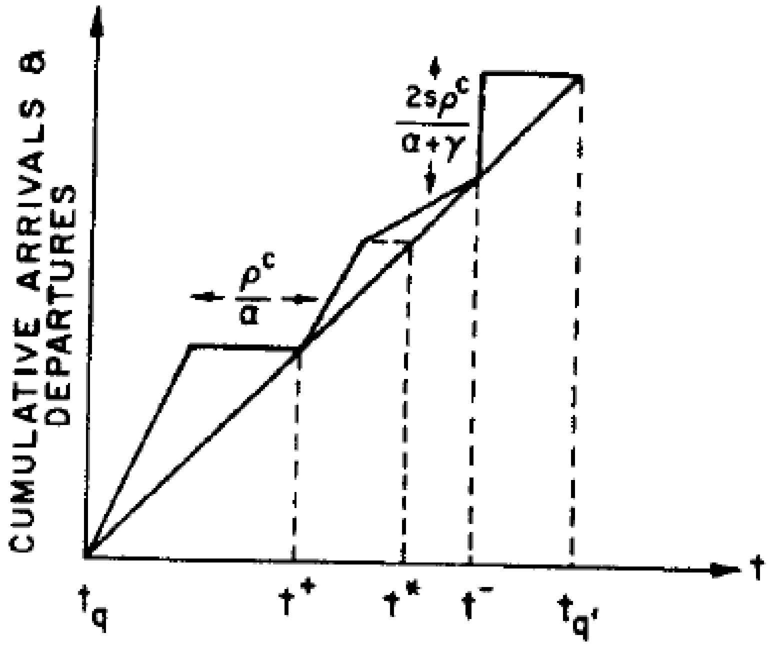


Figure B.3: Cumulative for arrival and departure times
(Figure From the Original Paper "Economics of a bottleneck"-Arnott de Palma Lindsey 1987)

Specifically, authors show how the departures/arrivals drop/start immediately before/after the toll (vertical and horizontal departure rate) in order to avoid the toll. Many problems occur when we try to apply these models in real practice.

1) *Heterogeneity of the demand:* Drivers behave in a different way during the rush hours. According to the activities, the wage, the destination and other elements we can have differences in the travel behavior, which leads to different value of t^0, α, β and γ. These differences should be considered in the model.
2) *Stochastic Demand and Capacity:* This model considers a specific value for the demand (*N*) and the road capacity (S). In reality, both this term could be dynamic. As a consequence of the elasticity of the demand, N can increase or decrease when the supply side change (i.e. when a toll is applied). On the other side, supply is dynamic too. Capacity can temporary increase or decrease as a consequence of traffic policy (i.e. opening the emergency lane during the rush hour).
3) *Road Network Complexity:* The model could be too simple to represent multiple origins, destinations and routes.
4) *Hyper congestion:* Considering urban-complex dynamics, such gridlock and stop and go phenomena.

5) *Nash equilibrium:* Concept of equilibrium in the basic model is based mainly on the assumption that users are informed. This is partially considered in more complex works.

All the elements have been evaluated during the past decades by several authors. Many of these works are extremely relevant for the problem of the demand estimation.

B.2 The Activity-Based Bottleneck Model

In this section, we talk about extensions of the classic bottleneck model which takes into account the all-day scheduling of the home-work-home activity pattern. The model here presented will be called ZL model, to specify that many of the assumptions are derived from two different papers. Zhang et al. (2005) have been the first one to extend the bottleneck model in order to consider the activity duration and the utility of the time spent performing a certain activity. Li et.al (2014) presented a similar methodology, investigating also the mathematical properties of the model. The second one will be presented in this section, but the main properties of the two models are similar. While both the methods find an equilibrium by maximizing the utility of all the users according to an MSA methods, the main difference is how to calculate the auxiliary flow. While Li et al. do an "all or nothing" assignment, pushing all the demand for the best alternative, Zhang et al. use a logit model to estimate auxiliary flows based on the utility function obtained in the previous iteration. Except for this difference, the two models look very similar. Anyway, the mathematic presented in this section is the one used in Li et.al (2014).

In the ZL model, the main consideration is related to the dependency between the morning and evening peak for the commuting demand. While this consideration seems very logical, this correlation has not been investigated until now. Considering a commuting home-work-home demand, it is clear that the connection between the morning and evening commuting depends mainly on the duration of the activity "*work*", the arrival time at work and the travel time to reach the destination. Considering these three elements, each user will try to maximize his own utility.

Consider again the network presented in Figure B.1. This time, we have two directions, not only A-B but also B-A. Keeping all the assumptions we made for the classic bottleneck model, we have again that, given a certain departure time *t*, the travel time is a function of a constant term, plus another term depending on the length of the queue, both for the morning and evening peak. In general, the system of equation B.1-B.5b still fully describes the system, but this time we have the same equations both for the morning and evening peak. The main difference is in the equation B.5c. While in the classic model considers only the scheduling delay, now it considers morning and evening scheduling delay together with the Activity Utility. We can define the scheduling delay as:

$$\begin{aligned} U^T = \alpha\left(T^M\left(t_h^d\right)\right) + \beta \cdot max\left(0; t_w^{a^0} - t_h^d - T^M\left(t_h^d\right)\right) + \gamma \cdot max\left(0; t_h^d + T^M\left(t_h^d\right) + t_w^{a^0}\right) \\ + \alpha\left(T^E\left(t_w^d\right)\right) + \mu \cdot max\left(0; t_w^{d^0} - t_w^d\right) + \lambda \cdot max\left(0; t_w^d - t_w^{d^0}\right) \end{aligned} \quad \text{(B.12}$$

Where:

t_h^d is the departure time from home to go to work in the morning.

$t_w^{a^0}$ is the preferred arrival time at work in the morning

$T^M\left(t_h^d\right)$ is the travel time given a certain departure time

$t_h^d + T^M\left(t_h^d\right)$ is the arrival time at work

t_w^d is the departure time from work to go home in the evening.

$t_w^{d^0}$ is the preferred departure time at work in the evening

$T^E(t_w^d)$ is the travel time given a certain departure time in the evening

$t_w^d + T^E(t_w^d)$ is the arrival time at work

The utility of performing an activity depends on by the time in which we begin that activity and the duration. Li et. al propose the following "time of the day" based utility function:

$$U^A = \int_0^{t_h^d} U_h(t)dt + \int_{t_h^d+T^M(t_h^d)}^{t_w^d} U_w(t^*)dt + \int_{t_w^d+T^E(t_w^d)}^{24} \widehat{U_h}(t)dt \tag{B.13}$$

Where:

U_h is the utility to stay home in the morning

U_w is the utility to stay at work

$\widehat{U_h}$ is the utility to stay home in the evening

The utility of work is function of t^*. According to Ettema and Timmerman [53] the activity utility is function of two elements, the time clock dependent utility, function of t, and the duration of the activity (t-t_w^a):

$$U_w(t^*) = U_w(t) \cdot (1-\zeta) + U_w(t - t_w^a) \cdot \zeta \tag{B.14}$$

Equation 14 means that for $\zeta = 1$, the utility of work is only dependent by the duration, while for $\zeta = 0$, the utility is only clock dependent. The final utility is a function of the departures time from home and from work:

$$U(t_h^d, t_w^d) = U^A - U^T \tag{B.15}$$

The elements of Equation B.15 are represented in Table B.1

	U^T		
Morning	$\alpha\left(T^M(t_h^d)\right)$ *Travel time*	$\beta \cdot max\left(0; t_w^{a^0} - t_h^d - T^M(t_h^d)\right)$ *Early Arrival*	$\gamma \cdot max\left(0; t_h^d + T^M(t_h^d) + t_w^{a^0}\right)$ *Late Arrival*
Evening	$\alpha\left(T^E(t_w^d)\right)$ *Travel time*	$\mu \cdot max\left(0; t_w^{d^0} - t_w^d\right)$ *Early Arrival*	$\lambda \cdot max\left(0; t_w^d - t_w^{d^0}\right)$ *Late Arrival*

Table B.1a: dis-Utility of traveling

U^A		
$\int_0^{t_h^d} U_h(t)dt$ *Stay at Home in the morning*	$\int_{t_h^d+T^M(t_h^d)}^{t_w^d} U_w(t^*)dt$ *Work*	$\int_{t_w^d+T^E(t_w^d)}^{24} \widehat{U_h}(t)dt$ *Stay at Home in the evening*

Table B.1b: Utility of performing an activity

Given Equation B.15, once the departure times t_h^d, t_w^d are defined it is possible to define his utility, and each user will try to maximize his own utility:

$$\max_{t_h^d, t_w^d} U\left(t_h^d, t_w^d\right) \quad \text{(B.16}$$

The solution of this problem, as for the standard Bottleneck model, is a Nash equilibrium in which each user maximize his utility and, at the equilibrium, all the users have the same utility. According to this equilibrium concept, the equilibrium condition is:

$$\begin{cases} q\left(t_h^d, t_w^d\right)\left[U\left(t_h^d, t_w^d\right) - U^*\right] = 0 \\ q\left(t_h^d, t_w^d\right) \geq 0, \left[U\left(t_h^d, t_w^d\right) - U^*\right] \geq 0 \end{cases} \quad \text{(B.17}$$

This equation recalls the famous variational inequality in transportation. Specifically, $q\left(t_h^d, t_w^d\right)$ is the demand for a certain combination of $\left(t_h^d, t_w^d\right)$, while $U^*$is the optimum value of the utility according to equation 15, which is the same for all users at the equilibrium. The set of equations B.12-B.17 represent an extension of the classic bottleneck model, where the daily based scheduling is considered. This system, exploiting equations 1-5, can be solved mathematically when utilities are constant, and iteratively when they are not.

In case of not constant utility functions, which is the general condition, the following MSA procedure is suggested:

Step 0: Choose a starting flow pattern $q\left(t_h^d, t_w^d\right)$ for each combination $q\left(t_h^d, t_w^d\right)$

Step 1: Calculate the net Utility according to equation B.15

Step 2: Obtain the auxiliary flow patterns $\hat{q}\left(t_h^d, t_w^d\right)$ by assigning all the demand to the best option (all or nothing assignment)

Step 3: Use MSA algorithm to obtain new demand flows:

$$q\left(t_h^d, t_w^d\right)^{i+1} = q\left(t_h^d, t_w^d\right)^i + \frac{1}{i+1}\left(\hat{q}\left(t_h^d, t_w^d\right) - q\left(t_h^d, t_w^d\right)^i\right)$$

Step 4: Check the convergence

They also show that the correlation between morning and evening commute exists only when the utility of the activity “work” is not constant.

Appendix C

Utility-Based DTA

This section introduces the appendices related to the results presented in Chapter 7.

C.1 Log-Likelihood

In this appendix, we test the proposed utility function (Equation 7.4-7.9) to evaluate their capability of reproducing a realistic behaviour. We exploit the data collected within the BMW (Behaviour and Mobility within the Week – Chapter 2) project.

We assume that our population behaves according to the bottleneck model introduced in sub-section 7.3.3, in the condition in which $N<S$ and $T^f \neq 0$. This assumption leads to a condition in which congestion does not occur, while users experience some travel time while travelling from A to B. Figure C1 shows the typical cumulative representation of departure/arrival time, if we assume that the users in our database travel on the network showed in Figure 7.2.

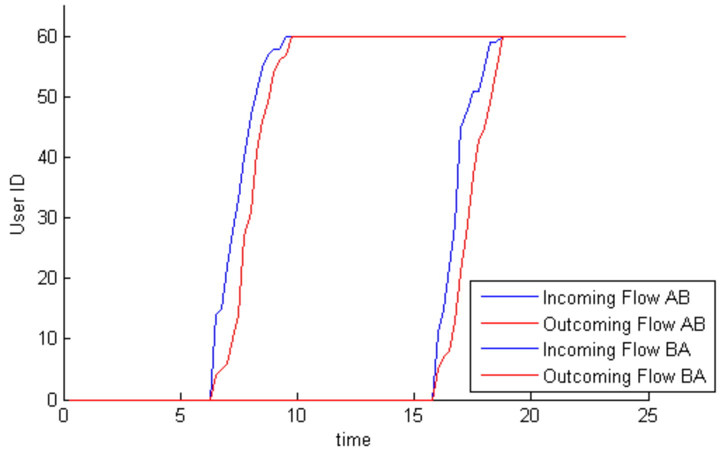


Figure C.1: Cumulative ingoing/outgoing flow derived from the data

After some data cleaning, and after selecting only the home-work-home commuting tours, we obtain an average of 60-80 users per day. The region of Ghent is not heavily congested, and, as shown in Figure C1, the assumption of having constant travel time should not lead to biased results for this database. For the working days, the following models are tested:

- Model1: All the activities Home-Work-Home are modelled through the *clock-based* utility function presented in equation (7.4). We assume a *rigid* activity scheduling for the users ($\tau_p = 0$);
- Model2: All the activities Home-Work-Home are modelled through the utility function presented in equation (7.9);
- Model3: All the activities Home-Work-Home are modelled through the utility function presented in equation (7.6);

Our goal at this stage is to estimate a realistic value for the utility function parameters. A common procedure to calibrate these models is the *maximum likelihood* approach. For each activity *p,* the model estimates the set of parameters $\boldsymbol{X}_p \ \{\beta_p, \alpha_p, \gamma_p, \tau_p, \dots\}$ that maximizes the probability of observing the choices made by the sampled users. In this work, we apply the *maximum likelihood* to estimate a realistic value for the utility function parameters. Then, we simulate the behaviour of a synthetic (homogeneous) population in order to check the difference with respect to the behaviour of the real (heterogeneous) population. It should be stressed that the goal of this appendix is not to perform an exhaustive calibration of the utility functions, since the dataset is too small, but to test the capability of the utility function of reproduce a realistic behaviour. The *maximum log- likelihood* algorithm applied in this study is the same presented in (Ettema and Timmermans 2003).

Table C1 shows the *Log-likelihood* results. *Model 1* represents the "standard" *clock-based* MU function introduced in Equation 7.4. Based on Table A1, the main conclusion is that considering both *clock-based* and *duration-based* MU (Model 2-3) leads to better results. However, for the proposed dataset, the difference in terms *log-likelihood* is not relevant. Thus, if we were to extrapolate from what the *log-likelihood* values, we would end up with expecting that a simple *clock-based* MU leads to reasonable results in most of the situations, although might not be the best solution.

However, the log-likelihood evaluates the capability of the model of reproducing the real choice for the observed users, when real travel time and departure time are given. Instead, we are interested in investigating how good these models are in reproducing a realistic behaviour when paired with a *homogeneous* – and unrealistic - class of users.

Table C1: Log-likelihood and ratio test ;

	Model1	*Model2*	*Model3*
Monday $L(\boldsymbol{X})^{Mon}$	***-192.395***	*-191.782*	*-192.996*
Tuesday $L(\boldsymbol{X})^{Tue}$	***-178.64***	*-175.438*	*-187.04*
Wednesday $L(\boldsymbol{X})^{Wed}$	***-173.345***	*-184.184*	*-176.98*
Thursday $L(\boldsymbol{X})^{Thu}$	***-139.991***	*-138.176*	*-146.605*
Friday $L(\boldsymbol{X})^{Fri}$	***-125.193***	*-133.238*	*-124.806*

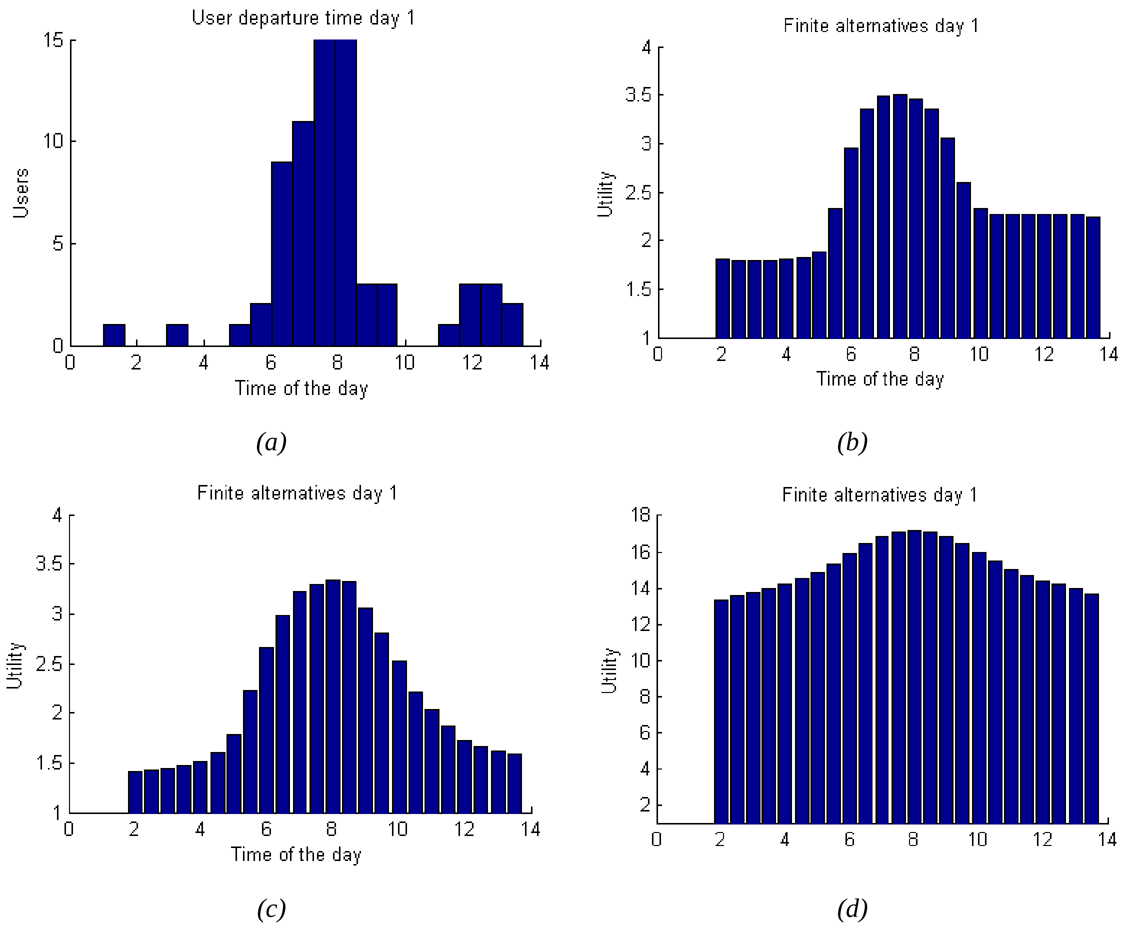


Figure C2: (a) observed departure time for real users; Utility related to different departure time interval according to Model 1 (b), Model 4 (c) and Model 5 (d).

We applied the estimated utility functions to model the perceived utility of a synthetic population, travelling on the network shown in Figure 7.2. We simulated the *home-work* commute for *N* users, which can choose their departure time for going work in the morning and returning home in the evening. These users will choose the departure time, which maximizes their own utility, thus we expect the profile of the utility over time to be similar to the profile of the departure time for the real population. Figure C2 shows the real departure time distributions for the observed users in our database (on Monday). Figures C2b-C2d show the time profile of the utility for a given (constant) travel time and preferred activity duration. As we can see, when Equation (7.9) is used to model the utility (Figure C.2c), the profile is similar to the expected departure time. If Equation (7.4) – Figure C2b- or Equation (7.6) – Figure C2d - are used the profile of the utility is more “symmetric”. The reason is that Equation (7.7) considers that different users might have a different duration for the same activity. In reality, we have that users have different departure times and different activity duration. *Clock-based* MU properly represents the utility of departing at a certain time for a single user, but they do not capture the dynamics related to activity duration. Moreover, if we use Equation (7.6) to model the utility, we might overestimate the utility during most of the time intervals. This is because of the second term of the equation –which takes into account the fatigue effect– is constant during all time intervals.

C.2 Demonstration of Proposition 1

Hereafter we provide the proof that Equation (7.22) is a necessary and sufficient condition to have a queue on the bottleneck, under the assumption that the last user can face congestion.

Necessity: First, we show that Equation (7.22a) is a necessary condition for having a queue. The Utility of the first user travelling from location 1 to location 2, which faces no congestion, can be calculated as:

$$U^{\varepsilon_1} = \int_{\varepsilon_1}^{t^0} U_2^a(t)dt - VoE(t^0 - \varepsilon_1) \qquad \text{(C2.1)}$$

While the overall utility for the user leaving at time $t = \hat{t}$ and arriving at the destination at time $t = \varepsilon_2$, will face both congestion and a late arrival penalty, can be calculated as:

$$U^{\varepsilon_2} = \int_{\varepsilon_1}^{\varepsilon_2 - T^b(\hat{t})} U_1^a(t)dt - VoL(\varepsilon_2 - t^0) - VoT(T^b(\hat{t})) \qquad \text{(C2.2)}$$

At the equilibrium, if the last user faces some congestion, the following equation has to hold:

$$\int_{\varepsilon_1}^{\varepsilon_2} U_2^a(t)dt - VoE(t^0 - \varepsilon_1) = \int_{\varepsilon_1}^{\varepsilon_2 - T^b(\hat{t})} U_1^a(t)dt - VoL(\varepsilon_2 - t^0) - VoT(T^b(\hat{t})) \qquad \text{(C2.3)}$$

By considering that:

$$\begin{aligned} &\int_{\varepsilon_1}^{\varepsilon_2} U_1^a(t)dt - VoL(\varepsilon_2 - t^0) \\ &\qquad > \int_{\varepsilon_1}^{\varepsilon_2 - T^b(\hat{t})} U_1^a(t)dt - VoL(\varepsilon_2 - t^0) - VoT\left(T^b(\hat{t})\right) \quad \forall\, T^b(\hat{t}) > 0 \end{aligned} \qquad \text{(C2.4)}$$

We can write that, at the equilibrium we have that:

$$\begin{cases} \int_{\varepsilon_1}^{\varepsilon_2} U_2^a(t)dt - VoE(t^0 - \varepsilon_1) < \int_{\varepsilon_1}^{\varepsilon_2} U_1^a(t)dt - VoL(\varepsilon_2 - t^0) \quad if\ T^b(\hat{t}) > 0 & \text{(C2.5)} \\ \int_{\varepsilon_1}^{\varepsilon_2} U_2^a(t)dt - VoE(t^0 - \varepsilon_1) = \int_{\varepsilon_1}^{\varepsilon_2} U_1^a(t)dt - VoL(\varepsilon_2 - t^0) \quad if\ T^b(\hat{t}) = 0 & \text{(C2.5)} \end{cases}$$

Since equation C2.5a holds only if $T^b(\hat{t}) > 0$, it is a necessary condition to observe congestion on the bottleneck. Equation (B5a) can be reformulated as Equation (22b):

$$\int_{\varepsilon_1}^{\varepsilon_2} \left(U_2^a(t) - U_1^a(t)\right)dt < VoE \cdot (t^0 - \varepsilon_1) - VoL \cdot (\varepsilon_2 - t^0) \qquad \text{(C2.6)}$$

Sufficiency: Equation C2.6 is a necessary but not sufficient condition. In fact, we might have that the first and last users do not face congestion, which satisfies Equation C2.5, but still have all the other users queuing at the bottleneck, as in the case described in (Li, Lam, and Wong 2014b). However, we can formulate a similar equation for any user travelling in the system. If we consider the user leaving in time interval $t = t^0 - T^b(\hat{t})$, we can calculate that at the equilibrium the following condition has to be satisfied:

$$\int_{\varepsilon_1}^{t^0} U_2^a(t)dt - VoE(t^0 - \varepsilon_1) = \int_{\varepsilon_1}^{t^0 - T^b(\hat{t})} U_1^a(t)dt - VoT(T^b(\hat{t})) \tag{C2.7}$$

Again, we can argue that, if $T^b(\hat{t})$=0, then:

$$\int_{\varepsilon_1}^{t^{t^0}} U_1^a(t)dt = \int_{\varepsilon_1}^{t^0 - T^b(\hat{t})} U_1^a(t)dt - VoT\left(T^b(t^0)\right) \tag{C2.8}$$

Which implies that if $T^b(t^0) > 0$ then

$$\int_{\varepsilon_1}^{t^0} U_1^a(t)dt > \int_{\varepsilon_1}^{t^0} U_2^a(t)dt - VoE(t^0 - \varepsilon_1) \tag{C2.9}$$

Which leads to:

$$\int_{\varepsilon_1}^{t^0} \left(U_2^a(t) - U_1^a(t)\right)dt < VoE(t^0 - \varepsilon_1) \tag{C2.10}$$

Equation C2.10 is a generalization of the one proposed in Lam, which was demonstrated to hold under the assumption that the first user travelling does not face congestion. Under the assumption that $T^b(\hat{t}) < T^b(t^0)\ \forall\ \hat{t}\ \in (\varepsilon_1, \varepsilon_2)$, Equation C2.7 contains Equation C2.6. However, in the case of special events, this condition does not hold (as in the example shown in Figure 7.4). Thus, in order to observe congestion, at least one of the two has to be satisfied.

C.3 Experiment Details

Hereafter, the details of the setting of each experiment and each utility functions are listed. Case 1 does not consider the utility at the destination, while for Case 2 we assumed constant utilities, specifically: $U_{Home-Morning} = 8$, $U_{Work} = 11$ and $U_{Home-Evening} = 10$. Table C2 shows the parameters for all the other cases.

Table C2	*Case 3*	*Case 4*	*Case 5*	*Case 6*	*Case 7*	*Case 8*	*Case 9*
Equation Home in the Morning	4	4	4	9	9	4	6
Equation Work Activity	4	4	4	9	9	5	6
Equation Home in the Evening	4	4	4	9	9	4	6
$U^{o}_{Home-Morning}$	8	8	8	8	8	8	130
U^{o}_{work}	11	11	22	11	22	11	160
$U^{o}_{Home-evening}$	10	10	10	10	10		150
$\alpha_{Morning-home}$	6.5	6.5	6.5	6.5	6.5	6.5	6.5
α_{work}	13	13	13	13	13		13
$\alpha_{Evening-home}$	19.5	19.5	19.5	19.5	19.5	19.5	19.5
$\beta_{Morning-home}$	0.02	0.02	0.02	0.02	0.02	0.02	1.5
β_{work}	0.01	0.01	0.02	0.01	0.02		1.5
$\beta_{Evening-home}$	0.02	0.02	0.02	0.02	0.02	0.02	1.5
$\gamma_{Morning-home}$	1	1	1	1	1	1	4
γ_{work}	1	1	1	1	1		4
$\gamma_{Evening-home}$	1	1	1	1	1	1	4
$\eta_{Morning-home}$							
η_{work}						18	
$\eta_{Evening-home}$							
$\tau_{Morning-home}$	0	0	0	0	0	0	
τ_{work}	0	0.5	0	0	0		
$\tau_{Evening-home}$	0	0	0	0	0	0	
$G_{Morning-home}$				0.2	0.2		
G_{work}							
$G_{Evening-home}$							
w						4	0.5

For all the above-mentioned cases, the same utility functions have been used for calculating the dis-utility of travelling. Details are reported in Table C3:

Table C3	α	β	γ	t^0	t_d^0
Morning Commute	*10*	*6*	*19*	*9*	
Evening Commute	*10*	*19*	*6*		*17*

Appendix D

Utility-Based OD Estimation: properties

This Appendix discusses the Properties of the Utility Based OD estimation.

D.1 Spatial Distribution of the demand

Equation (8.9) properly approximates the real gradient if the numerical perturbation c is small. For a given path r, the generalized route cost for departure time $t = \theta$ is gc_r^{θ} . This cost is function of the (dynamic) link flow, speed and density. The relation between link and OD flows is

$$f_l^{\theta^l} = \sum_{\theta=\theta^l}^{n_\theta} \sum_{od=1}^{n_{od}} M_{od,l}^{\theta^l,\theta}(\boldsymbol{x}) x_{od}^{\theta} \tag{D.1}$$

Where x_{od}^{θ} is the demand flow for a certain OD pair and a certain time interval and $f_l^{\theta^l}$ is the link flow on a specific link l during a certain time interval θ^l. The assignment matrix $M_{od,l}^{\theta^l,\theta}(\boldsymbol{x})$ can be further divided in *link-path* incidence matrix Δ and the route choice probability **Φ**:

$$M_{od,l}^{\theta^l,\theta}(\boldsymbol{x}) = \delta_{l,r}^{\theta^l,\theta} \cdot \phi_{r,od}^{\theta}(\, gc_r^{\theta}(\boldsymbol{x})) \tag{D.2}$$

Where $\delta_{l,r}^{\theta^l,\theta}$ is a proportional factor between 0 and 1 taking into account the dynamic component of the assignment and $\phi_{r,od}^{\theta}$ is the probability for the OD pair of using route r. Since the link flow is a function of the new generalized route cost gc_r^{θ}, some rerouting for all the OD pairs in the network passing through *l* will occur. This suggests that the numerical approximation overtakes the issue presented in (i). However, the effect on the link-flow depends directly on $\phi_{r,od}^{\theta}(\boldsymbol{x}, gc_r^{\theta})$, and specifically on how large is the set of OD pairs $od \in OD_l$, where OD_l is the set of OD pairs passing for link *l*. By assuming that the numerical perturbation is $c = \rho \cdot x_{od}^{\theta}$, the magnitude of the effect of the perturbation on network can be expressed as in Equation (8.10). It is intuitive to realize that for urban networks the set of OD pairs crossing the same link can be quite large. As a consequence, the only option to increase the impact of the perturbation on the network performance is to increase the parameter ρ, which however leads to an incorrect approximation of the gradient in Equation (8.9). On the other hand, by assuming an UB-DODE formulation, for a given value of ρ, the perturbation is larger.

D.2 UB-DODE: Proof or Property II

The MPRE as formulated in (Yang et al., 1991) is an OD coverage problem, which has been formulated for the static demand estimation, but can be applied to the dynamic case by assuming that the OD and link flows during different time interval are independent, which is usually the assumption behind the AB_DODE formulation. When the UB_DODE is applied instead of the standard AB_DODE, under the assumption of User Equilibrium (UE), any used set of departure times *t* leads to the same utility. Using the same notation as in the original paper, the MPRE is formulated as in Equations (D.3, 3a, 3b):

$$maximize\ (\boldsymbol{\lambda} \cdot \boldsymbol{\lambda^T}) \tag{D.3}$$

Subject to:

$$\boldsymbol{P^d X^d \lambda^T} = 0 \tag{D.3a}$$

$$\lambda_i \geq -1, \quad i = 1,2, \ldots . k \cdot n_{od} \tag{D.3b}$$

$$x_i^t [U(t) - U_i^*] = 0, \quad i = 1,2, \ldots . k \cdot n_{od} \ \ t = 1,2, \ldots, k \tag{D.3c}$$

Where, $\boldsymbol{P}$ is the matrix including the proportion of OD trips using each observed link on the network, $\boldsymbol{\lambda}$ is the column vector describing the relative deviation between estimated traffic volume and real traffic volume for each OD pair, and $\boldsymbol{X^d}$ is a diagonal matrix, whose i-j element is equal to x_i when i=j. By considering a utility-based departure time choice model, under the assumption of user equilibrium, we add the constraint D.3c to the MPRE. By assuming uncapacitated links, the contradiction method is applied by assuming Property 2 not to hold. This implies that the formulation automatically satisfies the constraint (D.3c). Under the assumption of having a continuous concave utility function, constraint (D.3c) implies that all the points within the MPRE solution will satisfy the following condition:

$$\begin{cases} x_{od}^{\theta} = 0 & \forall\, \theta < t_{start,od}, & \forall\, od \in OD \\ x_{od}^{\theta} > 0 & t_{start,od} \leq \theta \leq t_{end,od} & \forall\, od \in OD \\ x_{od}^{\theta} = 0 & \forall\, \theta > t_{end,od} & \forall\, od \in OD \end{cases} \tag{D.4}$$

Where $t_{start,od}$ and $t_{end,od}$ are the first and last time intervals in which one user has been observed travelling for a certain OD, respectively. According to MPRE, all the OD flows matching the observed flows are feasible solutions for the MPRE:

$$\hat{f}_l^{\theta^l} = f_l^{\theta^l} = \sum_{\theta=\theta^l}^{n_\theta} \sum_{od=1}^{n_{od}} M_{od,l}^{\theta^l,\theta}(\boldsymbol{x}) x_{od}^{\theta} \tag{D.5}$$

As a consequence, keeping the notations introduced in Appendix A, any values of $\phi_{r,od}^{\theta}$ that satisfy Equation (D.5) satisfy the system of Equations (B.3), including:

$$\begin{cases} \phi_{r,od}^{t_{start,od}-1} = 0 \\ x_{od}^{t_{start,od}-1} > 0 \end{cases} \tag{D.6}$$

Since $\phi_{r,od}^{t_{start,od}-1} = 0$, this solution matches the observed flow $\hat{f}_l^{\theta^l}$, while breaking the condition formulated in Equation (D.6). Since this is impossible, Property 2 holds.

D.3 UB-DODE: Proof or Property III and IV

Property 3: The MPRE (defined in Appendix D.2) has a finite value if and only if the trips between any OD pair are observed at at least one counting station. Constraint (D.3a) can be written as:

$$\boldsymbol{X^d}\boldsymbol{\lambda^T} = \begin{bmatrix} x_1^1 & & 0 \\ & \ddots & \\ 0 & & x_{n_{od}}^{n_\theta} \end{bmatrix} \begin{bmatrix} \lambda_1^1 \\ \vdots \\ \lambda_{n_{od}}^{n_\theta} \end{bmatrix} = \boldsymbol{HX} = \begin{bmatrix} \lambda_1^1 & & 0 \\ & \ddots & \\ 0 & & \lambda_{n_{od}}^{n_\theta} \end{bmatrix} \begin{bmatrix} x_1^1 \\ \vdots \\ x_{n_{od}}^{n_\theta} \end{bmatrix} \tag{D.7}$$

Where $\boldsymbol{H}$ is a square matrix $(n_\theta \cdot n_{od}) \times (n_\theta \cdot n_{od})$ containing the deviation λ_{od}^{θ}, as defined in (Yang et al., 1991). λ_{od}^{θ} describes the maximum relative deviation between estimated traffic volume and real traffic volume for the OD pair x_{od}^{θ}. If $\boldsymbol{H}$ is diagonal, as for the AB_DODE, then the maximum deviation for each OD flow in each time interval is independent. If, for instance, an OD pair is observable for n_θ-1 time intervals, then the MPRE goes to infinite. However, if a relation exists between OD flow belonging to different time intervals –i.e. the DTM – than the flow during the missing time interval can be estimated. We can now define $\lambda_{\text{od}}^{\theta^1|\theta}$ the maximum error for an OD pair “od” with departure time θ^1, which is observable through the flow belonging to the same spatial OD pair, but with departure time θ. Simply stated, $\lambda_{\text{od}}^{\theta^1|\theta}$ represents the co-variance of the error over different time intervals for the same OD pair $x_{od}^{\theta^1}$, given a certain utility function.

$$\boldsymbol{H_{T\_DODE}} = \begin{bmatrix} \begin{bmatrix} \lambda_1^1 & & 0 \\ & \ddots & \\ 0 & & \lambda_{n_{od}}^1 \end{bmatrix} & \begin{matrix} \lambda_1^{1|\theta} & & 0 \\ & \ddots & \\ 0 & & \lambda_{n_{od}}^{1|\theta} \end{matrix} & \begin{matrix} \lambda_1^{1|n_\theta} & & 0 \\ & \ddots & \\ 0 & & \lambda_{n_{od}}^{1|n_\theta} \end{matrix} \\ \begin{matrix} \lambda_1^{\theta|1} & & 0 \\ & \ddots & \\ 0 & & \lambda_{n_{od}}^{\theta|1} \end{matrix} & \begin{bmatrix} \lambda_1^{\theta} & & 0 \\ & \ddots & \\ 0 & & \lambda_{n_{od}}^{\theta} \end{bmatrix} & \begin{matrix} \lambda_1^{\theta|n_\theta} & & 0 \\ & \ddots & \\ 0 & & \lambda_{n_{od}}^{\theta|n_\theta} \end{matrix} \\ \begin{matrix} \lambda_1^{n_\theta|1} & & 0 \\ & \ddots & \\ 0 & & \lambda_{n_{od}}^{n_\theta|1} \end{matrix} & \begin{matrix} \lambda_1^{n_\theta|\theta} & & 0 \\ & \ddots & \\ 0 & & \lambda_{n_{od}}^{n_\theta|\theta} \end{matrix} & \begin{bmatrix} \lambda_1^{n_\theta} & & 0 \\ & \ddots & \\ 0 & & \lambda_{n_{od}}^{n_\theta} \end{bmatrix} \end{bmatrix} \tag{C.2}$$

In general, we can say that the observability of the OD flows increases with the number of non-zero elements in the $\boldsymbol{H}$**.** In fact, the more non-zero elements we have in $\boldsymbol{H}$, the less non zero elements in $\boldsymbol{P}$ we need to avoid that $\lambda_{od}^{\theta} = \infty$ satisfies Constraint (D.3a). This proves Property 3 to hold.

Property 4: If activity duration is considered, as in Equation (8.15), a correlation between demand flows belonging to different spatial OD exists. Given this condition, we can further extend the matrix $\boldsymbol{H}$, by considering the deviation $\lambda_{od^1|od}^{\theta^1|\theta}$. This element represent the deviation related to the OD pair $x_{od^1}^{\theta^1}$ observable through the OD pair x_{od}^{θ}. In general, $\lambda_{od^1|od}^{\theta^1|\theta}$ represents the covariance in time and space for different OD pairs, and $\lambda_{od^1|od}^{\theta^1|\theta} \neq 0$ if the two OD pairs belong to the same activity pattern ν. Since the number of non-zero number in the matrix $\boldsymbol{H}$ increases, Property 4 holds.

Author’s Publications

Journal Articles:

Cantelmo, Guido, Francesco Viti, Ernesto Cipriani, and Marialisa Nigro. 2017. “A Utility-Based Dynamic Demand Estimation Model That Explicitly Accounts for Activity Scheduling and Duration.” *Transportation Research Part A: Policy and practice (***in press***)*

Cantelmo, Guido, and Francesco Viti. 2016. “Effects of Incorporating Activity Duration and Scheduling Utility on the Equilibrium-Based Dynamic Traffic Assignment,” Transportation Research Part B: Methodological journal (**Second round of review – more than 300 days)**

Cantelmo, G., E. Cipriani, A. Gemma, and M. Nigro. (2014). “An Adaptive Bi-Level Gradient Procedure for the Estimation of Dynamic Traffic Demand.” *IEEE Transactions on Intelligent Transportation Systems* 15 (3): 1348–61. doi:10.1109/TITS.2014.2299734.

Cantelmo, Guido, Francesco Viti, Chris Tampêre, Ernesto Cipriani, and Marialisa Nigro. (2014). “Two-Step Approach for Correction of Seed Matrix in Dynamic Demand Estimation.” *Transportation Research Record: Journal of the Transportation Research Board (December): 125–33.* doi:10.3141/2466-14.

Peer reviewed conference papers:

Cantelmo, Guido, and Francesco Viti. (2018). “Assessing the Performances of a Two-Step Prediction Model on a Large Scale Congested Network Using Real Traffic Data.” *Transportation Research Board 2018.*

Cantelmo, Guido, Francesco Viti, Ernesto Cipriani, and Marialisa Nigro (2017). “A Utility-Based Dynamic Demand Estimation Model That Explicitly Accounts for Activity Scheduling and Duration.” *Transportation Research Procedia, Papers Selected for the 22nd International Symposium on Transportation and Traffic Theory Chicago, Illinois, USA, 24-26 July, 2017., 23 (Supplement C): 440–59.* doi:10.1016/j.trpro.2017.05.025.

Cantelmo, G., F. Viti, and T. Derrmann. (2017). “Effectiveness of the Two-Step Dynamic Demand Estimation Model on Large Networks.” In *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS),* 356–61. doi:10.1109/MTITS.2017.8005697.

Cantelmo, Guido, and Francesco Viti. (2016). “Effects of Incorporating Activity Duration and Scheduling Utility on the Equilibrium-Based Dynamic Traffic Assignment,” *Proceeding of the 6th Symposium on Dynamic Traffic Assignment*

Cantelmo, G., F. Viti, E. Cipriani, and N. Marialisa. (2015) "A Two-Step Dynamic Demand Estimation Approach Sequentially Adjusting Generations and Distributions." *In 2015 IEEE 18th International Conference on Intelligent Transportation Systems, 1477–82.* doi:10.1109/ITSC.2015.241

Cantelmo, G., F. Viti, E. Cipriani, and N. Marialisa. (2015). "A Two-Step Dynamic Demand Estimation Approach Sequentially Adjusting Generations and Distributions." *Transportation Research Board 2015*.

Cantelmo, G., F. Viti, M. Rinaldi, C. M. J. Tampère, M. J. Smith, and K. Nökel. 2015. "Systematic Assessment of Local Amp; Global Signal Control Policies: A Methodological Perspective." In *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 156–63. doi:10.1109/MTITS.2015.7223251

Cantelmo, G., F. Viti, E. Cipriani, and N. Marialisa. (2015). "A Two-Step Dynamic Demand Estimation Approach Sequentially Adjusting Generations and Distributions." In *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, 1477–82. doi:10.1109/ITSC.2015.241.

Cantelmo, Guido, Francesco Viti, Chris Tampêre, Ernesto Cipriani, and Marialisa Nigro. 2014. "Two-Step Approach for Correction of Seed Matrix in Dynamic Demand Estimation." Transportation Research Board 2014

Sceffer, Ariane, Guido Cantelmo, and Francesco Viti. 2017. "Generating Macroscopic, Purpose Dependent Production Factors through Monte Carlo Sampling Techniques." *Transportation Research Procedia*, 19th Euro Working Group on Transportation, EWGT 2017, 17-19 September 2018, Budapest, Hungary.

Viti, Francesco, Guido Cantelmo, Francesco Corman, and Marco Rinaldi (2015). "Improving the Reliability of Demand Estimation Using Traffic Counts by Including Information on Link Flow Observability." *INSTR2015* Conference Proceedings doi:10.3929/ethz-b-000183287.

Cipriani, E., A. Del Giudice, N. Marialisa, F. Viti, and G. Cantelmo. (2015). "The Impact of Route Choice Modeling on Dynamic OD Estimation." In *2015 IEEE 18th International Conference on Intelligent Transportation Systems (ITSC)*, 1483–88. doi:10.1109/ITSC.2015.242

Donna, S. A. Di, G. Cantelmo, and F. Viti. 2015. "A Markov Chain Dynamic Model for Trip Generation and Distribution Based on CDR." In *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 243–50. doi:10.1109/MTITS.2015.7223263

Hadjidimitriou, S. N., M. Dell'Amico, G. Cantelmo, and F. Viti. 2015. "Assessing the Consistency between Observed and Modelled Route Choices through GPS Data." In *2015 International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 216–22. doi:10.1109/MTITS.2015.7223259.

Presentation at international conferences:

Cantelmo, Guido, Francesco Viti, E. Cipriani, and M. Nigro. (2016). "A Utility-Based Framework for Macroscopic Dynamic Demand Estimation." In *5rd HEART Conference 15-16th September Delft, the Netherlands*.

Cantelmo, Guido, and Francesco Viti. (2015). "Activity Demand: An Empirical Analysis on the Influence of the Activities on the Traffic within-Day Demand Profile." In *4rd HEART Conference 9-11th September Copenhagen Denmark*.

Cantelmo, Guido, Francesco Viti, and Chris MJ Tampere. 2014. "Exploiting the Relation between Activity Data and Traffic Data within the Dynamic Demand Estimation Problem." In *3rd HEART Conference 10-12th September Leeds UK*.