A Within Day Dynamic Gravity Model that Integrates Data Driven and Time Geography approaches

Filippo Carrese
Matricola 1526949

Relatore   Correlatore
Gaetano Fusco   Francesco Viti

_____

Preface

For decades, researchers and practitioners have commonly relied on socio-demographic statistics and actively collected data to infer human mobility. For example, the gravity model predicts people movements based on coarse statistics about the population. However, this approach has the tendency to oversimplify human behaviour and to predict biased mobility patterns. Another option is to use Activity Based models to recreate individual mobility patterns. While this approach provides predictions that are more realistic, it relies on individual data, which are usually limited in sizes.

Recent advances in sensing technology and data collection offer an unprecedented opportunity to enhance travel demand models by leveraging more complete and abundant information. Regardless of their source, these data can be broadly classified into two main categories: small and passive data. The first class includes information at a user level that is actively collected through self-reporting processes and usually captures only a sample of the population. The second involves the collection of an immense amount of information (big-data), which is less accurate but more representative of the entire population.

The goal of this paper is to bridge the fundamental gap between individual and aggregate travel demand models by exploiting the wide range of GIS (Geographic Information Systems) data freely available on the internet. Specifically, we propose to use a conventional gravity model to capture the home-work commute and then to exploit temporal and spatial constraints derived from spatial geography to calculate the most likely distribution of secondary activities. The choice for leisure activities is thus based on topological characteristics of the network and some initial guess of the individual human behaviour for the given population.

In the proposed case study, we use activity patterns observed in Belgium to model human mobility in Luxembourg and, in addition, to test the sensitivity of the model, substantial changes have been applied to the inputs, simulating a workplace relocation, and obtaining consistent results.

_____

Index

_____

List of Figures

1. Introduction

1.1.   Transport Modelling

In the last decades the development of technology and globalization has caused a strong advancement in accessibility of urban realities, generating a notable rise in mobility demand. This, in addition to the growth of urban population, has led to always more congested cities, with severe economic and environmental repercussions on users and communities. In the same time period, the development of the infrastructural network did not manage to follow the same pace, and since it is unlikely it ever will, the efforts to a solution to these problems are always more directed to a more adequate operation management (Cascetta, 2009). In Transportation Planning, in order for the modeller to intervene in the best possible way, it is necessary not only to predict and estimate actual traffic states with significant accuracy, but also to see the effects that any modifications or interventions will have on them. For an optimal management and planning, modellers can rely on decisional support tools known as Traffic Assignment Models.

TA models are very sensitive to their inputs, even a small variation may generate relevantly different results, due to the delicateness of the phenomena these models want to simulate. The essential inputs are the network and the mobility demand, and if any of them do not respect reality TA will probably lead to biased congestion patters. Since achieving the network is not a complex task, estimating the demand is the most influential factor for precise modelling. Typically, mobility demand is represented as a set of Origin-Destination (OD) demand matrixes, where each cell will contain information about the number of trips from one traffic zone to another. One OD matrix can be related to a certain time period, a set of users, a mode of transport and even more, depending on the desired representation. To have the dynamic case, time is usually discretised in a determined number of intervals, to each of which a matrix is dedicated.

Measuring directly the mobility demand is normally not a pursuable direction, since it implies enormous resources: this is why Demand Modelling has achieved such a momentum in Transportation Engineering. Classical modelling, also known as Four-Step Modelling, tries to estimate travel demand as an outcome of aggregate attributes of each zone. The characteristics of this approach will be discussed in section 2.1.

However, this process is not able to account for the complexity of mobility patterns, which reflect traffic flow variability and congestion issues, which may appear in different points in time and space. Travel behaviour is influenced not only by commuting trips, on which most of the work focuses, but also in large measure by the presence of secondary activities, how the interact between them and between commuting. Trip-Based approaches often overestimate the commuting demand, which instead is only around 20% of the overall mobility demand. Activity-Based models try to overcome these limitations estimating demand as an outcome of the need of users to participate in different activities: in this case secondary activities play a fundamental role in the determination of travel demand.

_____

1.2.   BigData

For many decades, transport research has been mainly based on actively collected data, such as travel survey, and census information. With the late exponential growth of technology, ubiquitous sensing of human mobility flows became possible through mobile phones, public transport smart card transactions or Global Positioning System (GPS)-enabled devices (Anda, Erath, & Fourie, 2017). These Big Data sources allow to observe mobility at an incredibly higher level of detail, and overcome the difficulties and the costs of gathering large scale samples for the development of travel demand models.

However, since these new sources are not actively collected, the available information must be understood, contextualized, and accordingly inferred and processed in order to be valuable for demand modelling. To achieve this, according to (Chen C. M., 2016), the joint work of travel behavior researchers, who have long relied on household travel surveys, and big data researchers, has the potential to drive fundamental advances in human mobility studies. The paper stresses that cross-disciplinary research should concentrate on three areas: identification of behavioural factors, modelling of travel behaviour, and recognition of human mobility patterns.

1.3.   Objective

Considering the difficulties of estimating travel demand in terms of individual behaviours, and the cost at which the generally required data is available, the following work emphasizes the need of directing transport research by chasing the opportunities that the development of technology has brought to.

The scope of this thesis is to estimate travel demand accounting for activity and spatial behaviours by exploiting the wide range of possibilities that can be found on internet. The proposed methodology relies on opensource data, network topology and spatial relationships to develop a within a day dynamic gravity model that includes activity tours and captures behavioural aspects in terms of spatial dependancies.

In addition, one of the goals that have inspired the following work has been to exploit the features of the two fundamental modelling approaches presented in chapter two. On one side, the aggregated macroscopic approach of gravity models, which is more suited for simulating flows and for demand modelling of large congested networks, and, on the other, the fine-grained properties of activity based models for modelling trip chains and activity-travel behaviour. While an intensive work has been done to achieve a more comprehensive understanding of human behaviour, we believe that the difference between aggregate and disaggregate models is still significant. Disaggregate approaches requires an immense volume of small data to provide realistic predictions, while aggregate approaches often provide a representation of the demand which is too coarse. The methodology presented in this paper tries to fill this gap by providing an aggregate representation of the demand that still takes into account individual mobility on a daily basis.

_____

_____

1.4.   Summary

The following thesis is structured in five chapters. The first one is the introduction, it ends with this paragraph and it aimed to present the context in which the work started in the first place. Chapter two is dedicated to a literature Review: two big families of demand modelling are introduced, discussing the theorical fundamentals that underline all the assumptions and considerations that hold this thesis. In Chapter three the methodology is finally presented, structured in three parts and with some theorical premises from which the model begins. In order to test the model on a real case and have a feedback on its accuracy, in chapter four it will be applied to estimate the total demand of Luxemburg city, and to test the sensitivity of the model, a new scenario with a workplace relocation will be proposed, analysing all the results. The effects of considering secondary activities will be compared to Chapter 5 concludes the thesis with a discussion on the results and an overview of the future developments.

2. Literature Review

We aim in this section to provide an overview of the theory underlying the methodology presented in this thesis by describing some relevant aspects of the state of the art of demand modelling.

Travel demand models provide information about the need for transportation of a given population around a chosen study area; they are used to predict travel characteristics and usage of transport services under alternate socio-economic scenarios, and for alternative transport service and land use configurations (Bhat & Koppelman, 1999). Forecasting traffic flows, they are an essential tool to support the decision-making process.

This review is structured as follows. First, all the steps of conventional demand modelling will be discussed, and particular attention will be paid to Gravity Models and their state of the art; the second part of this section will be dedicated to Activity Based Approaches, with a special focus on synthetic populations and space-time behaviour. To conclude, a list of key points summarizes all the notions and concepts that have inspired this methodology.

2.1. Conventional Models

During the late 1950s, the pioneering research and experimentation developed the fundamentals of transport modelling (Bates J. , 2007), which was destined to remain the "classic" transport model. It quickly assumed the structure of the so-called Four Step Model, due to its sequential procedure that was introduced piecewise (McNally, 2007a). Even though in the past 60 years it has been significantly enhanced, the framework deriving from those years has remained more or less unaltered (Ortúzar & Willumsen, 2011). These models follow a trip-based approach, which relies on the simplified assumption that the choices concerning to each trip are made independently of the choices for other trips made by the same user within the same journey (Cascetta, 2009).

The Four Step Model predicts travel demand by aggregating trips with given characteristics into average trip flows between zones, and decomposing global demand into a product of submodels, each of which relates to one or more choice dimensions (Cascetta, 2009). The sequence in which it is presented (Figure 1) addresses four specific questions, one for each stage of the model: how many travel movements will be made, where will they go, by what mode will the travel be carried out, and what route will be taken? (Bates J. , 2007)

The first trip generation model estimates the number of trips produced and attracted by each zone, and effectively serves to scale the problem. Next, trip distribution models allocate the generated trips between pairs of zones, reflecting travel impedance and yielding Origin-Destination matrixes. The third step is mode choice, where trip tables are divided according to the proportions of use of alternative modes. In the last submodel, assignment is performed for the identified demand, introducing route choice issues and congestion effects. The delays deriving from congestion may affect not only the route choice model, but have influence also on mode choice and trip distribution. Feedback loops are usually introduced for these

stages, so as to address the issues regarding the fixed sequential order in which stages are applied (McNally, 2007a).
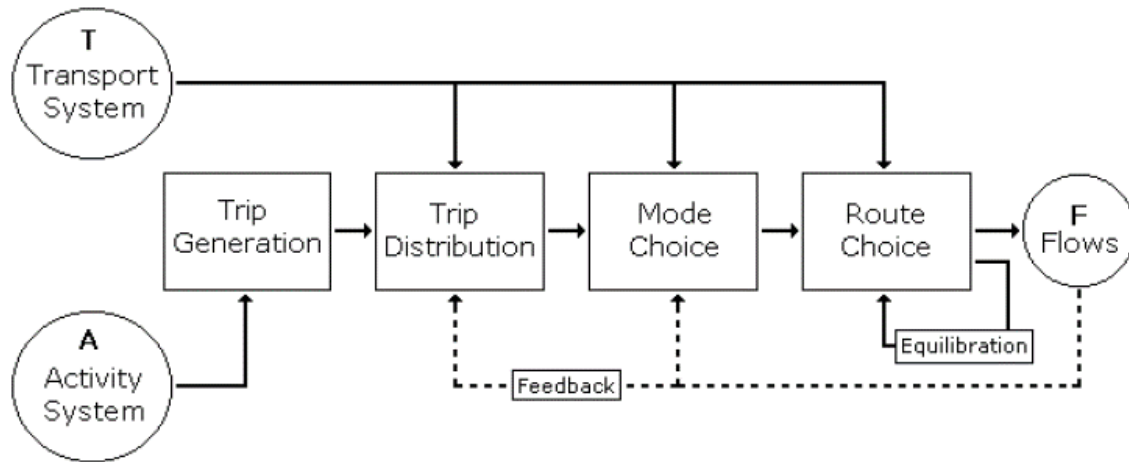
Figure 1: McNally: The Four Step Model

Data for Four Step modelling mainly consists in household travel surveys, with travel-activity diaries and census data, along with a representation of transportation networks. These data, and observed traffic count and speed, provide also much of the data needed for the validation of the representativeness of the sample (McNally, 2007a).

Despite the fortune that this classic modelling approach has encountered, it was clear from the beginning that the derived nature of the demand for transportation was understood and accepted, but not reflected by it (McNally, 2007b). It is generally recognized that travel decisions are not actually taken in this type of sequence, since attention is concentrated only a limited range of travellers' responses (Ortúzar & Willumsen, 2011). The focus on individual trips ignores the spatial and temporal interrelations between the different trips of an individual activity pattern, misrepresenting overall behaviour as an outcome of a choice process, rather than as a differentiated set of choices delimitated by various constraints (McNally, 2007b).

2.1.1   Gravity Models

Due to the complexity of the interactions between origins and destinations, distribution models have been a main focus of transportation research, but still represent a point of weakness in the Four Step Model, and while this weakness has always been recognized, it has probably been underestimated (Bates & Dasgupta, 1990). The objective of distribution models is to recombine and match the trip ends from the generation process into trips with origin and destination, expressing the probabilities of reaching a certain destination given an origin.

Distribution models of a different kind have been developed to assist in forecasting future trip patterns when important changes in the network take place. They start from assumptions about group trip making behaviour and the way this is influenced by external factors such as total trip ends and distance travelled (Ortúzar & Willumsen, 2011). Gravity models are the most known distribution models; rather than

providing the destination shares, their modelling process can be seen as the prediction of a trip matrix consisting of elements T_{od}, where the number of trips from o to d are likely to be related to: (Bates J. , 2007)

- *The characteristics of the origin/production zone o,*
- *The characteristics of the destination/attraction zone d,*
- *The characteristics of the "separation", or "cost" of travel, between zones o and d.*

Gravity models originally derived their name according to these three features, since they suggested the main factors of the first formulations:

$$T_{od} = a_o \cdot b_d \cdot f_{od}$$

where f_{od} can be characterised as the inverse of the square of the distance between *o* and *d*. Even though this formulation changed over the years, the initially obvious analogy with Newton's law of universal gravitation established the name with which they are most known nowadays.

Cascetta (2009) expresses its typical formulation as:

$$d_{od}[sh] = \alpha \cdot d_o[sh] \cdot d_d[sh] \cdot f(C_{od})$$

where α is a constant, $d_o[sh]$ and $d_o[sh]$ represent, respectively, the total trip production from o and total trip attraction to d for purpose s in period h. C_{od} represents the notion of generalized transportation cost and $f(C_{od})$ is an impedence function that decreases with C_{od}. Popular versions for this function are: (Ortúzar & Willumsen, 2011)

$$f(C_{od}) = \exp(-\beta C_{od}) \qquad \text{exponential function}$$

$$f(C_{od}) = C_{od}^{-\beta} \qquad \text{power function}$$

$$f(C_{od}) = C_{od}^{-\beta} \exp(-\beta C_{od}) \qquad \text{combined function}$$

These models can be increasingly developed by a better fitting impedance function, or by further defining α. The substitution of α with two factors, which depend on the origin and destination zones, gives rise to the doubly constrained gravity model:

$$d_{od}[sh] = A_o \cdot B_d \cdot d_o[sh] \cdot d_d[sh] \cdot f(C_{od})$$

where the following equations are mutually dependent and A_o and B_d can be solved by an iterative procedure.

$$A_o = 1/\sum_{d\prime} B_{d\prime} \cdot d_{d\prime}[sh] \cdot f(C_{od\prime}) \qquad\qquad B_d = 1/\sum_{o\prime} A_{o\prime} \cdot d_{o\prime}[sh] \cdot f(C_{o\prime d})$$

The singly constrained gravity model comes from only one of the two equations being verified, such as $B_d = 1$ and

$$d_{od}[sh] = \frac{d_o[sh] \cdot d_d[sh] \cdot f(C_{od})}{\sum_{d\prime} d_{d\prime}[sh] \cdot f(C_{od\prime})}$$

Gravity models have also been furtherly developed in many directions. One is surely the Entropy Maximising Approach (Wilson, 1974), that is generally acknowledged as one of the important contributions to improved modelling in transport (Ortúzar & Willumsen, 2011). A stochastic process which captures local

_____

mobility decisions (Simini, González, Maritan, & Barabási, 2012) was implemented in order to overcome the need of traffic data for the validation of the model's parameters. It analytically derives commuting and mobility fluxes requiring as input only information on the population and employment opportunities distribution. Contrarily, other studies have concentrated in particular on the use of Big Data. Di Donna et al. (2015) has exploited Call Detail Records to dynamically model trip generation and distribution among different zones within the day. The modelled variations of CDR in time and space have been used as a proxy for trip distribution. Kings et al. (2009) instead have worked on the anonymous communication patterns between 571 cities in Belgium. They arrived to show that inter-city communication intensity is characterized by a gravity model: the communication intensity between two cities is proportional to the product of their sizes divided by the square of their distance.

2.2.   Activity-Based Models

The link between travel and activities was firstly established by Mitchell & Rapkin (1954), who also called for a comprehensive framework and inquiries into travel behaviour. The pioneering work of Chapin (1974) and Hägerstraand (1970) based this approach on the analysis of travelactivity patterns. Chapin focused on understanding the factors that lead individuals to participate in various activities, by analysing activity patterns at the level of daily routines, while Hägerstrand directed his work in the determination of constraints that limit individuals' opportunities in time and space (Ettema, 1996). Their works will be better discussed in the following section for their importance in space-time relations.

Activity-Based Models were introduced to overcome the limitations of the Four Step Model, since looking at trips independently misses some of the behavioural richness of linking activities in different locations and with different time windows and constraints (Ortúzar & Willumsen, 2011). The interest in these models is due to their capability of representing the complex trip making behaviour that is always more frequent in urban realities. Travel is considered as a derived demand, manifested by the necessity of participating in activities that take place at different times and in different locations. This permits a better representation of travel behaviour, since all trips are considered in the context of activity patterns, and decisions regard a higher number of choice dimensions in respect to trip-based models (Ettema, 1996). Travel demand is estimated in function of the need of users to participate in different activities, where the relevant unit of analysis are the sequences, or patterns, of activities, and not individual trips (McNally, 2007b).

Even with the absence of a widely accepted general framework for Activity-Based Models (McNally, 2007b), Ben-Akiva and Bowman (1998) distinguish two most common applications. Econometric activity-based models focus on representing the choice process in a multidimensional way and are based on random utility theory: the choice probabilities result from a system of mathematical expressions and systematic utility functions. Alternatively, simulation models work by generating a discrete choice set, aiming to reduce it as much as possible through the implementation of sequential decisional rules, aiming to predict the results of the decision-making process.

In general, activity-based models are implemented at the level of individual households, emphasising on the household as the decision-making unit (Ettema, 1996). Representing household dynamics and

_____

_____

characteristic is necessary to determine feasibility constraint and interpersonal linkages that influence activity patterns. Practical examples regard allocation of resources, such as car ownership, or appointments, such as having to eat at the same time. This requires a highly detailed information about geographic and socioeconomic characteristics of the households, and often determines the need of the generation of a synthetic population, to which the following section is dedicated.

However, despite the strong emphasis of Activity-Based Models on household behaviour, real activity scheduling is still hard to achieve with microsimulation tour-based models using a random utility choice-modelling framework (Ortúzar & Willumsen, 2011). Trying to include activity decisions spanning a day or more is difficult: the variety of available schedules is immense, and the factors underlying household decisions are still not well understood (Bowman & Ben-Akiva, 2001). Accordingly, state of the art under this aspect is still at experimental level, where the most promising research consists in Computation Process Models (Ortúzar & Willumsen, 2011).

The following paragraphs are dedicated to two aspects of activity-based modelling which had a strong influence in the determination of the methodology presented in this thesis: population synthesis and space-time behaviour.

2.2.1   Synthetic Population

The generation of a Synthetic Population is not exclusively used for activity-based modelling. It is needed for a wide range of microsimulation models who necessitate extensive disaggregated data, that is prohibitive to collect directly due to the significant cost of travel surveys, and the restrictive privacy policies in certain countries. Population synthesis consists in generating an artificial population by expanding disaggregate sample data to match known aggregate distributions of household and person characteristics, typically extracted from existing census sources (Ortúzar & Willumsen, 2011).

A synthetic population can be generated either through Synthetic Reconstruction (SR), or with Combinatorial Optimization (CO) (Barthelemy & Toint, 2013). In the SR method two steps are involved: first the attributes' joint distribution is estimated preserving the known marginal distribution; then agents from the sample are assigned to the population generally using an Iterative Proportional Fitting Procedure (IPFP) according to the previously computed distributions. With IPFP, a Contingency Table (CT) is estimated using an iterative procedure to minimize the deviation between estimated and observed marginal distributions. The CO method, which is not as common, divides the area of interest in zones of known marginal distributions, and the sample is fitted for each zone according to the given set of margins. Selecting iteratively a combination of househoulds from the sample, replacing the chosen set only if an improvement in the goodness of fit is observed (Saadi, Mustafa, Teller, Farooq, & Cools, 2016). During the last years, a third family of population synthesis has emerged with outstanding results: Monte Carlo Markov Chain (MCMC) simulations (Sceffer, Cantelmo, & Viti, 2017).

However, all approaches rely on very strong assumptions on the requirements for data, and research is trying to obviate this limitation. Barthelemy and Toint (2013) managed to overcome the need of a Sample developing a Synthetic Reconstruction Method, while Saadi et al (2016) presented an extended Hidden Markov model, which is able to reproduce the structural configuration of a given population from an unlimited number of micro-samples and a marginal distribution.

_____

_____

2.2.2   Space - Time Behaviour

Since the 1960s, research in geography and urban planning has studied how human activity patterns are associated to transportation, focusing on the central role of the individual, and developing what will from there be known as behavioural research.

One of the first researchers to recognize the importance of time and space in urban planning was Chapin (1974). His theory firmly stresses the importance of activity patterns, recognising the fundamental role that space and time play in their definition. One key factor is opportunity, which reflects the role of time and space in the definition of activity choice probabilities, underlining how the location and availability in time of determined facilities influence the decision-making. He also introduces the concept of appropriateness of timing and circumstances, which may depend on many factors, going from past involvements in that certain activity, to commitments made to participate in that or other activities, concluding with opening hours of the visited places. It also takes into account that the activity must be performed for its full duration, accordingly to the existence of other activities.

However, Chapin's theory does not process the spatial and temporal dimensions in the same framework. This issue was therefore addressed in the space-time geography of the Lund school, led by Hägerstraand (1970). Space-time geography systematically explores the opportunity to unfold activity patterns in a specific spatio-temporal environment. It is assumed that time and space are scarce goods and that, consequently, daily activity patterns are largely determined by space-time constraints (Ettema, 1996). It is possible to visualize the interdependency between time and space in the daily schedule of a user in a three-dimensional space (Figure 2).
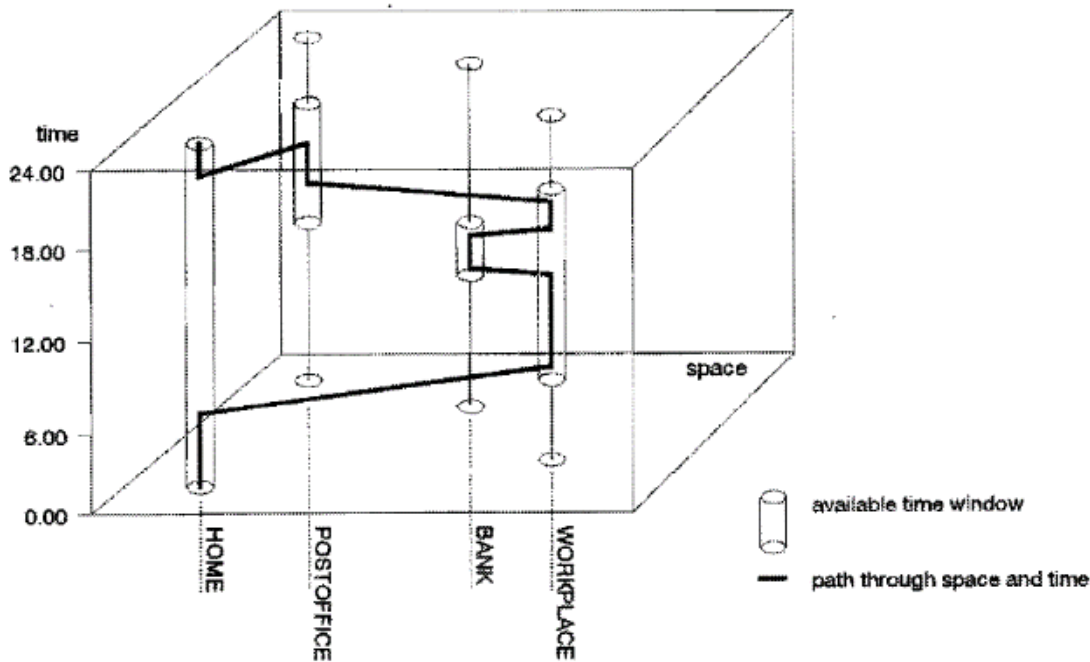
Figure 2: Example of a Space-Time Path (source: Ettema, 1996)

_____

However, the space time path of each user is subject to three types of limitations:

- Capability constraints refer to physical limitations, such as the need to sleep at home or the maximum speed a vehicle can travel to.
- Coupling constraints define where, when and for how long the individual has to join other individuals, tools and materials in order to produce, consume and transact (Hägerstrand, 1970)
- Authority constraints, finally, define the periods at which certain facilities or activities will be available.

The effects of these constraints result in the definition of the time-space prism (Figure 3).
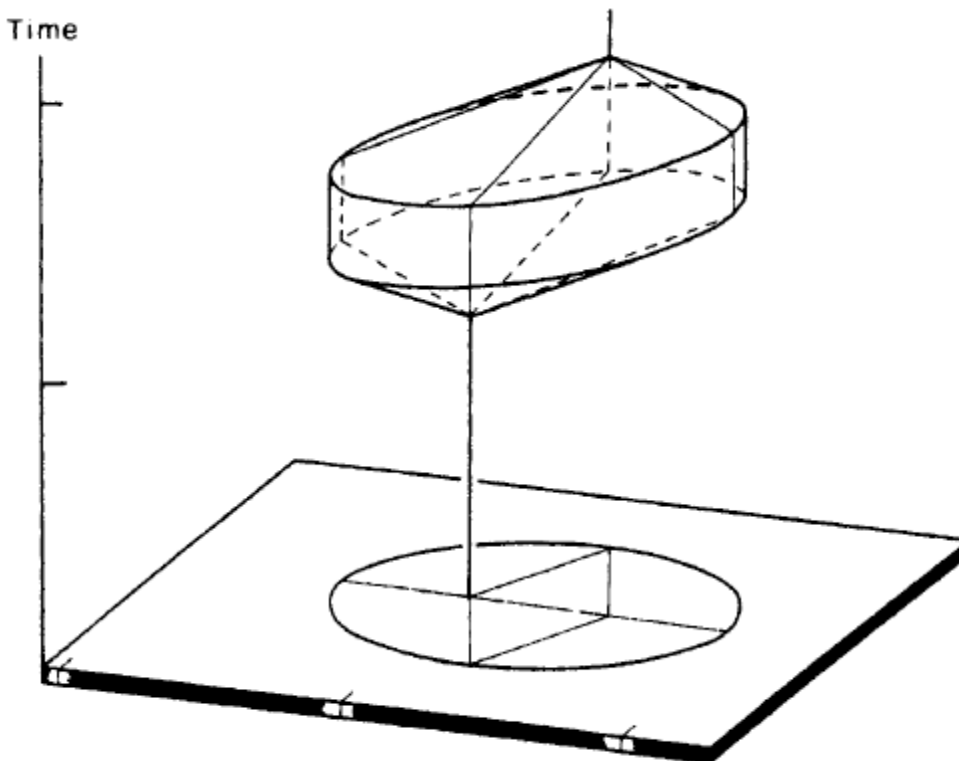
Figure 3: Space time prism (Source: (Lenntorp, 1976))

We will now discuss how the concept of space time prism, and more in general space time geography, has been widely used and expanded in the last years for transportation issues.

Since also most activity-based models take residence and work location as given, also many studies that exploited a time-geography approach benefited from the application of one of the main notions from trip chaining, where each journey is associated with a primary activity (or purpose), and that this activity is conducted in a particular place, known as the primary destination (Cascetta, 2009).

Dijst and Vidakovic (2000) have used these concepts to introduce the travel time ratio, that expresses the relation between travel time and idle time, focusing on the time people are willing to spend on reaching activity places, and arrive to show that values of this ratio are in accordance for similar types of activity

_____

places. Time geography has been applied conditionally to individual's basic places, such as home, work or school. The time available for visits to other activity places is bounded by the departure from a base and arrival on the same or another base. The space time prism led to the definition of the potential action space: the area containing all activity places which are reachable, subject to a set of temporal and spatial conditions (Dijst & Vidakovic, 2000).

Schönfelder and Axhausen (2003) have also pursued in the study of activity space, defining three approaches of increasing complexity with the aim of computing a more precise measure to the size of human activity spaces. The first method concerned the definition of confidence ellipse, initially relying on the recognition of the home location as the peg of daily mobility, and then adjusting to the aforementioned geometrical figure due to the revelation of two focal points for most travellers. The paper then perfections the achieved results by proposing kernel densities estimates, and minimum spanning trees.

Also Sprumont and Viti (2017) have focused on the consideration of homes and workplaces as anchor locations of daily mobility patterns. They have seen how the entire daily activity travel chain has been affected by a workplace relocation, using descriptive statistics as well as Standard Deviational Ellipse (SDE) theory.

2.3.   Key Points of Review

The following key points have been the landmarks of this work, and the integration of these three issues is the gap that we want to fill with this research.

- Simplicity and effectiveness of trip-based macroscopic approach of Gravity Models for evaluating large congested networks

- Importance of Mobility Patterns for travel demand estimation, that can outcome from daily scheduling, trip chaining or activity participation

- Users resent of constraints in time and space: home and work location represent two pegs of daily mobility

_____

3. Methodology

This section will present the formulation of the model, whose development has been driven by the goal of including activity tours and activity locations within a dynamic gravity model. The classical gravity model will be extended to include this information, explaining how the use of internet data, socio-demographic data and a travel survey bring to the definition of dynamic OD matrices that take account of activity patterns.

The whole process will be exhaustively explained in the paragraphs of this section, but it can be summarized in the following way. In order to generate and distribute purpose-dependent flows in a dynamic way, one of the activity patterns extracted from e.g. a travel diary will be assigned to the population of the study area generated from the sociodemographic information. The location where each user will perform its activities will then be assigned through the exploitation of spatial relationships deriving from the coordinates of all the possible places where activities can be performed. However, one single trace is not going to reproduce any real user, instead, constraints work at an aggregate level to capture phenomena like activity relocation and the correlation between demand flows belonging to different traffic zones. The logic behind this approach is that the temporal distribution of activity-travel patterns does not significantly depend on the study area, while the spatial distribution depends on the geographical distribution of the different activity locations.

This section starts with a premise, which consists in a brief description of how the theories presented in the literature review have been adapted to our means. It precedes the actual methodology, that is broken up in three parts. Starting from the end, the real aim and heart of the methodology are immediately explained. The following two sections introduce the two initial parallel processes, respectively reflecting the spatial and temporal aspects of the model. Section 3.5 concludes with an overview on the whole process.

3.1.   Premise

This section addresses the issue of how two different research fields, gravity models and space-time geography, have been exploited in order to achieve a better representation of travel behaviour. In brief, we aim to extend the classical gravity theory, where flows are spatially distributed following the gravity force principles, with time geography principles, where the space-time distribution of activity patterns is modelled, but at an individual level.

3.1.1   Gravity Model extension

As stated in chapter 2.1.1.1., OD matrixes deriving from gravity models are a direct function of generation and attraction parameters, of the total production of the origin, of the total attraction of the destination, and

_____

of the impedance. This means that the distance between two zones will directly influence the number of trips from a zone to another.

The key difference in the proposed methodology, stands essentially in the level at which these factors are introduced. Generation and attraction are almost exclusively used to define how many of the people living in one zone go to work in each zone. The impedance function steps in subsequently: it will not have a central role in defining the exchange between zones, but, on the contrary, will be the main parameter in the choice of secondary activities, for users going from home to work with predetermined activity patterns.

The concept of attractiveness depending principally on the masses and on the distance between them is still present, but it has been reassembled in successive steps, as showed in Figure 4, in order to include activity patterns and replicate behaviours that conventional gravity models are not able to capture.

Generation and Attraction
- create population of study area

Impedence function
- choose location of secondary activities
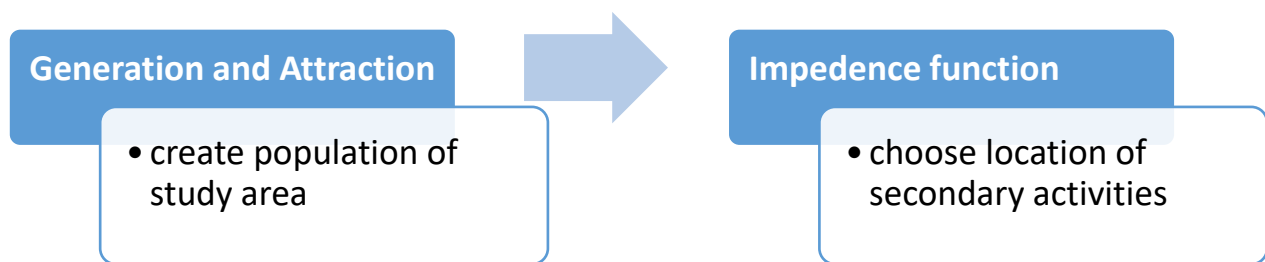
Figure 4: splitting of Gravity Model

3.1.2   Ellipse Theory

From the enormous amount of research compiled on space time geography, of which a brief explanation is given in the literature, only a very simple assumption has been taken into consideration for the development of the model.

According to Hägerstraand's theory of the space-time constraints affecting daily activity patterns, individuals need to trade off between travel time and activity duration (Timmermans, Arentze, & Joh, 2002). It is supposed that people going to work have limited time budget to perform secondary activities, and are not keen on wasting too much time for travelling. Dijst and Vidakovic (2000) note that people do have choices: for example, they can choose between a long stay in an activity place in the neighbourhood, to perform one or more activities, or a short stay in an activity place at a large distance from home. This methodology wants to represent this choice dimension assuming that workers will not likely take part in secondary activities that deviate them too much from their home-work journey.

This results in the definition of an ellipse of influence for each home-work zone couple, with the two centroids of the zones representing its focal points. Each user of the network, living in zone h and working in zone w, will then consider performing his/her tasks choosing between the locations inside an ellipse, whose two focal points are the centroids of zone h and w. The limitation will obviously be posed only upon the users that go to work, and it will impede them from considering to perform secondary activities outside of the ellipse they belong to.

_____

_____

By aggregating individuals' ellipses at a zone level, the total number of ellipses will be equal to the number of OD couples. For n zones, there will be n^2 ellipses, including the OD couples where origin and destination coincide, for which the ellipses will collapse in a circle. The first consideration to do before the creation of these ellipses will regard their shape how much they differ in size. Since for each one of them the distance between the focal points is already set, the only parameter on which their shape depends on is the eccentricity (Figure 5).

$$b^2 = a^2 (1 - e^2)$$

$e = 0$ $e = 0.5$ $e = 0.75$ $e = 0.95$
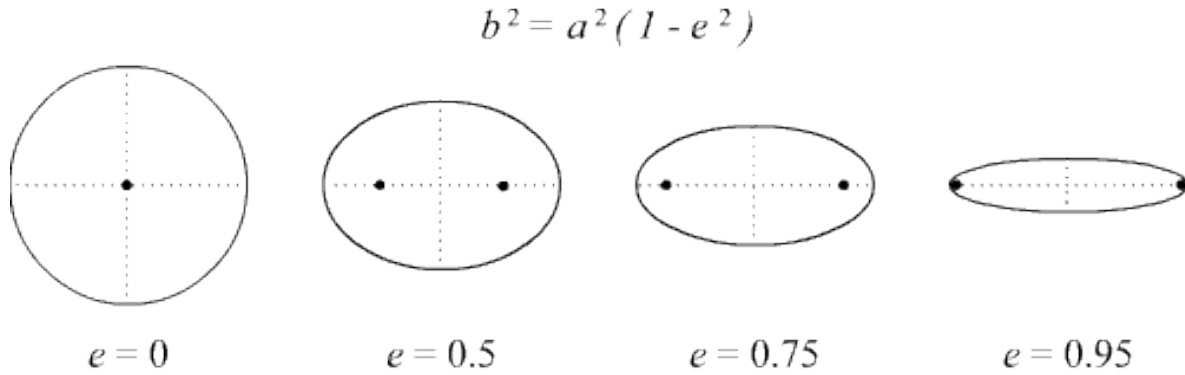
Figure 5: Difference in shape for values of eccentricity

Imposing the same eccentricity value to all of the OD couples, produces ellipses of the same precise shape, but with surface values completely different. This has been supposed to respect reality poorly, since we expect users to perceive their influence areas of more or less the same size. In order to account for this behaviour, a different value of eccentricity should be taken for each ellipse, and since the only available parameter denoting them is the distance between centroids, it is logical to assume the eccentricity to be a function of this distance. For example, two centroids very close to each other will produce an ellipse with a very low eccentricity, going towards a circular shape; while if the two centroids are very far apart, the eccentricity will be very close to 1, resulting in a very flat ellipse. These properties derive not only from logical assumptions on users' behaviours, but it represents dynamics that appear often in literature. For example, in (Sprumont & Viti, 2017), the spatial distribution of activities resulted in the generation of the ellipses shown in Figure 6: the further the focal points, the thinner the ellipse.

This means that users who spend significant amount of time commuting, will try to find a secondary activity very close to their home-work path. Contrariwise, people who live very close to their working location have more time to dedicate to travel, also because the shorter the path, the lower the chances of finding a location that gives the opportunity to perform a specific activity.

_____
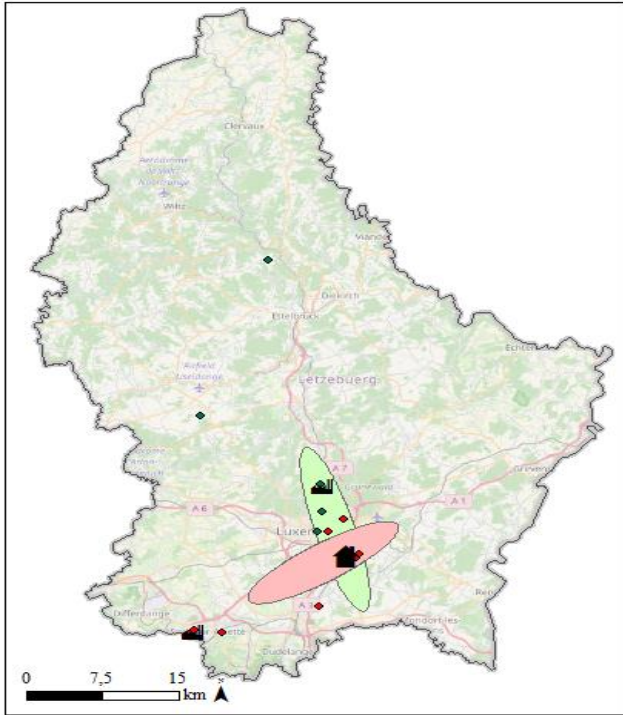
a) Representative of cluster 1

b) Representative of cluster 2

Legend

🏭 Walferdange (North, Old campus)

🏭 Belval (South, New campus)

🏠 Respondent's living place

• 2016 Activity Locations

• 2015 Activity Locations

2016 Activity Space

2015 Activity Space

National Border

Realization: Sprumont Francois
MobiLab, University of Luxembourg
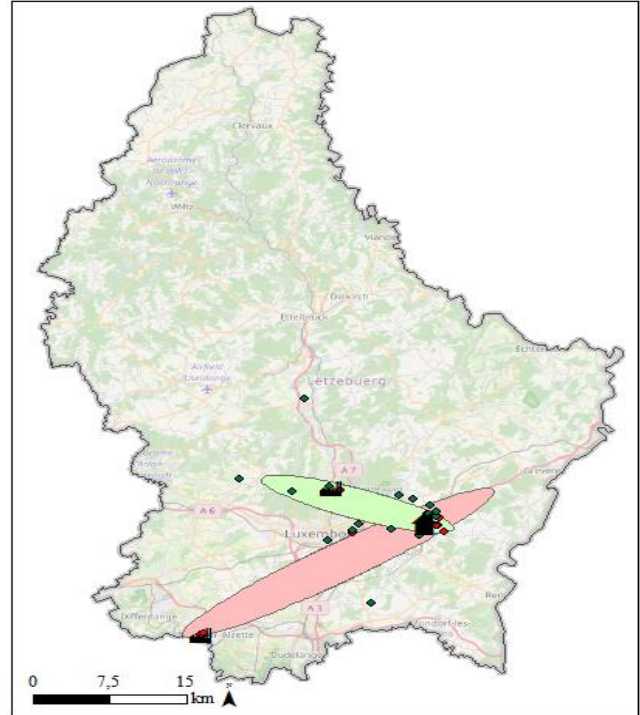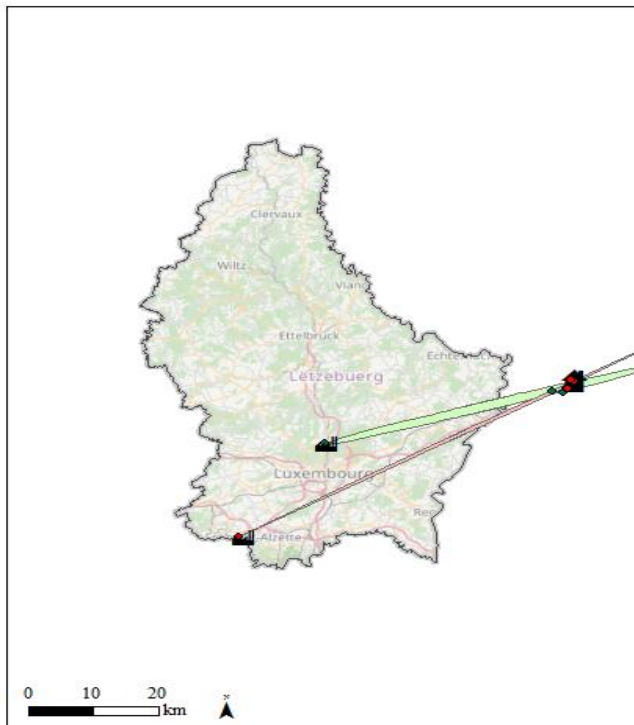
c) Representative of cluster 3
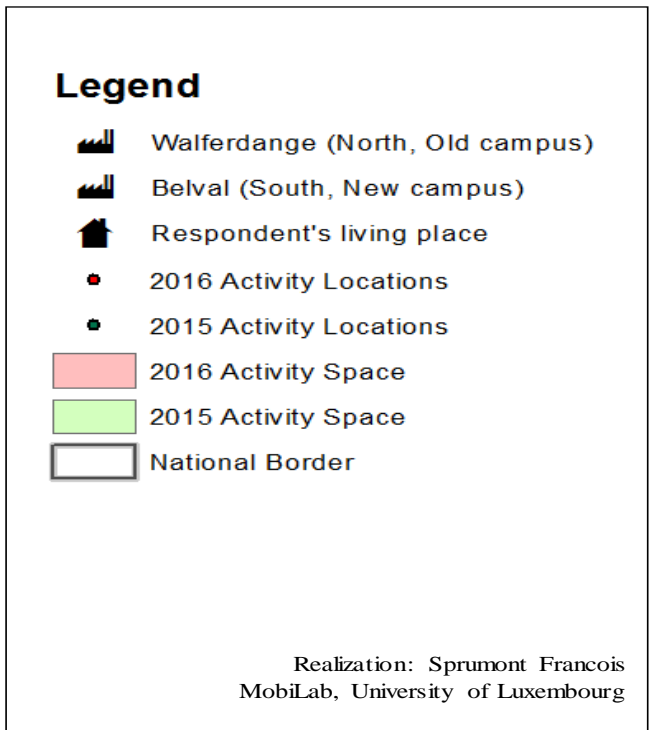
Figure 6: Ellipses for representatives of different clusters (source: Sprumont and Viti, 2017)

_____

Practically, it's through the manipulation of this geometrical parameter that we can represents this very intuitive behaviour. The eccentricity can be seen as an indicator of how much each user is prone to extend his/her home-work trip in order to include a secondary activity. Mathematically, it is equal to the ratio of the distance between centroids, over the sum of the maximum distance a user is willing to cover to go from one centroid to the other passing through a certain activity location.

$$e = \frac{dist\,(centroid - centroid)}{\max dist(\,centroid - activity - centroid)}$$

For example, an eccentricity of ½ will determine that users will not consider performing secondary activities in places for which the time to go from home to that location to work is more than the double of the home work commute time.

Regarding the formulation of the eccentricity, the only variable it should depend on is the distance between focal points and, more precisely, it must increase as this distance increases. There are many functions that are suitable for this purpose: exponential, parabolic or even linear. However, the equation used in this methodology, derives from assuming a different point of view: the eccentricity is not directly explicated. It is in fact easier to express directly the maximum distance that can be travelled as a linear function of the distance between centroids. Graph 1 shows an example of this relationship and the consequent trend of the eccentricity.
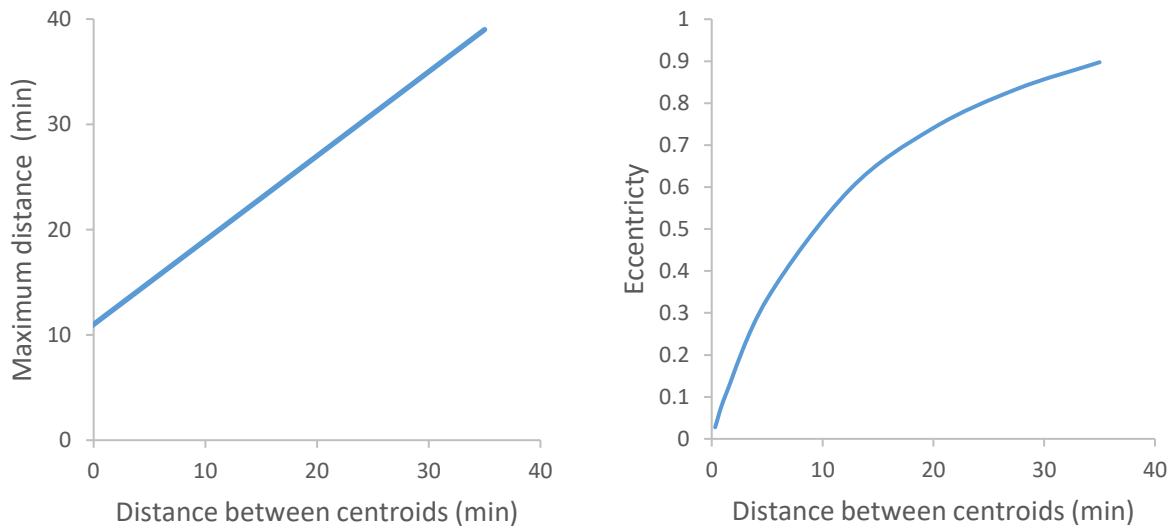
Graph 1: Characteristics of Ellipses depending of distance between centroids

_____

_____

3.2.   Core of the model

The scope of the research done in this thesis can be intuitively shown through the description of the final steps of the model. Figure 7 shows how the dynamic OD matrixes are finally estimated: first all the trips that have been generated are stored into a matrix, that will then be processed by aggregating all these trips given their origin, their destination, and the time of day at which they have been performed.
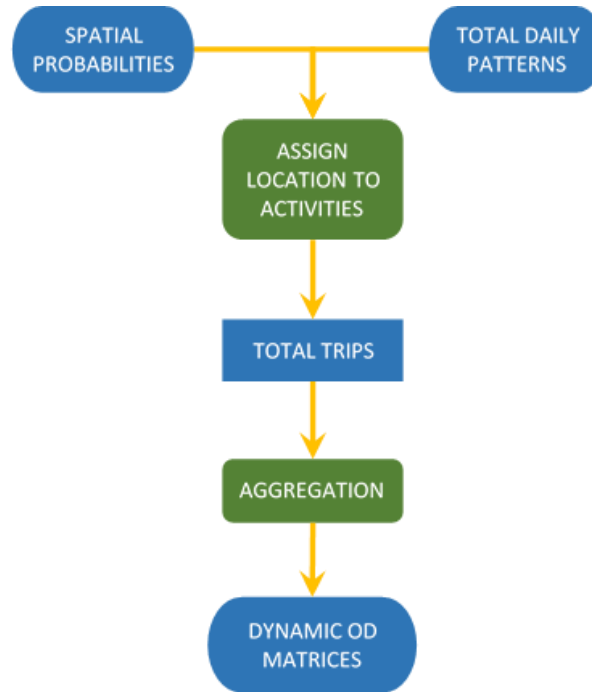
Figure 7: Determination of Dynamic OD matrixes

These final processes start from two important set of variables, the explanation of how they are determined will be addressed in the next two sections. The first one accounts for the temporal aspect and for the magnitude of the scenario: the list of the daily activity pattern for all of the required population. An activity pattern is expressed as the sequence of the activities that are done in a day, Figure 8 reports some examples. In our case, the daily activity pattern of each user must also contain information about the time at which each activity is performed.
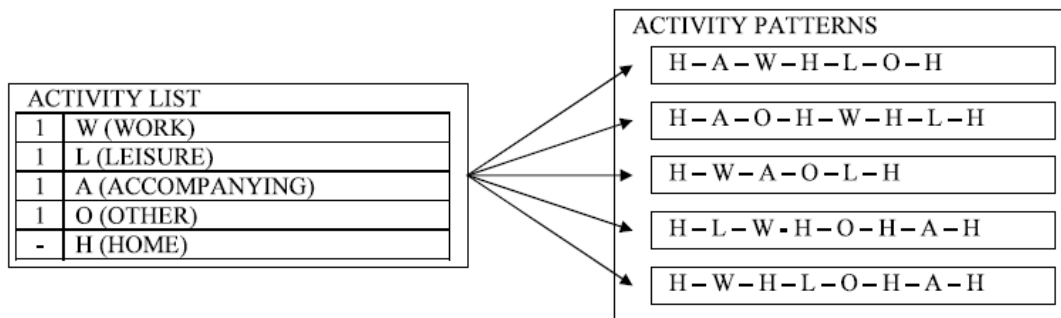
Figure 8: Activity List and Patterns (source: Cascetta, 2009)

_____

The second set of variables provides the model with the spatial information needed to transform the activity patterns in a travelled path: a series of successive trips where the destination of one coincides with the origin of the following. These variables are under the form of probabilities, which express the likelihood of choosing a determined location for carrying out an activity of a certain kind. These probabilities depend mainly upon the distance between the respective location and the zone of home and work. The further the location from the home-work commute of a user, the less likely the user will choose this location.

According to these probabilities, a location is assigned to each activity from the activity pattern of each user. The results can directly be stored as individual trips, creating the list of all the trips performed by all users during one day. With each trip characterized by zone and time of departure, zone of arrival, and purpose of the trip, it is possible to proceed to the aggregation process, creating a set of dynamic OD matrixes for each activity cluster.

3.3.   Probabilities (spatial)

This section describes the process that leads to the definition of probabilities, starting from the basic inputs, thoroughly explaining the involved procedures, and clarifying the theoretical constraints that have been considered. Each relation featured in Figure 9 will be extensively described in the following paragraphs.
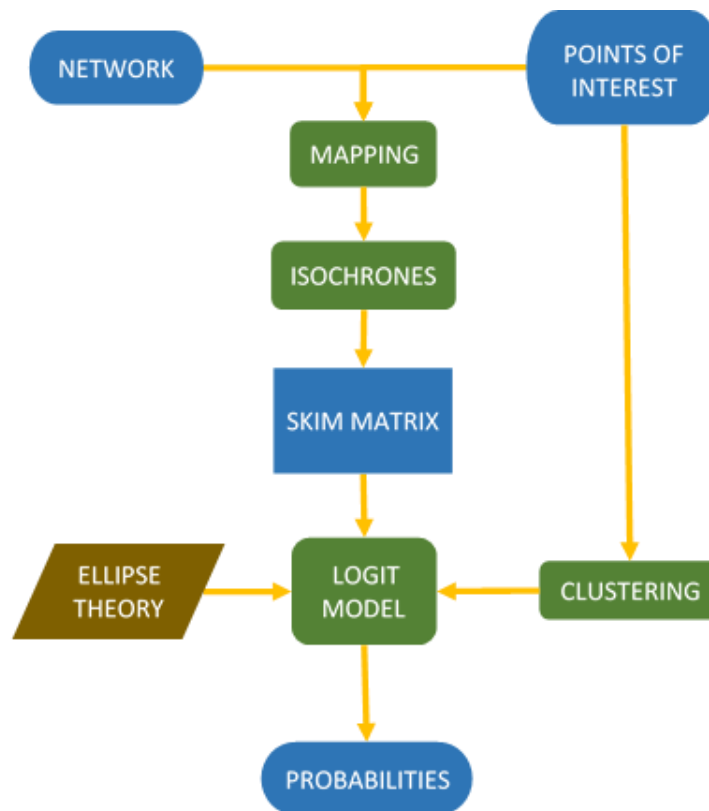
Figure 9: Determination of Spatial probabilities

3.3.1   Inputs

As previously stated, the model is based on the exploitation of internet data. This part of the methodology relies on two fundamental inputs:

- The list of POI inside the study area

Google gives us the opportunity to download, for any POI, information about its position (address and coordinates), phone number, users' rating, the searchterms (that represent the kind of business, the category it belongs to) and the flows during the hours the activity is open. In our model, we have mainly benefited from the use of coordinates and searchterms, but the hourly flows remain available for a wide range of possibilities, mostly for validation issues.
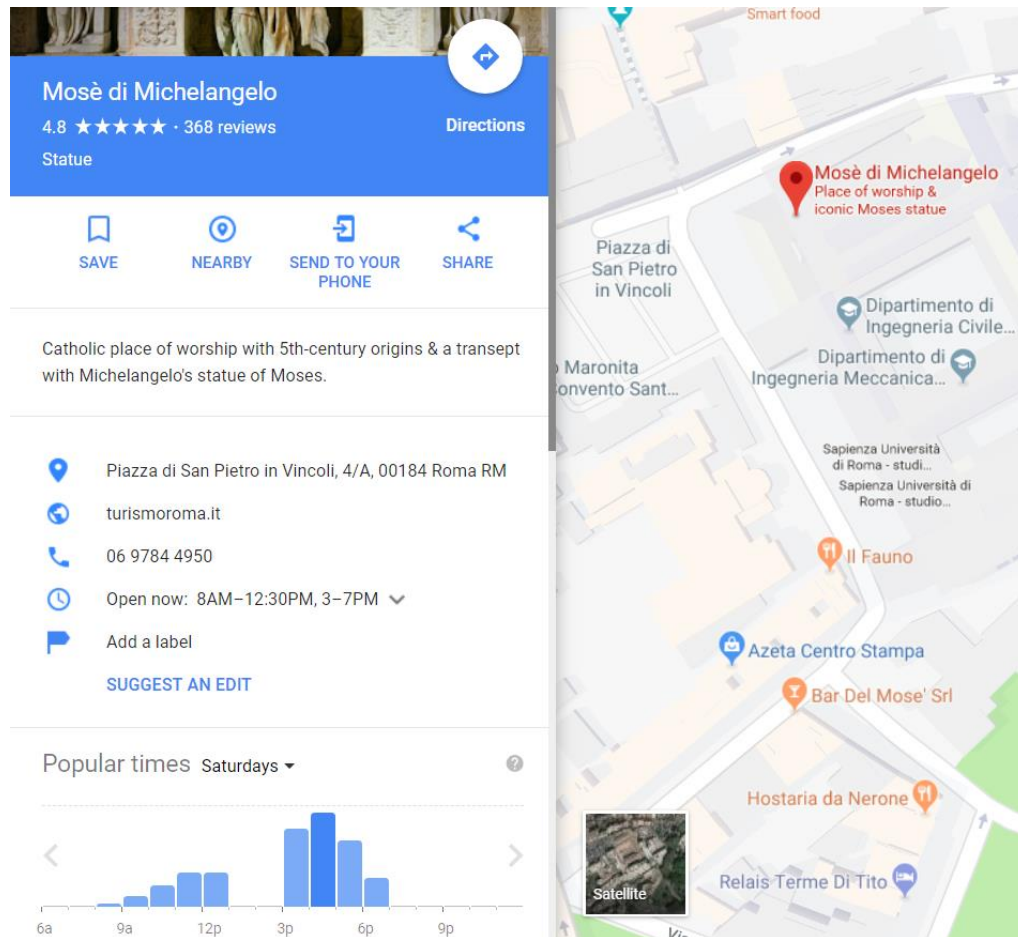
Figure 10: Example of Point Of Interest

- Network

The goal of this part of the methodology is to use geographical characteristics to estimate the probability of choosing a certain zone instead of another for performing a certain activity. In order to exploit the spatial relationships deriving from the knowledge of the coordinates of the POIs, appropriate topological data is necessary. This topological information has to be detailed enough to capture the geographical differences between POIs: it cannot be too much approximate since it will have to include also relatively small roads.

_____

However, since our objective is not to enter in microsimulation fields, the network does not need an incredibly high level of accuracy.

Hence, an open source such as OpenStreetMaps, which has generated data for almost all the roads and networks of the world, is perfect for our scope. It uses a topological structure, where one layer for nodes, one for edges and one for the cartographic map are the only needed for representation issues. The list of nodes is very important for the first process of the model, while information about relations between nodes and edges, such as restrictions on roads or permitted turns, will be also used further on in this methodology, but without the need of downloading them. The QGIS software facilitates the visual representation of the desired network, and permits to manage the desired elements across layers, or even to upload external features.

3.3.2   Mapping POIs on Network

The first step consists in the merging of the two datasets. Since POIs are represented through a pair of coordinates, it may happen that some of these points do not belong to the network from OSM. To overcome this problem, each POI is matched to the closest node of the network using a Python code. From this moment on, each POI's location will be considered the one of the node to which it has been matched. The error given by this approximation will not interfere with the effectiveness of this model due to its small order of magnitude since intersections are very dense in urban networks. Additionally, more than one POI can be matched to the same node; this neither will affect our model.

3.3.3   Skim Matrix

The network will be now including all POIs, making it possible to derive spatial relations between them and the zones. This has been done with the use of a tool called OSMnx created by Geoff Boeing and available on his website and on GitHub. In (Boeing, 2017) OSMnx is presented as "*a free, open-source Python package that downloads political/administrative boundary geometries, building footprints, and street networks from OpenStreetMap. It enables researchers to easily construct, project, visualize, and analyze non-planar complex street networks consistently by constructing a city's or neighborhood's walking, driving, or biking network with a single line of Python code.*"

This library was used to extract and assemble an algorithm that, given a starting point, graphs the network around it and plots the nodes belonging to it in various colours. Each colour represents a precise distance from the starting point, and each set of equidistant nodes having the same colour forms an isochrone (Figure 11).

Different networks are available: one for cars, one for pedestrians, and one for bicycles. The choice of which one to use must relate to the dimensions of the study area and to the behaviours the modeller wants to capture. An important parameter to choose is the average speed for the users: it will strongly affect the number of isochrones, since each isochrone is matched to a certain time interval. The resulting number of

_____

isochrones must be sufficiently large to permit an accurate representation and to cover all the nodes of the study area.
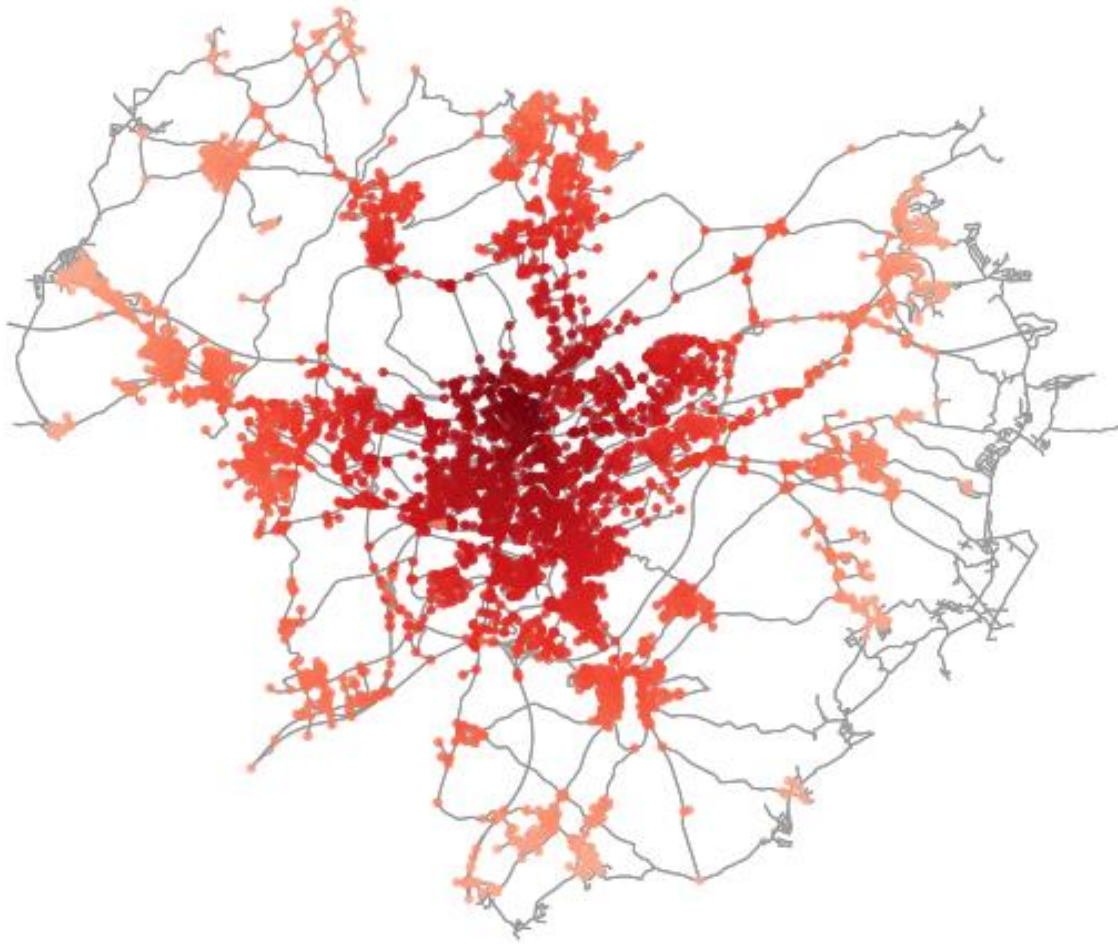
Figure 11: Isochrones

By extracting the time referred to each colour, the algorithm is finally able to extract the array of distances from any node to all the other nodes of the network. This resulting algorithm has then been run for the centroids of all zones, finding the distances between each centroid and all nodes of the network. Since all centroids, all POIs and all schools are linked to a node, it was possible to create the needed skim matrixes.

The final output consists of two matrixes: one referred to the distance between the centroids of each zone, and the other one to the distances between each POI and each centroid. These two matrixes will be the major inputs for our following process: the distances from POI will be the primary attribute in determining probabilities, while the comparison between these and distances between centroids will be used for the definition of the ellipses.

_____

3.3.4   Clustering POI categories

In order to use an appropriate location for a certain activity, that means going to a restaurant for "eat" or going to a store for "shop", it is fundamental to cluster each POI according to what kind of activities people perform there. Each POI comes with various searchterms that define what category it belongs to. The complete list of types and information about the logical order in which these are divided can be found here: https://developers.google.com/places/supported_types. However, the searchterms, after the first one, do not bring further information and, consequently, only the first column can be used for the clustering.

This step is very important for a correct geographic distribution of activities, since an error in the searchterm clustering will bring to a biased database. Even though an automatic clustering can be done, the results to which it will bring might not be reliable enough. The induced error could still be reduced through the utilization of through machine learning techniques, but it is not unlikely to find cases where this might still not be enough. Since the insertion of details of Google Places are left to the everyday users of this platform, and there is not a certified or reliable source, for some POI information is not precise and it should be manually added or modified.

3.3.5   Probabilities

With the final skim matrixes, a correct clustering, and the theory related to the choices of secondary activities presented in section 3.1.2., it is possible to compute the probabilities of choosing a certain POI for a certain type of activity for each OD couple. All users living in zone h and working in zone w, must have a set of possibilities for each activity type, that accounts for the propensity of choosing one out of the POIs belonging to their predetermined ellipse.

For each OD couple and for each activity, probabilities relative to a POI have been assigned using a Multinomial Logit Model, where the only attribute defining each POI is the sum of the travel time to reach that location from home and from work. The inverse of these distances will represent the utility V_j of each POI[j], while the whole model is governed by the only parameter θ.

$$V_j = {}^{1}/_{dist} \qquad\qquad p[j] = \frac{\exp(V_j/\theta)}{\sum_i \exp(V_i/\theta)}$$

The influence of θ is very strong: a high value raises strong stochasticity; lower values make the choices become deterministic. This trend is explained in Figure 12 (Cascetta, 2009), where three curves report the values of the probability of choosing alternative A or alternative B, in function of their difference in utility, for three different θ values.

_____

_____
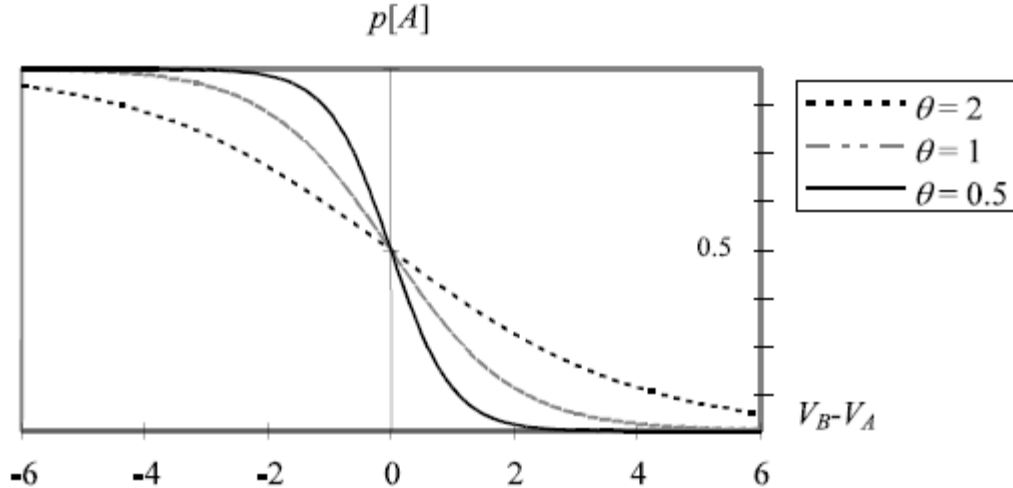
Figure 12: Diagram of choice probability p[A] of a binomial logit model (source: Cascetta, 2009)

The choice of this parameter is the most powerful tool for the modeller to define the expected behaviour of the users, but bearing in mind that it also strongly depends on the size chosen for the ellipses and on the characteristics of the study area. If a network is highly dispersive and many users are actually obliged to endure long journeys to take part in different activities, the model should have a greater θ than if the network was small and dense.

People who do not go to work, or people who work and live in the same zone, represent a particular case. For them, the ellipse of influence collapses in a circle, since the distance between focal points will result to be 0. This circular area can still result from the parametrization of the ellipse in function of the distance of centroids, if not, it is possible to choose a different radius that must be coherent with the size of the other ellipses. Another possibility is to rely on the logit model to reflect this behaviour: ignoring any geographical limitation, since locations at a big distance from home would have very small probabilities, not many users would be directed so far away from the starting point.

The final arrays of probabilities come out structured in two sets. One set refers to people who do not work: it contains one matrix for each secondary activity, each matrix containing an array of probabilities of POIs for each zone. The second set considers the ellipses, with again one matrix for each activity, but this time containing an array of probabilities for each OD couple. For couples where origin and destination coincide, the array will be taken from the array of the respective zone of the first set.

3.4.   Total Daily Patterns (magnitude and temporal)

This branch on the methodology addresses two key aspects: it accounts for the magnitude of the users and for the temporal dimension. Again, two are the main steps that bring to the definition of the daily patterns of all the users of the network (Figure 13).

_____
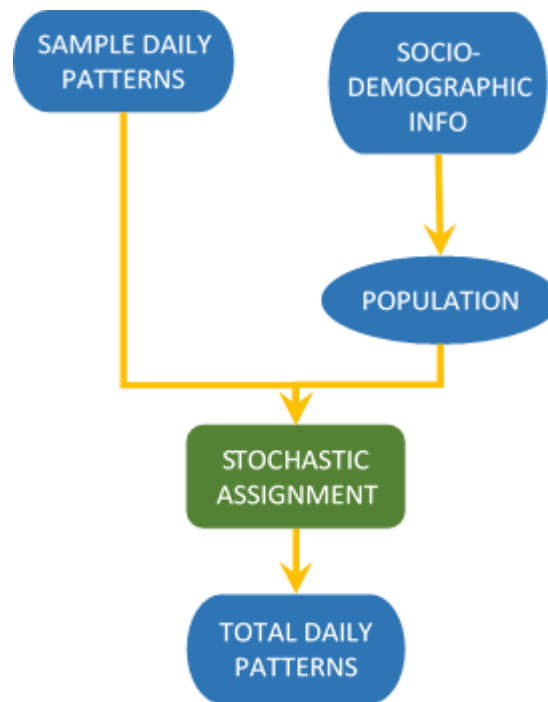
Figure 13: Creation of daily patterns for total population

3.4.1   Inputs

- Daily Patterns

This is the only set of data which is not available in opensource format, and is probably the most limiting weakness of this methodology. The easiest way of achieving these patterns is using travel diaries, in order to have a reasonable set of activity patterns and attempting to generate a heterogeneous behaviour on the network. Nowadays, other methods of achieving daily patterns is subject of a huge amount of research. Especially with the surge of Big Data in transportation, many researchers have worked on the possible ways to infer activity patterns from users' trajectories (González, Hidalgo, & Barabasi, 2008), that could derive from GPS (Xie, Deng, & Zhou, 2009) (Huang, Li, & Yue, 2010) or mobile phone (Phithakkitnukoon, Horanont, Di Lorenzo, Shibasaki, & Ratti, 2010) (Alexander, Jiang, Murga, & González, 2015) data. In addition, many papers (Spinsanti, Celli, & Renso, 2010) (Chen, Bian, & Ma, 2015) also exploit the same POI information used in this thesis, therefore research to integrate these models inside this methodology is strongly encouraged.

- Socio-demographic information

The first step of this process relies only on this data. For each zone, the number of inhabitants and the number of workers is required. This should be done after or together with the zoning, since the data on official websites is structured according to administrative boundaries.

_____

3.4.2   Population Synthesis

Along with the choice of the study area, a stage that is fundamental for the proposed methodology is the creation of a desired population to assign the daily patterns to. A great deal of research has been spent on the topic of synthetic population generation, and a brief explanation of the state of the art has been presented in section 2.2.2.. However, this methodology does not require an excessively sophisticated population: there is no need to represent the multiple facets of society distribution by extracting highly detailed socioeconomic and demographic attributes.

The information that the final population must reflect is how many people live in each zone, how many of them go to work and where they go to work. Addressing the three issues in sequence, the number of people living in each zone is usually always available as inhabitants' information on official websites. Deciding how much of the population actually has a job can be done or according to employment information again from the census, or to the ratio of workers that results from travel survey. Generally, since the total number of inhabitants rarely matches the total number of employees, the number of workers will just function as an attraction parameter, and workers will be distributed proportionally to them.

3.4.3   Stochastic assignments

With an appropriate number of daily patterns, we then proceed to randomly assign one of these daily patterns to all of our population.

The sample of daily patterns is split into two sets, the first one containing patterns that have at least one work activity during the day, while the other one contains the rest and will have no work activity in it.

Randomly, one pattern belonging to the first group will be assigned to each worker, while the rest of the population will have a pattern assigned from the second set.

These procedures will result in a matrix with the daily activity pattern of all the users of the population.

3.5.   The Model at a glance
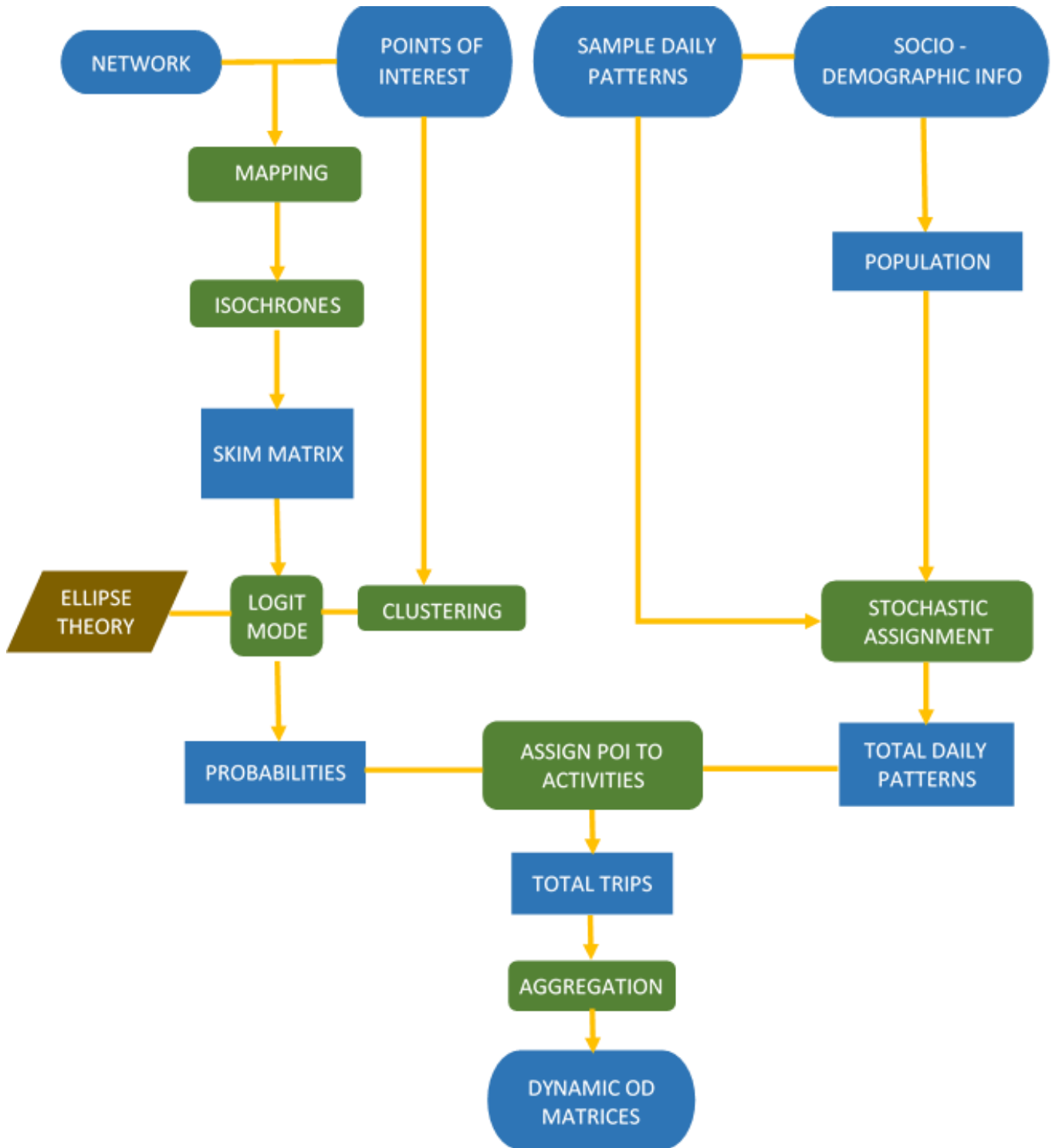
Figure 14: The whole model

Figure 14 captures the whole development of this methodology: starting from the four main inputs, arriving to the final OD matrices. Inputs and Outputs are the ones represented in the blue squared oval.

In green, the processes are presented, and to each of these green boxes, a Matlab or Python code is correlated, and all can be found in appendix. The mapping of the POIs on the network (A.1.) and the creation of isochrones (A.2.) are both achieved on python. Matlab was used firstly for the simple activity clustering (B.1.) and then, along with the skim matrix deriving from the isochrones, was used to include the ellipse theory and perform the logit model, determining the probabilities (B.2.). Parallely, working on the population, other Matlab codes have been written to extract the sample daily patterns from the BMW survey, and then stochastically assigning one of them to each user (B.3.). Finally, also the last two procedures, the assignment of locations to each trip (B.4.) and the aggregation process (B.5.), are run on Matlab.

4. Case Study

The methodology presented in the previous chapter will here be applied to the population of Luxembourg City. First, results are analysed and compared with official statistical data. Then, a case study is presented to show that the model properly captures complex macro-dynamics that take place in the city.

Firstly, a brief description of the case study area is presented, underlining the most important details that must be taken into consideration when applying the methodology. Section 4.2. will then present the results of this application, and a primary validation can be found. In the last section, the sensitivity of the methodology was tested by changing some input data, studying the comparison with the initial scenario.

4.1.   Luxembourg City

4.1.1   Overview of city and state

The city of Luxembourg is the capital of the Grand Duchy of Luxembourg, and with 115000 inhabitants is the largest city of the nation. Some characteristics of the state, such as its small size, a high level of prosperity, and its centrality and importance in Europe, have determined the city to be a very peculiar reality. Luxembourg has become an extremely multi-ethnic city, attracting not only a high number of foreigners from the bordering countries of France, Germany and Belgium, but also many people from all parts of Europe and all across the world. One of the peculiarities of this region stands indeed in its share of non-Luxembourgers: 48% of the total population of the Grand Duchy is foreign, reaching a peak of 70% in the city. Graph 3 shows how much the share between local and foreigners has risen in the last 30 years.
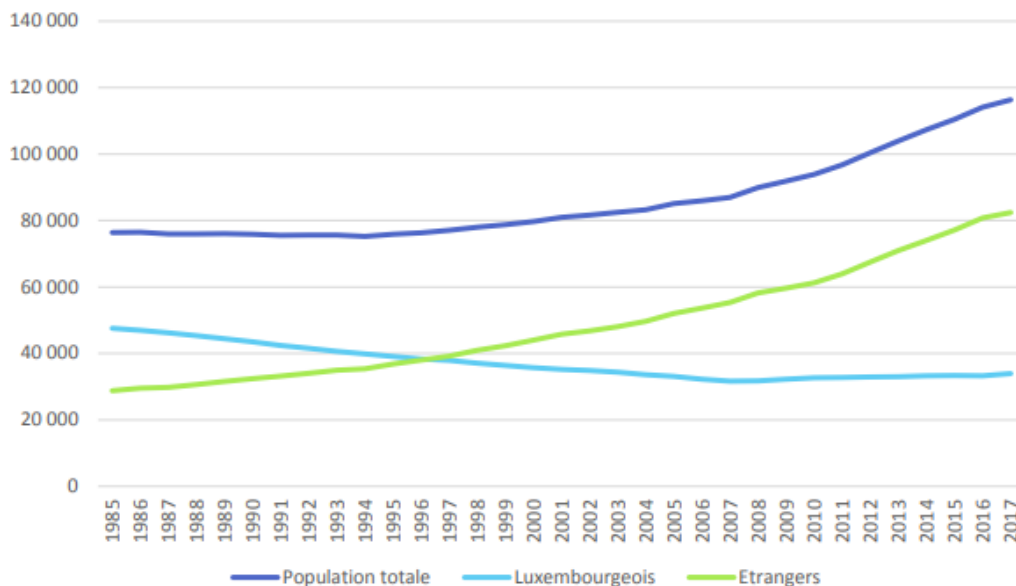
Graph 2: Population in Luxembourg city

Such a higher percentage of foreigners is principally explained by the fact that most of the job offers that attract people from different countries come from businesses and industries located inside or close to the city. With a strong lead of the financial sector, that alone amounts to 36% of Luxembourg's GDP, services drive Luxembourg's economy, demanding many more work places that the ones its workforce can cover.

The state in fact accounts for more than 430 000 available jobs, that far exceed the number of employees the nation can produce. This is due to the presence of a significant number of cross border workers, who enter every workday in the city from the confining France, Germany and Belgium (respectively around 50%, 25% and 25%). The number of cross border workers has arrived in 2017 to 190 000, representing the 43% of the total, and two main causes explain the magnitude of this phenomenon. "Transfrontaliers" are attracted to Luxembourg by the higher wages and greater working opportunities, but in the same time find their convenience in this long commute due to the high cost of housing in the Grand Duchy, that increases drastically in the city. Figures show the evolution in time since 2000 of national and cross border workers, and the distribution of transfrontaliers between confining countries.
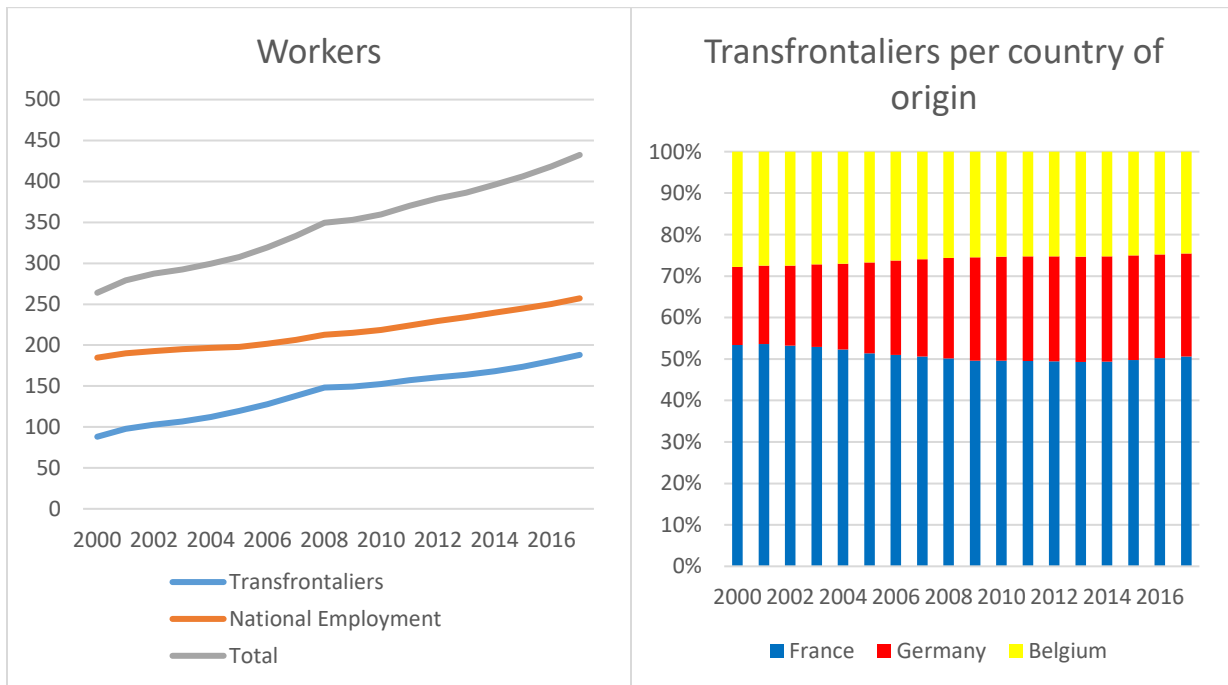
Graph 3: Distribution of workers in Luxembourg

To conclude, the city of Luxembourg, despite being a small city, is a very attractive site and the area from which it attracts employees goes much beyond its borders, and furtherly beyond the ones of the country. This must be one important aspect to take into consideration during the following validation and analysis of results.

4.1.2   Network and study area

The area that has been considered embraces the whole city of Luxembourg, in addition to some peripheral neighbourhoods. These have been included due to the absence of a clear discontinuity between them and

_____

the city centre. Furthermore, one of these areas contains an important shopping centre, which is expected to be an important factor in the distribution of activities.

Firstly, the network of the interested study area was downloaded from OpenStreetMap, exporting the list of the all the nodes, each one expressed by an ID and its coordinates.

Zoning has been performed considering the structure in which socio-demographic information was available: it has been downloaded from two sources. Data about the city population comes from a brochure from the official site of the city (https://www.vdl.lu), while the rest from a new open-data portal (https://data.public.lu) developed from the Grand Duchy. The first source lists the total number of inhabitants for each one of the 24 neighbourhoods, while the portal gives information about each canton, but only the communes contiguous to the city have been considered. In order to have regions of reasonable sizes, some of the city neighbourhoods have been merged, resulting in the definition of 22 traffic zones, as it can be seen in Figure 15. The creation of the zones was performed on Visum software, but OSM was used for the choice of the centroid of each zone, by identifying the node closest to a hypothetical geographical barycentre of the respective zone.
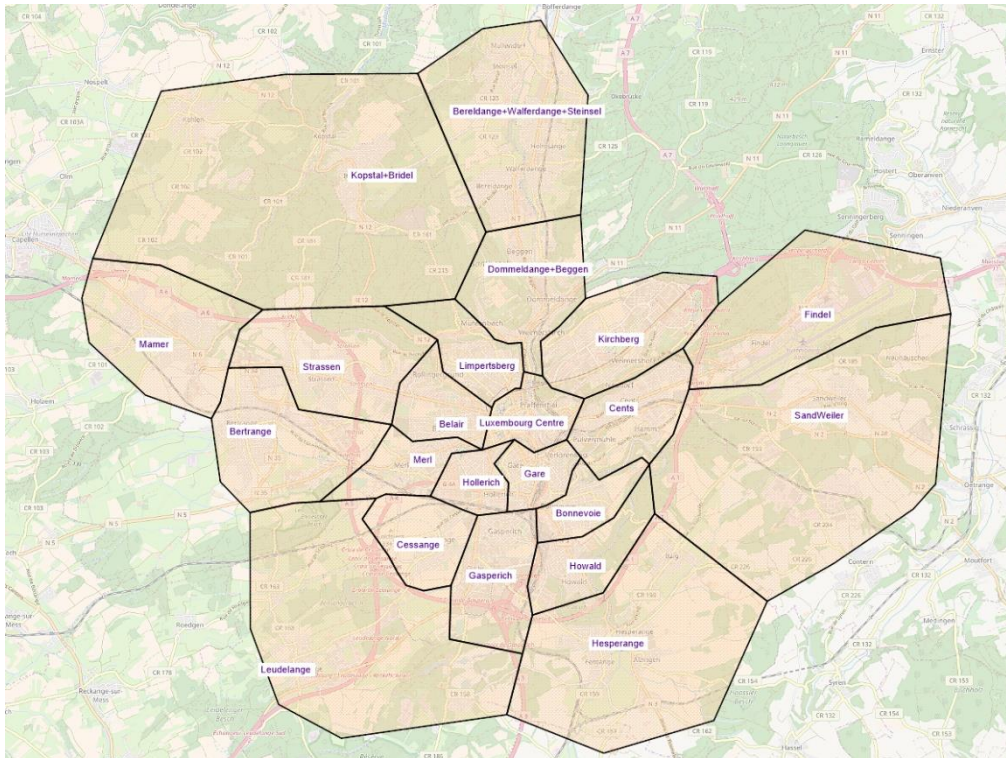
Figure 15: Zoning of study area

Regarding socio-demographic information, for each zone the number of residents was calculated merging and splitting the population of the neighbourhoods according to the zoning process. An attraction factor was determined from a presentation from LuxTram (http://www.luxtram.lu), that is proportional to the amount of employees who worked in each zone. Finally, the 18 biggest and most important schools of the city have been manually chosen and mapped.

_____

_____

For the same study area, more than 700 POIs have been downloaded, covering around 60 different categories. After a manual filtering, in order to delete doubles or errors, the total list of locations where users could perform secondary activities consisted in 652 POIs, distributed as in Figure 16.
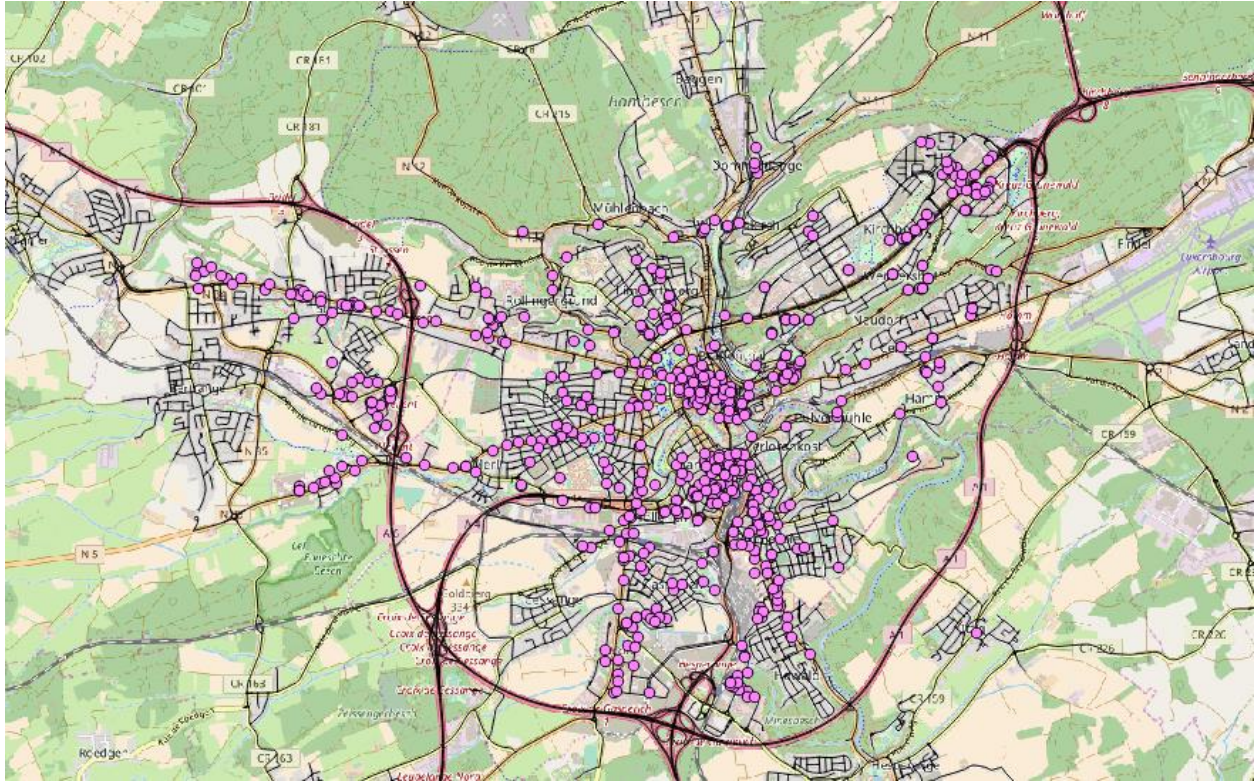
Figure 16: POIs inside the study area

4.1.3   Travel patterns

As for the daily patterns the user of our network had to carry out, we relied on the BMW dataset (Castaigne et al., 2009) which was collected by the University of Namur in 2008. 717 users have been traced for an entire week (not same week for everyone), registering times of departure and arrival, purpose of trips, mode of transport, day of the week and of the year. The most precious data for our model consisted in the purpose, the time of departure and the sequence of each trip during a user's day. The survey revealed more than 3500 daily patterns, containing trips of more than 700 users for 7 days. After filtering only the workdays Monday to Friday, we had a list of 2624 patterns, of which 996 did not have the activity work among their trips. We also kept note of the time of departure of each trip.

The decision of using the travel survey of a different city for our study area, was made assuming small or no difference in daily scheduling between the two populations, considering the similar culture and socio-demographic characteristics of Ghent and Luxembourg. A second travel survey, containing two weeks of trips from users working at the university of Luxembourg, in the Kirchberg district, was used to confirm this hypothesis. A comparison between the two surveys showed that activity patterns performed by citizens of Luxembourg City, can be considered analogous to the ones of Ghent.

_____

4.2.   Results

With the network and the POIs, probabilities have been assigned as described in section 3.3.; in parallel, travel patterns have been assigned to the population created from the socio-demographic data, following the method of section 3.4.. The results coming from this first application are analysed in this section, that represents the first step toward the proof of the effectiveness of this methodology.

Firstly, the total generation of trips around the city will be put in comparison with data from a document of the ministry of Infrastructures of Luxembourg, where graphs also represent the share of different activities. Following, to see if the distribution of trips across the zones has worked properly, generation and attraction of relevant zones will be analysed for each activity.

4.2.1   Total Generation of Trips

The following graphs show the hourly distribution of the total number of trips according to the achieved results (graphs on the left), and to the results of the Enquête Luxmobil 2017 (graphs on the right).

The strong resemblance of results, that both for the magnitude and for the profile matches the official data from the ministry, validates two strong assumptions:

1. It is appropriate to use the information of the BMW survey, that has been collected on users from Belgium, to represent the activity-specific demand in Luxembourg (data transferability).
2. Since the demand has been scaled using population from census data, it is reasonable to assume that the methodology will bring reliable results for each traffic zone, for each of which the population has been created following the same approach.

The graphs released from the ministry are divided per main activity purposes and do not follow the same activity clustering that has been done in BMW travel survey. The enquête had trips divided in "work", "education", "leisure" and "other", while the results of the applied methodology had also "home", "eat" and "shopping". To make a coherent comparison, the trips according to these last three clusters could not have been ignored, so in the following graphs on the left, eat activities were included in leisure, shopping was included in other, while for home activities a more complex decision had to be made. In the Enquête, all the trips directed to home have been included in the graphs relative to the activity in which the user was occupied before. For example, an activity pattern such as home-work-leisure-home, would generate in order, a trip for work, one for leisure, and another one for leisure, even though the destination of the last activity would have been home. For this, all the trips with home purpose have been distributed among the four activities of the Enquête, proportionally to how often a trip for that certain activity appears among the results of the application of our model.

In addition, the data of the ministry included the trips of all the cross borders, but the model is able to predict trips only for the inserted population. Thus, for this validation, we decided to assign travel patterns also to 160 000 other workers, assigning randomly among the patterns that included at least one work activity. It is not important to model their behaviour geographically, since the graphs show the overall generation of trips of all the study area, with no distinction in zones.

The comparison between graphs is firstly presented by pairing the graphs of the single activities education, work and other (Graph 5). We can notice that the model was able not only to reproduce the general trend of the day, but also to recognise the time at which each activity has a peak and to accurately predict the magnitude of most of these peaks. Despite the interesting results, two differences can be found in these graphs: our model strongly underestimated the morning peak of trips with education purpose; and inside the afternoon peak some of the work-related trips have been counted as trips belonging to the other activity.
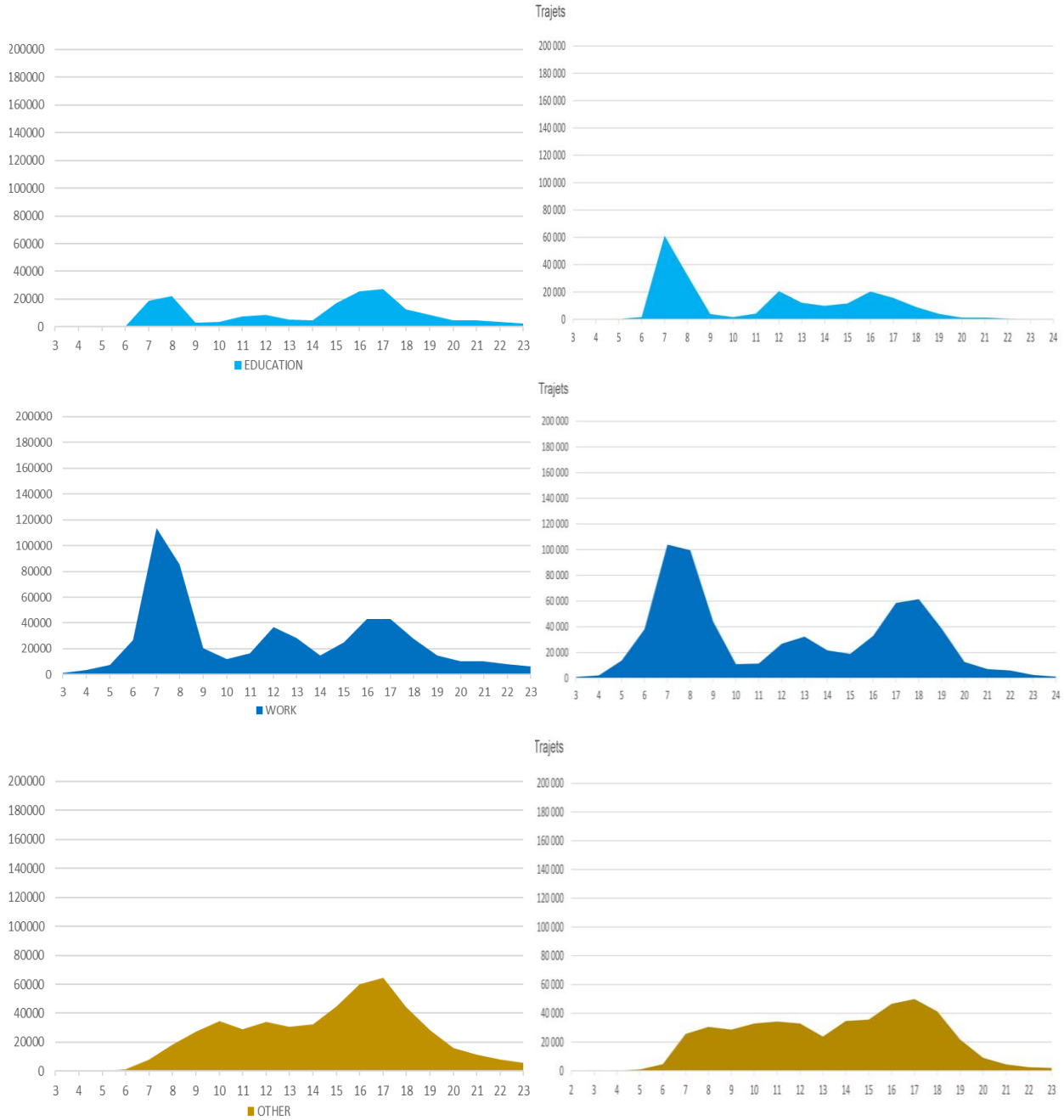
Graph 4: Comparison of trips for Education, Work and Other purpose

The sum of the trips for all activities gives the general idea of the effectiveness of the model (Graph 6). Apart from the absence of the morning education peak, we can appreciate its ability of representing reality, such as the recognition of the three different daily peaks, and the high similarity in terms of magnitude.
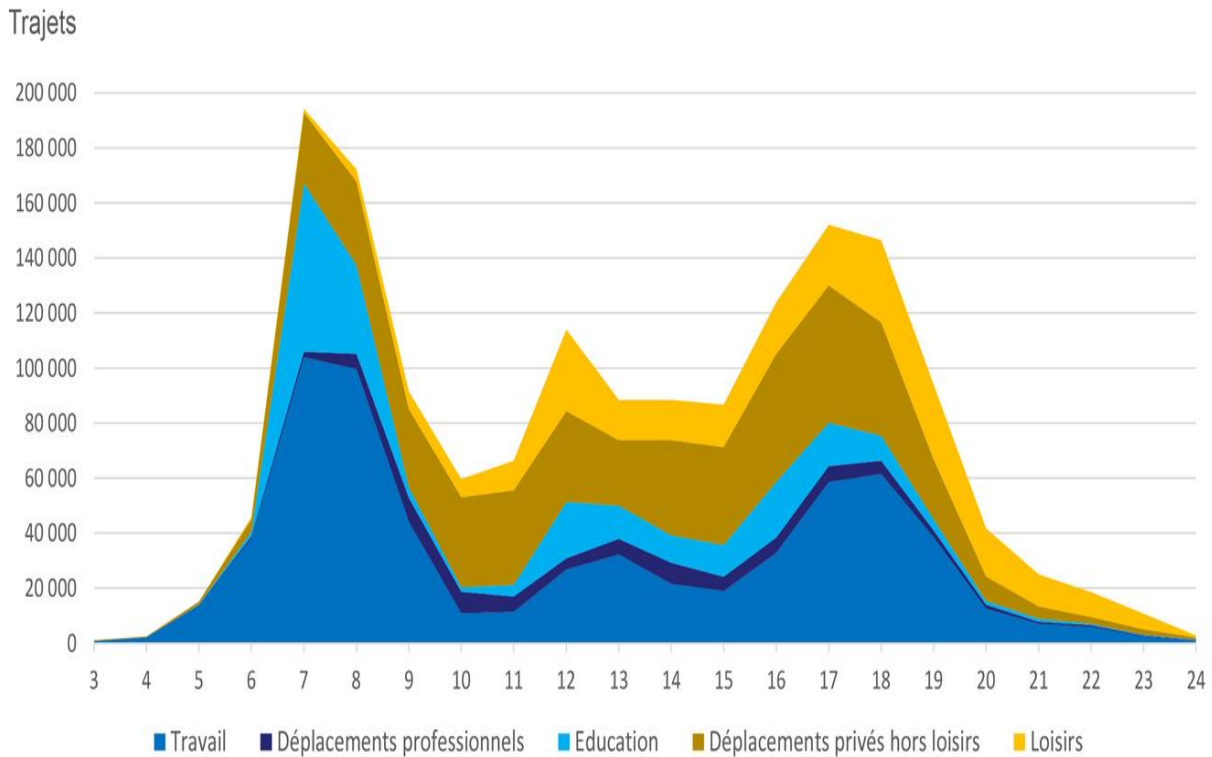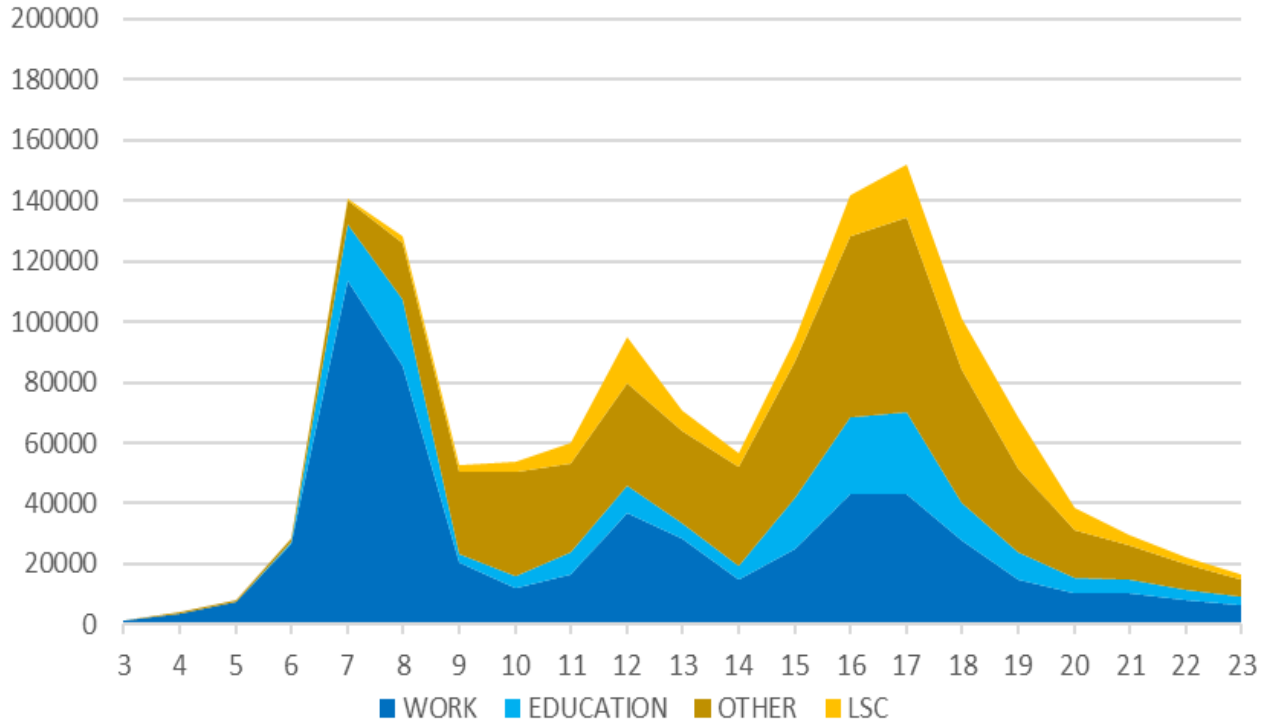
Graph 5: Comparison of total trips

4.2.2   Generation and attraction of single zones

The previous graphs were used to confirm that the model was able to capture the overall number of trips for each activity and that they were correctly spread throughout the day. Since the temporal distribution has then be validated, this section concentrates instead on the single zones, analysing the goodness of the approach for modelling the spatial distribution of flows. In general, we want to see if activities are distributed among zones according to our predictions: we will analyse an industrial zone (Kirchberg), a residential zone with a high number of schools (Limpetrsberg) and the city centre. Graph 7 introduces their total flows during the day, where we can see a strong difference between these zones. Following, these differences will be explained by analysing the repartition of trips among activities for each of the three zones.
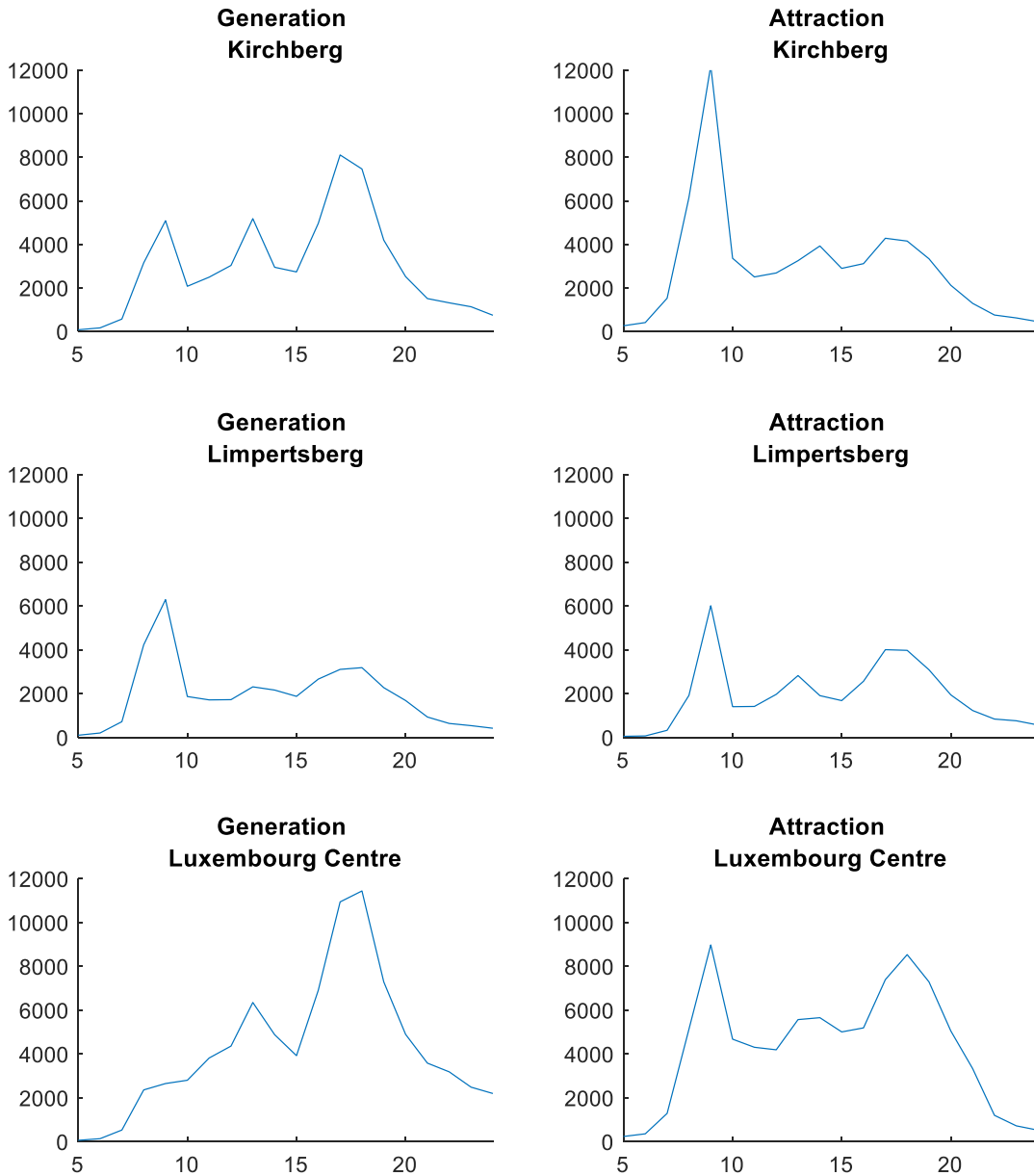
Graph 6: Temporal distribution of Generation and Attraction of different zones

   The industrial zone with the highest number of employees was found to be Kirchberg. Graph 8 shows the generation and the attraction of trips of this zone, with a curve representing each activity. The two images well represent what we expected: a strong attraction of workers in the morning, and a more dispersed peak in generation of home trips for people who go back from work in the afternoon.
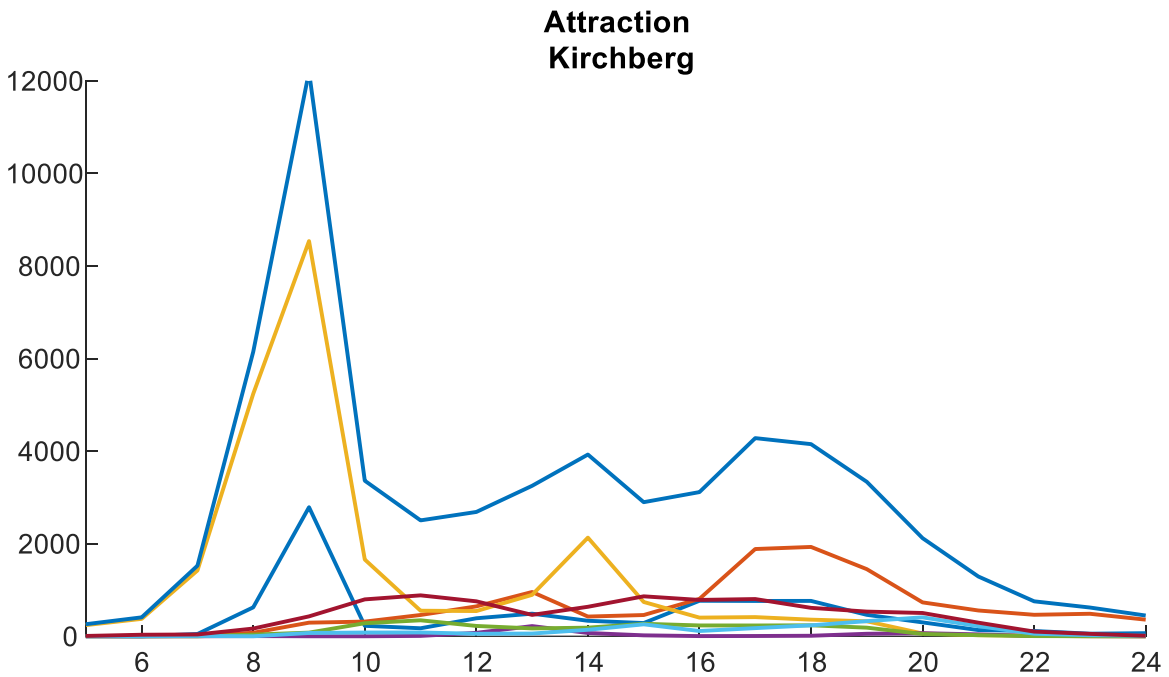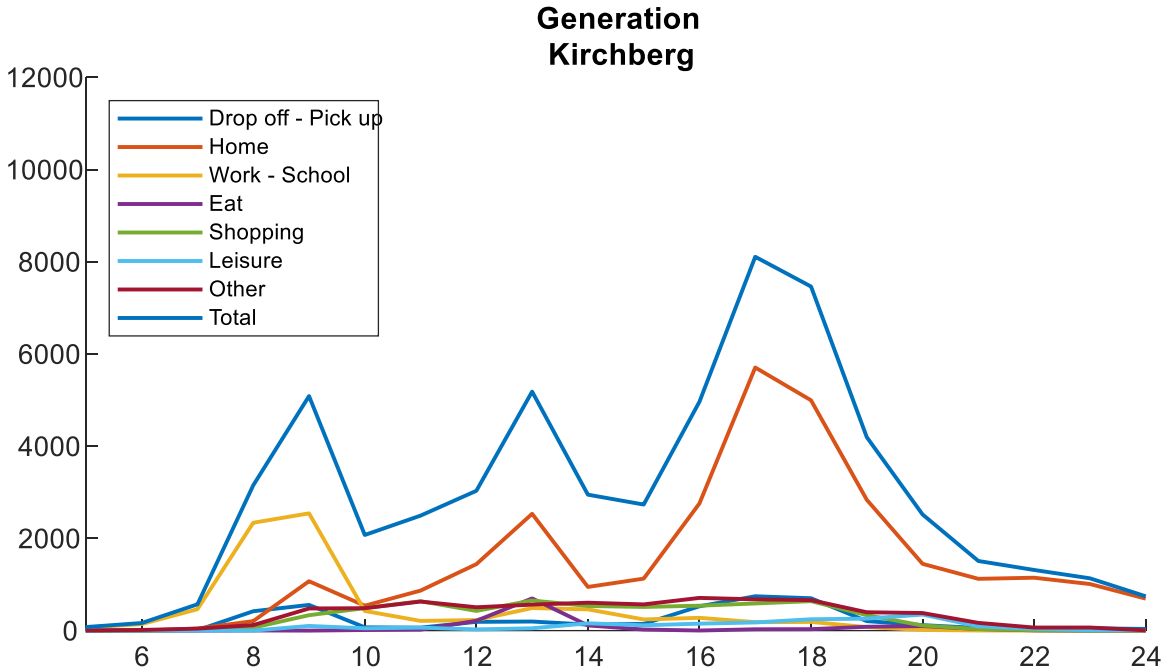
Graph 7: Temporal distribution of activity-specific Generation and Attraction in Kirchberg

Limpetrsberg, with its strong residential nature and its numerous big schools, is ideal to verify other dynamics. As we can see in Graph 9, there are several peaks. In the morning, generation shows people going to work, while in attraction it captures the users who bring their children to school. The afternoon peaks are also present, because in attraction there are the people coming back home from work, but also people accumulated during the day for education, work or secondary activities that leave the zone to go back home.

Generation
Limpertsberg

Attraction
Limpertsberg
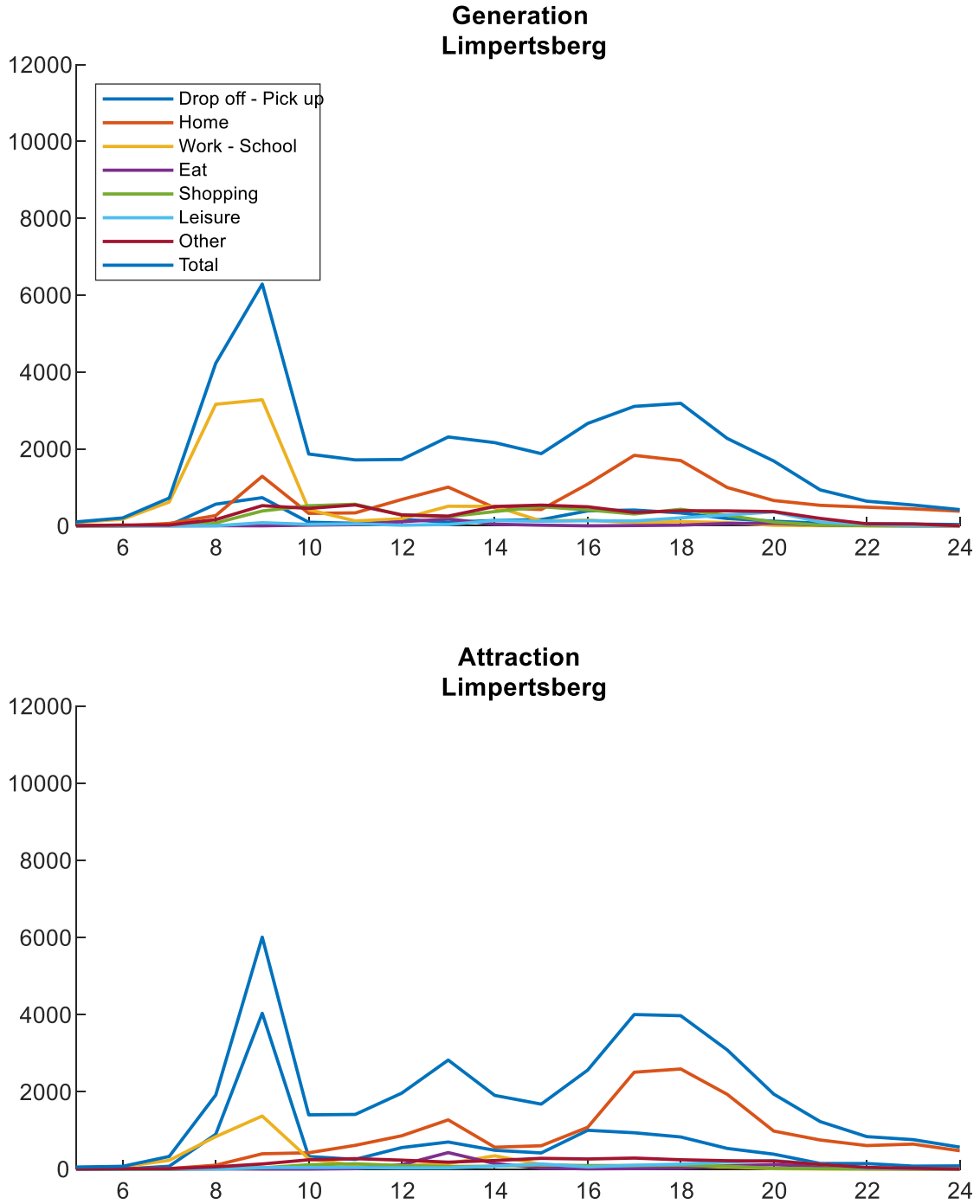
Graph 8: Temporal distribution of activity-specific Generation and Attraction in Limpertsberg

The last zone that has been considered also shows significant results. The city centre is in fact very important for the dynamics of the city, due to the high presence of POIs, its centrality, and the cultural relevance for the whole state. Graph 10 indeed captures interesting details: a high presence of secondary activities during the afternoon, confirmed by the huge amount of people leaving the zone to go back home in the evening; it also has a relevant attraction of workers in the morning, since the city centre contains many businesses, shops and activities.
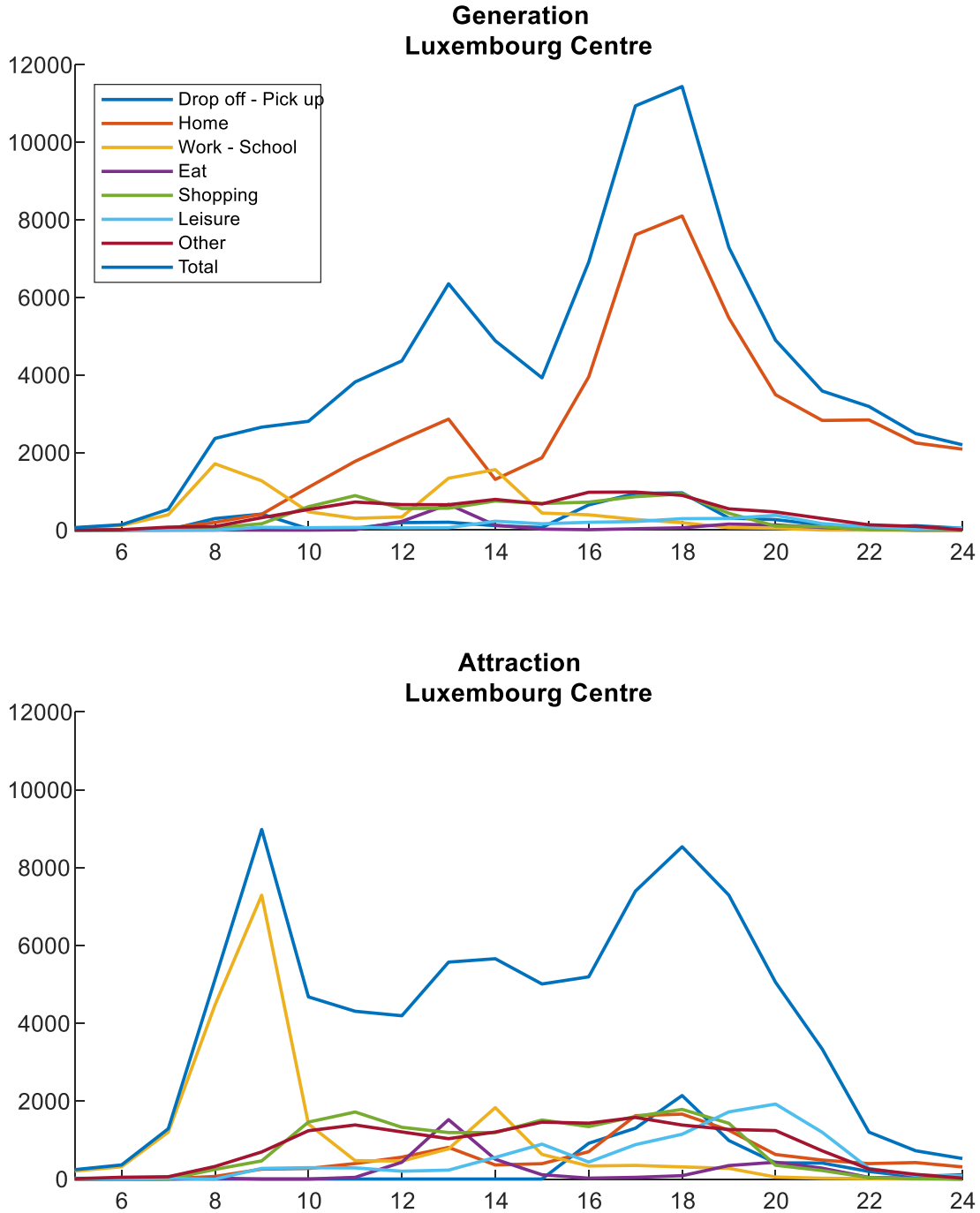
Graph 9: Temporal distribution of activity-specific Generation and Attraction in Luxembourg Centre

_____

4.3. Application of the Model to a Demand Forecasting future scenario

Now that the results have shown how the model is able to capture interzonal movements according to the different kinds of zones, this section aims at testing and verifying the sensitivity of the model to substantial changes in the input data. The model was tested by shifting a significant number of employees from one part of the city to the opposite side, in this case 20 000 employees have been relocated from Kirchberg all across to Howald and Gasperich (10 000 employees more for each zone). This was simulated by changing the attraction parameters of the single zones, increasing it for Howald and Gasperich of the amount for which Kirchberg has been decreased.
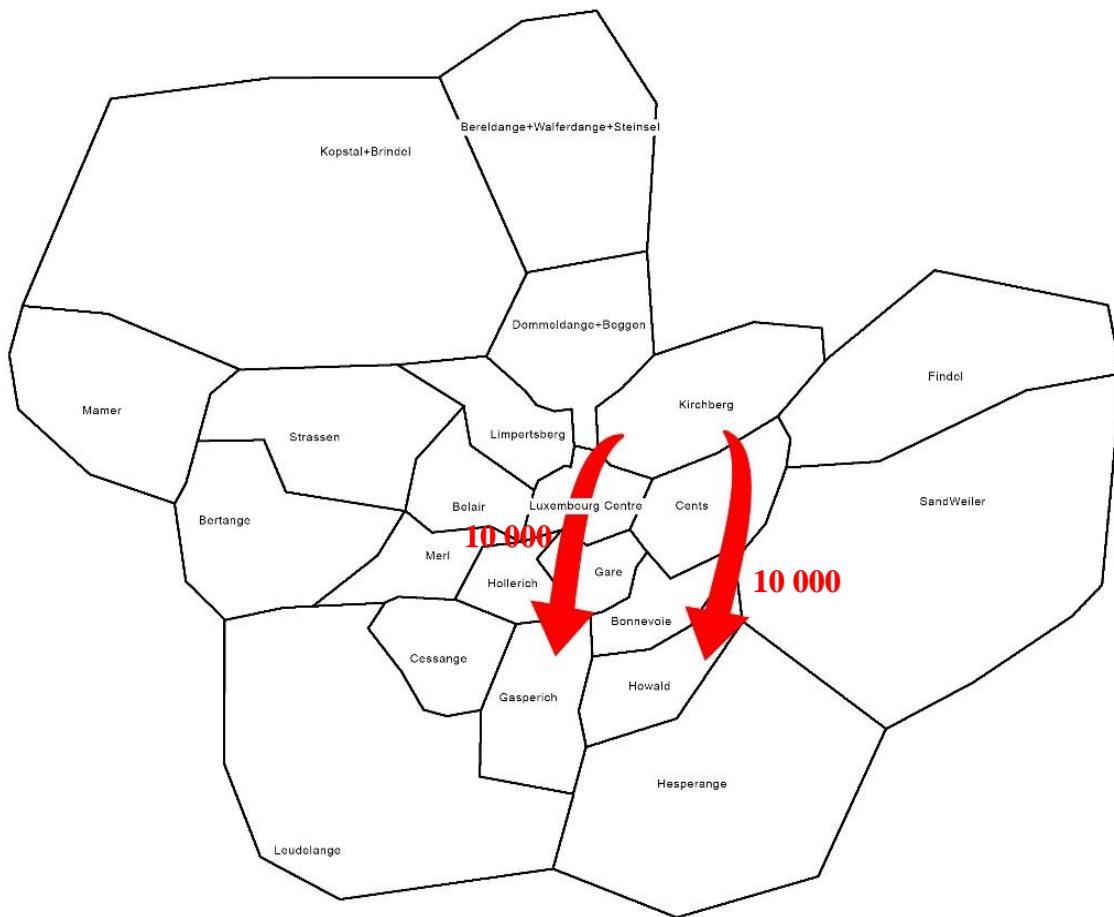
Figure 17: Workplace relocation of 20 000 users

The relevancy of this new scenario will not be in the variation of the generated and attracted commuting trips regarding the interested zones, since these will be a direct product of the attraction parameter. What would be far more interesting to achieve, is to prove that our model is able to capture the dynamics regarding how and where workers perform secondary activities. This is where this methodology majorly benefits from the knowledge of activity patterns, and how we want to exploit the information about the tours of our users.

_____

The main objective will then be to recognise a decrease in the secondary activities performed around Kirchberg, and their increase between Howald and Gasperich (section 4.3.3.).

This will be due to the variation of the number of users who perform their secondary activities in a certain ellipse. For these 20 000 hypothetical employees, the focal point referred to their zone of residence does not change, but the other one most certainly does, leading to 20 000 activity tours that will then be performed in ellipses located in a totally different part of the city.

Two logical steps have directed us towards the final demonstration of these dynamic: section 4.3.1. contains considerations about the teta parameter and explains how the final value was chosen; instead in section 4.3.2. a first analysis of the variations of OD flows has been performed.

4.3.1   Dependence of probabilities on theta

The most effective way for the modeller to manage the sensitivity of the model is to work on the parameter theta. As explained in section 3.3.3., the lower the theta, the shorter the distance to the POI that the users will choose. To have a clear idea of the effect of this parameter, Graph 10 presents the probabilities of choosing a certain POI change according to it.

The two sets of graphs each belong to one out of the 652 alternative locations, and each graph represents the probabilities of choosing that precise location for each of the 22 ellipses of which Kirchberg is one of the focal points.

_____
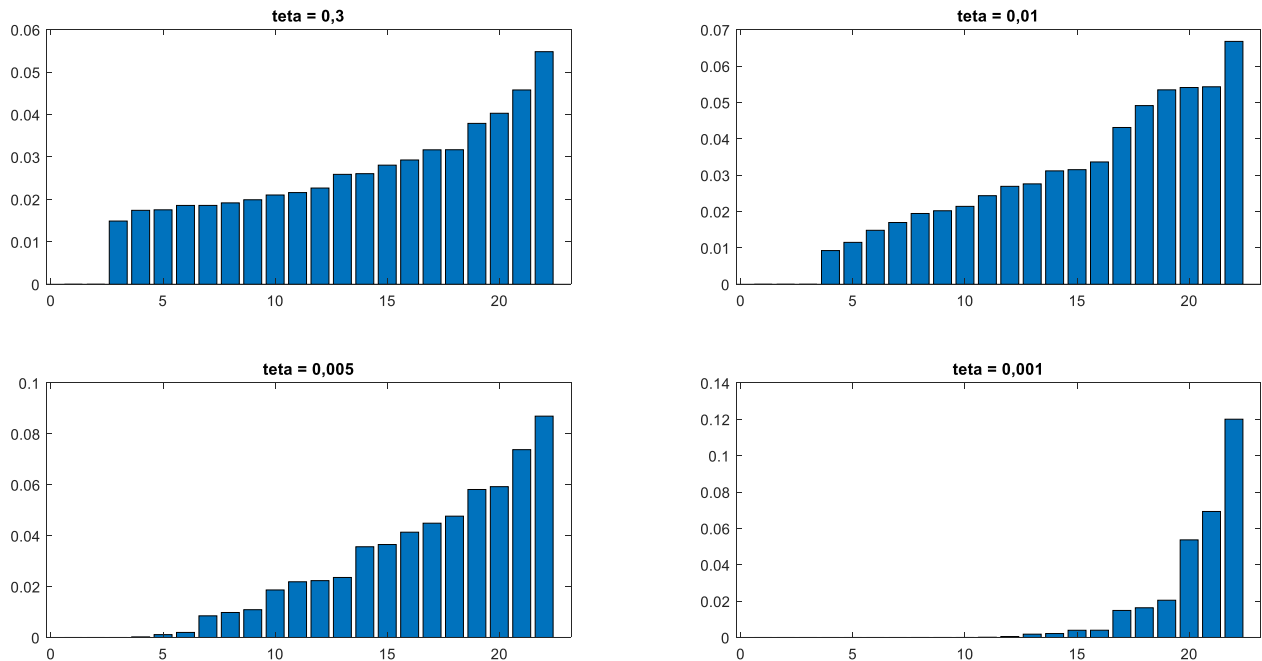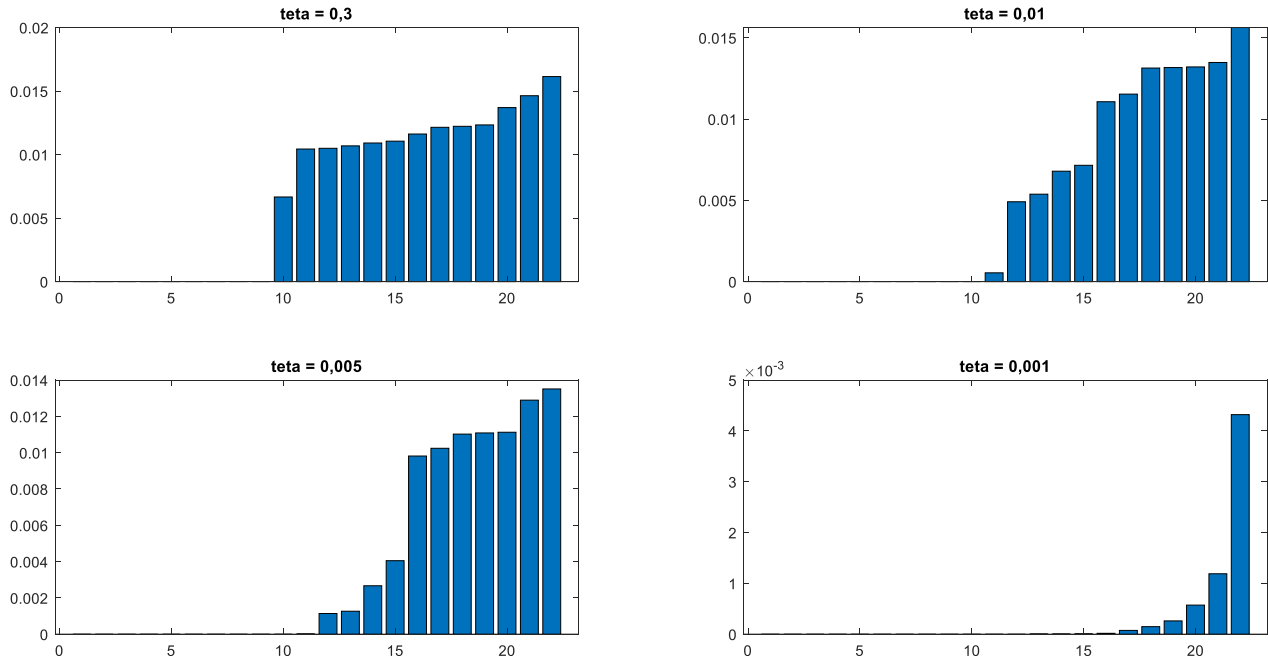
Graph 10: Probabilities of two POIs according to 22 different ellipses

The behaviour of theta is clearly explained by the differences in height of the singular bars, since higher thetas bring to a stochastic choice and low thetas narrow down the possible choices. In particular we can see that decreasing theta, for some ellipses the probabilities of choosing the first POI decrease and for others strongly increase, while for the second set of graphs all the probabilities decrease and many tend to zero.

The choice of which value of theta to use, should derive from a calibration process, where users are traced and their choices of locations for secondary activities are compared to the home and work pegs, as it has been done in (Sprumont & Viti, 2017). However, for a qualitative validation of this methodology, the theta that has been fixed for the future analyses is 0.005, considering that higher values still had a strong stochastic weight, and 0.001 made the model too much deterministic, not allowing enough choice for longer alternatives to users.

4.3.2   Scatter diagrams, differences in OD flows

The first analysis that has been performed to determine the sensitivity of the new scenario regard the variation of OD flows. The following scatter diagrams (Figure 19 and 20) represent how many and how much the hourly flows have changed due to the relocation. Each point represents the flow from one zone to another in a certain hour of the day. Since our interest mainly concerns secondary activities, only the afternoon hours from 15PM to 20 PM have been considered.

Figure 19 shows the change of the flows originated from only the primarily interested zones: Kirchberg, Howald and Gasperich

_____

Kirchberg - red
Gasperich and Howald - green
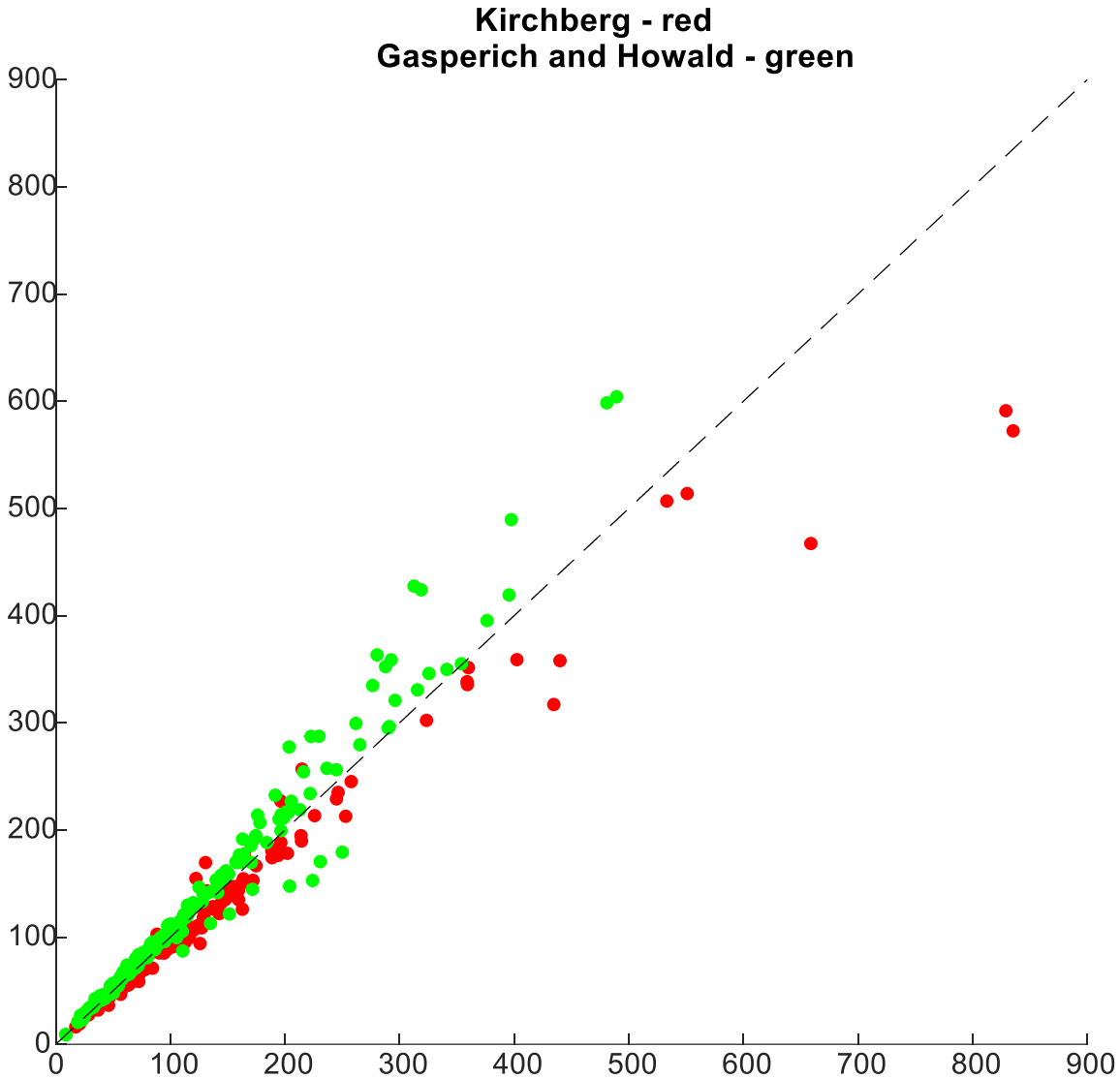
Figure 18: Variation of hourly OD flows generated from Kirchberg, Howald and Gasperich

These results clearly show that there are big variations, but this is normal since 20 000 workers have been relocated from Kirchberg to the other two zones.

What really states that our model is able to capture the variations in secondary activities and urges us to go deeper with the analyses is the following Figure 20. The points represent the variation of flows originated from all the other zones, where the total number of employees has stayed the same, and consequently work related trips have not changed much. This means that all the dots that are not close to the central line are due to people changing the location of their secondary activities.
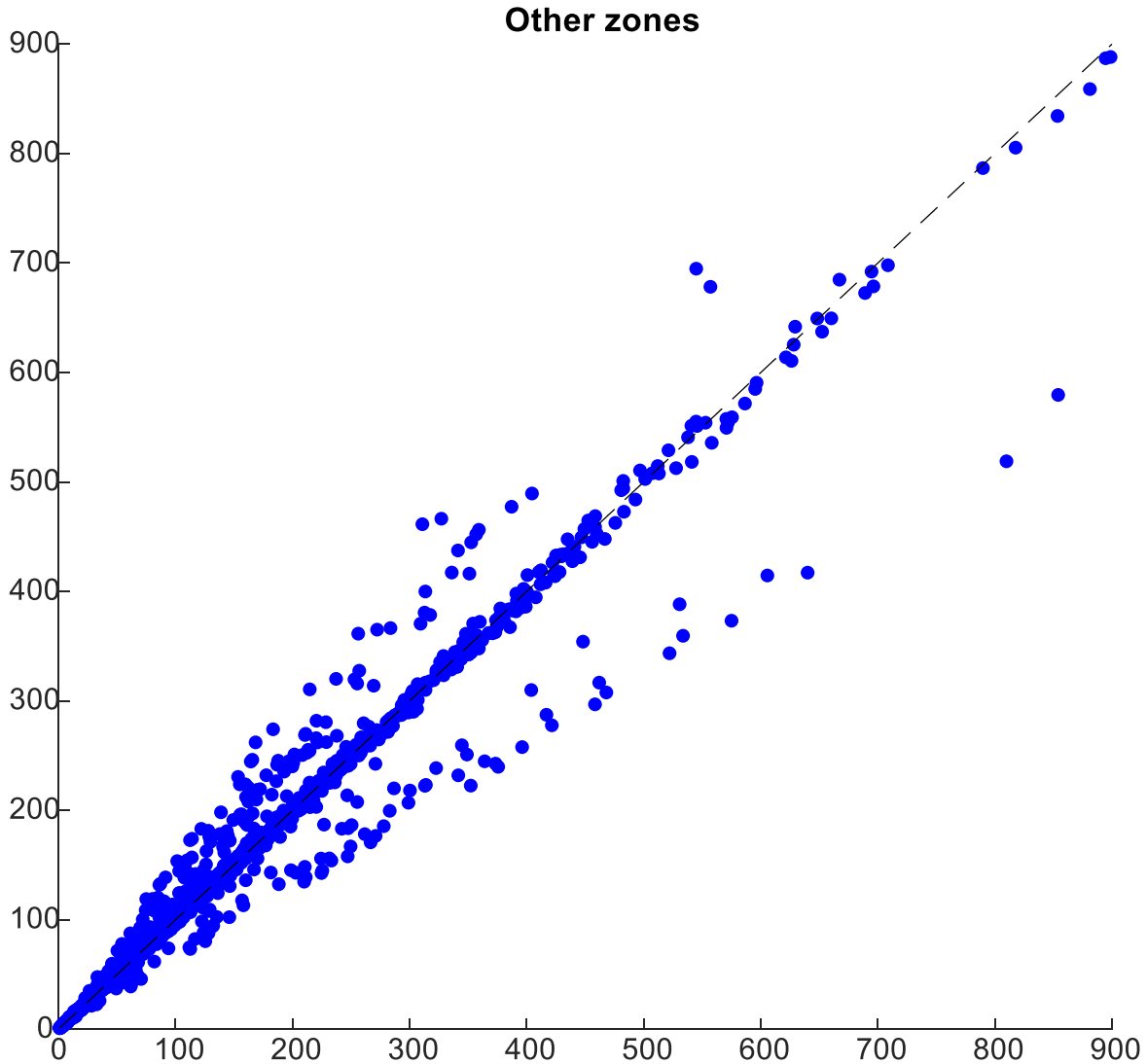
Figure 19: Variation of hourly OD flows generated from all zones except Kirchberg, Howald and Gasperich

A more accurate analysis on the individual zones and on the single secondary activities will better explain the variation in flows with the old scenario, but this is a first important step in realizing how sensitive the model is.

4.3.3   Secondary activities (map of flow)

To conclude the analysis of results, this section aims to demonstrate the effectiveness of the methodology by showing how much this job relocation has influenced the daily mobility. We expect that with a significant share of workers being transferred from one side of the city to the other, many zones will perceive a substantial variation of flows not only due to commuting, but also in relevant part due to secondary activities. With the conventional gravity model, differences in flows would be observable only

for areas directly interested from the relocation, proportionally to how much the attributes of these zones would change. This methodology aims at better representing the complex spectrum of urban variations that would consequence from this shift of workplace destination.

The following figures represents how the attractiveness for secondary activities has changed for all the zones. The stronger the colour, the higher the variation of trips attracted for shopping, eating, leisure or other reasons, excluding work and going back home.
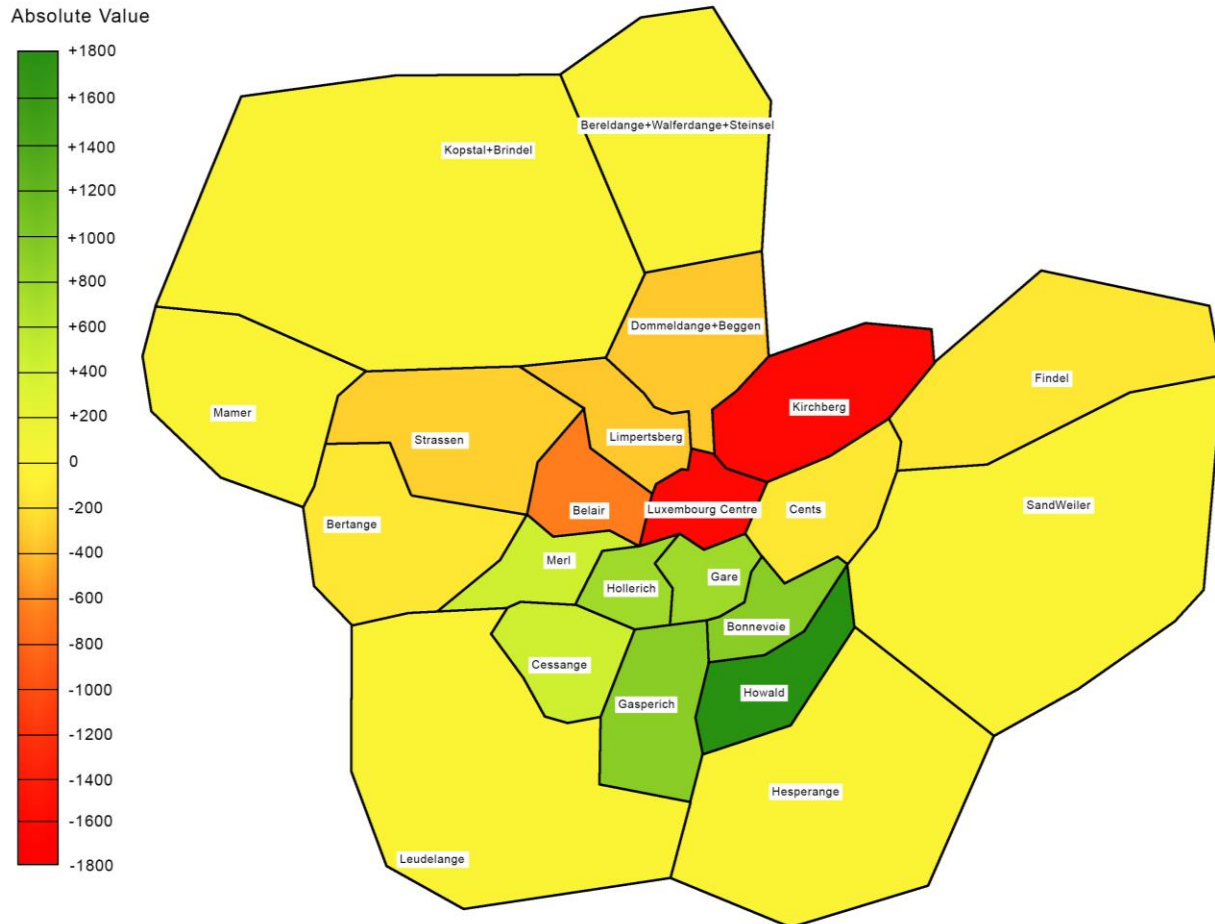
Figure 20: Variations of Attraction for secondary activities in each zone in absolute value

The map from Figure 21 clearly shows that the zones which have lost more in terms of attractivity for secondary activities are Kirchberg and Luxembourg Centre, while Howald and Gasperich have attracted much more than the first scenario. It seems that all other zones have instead been affected much less than these four zones: however, the magnitudes of these variations should be also put in comparison with the whole daily flows. Therefore, Figure 22 shows the same map but with the values taken as percentage of the total number of trips for secondary activity in each zone.
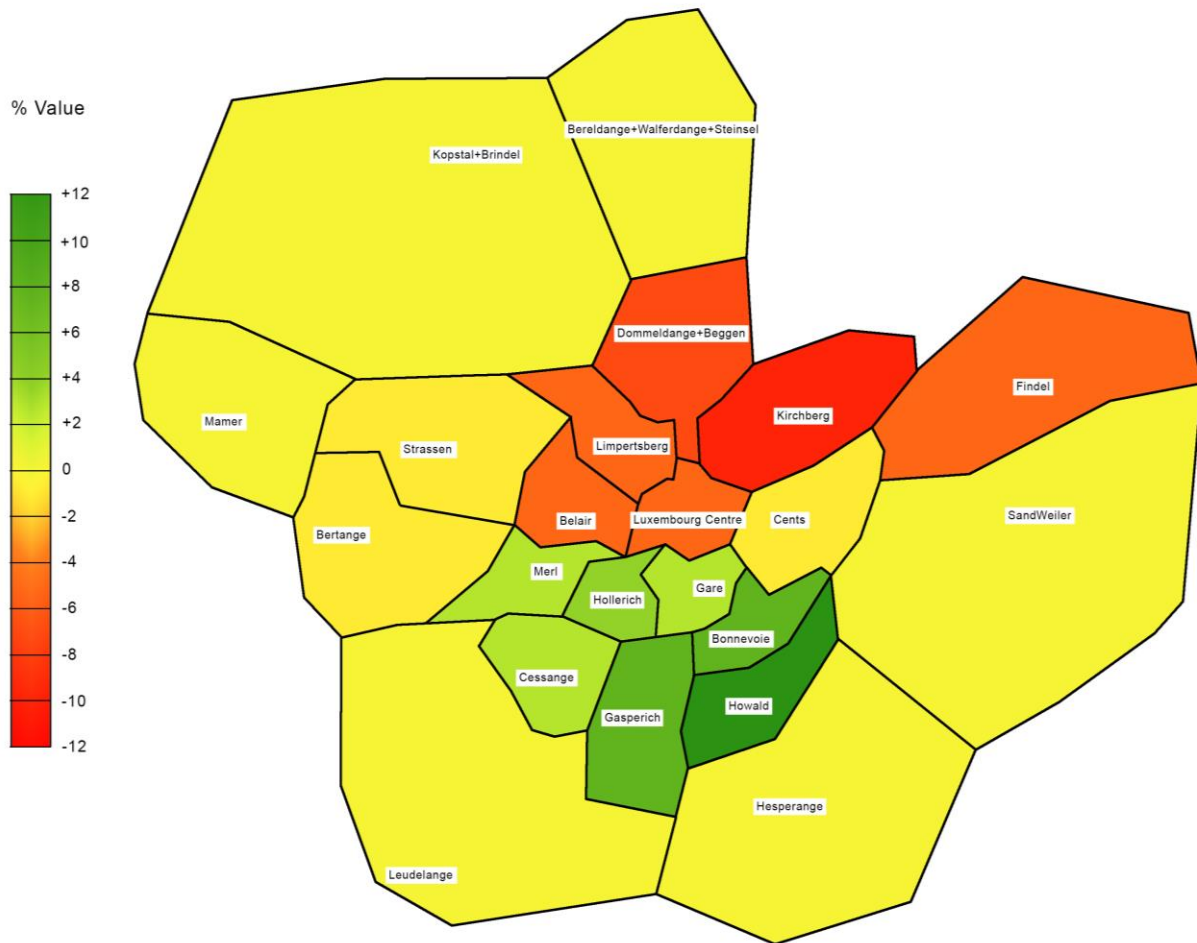
Figure 21: Variations of Attraction for secondary activities in each zone in percentual value

This makes us appreciate the extent of the effects of the relocation: more than 10 zones, almost all of the central ones, change of more than 5%. The zones around Gasperich and Howald, the ones close to the centre, also benefit of a high increase in terms of traffic of secondary activities. This is given to the fact that the POIs in these zones belong to the ellipses of many more users, even though maybe not all of the 20 000 workers. 5 000 trips for secondary activities that were performed in the zones around Kirchberg are now distributed among the southern zones of the city.

5. Conclusions and Future Developments

This thesis had the main goal of presenting a methodology that relied on three assumptions. First, the surge of BigData increased the set of opportunities for modellers, and research must follow the exploitation of the data as much as possible. Second, an emphasize on secondary activities is fundamental to account for travel behaviour. Finally, the theory behind Time Geography donates many chances to enhance the use of spatial relationships as an element driving travel behaviour. The results to the applied methodology showed that these assumptions definitely increase the possibilities of modellers to estimate travel demand.

The opensource format in which the list of POIs and the topology of the Network have been downloaded revealed to be an extremely powerful tool, and the amount of opportunities that still have to be seized is not quantifiable. Basically any kind of spatial relationship can be extrapolated from these data, permitting infinite applications.

The emphasis on secondary activities turned out to be very profitable, as much as the initial guess on individual human behaviour : Figures 20 and 21 visibly explicate how much the conventional gravity model has improved. Behaviours that conventional models were unable to capture are now significantly represented. Detaching the systematic demand from secondary activities, and assigning the location of the latter ones in terms of spatial relations, permitted to spread the total demand not only in one precise zone, but in all the surroundings of the home and work locations.

5.1. Future Developments

Even though the applications of this methodology arrived to notable results, in order to achieve significant scientific relevance, the model is wide open to future developments. Future research is suggested in different directions:

- Getting rid of the dependency on any kind of travel surveys would be an incredibly interesting target. Fortunately, BigData might resolve these issues, since as stated in section 3.4., many researchers are trying to infer activity patterns from Floating Car Data. If this methodology was enhanced with a set of activity patterns deriving from a complete opensource database, then the model would become much more powerful and applicable to any part of the world.

- An application at commercial level would need a calibration process: tuning the parameter theta assumes a very important role in the context of user behaviour. To accomplish this, a reliable dataset must be obviously available, and considering also the previous point, if individual traces were available they would be the input for both enhancements.

_____

- Population synthesis has been carried out through very basic procedure. Even though individual traces do not reproduce real users, there might be macro-dynamics that this population would not be able to represent. The use of more detailed socio-economic attributes, or the explicit representation of known commuting demand would indeed improve the representativeness of the derived OD flows.

In conclusion, the idea of using the extremely detailed disaggregated data from daily mobility of individuals and applying it to model macroscopic demand flows, hopes to bring many innovations in the field. Space time behaviour is rarely considered in aggregate demand models, but the work carried out in this thesis might inspire copious research in the integration between these two modelling approaches.

Appendix

A. Python Codes

A.1. Mapping POIs to Network

```python
exacts = xlrd.open_workbook('C:/Users/Administrator/Desktop/activities.xlsx')
sheet = exacts.sheet_by_index(0)
#data = [[sheet.cell_value(r, c) for c in range(sheet.ncols)] for r in
range(sheet.nrows)]
#act_lat = [sheet.cell_value(r, 1) for r in range(1,sheet.nrows)]
#act_lng = [sheet.cell_value(r, 2) for r in range(1,sheet.nrows)]

acts = [[sheet.cell_value(r, c) for c in range(2,4)] for r in
range(1,sheet.nrows)]

exnodes = xlrd.open_workbook('C:/Users/Administrator/Desktop/nodes.xlsx')
sheet = exnodes.sheet_by_index(0)
#data = [[sheet.cell_value(r, c) for c in range(sheet.ncols)] for r in
range(sheet.nrows)]
#nod_lat = [sheet.cell_value(r, 0) for r in range(1,sheet.nrows)]
#nod_lng = [sheet.cell_value(r, 1) for r in range(1,sheet.nrows)]
#nod_osmid = [sheet.cell_value(r, 3) for r in range(1,sheet.nrows)]

nods = [[sheet.cell_value(r, c) for c in [1, 0, 3]] for r in
range(1,sheet.nrows)]

#activities = [[ra,ca] for ]
#nodes = []

#for lata in act_lat:
#    for lnga in act_lng:
#        activities.append([lata,lnga])
#
#for latn in nod_lat:
#    for lngn in nod_lng:
#        nodes.append([latn,lngn])
dists = []
#inds = []

rnodes = []
r1 = []
r2 = []
r3 = []

for act in acts:
    ci = []
    import math
    for node in nods:
        dist = math.sqrt((act[0]-node[0])**2+(act[1]-node[1])**2)
        ci.append(dist)
#    dists.append(min(ci))
    import numpy as np
    ind = np.argmin(ci)
#    inds.append(ind)
```

_____

```
    rnodes.append(nods[ind])

for r in rnodes:
    r1.append(r[0])
    r2.append(r[1])
    r3.append(r[2])
```

A.2. Isochrones

```
import osmnx as ox, networkx as nx, geopandas as gpd, matplotlib.pyplot as
plt
from shapely.geometry import Point, LineString, Polygon
from descartes import PolygonPatch
ox.config(log_console=True, use_cache=True)

# configure the place, network type, trip times, and travel speed
Network='MiddleAugmentedLuxCity2.graphml'
Folder='C:/Users/filip/Desktop/Dati OSM/Test'

# Laod the street network
G = ox.load_graphml(Network, folder=Folder)
G_map = ox.project_graph(G)


network_type = 'drive'
trip_times = range(20) #in minutes
travel_speed = 25 #driving/walking speed in km/hour


##ADDRESS find the centermost node and then project the graph to UTM
#address='157 avenue Pasteur 2311 Luxembourg'
#Where = ox.graph_from_address(address, network_type=network_type)
##ox.plot_graph(Where)
#gdf_nodes = ox.graph_to_gdfs(Where, edges=False)
#x, y = gdf_nodes['geometry'].unary_union.centroid.xy
#center_node = ox.get_nearest_node(G, (y[0], x[0]))
#COORDINATES find the centermost node and then project the graph to UTM
y,x = 49.6038746, 6.0975787

Where = ox.graph_from_point((y, x), distance=750, network_type='all')
#ox.plot_graph(Where)
gdf_nodes = ox.graph_to_gdfs(Where, edges=False)
center_node = ox.get_nearest_node(G, (y, x))


# add an edge attribute for time in minutes required to traverse each edge
meters_per_minute = travel_speed * 1000 / 60 #km per hour to m per minute
for u, v, k, data in G_map.edges(data=True, keys=True):
    data['time'] = data['length'] / meters_per_minute

# get one color for each isochrone
iso_colors = ox.get_colors(n=len(trip_times), cmap='Reds', start=0.3,
return_hex=True)

# color the nodes according to isochrone then plot the street network
node_colors = {}
node_time = {}
for trip_time, color in zip(sorted(trip_times, reverse=True), iso_colors):
```

_____

_____

```
    subgraph = nx.ego_graph(G_map, center_node, radius=trip_time,
distance='time')
    for node in subgraph.nodes():
        node_colors[node] = color
        node_time[node] = trip_time
nt = [node_time[node] if node in node_colors else '1000' for node in
G_map.nodes()]
nc = [node_colors[node] if node in node_colors else 'none' for node in
G_map.nodes()]
ns = [20 if node in node_colors else 0 for node in G_map.nodes()]



fig, ax = ox.plot_graph(G_map, fig_height=8, node_color=nc, node_size=ns,
node_alpha=0.8, node_zorder=2)
```

B.   Matlab Codes

```
clear

for scenario = 3 % 1 = no relocation ; 2 = 50% from kirchberg ; 3 = 25% from
kirchberg
    for teta_numb = 1:3

        tic

        time_interval = 60; %in minutes
        n_intervals = 1440/time_interval;
        inter = 60;
        iterations = [1 49];
        tetalist = [0.3 0.01 0.005 0.001];



        teta = tetalist(teta_numb);
        teta1 = 0.3;

        %import data
        %import survey
        load data.mat

        idx = Survey.cleanSurvey(:,6)<0;
        Survey.cleanSurvey(idx,6)=0;
```

B.1. Activity Clustering

```
        %ACTIVITY CLASSIFICATION AND CLUSTERING
        type = data.type;
```

_____

```matlab
        iEat = type == 'food' | type =='restaurant'| type =='bakery'| type
==' bar'| type =='cafe'| type =='meal_delivery'| type =='meal_takeaway';
        iShop = type == 'clothing_store'| type == 'convenience_store'| type
== 'department_store'| type == 'electronics_store'| type ==
'furniture_store'| type == 'grocery_or_supermarket'| type ==
'hardware_store'| type == 'home_goods_store'| type == 'jewelry_store'| type
== 'liquor_store'| type == 'shoe_store'| type == 'shopping_mall'| type ==
'store'| type == 'supermarket';
        iLSC = type == 'amusement_park'| type == 'culture'| type == 'gym'|
type == 'library'| type == 'movie_theater'| type == 'museum'| type ==
'night_club'| type == 'park'| type == 'sport';
        iOther = type == 'bank'| type == 'beauty_salon'| type ==
'car_dealer'| type == 'car_rental'| type == 'car_repair'| type == 'car_wash'|
type == 'city_hall'| type == 'dentist'| type == 'doctor'| type == 'embassy'|
type == 'finance'| type == 'gas_station'| type == 'hair_care'| type ==
'health'| type == 'laundry'| type == 'lawyer'| type ==
'local_government_office'| type == 'pet_store'| type == 'pharmacy'| type ==
'physiotherapist'| type == 'post_office'| type == 'train_station'| type ==
'transit_station'| type == 'travel_agency';
```

B.2. Determination of Probabilities

```matlab
        %PROBABILITIES - For each zone, and for each activity type(eat, shop,
LSC,
        %Other), a logit model has been used. Utility is the inverse of the
        %distance in time from zone to activity as utility and teta is 0.3.

        act_zon = data.act_zon;
        zon_zon = data.zon_zon;

        %CIRCLE - probabilities of each activity for patterns not including
work.

        Cir_Prob = {3,4};

        for x = 1:4
            Cir_Prob{1,x} = zeros(652,22);
            Cir_Prob{2,x} = zeros(652,22);
            Cir_Prob{3,x} = zeros(652,22);
            for y = 1:22
                for z = 1:652
                    %create array for each activity set
                    Cir_Prob{1,1}(z,y) = act_zon(z,y)*iEat(z);
                    Cir_Prob{1,2}(z,y) = act_zon(z,y)*iShop(z);
                    Cir_Prob{1,3}(z,y) = act_zon(z,y)*iLSC(z);
                    Cir_Prob{1,4}(z,y) = act_zon(z,y)*iOther(z);

                    %create exp(v/teta)
                    if Cir_Prob{1,x}(z,y) == 0
                        Cir_Prob{2,x}(z,y) = 0;
                    else
```

_____

```matlab
                        Cir_Prob{2,x}(z,y) =
exp(1/teta1/(Cir_Prob{1,x}(z,y)));
                end
            end
        end
    end

    %assign probabilities with logit
    for x = 1:4
        for y = 1:22
            for z = 1:652
                Cir_Prob{3,x}(z,y) =
Cir_Prob{2,x}(z,y)/sum(Cir_Prob{2,x}(:,y));
            end
        end
    end

    %ELLIPSIS - probabilities with patterns including work

    ellipse = {3,22};
    %ellyssis = zeros(652,22);

    %find logical array if activities that make part of ellypsis
    for z = 1:22 %for every origin/home
        for y = 1:22 %for every destination/work
            for x = 1:652 %for every activity
                if zon_zon(z,y) > 5/4*(act_zon(x,z)+act_zon(x,y))-55/4 &&
zon_zon(z,y) > 0 %(5:15-34:39)
                %2*zon_zon(z,y) > act_zon(x,z)+act_zon(x,y) (isoscile)
                    ellipse{1,z}(x,y) = 1;
                else
                    ellipse{1,z}(x,y) = 0;
                end
            end
        end
        %ellipsis{1,z} = ellyssis;
    end

    %find time for each activity for each H-W pattern
    for z = 1:22 %for every origin/home
        for y = 1:22 %for every destination/work
            for x = 1:652 %for every activity
                ellipse{2,z}(x,y) = act_zon(x,z)+act_zon(x,y);
            end
        end
        %ellipsis{2,z} = ellyssis;
    end

    %take only times inside ellipsis
    for z = 1:22 %for every origin/home
        for y = 1:22 %for every destination/work
            for x = 1:652 %for every activity
                ellipse{3,z} = ellipse{2,z}.*ellipse{1,z};
            end
        end
```

_____

```matlab
        %ellipsis{3,z} = ellyssis;
    end



    EllEat = {3,22};
    EllShop = {3,22};
    EllLSC = {3,22};
    EllOther = {3,22};

    %create array for each activity set
    for x = 1:22
        EllEat{1,x}=ellipse{3,x}.*iEat;
        EllShop{1,x}=ellipse{3,x}.*iShop;
        EllLSC{1,x}=ellipse{3,x}.*iLSC;
        EllOther{1,x}=ellipse{3,x}.*iOther;
    end



    no_ell = zeros(22);
    %create exp(v/teta) and filter columns with no activities
    %variable no_ell is to show which od pair has no activities
    %1 = no food, 10 = no shop, 100 = no LSC, 1000 = no Other, 1011 =
comb
    for x = 1:22 %home
        for y = 1:22 %work
            for z = 1:652

                if sum(EllEat{1,x}(:,y)) == 0
                    no_ell(x,y) = 1;
                    EllEat{1,x}(:,y) = Cir_Prob{1,1}(:,y);
                else
                    if EllEat{1,x}(z,y)== 0
                        EllEat{2,x}(z,y) = 0;
                    else
                        EllEat{2,x}(z,y) =
exp(1/teta/(EllEat{1,x}(z,y)));
                        EllEat{2,x}(EllEat{2,x}==Inf) = exp(700);
                    end
                end
                if sum(EllShop{1,x}(:,y)) == 0
                    no_ell(x,y) = no_ell(x,y) + 10;
                    EllShop{1,x}(:,y) = Cir_Prob{1,2}(:,y);
                else
                    if EllShop{1,x}(z,y)== 0
                        EllShop{2,x}(z,y) = 0;
                    else
                        EllShop{2,x}(z,y) =
exp(1/teta/(EllShop{1,x}(z,y)));
                        EllShop{2,x}(EllShop{2,x}==Inf) = exp(700);
                    end
                end
                if sum(EllLSC{1,x}(:,y)) == 0
                    no_ell(x,y) = no_ell(x,y) + 100;
                    EllLSC{1,x}(:,y) = Cir_Prob{1,3}(:,y);
                else
```

```matlab
                            if EllLSC{1,x}(z,y)== 0
                                EllLSC{2,x}(z,y) = 0;
                            else
                                EllLSC{2,x}(z,y) =
exp(1/teta/(EllLSC{1,x}(z,y)));
                                EllLSC{2,x}(EllLSC{2,x}==Inf) = exp(700);
                            end
                        end
                        if sum(EllOther{1,x}(:,y)) == 0
                            no_ell(x,y) = no_ell(x,y) + 1000;
                            EllOther{1,x}(:,y) = Cir_Prob{1,4}(:,y);
                        else
                            if EllOther{1,x}(z,y)== 0
                                EllOther{2,x}(z,y) = 0;
                            else
                                EllOther{2,x}(z,y) =
exp(1/teta/(EllOther{1,x}(z,y)));
                                EllOther{2,x}(EllOther{2,x}==Inf) = exp(700);
                            end
                        end
                    end
                end
            end
        end

        %assign probabilities with logit
        for x = 1:22
            for y = 1:22
                for z = 1:652
                    EllEat{3,x}(z,y) =
EllEat{2,x}(z,y)/sum(EllEat{2,x}(:,y));
                    EllShop{3,x}(z,y) =
EllShop{2,x}(z,y)/sum(EllShop{2,x}(:,y));
                    EllLSC{3,x}(z,y) =
EllLSC{2,x}(z,y)/sum(EllLSC{2,x}(:,y));
                    EllOther{3,x}(z,y) =
EllOther{2,x}(z,y)/sum(EllOther{2,x}(:,y));
                end
            end
        end

        for y = 1:22
            idx = isnan(EllEat{3,y});
            EllEat{3,y}(idx)=0;
            idx = isnan(EllShop{3,y});
            EllShop{3,y}(idx)=0;
            idx = isnan(EllLSC{3,y});
            EllLSC{3,y}(idx)=0;
            idx = isnan(EllOther{3,y});
            EllOther{3,y}(idx)=0;
        end

        clear x y z ellyssis cats type act_zon zon_zon teta iEat iShop iLSC
iOther ellipsis idx
```

B.3. Extracting daily Patterns from dataset

```matlab
% BMW Activity patterns

n_users=unique(Survey.cleanSurvey(:,1));
day_of_year = unique(Survey.cleanSurvey(:,5));

Activity_Pattern=[];
r=1;
for u = 1:numel(n_users)
    for d = 1:numel(day_of_year)
        index=find(Survey.cleanSurvey(:,1)==n_users(u) &
Survey.cleanSurvey(:,5)==day_of_year(d));
        if isempty(index)==0
            % select only workdays (1:5)
            if Survey.cleanSurvey(index,8) <= 5
                data_tmp=Survey.cleanSurvey(index,3);
                Activity_Pattern(r,1:numel(data_tmp))=data_tmp;
            end
            if Survey.cleanSurvey(index,8) <= 5

data_tmp2=Survey.cleanSurvey(index,6)*3600+Survey.cleanSurvey(index,7)*60; %
time dep
                Activity_Pattern_dep(r,1:numel(data_tmp2))=data_tmp2;
            end
            if Survey.cleanSurvey(index,8) <= 5

data_tmp3=Survey.cleanSurvey(index,9)*3600+Survey.cleanSurvey(index,10)*60; %
time arr
                Activity_Pattern_arr(r,1:numel(data_tmp3))=data_tmp3;
            end
                r=r+1;
        end
    end
end
% cleaning people with just one trip with destination home

idx = Activity_Pattern(:,2)==0;
Activity_Pattern(idx,:) = [];
Activity_Pattern_arr(idx,:) = [];
Activity_Pattern_dep(idx,:) = [];

%Clustering
Activity_Clustered = Activity_Pattern;

idx = Activity_Clustered(:,:)==4;%school
Activity_Clustered(idx)=3; %work
idx = Activity_Clustered(:,:)==7;%long term shopping
Activity_Clustered(idx)=6; %short term shopping
idx = Activity_Clustered(:,:)==8; %personal business
Activity_Clustered(idx)=12; %other
idx = Activity_Clustered(:,:)==9; %visit to family and friends
Activity_Clustered(idx)=12; %other
idx = Activity_Clustered(:,:)==10; %walking riding etc
```

```matlab
        Activity_Clustered(idx)=12; %other


        %Creating Set of admissible patterns

        BMW_Cir_Act=zeros(size(Activity_Clustered,1),15);
        BMW_Cir_Dep=zeros(size(Activity_Clustered,1),15);
        BMW_Cir_Arr=zeros(size(Activity_Clustered,1),15);
        BMW_Ell_Act=zeros(size(Activity_Clustered,1),15);
        BMW_Ell_Dep=zeros(size(Activity_Clustered,1),15);
        BMW_Ell_Arr=zeros(size(Activity_Clustered,1),15);


        for x = 1:length(Activity_Clustered)
            for y = 1:15
                if Activity_Clustered(x,y) == 3
                    BMW_Ell_Act(x,:) = Activity_Clustered(x,:);
                    BMW_Ell_Dep(x,:) = Activity_Pattern_dep(x,:);
                    BMW_Ell_Arr(x,:) = Activity_Pattern_arr(x,:);
                else
                    BMW_Cir_Act(x,:) = Activity_Clustered(x,:);
                    BMW_Cir_Dep(x,:) = Activity_Pattern_dep(x,:);
                    BMW_Cir_Arr(x,:) = Activity_Pattern_arr(x,:);
                end
            end
        end

        idxE = BMW_Ell_Act(:,1)==0;
        BMW_Ell_Act(idxE,:) = [];
        BMW_Ell_Dep(idxE,:) = [];
        BMW_Ell_Arr(idxE,:) = [];

        idxC = (1-idxE);
        idxC = logical(idxC);
        BMW_Cir_Act(idxC,:) = [];
        BMW_Cir_Dep(idxC,:) = [];
        BMW_Cir_Arr(idxC,:) = [];

        clear Activity_Pattern_dep Activity_Clustered Activity_Pattern
Activity_Pattern_arr day_of_year idxC idxE index x y
        clear data_tmp data_tmp2 data_tmp3 d u r idx index n_users
day_of_year;
```

B.4. Assign location to each trip of each user

```matlab
        % define time and OD before p cycle;

        OD_temp{22,22} = zeros(iterations(2),n_intervals*inter);

        GEN_temp{iterations(2),22} = [];
        ATT_temp{iterations(2),22} = [];
```

_____

```matlab
        Employees = round(data.empl_prob(:,scenario)*data.pop_ell');
        % Employees(4,14) = Employees(4,14) + 40000;
        % Employees(4,15) = Employees(4,15) + 40000;
        % Employees(4,19) = Employees(4,19) + 80000;
        %
        for x = 1:22
            for y = 1:22
                for p = iterations(1):iterations(2)
                    OD_temp{x,y}(p,:)= zeros(1,n_intervals*inter);
                end
            end
        end


        for p = iterations(1):iterations(2)
            for x = 1:22
                GEN_temp{p,x} = zeros(12,n_intervals*inter);
                ATT_temp{p,x} = zeros(12,n_intervals*inter);
            end
            disp(p)
            Cir_Patt = {2,22};

            for x = 1:22
                Cir_Patt{1,x} = zeros(data.pop_cir(x),15);
                for y = 1:data.pop_cir(x)
                    z = randi([1 length(BMW_Cir_Act)]);
                    Cir_Patt{1,x}(y,:) = BMW_Cir_Act(z,:);
                    Cir_Patt{2,x}(y,:) = BMW_Cir_Dep(z,:);
                end
            end


            Ell_Patt = {22,22};
            Ell_Patt_Dep = {22,22};

            for h = 1:22
                for w = 1:22
                    Ell_Patt{w,h} = zeros(Employees(w,h),15);
                    Ell_Patt_Dep{w,h} = zeros(Employees(w,h),15);
                    for y = 1:Employees(w,h)
                        z = randi([1 length(BMW_Ell_Act)]);
                        Ell_Patt{w,h}(y,:) = BMW_Ell_Act(z,:);
                        Ell_Patt_Dep{w,h}(y,:) = BMW_Ell_Dep(z,:);
                    end
                end
            end


            %assign a specific activity location to each trip of every user's
pattern
            zon_lat = data.zon_lat_lng(:,1);
            zon_lng = data.zon_lat_lng(:,2);
            sch_lat = data.sch_lat_lng(:,1);
            sch_lng = data.sch_lat_lng(:,2);
            sch_zone = data.sch_lat_lng(:,3);
            act_lat = data.activities(:,2);
```

_____

_____

```matlab
            act_lng = data.activities(:,3);
            act_zone = data.activities(:,6);
            n_act = 1:size(data.activities,1);
            n_sch = 1:size(data.schools,1);

            Cir_Omnid = {1,22};
            Cir_Lat = {1,22};
            Cir_Lng = {1,22};
            Cir_Zon = {1,22};
            for z = 1:22
                Cir_Omnid{1,z} = zeros(data.pop_cir(z),15);
                Cir_Lat{1,z} = zeros(data.pop_cir(z),15);
                Cir_Lng{1,z} = zeros(data.pop_cir(z),15);
                for x = 1:15
                    for y = 1:size(Cir_Patt{1,z},1)
                        if Cir_Patt{1,z}(y,x) == 5 %food
                            idx =
n_act(find(rand<cumsum(Cir_Prob{3,1}(:,z)),1,'first'));
                            Cir_Omnid{1,z}(y,x) = data.acts(idx);
                            Cir_Lat{1,z}(y,x) = act_lat(idx);
                            Cir_Lng{1,z}(y,x) = act_lng(idx);
                            Cir_Zon{1,z}(y,x) = act_zone(idx);
                        elseif Cir_Patt{1,z}(y,x) == 6 %shop
                            idx =
n_act(find(rand<cumsum(Cir_Prob{3,2}(:,z)),1,'first'));
                            Cir_Omnid{1,z}(y,x) = data.acts(idx);
                            Cir_Lat{1,z}(y,x) = act_lat(idx);
                            Cir_Lng{1,z}(y,x) = act_lng(idx);
                            Cir_Zon{1,z}(y,x) = act_zone(idx);
                        elseif Cir_Patt{1,z}(y,x) == 11 %LSC
                            idx =
n_act(find(rand<cumsum(Cir_Prob{3,3}(:,z)),1,'first'));
                            Cir_Omnid{1,z}(y,x) = data.acts(idx);
                            Cir_Lat{1,z}(y,x) = act_lat(idx);
                            Cir_Lng{1,z}(y,x) = act_lng(idx);
                            Cir_Zon{1,z}(y,x) = act_zone(idx);
                        elseif Cir_Patt{1,z}(y,x) == 12 %other
                            idx =
n_act(find(rand<cumsum(Cir_Prob{3,4}(:,z)),1,'first'));
                            Cir_Omnid{1,z}(y,x) = data.acts(idx);
                            Cir_Lat{1,z}(y,x) = act_lat(idx);
                            Cir_Lng{1,z}(y,x) = act_lng(idx);
                            Cir_Zon{1,z}(y,x) = act_zone(idx);
                        elseif Cir_Patt{1,z}(y,x) == 2 %home
                            Cir_Omnid{1,z}(y,x) = data.zones(z);
                            Cir_Lat{1,z}(y,x) = zon_lat(z);
                            Cir_Lng{1,z}(y,x) = zon_lng(z);
                            Cir_Zon{1,z}(y,x) = z;
                        elseif Cir_Patt{1,z}(y,x) == 1 %pickup
                            idx =
n_sch(find(rand<cumsum(data.Schoolprob),1,'first'));
                            Cir_Omnid{1,z}(y,x) = data.schools(idx);
                            Cir_Lat{1,z}(y,x) = sch_lat(idx);
                            Cir_Lng{1,z}(y,x) = sch_lng(idx);
                            Cir_Zon{1,z}(y,x) = sch_zone(idx);
                        end
```

_____

```matlab
                end
            end
        end


        Ell_Omnid = {22,22};
        Ell_Lat = {22,22};
        Ell_Lng = {22,22};
        Ell_Zon = {22,22};
        for w = 1:22
            for h = 1:22
%               if h == w
%                   Ell_Omnid{w,h} = [];
%               else
                    Ell_Omnid{w,h} = zeros(Employees(w,h),15);
                    Ell_Lat{w,h} = zeros(Employees(w,h),15);
                    Ell_Lng{w,h} = zeros(Employees(w,h),15);
                    Ell_Zon{w,h} = zeros(Employees(w,h),15);
                    for x = 1:15
                        for y = 1:Employees(w,h)
                            if Ell_Patt{w,h}(y,x) == 5 %food
                                idx =
n_act(find(rand<cumsum(EllEat{3,w}(:,h)),1,'first'));
                                Ell_Omnid{w,h}(y,x) = data.acts(idx);
                                Ell_Lat{w,h}(y,x) = act_lat(idx);
                                Ell_Lng{w,h}(y,x) = act_lng(idx);
                                Ell_Zon{w,h}(y,x) = act_zone(idx);
                                %if x > 1
                                %Ell_Time{w,h}(y,x) =
data.act_zon(idx,Ell_Zon{w,h}(y,x-1))

                            elseif Ell_Patt{w,h}(y,x) == 6 %shop
                                idx =
n_act(find(rand<cumsum(EllShop{3,w}(:,h)),1,'first'));
                                Ell_Omnid{w,h}(y,x) = data.acts(idx);
                                Ell_Lat{w,h}(y,x) = act_lat(idx);
                                Ell_Lng{w,h}(y,x) = act_lng(idx);
                                Ell_Zon{w,h}(y,x) = act_zone(idx);
                            elseif Ell_Patt{w,h}(y,x) == 11 %LSC
                                idx =
n_act(find(rand<cumsum(EllLSC{3,w}(:,h)),1,'first'));
                                Ell_Omnid{w,h}(y,x) = data.acts(idx);
                                Ell_Lat{w,h}(y,x) = act_lat(idx);
                                Ell_Lng{w,h}(y,x) = act_lng(idx);
                                Ell_Zon{w,h}(y,x) = act_zone(idx);
                            elseif Ell_Patt{w,h}(y,x) == 12 %other
                                idx =
n_act(find(rand<cumsum(EllOther{3,w}(:,h)),1,'first'));
                                Ell_Omnid{w,h}(y,x) = data.acts(idx);
                                Ell_Lat{w,h}(y,x) = act_lat(idx);
                                Ell_Lng{w,h}(y,x) = act_lng(idx);
                                Ell_Zon{w,h}(y,x) = act_zone(idx);
                            elseif Ell_Patt{w,h}(y,x) == 2 %home
                                Ell_Omnid{w,h}(y,x) = data.zones(h);
                                Ell_Lat{w,h}(y,x) = zon_lat(h);
                                Ell_Lng{w,h}(y,x) = zon_lng(h);
```

```matlab
                                Ell_Zon{w,h}(y,x) = h;
                            elseif Ell_Patt{w,h}(y,x) == 1 %pickup
                                if Ell_Patt_Dep{w,h}(y,x) < 54000
                                    idx =
n_sch(find(rand<cumsum(data.Schoolprob),1,'first'));
                                    Ell_Omnid{w,h}(y,x) =
data.schools(idx);

                                    Ell_Lat{w,h}(y,x) = sch_lat(idx);
                                    Ell_Lng{w,h}(y,x) = sch_lng(idx);
                                    Ell_Zon{w,h}(y,x) = sch_zone(idx);
                                else
                                    idx =
n_act(find(rand<cumsum(EllLSC{3,w}(:,h)),1,'first'));
                                    Ell_Omnid{w,h}(y,x) = data.acts(idx);
                                    Ell_Lat{w,h}(y,x) = act_lat(idx);
                                    Ell_Lng{w,h}(y,x) = act_lng(idx);
                                    Ell_Zon{w,h}(y,x) = act_zone(idx);
                                end
                            elseif Ell_Patt{w,h}(y,x) == 3 %work
                                Ell_Omnid{w,h}(y,x) = data.zones(w);
                                Ell_Lat{w,h}(y,x) = zon_lat(w);
                                Ell_Lng{w,h}(y,x) = zon_lng(w);
                                Ell_Zon{w,h}(y,x) = w;
                            end
                        end
                    end
%                 end
            end
        end

        % X = [1 2 3 4]
        % P = [0.1 0.3 0.4 0.2]
        % f = zeros(100,1);
        % for z = 1:100
        %     f(z,1) = X(find(rand<cumsum(P),1,'first'));
        % end
        clear x y z h w zon_lat zon_lng act_lat act_lng sch_lat sch_lng
n_act n_sch idx

        %TRIPS - collect all patterns in a unique matrix containing all
trips
        %trips of pattern without work
        tripcir = 0;
        for h = 1:22
            for x = 1:15
                for y = 1:data.pop_cir(h)
                    if Cir_Omnid{1,h}(y,x)==0
                    else, tripcir = tripcir + 1;
                    end
                end
            end
        end


        tripell = 0;
        for h = 1:22
```

```
            for w = 1:22
                if h == w
                else
                    for x = 1:15
                        for y = 1:Employees(w,h)
                            if Ell_Omnid{w,h}(y,x)==0
                            else, tripell = tripell + 1;
                            end
                        end
                    end
                end
            end
        end

        Trips_cir = zeros(tripcir,5);
        Trips_ell = zeros(tripell,5);

        z = 1;
        for h = 1:22
            for y = 1:data.pop_cir(h)
                for x = 1:15
                    if Cir_Omnid{1,h}(y,x)==0
                    else, Trips_cir(z,1) = Cir_Patt{2,h}(y,x); %deptime
                        Trips_cir(z,8) = Cir_Patt{1,h}(y,x); %activity
                        Trips_cir(z,4)=Cir_Lat{1,h}(y,x); %lat of
destination
                        Trips_cir(z,5)=Cir_Lng{1,h}(y,x); %lng of
destination
                        Trips_cir(z,6)=Cir_Zon{1,h}(y,x); %zone of
destination
                        if x == 1
                            Trips_cir(z,2) = data.zon_lat_lng(h,1); %lat
of home
                            Trips_cir(z,3) = data.zon_lat_lng(h,2); %lng
of home
                            Trips_cir(z,7) = h; %lng of home
                        else
                            Trips_cir(z,2) = Trips_cir(z-1,4); %lat of
previous activity
                            Trips_cir(z,3) = Trips_cir(z-1,5); %lng of
previous activity
                            Trips_cir(z,7) = Trips_cir(z-1,6); %zone of
previous activity
                        end
                        z = z+1;
                    end
                end
            end
        end

        z = 1;
        for h = 1:22
            for w = 1:22
                for y = 1:Employees(w,h)
                    for x = 1:15
    %                   if h==w
```

```
        %                               else
                                            if Ell_Omnid{w,h}(y,x)==0
                                            else, Trips_ell(z,1) =
Ell_Patt_Dep{w,h}(y,x); %deptime

                                                Trips_ell(z,8) = Ell_Patt{w,h}(y,x);
%activity

                                                Trips_ell(z,4)=Ell_Lat{w,h}(y,x); %lat of
destination

                                                Trips_ell(z,5)=Ell_Lng{w,h}(y,x); %lng of
destination

                                                Trips_ell(z,6)=Ell_Zon{w,h}(y,x); %zone
of destination

                                            if x == 1
                                                Trips_ell(z,2) =
data.zon_lat_lng(h,1); %lat of home

                                                Trips_ell(z,3) =
data.zon_lat_lng(h,2); %lng of home

                                                Trips_ell(z,7) = h; %zone of home
                                            else
                                                Trips_ell(z,2) = Trips_ell(z-1,4);
%lat of previous activity

                                                Trips_ell(z,3) = Trips_ell(z-1,5);
%lng of previous activity

                                                Trips_ell(z,7) = Trips_ell(z-1,6);
%zone of previous activity

                                            end
                                            z = z+1;
                                        end
        %                           end
                            end
                        end
                    end
                end


            TRIPS = [Trips_cir; Trips_ell];
            clear x y z w h
```

B.5. Aggregation Process

```
        for y = 1:size(TRIPS,1)
            for x = 1:n_intervals*inter
                if TRIPS(y,1) < time_interval/inter*60*x && TRIPS(y,1)>=
time_interval/inter*60*(x-1)
        %                   time{p,x} = [time{p,x}; TRIPS(y,:)];
                    OD_temp{TRIPS(y,7),TRIPS(y,6)}(p,x) =
OD_temp{TRIPS(y,7),TRIPS(y,6)}(p,x) + 1;
                    GEN_temp{p,TRIPS(y,7)}(TRIPS(y,8),x) =
GEN_temp{p,TRIPS(y,7)}(TRIPS(y,8),x) + 1;
                end
                if TRIPS(y,1) < time_interval/inter*60*x - 1200 &&
TRIPS(y,1)>= time_interval/inter*60*(x-1) - 1200
                    ATT_temp{p,TRIPS(y,6)}(TRIPS(y,8),x) =
ATT_temp{p,TRIPS(y,6)}(TRIPS(y,8),x) + 1;
```

_____

```
                        end
                    end
                end

                for x = 1:22
                    GEN_temp{p,x}(4,:) = [];
                    GEN_temp{p,x}(6,:) = [];
                    GEN_temp{p,x}(6,:) = [];
                    GEN_temp{p,x}(6,:) = [];
                    GEN_temp{p,x}(6,:) = [];
                    ATT_temp{p,x}(4,:) = [];
                    ATT_temp{p,x}(6,:) = [];
                    ATT_temp{p,x}(6,:) = [];
                    ATT_temp{p,x}(6,:) = [];
                    ATT_temp{p,x}(6,:) = [];
                    for y = 1:n_intervals*inter
                        GEN_temp{p,x}(8,y) = sum(GEN_temp{p,x}(:,y));
                        ATT_temp{p,x}(8,y) = sum(ATT_temp{p,x}(:,y));
                    end

                end


                for z = 1:22
                    for a = 1:8
                        for x = 1:n_intervals
                            GEN{p,z}(a,x) = sum(GEN_temp{p,z}(a,((x-
1)*inter+1):(x*inter)));
                            ATT{p,z}(a,x) = sum(ATT_temp{p,z}(a,((x-
1)*inter+1):(x*inter)));
                        end
                    end
                end

            end

            for o = 1:22
                for d = 1:22
                    for i = iterations(1):iterations(2)
                        for x = 1:24
                            OD{d,o}(i,x) = sum(OD_temp{d,o}(i,((x-
1)*inter+1):(x*inter)));
                        end
                    end
                end
            end

            for o = 1:22
                for d = 1:22
                    for x = 1:24
                        OD{d,o}(iterations(2)+1,x) =
mean(OD{d,o}(1:iterations(2),x));
                    end
                end
            end
```

_____

```
        itr = iterations(2);
        for x = 1:22
            GEN{itr+1,x} = zeros(8,n_intervals);
            ATT{itr+1,x} = zeros(8,n_intervals);
            for i = 1:itr
                GEN{itr+1,x}= GEN{itr+1,x} + GEN{i,x};
                ATT{itr+1,x}= ATT{itr+1,x} + ATT{i,x};
            end
            GEN{itr+1,x}= round(GEN{itr+1,x}/itr);
            ATT{itr+1,x}= round(ATT{itr+1,x}/itr);
            for y = 1:24
                GEN{itr+1,x}(8,y) = sum(GEN{itr+1,x}(1:7,y));
                ATT{itr+1,x}(8,y) = sum(ATT{itr+1,x}(1:7,y));
            end
        end


        clear i itr x n_intervals iterations inter
        clear x y t sch_zone act_zone
        clear BMW_Cir_Act BMW_Cir_Arr BMW_Cir_Dep BMW_Ell_Act BMW_Ell_Arr
BMW_Ell_Dep
        clear Cir_Lat Cir_Lng Cir_Omnid Cir_Patt Cir_Prob Cir_Zon EllEat
EllLSC
        clear Ell_Lat Ell_Lng Ell_Omnid Ell_Patt Ell_Patt_Dep Ell_Zon ellipse
Employees
        clear tripcir tripell Trips_cir Trips_ell sch_zone act_zone p EllShop
EllOther
        clear OD_temp GEN_temp ATT_temp teta1 tetalist o d a

        save(['60sc' num2str(scenario) 'teta' num2str(teta_numb) '.mat'])
        toc
    end
end
```

Bibliography

Alexander, L., Jiang, S., Murga, M., & González, M. (2015). Origin–destination trips by purpose and time of day inferred from mobile phone data. *Transportation research part c: emerging technologies, 58*, 240-250.

Anda, C., Erath, A., & Fourie, P. (2017). Transport modelling in the age of big data. *International Journal of Urban Sciences, 21(sup1)*, 19-42.

Barthelemy, J., & Toint, P. (2013). Synthetic population generation without a sample. *Transportation Science, 47(2)*, 266-279.

Bates, J. (2007). Hsitory of demand modelling. In *Handbook of Transport Modelling: 2nd Edition* (p. 11-34). Emerald Group Publishing Limited.

Bates, J., & Dasgupta, M. (1990). *Review of Techniques of Travel Demand Analysis: Interim Report.* Transport and Road Research Laboratory.

Ben-Akiva, M., & Bowman, J. (1998). Activity Based travel Demand Model Systems. In *Equilibrium and Advanced Transportation Modelling* (p. 27-46). Boston, MA: Springer, US.

Bhat, C., & Koppelman, F. (1999). Activity-based modeling of travel demand. In *Handbbook of transportation Science* (p. 35-61). Springer, Boston, MA.

Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems, 65*, 126-139.

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

Cascetta, E. (2009). *Transportation systems analysis: models and applications (Vol. 29).* Springer Science & Business Media.

Chapin, F. S. (1974). *Human activity pattern in the city.* Wiley.

Chen, C. M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. *Transportation research part C: emerging technologies, 68*, 285-299.

Chen, C., Bian, L., & Ma, J. (2015). From traces to trajectories: How well can we guess activity locations from mobile phone traces? *Transportation Research Part C: Emerging Technologies, 46*, 326-337.

Di Donna, S., Cantelmo, G., & Viti, F. (2015). A Markov chain dynamic model for trip generation and distribution based on CDR. *Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 243-250.

Dijst, M., & Vidakovic, V. (2000). Travel time ratio: the key factor of spatial reach. *Transportation 27(2)*, 179-199.

Ettema, D. (1996). *Activity-based travel demand modeling.* Technische Universiteit Eindhoven.

_____

González, M., Hidalgo, C., & Barabasi, A. (2008). Understanding individual human mobility patterns. *Nature, 453(7196), 779*.

Hägerstraand, T. (1970). What about people in regional science? *Papers in regional science, 24(1)*, 7-24.

Huang, L., Li, Q., & Yue, Y. (2010). Activity identification from GPS trajectories using spatial temporal POIs' attractiveness. *In Proceedings of the 2nd ACM SIGSPATIAL International Workshop on location based social networks*, 27-30.

Krings, G., Calabrese, F., Ratti, C., & Blondel, V. (2009). Urban gravity: a model for inter-city telecommunication flows. *Journal of Statistical Mechanics: Theory and Experiment, 2009(07), L07003*.

Lenntorp, B. (1976). Paths in space-time environments: A time-geographic study of movement possibilities of individuals. *Lund Studies in Geography Series B Human Geography, (44)*.

McNally, M. G. (2007a). The four step model. In *Handbook of Transport Modelling: 2nd Edition* (p. 35-53). Emerald Group Publishing Limited.

McNally, M. G. (2007b). The activity-based approach. In *Handbook of Transport Modelling: 2nd Edition* (p. 56-68). Emerald Group Publishing Limited.

Mitchell, R., & Rapkin, C. (1954). Urban Traffic: A function of land use. *New York: Columbia University Press*.

Ortúzar, J., & Willumsen, L. (2011). *Modelling transport.* John Wiley & Sons.

Phithakkitnukoon, S., Horanont, T., Di Lorenzo, G., Shibasaki, R., & Ratti, C. (2010). Activity-aware map: Identifying human daily activity pattern using mobile phone data. In *International Workshop on Human Behavior Understanding* (p. 14-25). Springer, Berlin, Heidelberg.

Saadi, I., Mustafa, A., Teller, J., Farooq, B., & Cools, M. (2016). Hidden Markov Model-based population synthesis. *Transportation Research Part B: Methodological, 90*, 1-21.

Sceffer, A., Cantelmo, G., & Viti, F. (2017). Generating macroscopic, purpose-dependent trips through Monte Carlo sampling techniques. *Transportation Research Procedia, 27*, 585-592.

Schönfelder, S., & Axhausen, K. (2003). Activity spaces: measures of social exclusion? *Transport policy, 10(4)*, 273-286.

Simini, F., González, M., Maritan, A., & Barabási, A. (2012). A universal model for mobility and migration patterns. *Nature,484(7392)*, 96.

Spinsanti, L., Celli, F., & Renso, C. (2010). Where you stop is who you are: understanding people's activities by places visited. *In the proceedings of Behaviour Monitoring and Interpretation (BMI) workshop*.

Sprumont, F., & Viti, F. (2017). *The effect of workplace relocation on individuals' activity travel behaviour.*

Timmermans, H., Arentze, T., & Joh, C. (2002). Analysing space-time behaviour: new approaches to old problems. *Progress in human geography, 26(2)*, 175-190.

_____

Wilson, A. G. (1974). *Urban and regional models in geography and planning.* John Wiley & Sons Inc.

Xie, k., Deng, K., & Zhou, X. (2009). From trajectories to activities: a spatio-temporal join approach. *In Proceedings of the 2009 International Workshop on Location Based Social Networks*, 25-32.