SPSD II

SPATIAL ANALYSIS AND MODELLING BASED ON ACTIVITIES (SAMBA)

PH. TOINT, I. THOMAS, A. VERHETSEL, F. WITLOX, E. CORNELIS

GENERAL ISSUES

AGRO-FOOD

ENERGY

PART 1

SUSTAINABLE PRODUCTION AND CONSUMPTION PATTERNS

TRANSPORT

.be

SCIENTIFIC SUPPORT PLAN FOR A SUSTAINABLE DEVELOPMENT POLICY
(SPSD II)

Part 1:
Sustainable production and consumption patterns

FINAL REPORT

SPATIAL ANALYSIS AND MODELLING BASED ON ACTIVITIES (SAMBA)

CP/41

Philippe Toint, Eric Cornélis and Laetitia Legrain
Facultés Universitaires Notre- Dame de La Paix - Namur
Groupe de Recherche sur Les Transports

Ann Verhetsel, Dries Van Hofstraeten
Universiteit Antwerpen
Vakgroep Transport en Ruimte

Prof. Frank Witlox, Hans Tindemans
Universiteit Gent
Vakgroep Geografie

Prof. Isabelle Thomas, Hakim Hammadou
Université Catholique de Louvain-La-Neuve
Département de géographie

September 2005

BELGIAN SCIENCE POLICY

D/2007/1191/19
Published in 2007 by the Belgian Science Policy
Rue de la Science 8
Wetenschapsstraat 8
B-1000 Brussels
Belgium
Tel: +32 (0)2 238 34 11 – Fax: +32 (0)2 230 59 12
http://www.belspo.be

Contact person:
Mrs Hilde Van Dongen
Secretariat: +32 (0)2 238 37 61

TABLE OF CONTENTS

1. Introduction

1.1 Context and summary

SAMBA stands for Spatial Analysis and Modelling Based on Activities. This project aims to progress in travel behaviour understanding by means of the activity-based approach which is indisputably the most in vogue approach. Moreover by focusing on the spatial components of travel we challenge ourselves to fill the lack of research in spatial analysis of activity chains. Our last objective is to estimate the entire Belgian mobility demand for all trip modes and purposes. Indeed, the main objective of this project is to obtain an O/D matrix reflecting the mobility demand in Belgium for all trip purposes in 1999. Such a matrix is fundamental for transport planification. But, up to now, the only available matrix is one deduced from the data collected during the 1991 census. Besides the fact that the mobility has largely evolved since ten years and that, therefore, we can have doubt on the actual representativity of this data, it must be mentioned that, in this census, only "house-work" and "house-school" trips have been recorded. The national survey on households' mobility (MOBEL) clearly showed that these trips purposes contain less than 50% of the demand. And, if the matrix can be updated with data from the next census (2001), this will still be limited to the same trip purposes, once more neglecting a large part of the demand.

From this goal, a second one, necessary to obtain the first objective but of interest in itself, can be derived: enriching the databases from the national survey and from other regional (Flanders) or local (Antwerp, Ghent, Hasselt) surveys with a spatial dimension. Of course, studying the demand implies a spatialization. So, we will have to build it from the addresses collected within the surveys. Probably, this building will help discovering still hidden aspects of the results of these surveys.

It must be noted that the goal is not to design a complete modal choice model but to build a "snapshot" of the demand in 1999 for all trip purposes. This limit allows a less accurate spatial definition than the one obtained with geocoding but this technique remains the most coherent and probably the most efficient for the fixed objectives.

Finally, the researches conducted during this project for this objective will also allow determining the following results:

- Measuring the barrier effects on trips behaviour
- Evaluating the compatibility between a national and a regional views of the mobility

Increasing the knowledge in these two areas is also one of the goals aimed by this project.

The project, coordinated by GRT (Transportation Research Group, University of Namur), began in November 2002 and ended on June 30, 2005. It involves four research groups:

I. GRT, University of Namur (FUNDP): Prof. Toint Ph., Cornélis E. and Legrain L.
II. Department of Transport and Regional Economics, University of Antwerp (UA): Prof. Verhetsel A., Van Hofstraeten D.
III. Department of Geography, University of Louvain-La-Neuve (UCL): Dr. Thomas I., Hammadou H.
IV. Vakgroep Geografie, University of Ghent (UG): Prof. Witlox F., Tindemans H.

1.2 Objectives

Our analyses have been based on the activity chains collected from the Belgian national mobility survey, MOBEL (1999), as well as from the same exercise at regional level for the Flemish region, OVG (1999-2001). The researches achieved during this project were planned for four successive steps. First of all, all activity places reported in the surveys have been geocoded within a Geographical Information System (GIS). This process consists of assigning geographical coordinates to activity locations. Thanks to this geocoding step, trips can be reported on maps and can be matched with thematic maps such as land use so that we would be able to link spatial characteristics to trips origin and destination allowing therefore spatial analysis. Several descriptive analyses were then conducted on this travel data aiming at emphasizing characteristics affecting travel behaviour such as mode of transport, departure time, travel time, activity duration, purpose of trips, land use or else socio-demographics. A specific spatial analysis has been conducted on distance decay. In this analysis, distance, the main spatial component, is studied. Then because of our focus on spatial components of chains, destination choice models have been developed both on MOBEL data set and OVG travel data. Another approach to destination choice model was aimed at studying border effects. Parallel to these analyses, an additional explorative analysis has tried to use rough sets techniques in the analysis of activity chains. The next step has consisted on creating a synthetic population. The aim is to build a population statistically close to the actual Belgian population and then to assign to each synthetic individual an activity pattern. Thanks to the previous results on destination choice modelling, we would then be able to assign activity locations to all synthetic individuals and, from that, to build an origin-destination matrix to estimate the travel demand. But because of time limitation we were not able to fulfil that last destination assignment.

Briefly, the different tasks are:

a. Geocoding of all information collected about activity locations
b. Spatial activity chains modelling:
    - distance decay
    - destination choice models
    - border effects
c. Rough sets technique studying
d. Creating a synthetic population
e. Origin-destination matrix building

2. Activity based approach

The hypothesis on which this study relies is that individuals travel because they need to achieve their daily activities. Indeed McNally (2000) has already described travel as a physical mechanism to reach an activity site to participate in some activity. Few years before Pas (1996) had stated that the main idea of activity-based approach is that one first needs to understand activity behaviour before one can analyse travel behaviour, and that travel is generally not undertaken as such, but follows from taking part in activities at locations that are not the person's current location.

The scheduling of activity chains is very complex. It concerns a lot of various decision-making: Which is the activity? Who wants to participate? When is the activity scheduled? Where can this activity take place? Which are the other activities to be scheduled? Which is the time budget? Some variables can explain how a decision is taken; one of our concerns in this project is to discover which factors influence the spatial spreading of individual daily activity diaries.
The activity chains are often described in terms of number of stops or out-of-home activities, time (travel and activity duration), transport mode, distance and purposes. A first approach is generally to study how different the chains are when looking at those components. Secondly, we can be interested in identifying which socio-demographic characteristics of the individual are important in activities chaining decisions: e.g. are the incomes and the number of children determining in making an activity like shopping? Finally, we can focus more precisely on the destination where the activity takes place. We can wonder which spatial characteristics of the place where the individual decides to undertake an activity are important. This question is surely the most important in this project because we mainly search to fill the lack of research in spatial analysis of travel especially in activity-based analysis. Indeed a lot of research was conducted on temporal aspects of travel behaviour: it has been modelled in departure time choice models (Bradley and al., 2001) as well as in mode choice models (Cirillo and al., 2002). Nevertheless travel demand analysis is intrinsically spatial but seldom recognized (Bhat and Zhao, 2002). Bhat and Zhao (2002) highlight the need to accommodate spatial issues in travel modelling and advocate a spatial analysis of activity stop generation.

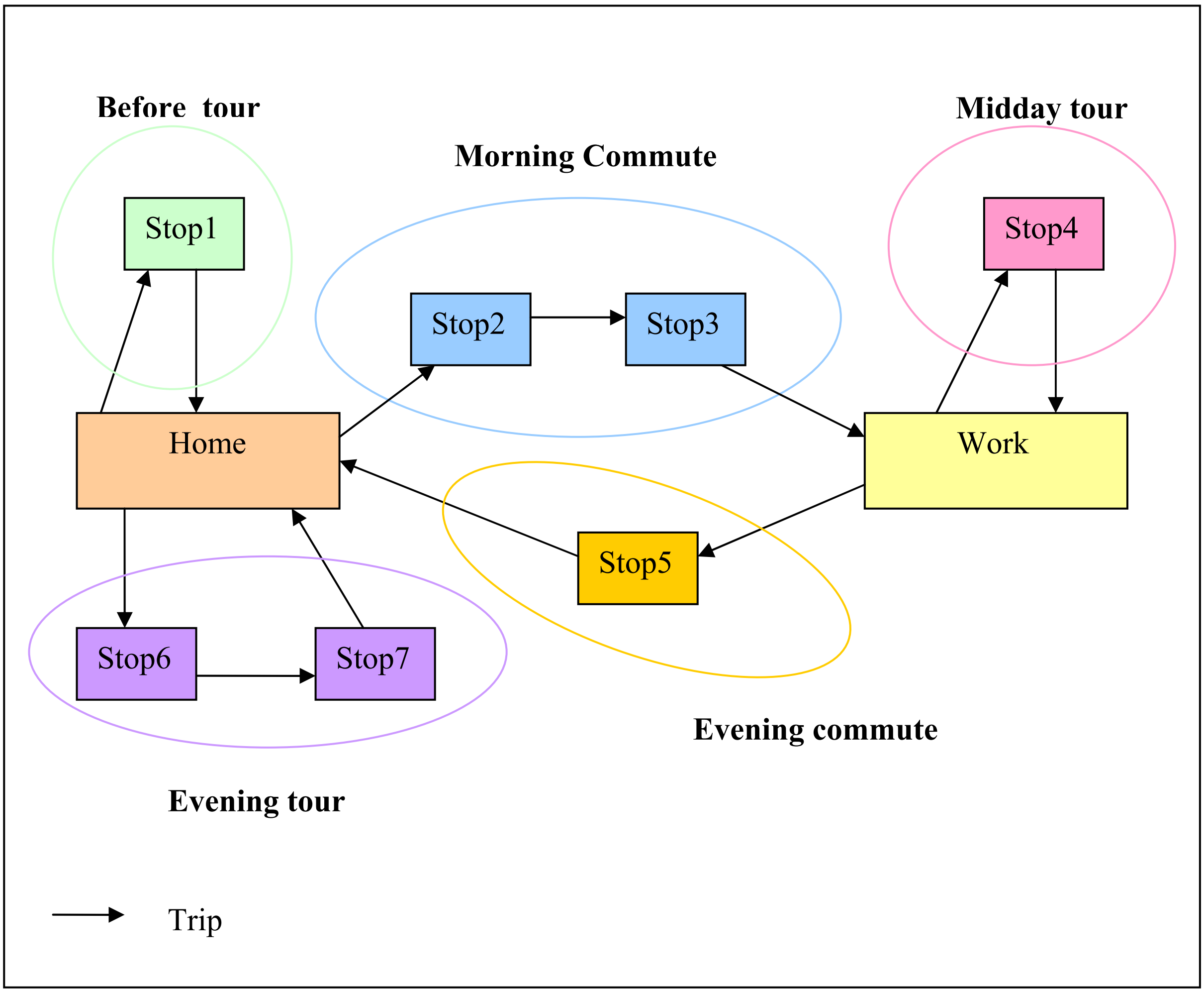
Our investigation in activity based approach first requires defining some useful concepts. The way activity chains have been described in this project results from an approach developed by Ben-Akiva and Bowman (1995). This framework was already adopted in Hubert and Toint (2001).

A *daily activity chain* (or *daily activity pattern*) is defined as the set of out-of-home activities (or *stops*) and *trips* made during a day. The out-of-home activity with the longest duration is called the *main activity*. If the main activity of an individual is work or school, the individual is therefore called *worker*, otherwise we talk about *non-worker*. The daily activity chain is therefore described around the main activity. This description also assumes that home and work are fixed locations.

In activity chains, we define a *tour* as a sequence of trips starting from home or work and ending respectively at home (*home-based tour*) or work (*work-based tour*).

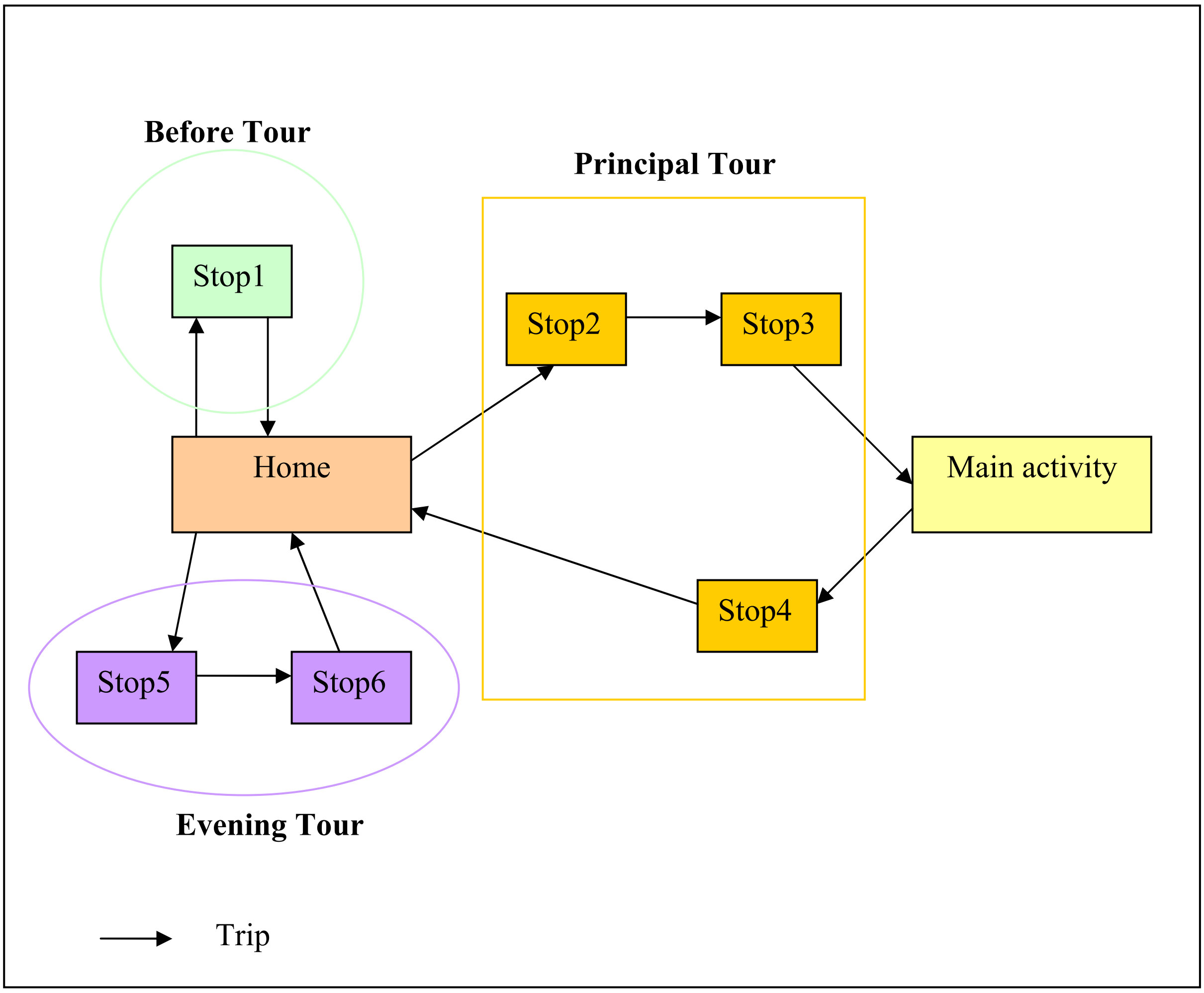
For workers (see Figure 2.1), the daily activity chain can be divided into five parts: the first one is a home-based tour before going to work (the *before tour*), the second one consists of leaving home to commute to the work place (the *morning commute*); the third represents a work-based tour at lunchtime (*midday tour*); the fourth part is the *evening commute* from work to the home place and the last one is constituted by a home-based tour after the working day (*evening tour*).

Figure 2.1 Worker daily activity chain



At variance, all non-workers' trips (Figure 2.2) are assumed to be home-based. Hence, the non-worker's daily activity chain can be divided into three parts: a tour during which the individual undertakes his/her daily activity with the longest duration (the *principal tour*); the *before tour* represents the tour before the individual begins his/her principal tour; and finally the *evening tour* consists of the tour made after the return from the principal tour.

Figure 2.2 Non-worker daily activity chain



We point here that depending on how the individual schedules his activities, some part of his daily activity pattern can appear missing (e.g. an individual can have no before tour nor evening tour).

3. Data description

This project uses data collected during several mobility surveys: one at the national level, the first Belgian mobility survey (1999), MOBEL, coordinated by the Transportation Research Group (University of Namur) and financed by the Belgian federal Science Policy (formerly OSTC) in the framework of SPSD 1 (Hubert and Toint, 2001) and some at regional level, Flemish travel behaviour research surveys (1999-2001), OVG, coordinated by the Provincial College of Limburg and the Flemish Government, and conducted on the city regions of Antwerp, Ghent, Mechelen, Aalst, Hasselt-Genk, Leuven and Halle-Vilvoorde.

These surveys collected one-day travel diaries at national level or two-day travel diaries at regional level. The individuals over the age of five reported characteristics for each of their trips: origin and destination locations (street name, zip code), departure and arrival times, duration, travel distance, transport mode, purpose etc. Moreover, these data sets contain a detailed socio-demographic description of each questioned household and individual such as household structure, incomes, home location, sex, age, professional activity, driving licence, school degree, etc. Tables 3.1 and 3.2 show a summary description of travel data sets.

Table 3.1 - MOBEL data overview

	Belgium
Survey year	1999
Number of households	3063
Number of persons	7037
Number of trips	**21 096**

Table 3.2 - OVG data overview

URBAN DISTRICT	ANTWERP	GHENT	HASSELT-GENK	LEUVEN	AALST	BRUSSELSE RAND	MECHELEN
Survey year	1999	2000	1999	2001	2001	2001	2001
Number of households	2527	2995	2782	3224	3187	2874	3240
Number of persons	5613	6785	6935	6758	7242	7196	7588
Number of trips	**29 778**	**35 878**	**37 976**	**37 167**	**33 857**	**34 910**	**37 956**

Before being suitable for geocoding through GIS software, the addresses collected from mobility surveys have required corrections. Address files presented a large amount of different and particular errors which often had to be manually corrected. The most common errors are incorrect spelling of street names, erroneous zip codes, or else *alias* been used to define destinations. These errors have been corrected by using search engines on Internet or by consulting street atlas maps. Moreover non responses and inconsistencies have been corrected when it was possible or otherwise removed as well as foreign addresses have been deleted from the travel data set.

4. Geocoding process

As our project focused on the spatial analysis of activity chains, all reported trips were first geocoded within GIS-software (Geographical Information System) framework. This work consists of assigning geographical coordinates to each address in our travel data set.

Geocoding process consists of a matching between a table of addresses (house number + street name + zip code) and a geo-referenced map of streets and administrative boundaries (more precisely municipalities). Obviously, the matching works better when the addresses to be geocoded are in a correct form (correct street name and correct zip code). When a matching is found, a list of candidates is proposed and each candidate has a score indicating how well the matching is performed (the higher the score, the better the matching).

Some problems related to the structure of the street name occur more specifically because of the geocoding service in itself. Indeed it is more complicated to geocode addresses from the French speaking part of Belgium than the Flemish ones because the geocoding service is more suitable for addresses which are built like English addresses. In particular, the first step in the geocoding process means "standardizing" the address. It consists of segmenting the address in different fields on which the matching is then based: one for the street name, another for the street type and a last one for the zip code. In English addresses as in Flemish addresses, the street type is at the end of the street name (Brussel"straat") whilst in French, it appears at the beginning of the street name ("Rue" de Bruxelles). The geocoding service does not distinguish the different cases and has therefore difficulties to segment French addresses. The structure of the reference table as well as the one for the set of addresses could be modified in order to improve standardisation but it had not been sufficient. Hence matching for "French" addresses often results in manual standardisation allowing the geocoding service to find some candidates. The relevant candidate is then manually selected.

Here is the description of each set geocoded by the different teams and geocoding results:

Table 4.1 Geocoding results

	Antwerp	**Mechelen**	**Ghent**	**Hasselt-Genk**	**Aalst**	**Halle-Vilvoorde**	**Leuven**	**MOBEL**
Number of trips	29 778	37 976	35 878	37 976	33 857	34 910	37 157	21 096
Trips successfully geocoded	25 338	33 088	29 851	31 147	28 820	27 687	32 243	16 679
Success (%)	**85.1**	**87.2**	**83.2**	**82**	**85**	**79**	**87**	**79.1**
GIS software	MapInfo	MapInfo	ArcView 3.2	ArcView 3.2	ArcMap	ArcMap	ArcMap	ArcGiS 8.1
Reference map	OC-GIS Vlaanderen	OC-GIS Vlaanderen	OC-GIS Vlaanderen	OC-GIS Vlaanderen	TeleAtlas' StreetNet	TeleAtlas' StreetNet	TeleAtlas' StreetNet	TeleAtlas' Multinet
Geocoded by	UA	UA	UG	UG	UCL	UCL	UCL	GRT

The difference in the success rate between MOBEL and OVG data comes from the fact that a MOBEL-trip has been considered as well geocoded if and only if the trip origin and destination are both successfully geocoded, while in the OVG data set, a trip is already considered well geocoded when the destination only has been geocoded.

5. Data activity based analysis: preliminary analysis

If the MOBEL data sets were largely explored previously, (see the conclusions in Hubert and Toint (2001)), on the other hand, results of OVG have never been explored in terms of activity chains framework. Louvain-La-Neuve, Antwerp and Ghent teams therefore first focused on an explorative analysis, on Antwerp and Ghent travel data sets, aiming at formulating hypotheses on the activity-related and spatial characteristics of trips. Mobility of socio-demographic groups, distribution of chains, trip purposes, distance and duration as well as transport mode have been examined. All results have been published in the following paper: '*SAMBA: Spatial Analysis and Modelling Based on Activities: A pilot study for Antwerp and Gent'*, Verhetsel *and al.* (2002).

Here is a description of the most interesting results. In OVG-Antwerp, we see that 65% of all chains are two-trips-chains (home-stop-home). Moreover, more than 20% of the trips are either in home-shopping-home or home-work-home chains. It emphasizes the importance of shopping and work purposes. Indeed, we found that shopping is the most important out-of-home activity, followed by work, entertainment-sports-culture or leisure, visit someone, (bring/collect someone) and school. The spatial link between shopping activities and work could also be of interest for study since shopping and work appear frequently combined. On the other hand, leisure activities seem to be more spread in space and time and more person-related, so it could be difficult to focus on it.

By looking at the relationships between trip purpose and mode of transport, we report that bring/get someone is generally made by car, as well as work related trips. For leisure activities, car as passenger is more frequent while public transport, bicycle and train are often used for school trips. For the shopping purpose, car and bicycle are frequently chosen.

The first exploration concerning the *distance* shows that individuals cover fewer kilometres (about 4 km) for school, shopping and service trips. It seems that those activities are available on a small area so that, most of time, individuals do not have to go further from their home location. On the other hand, business visits and work trips cover longer distance (about 13 km) as well as leisure trips (about 10 km).

The study also confirms the importance of socio-demographics. Indeed, gender, age, profession and education are decisive elements with respect to the activity-level. We found that people between 25 and 45 of age, with higher education, employees and executives are most active, resulting in a larger proportion of trips. Women are more related to shopping, bring or get someone, visit someone and services while men do more trips for work, leisure and business visits. People with higher monthly income are likely to do more trips for work, business visit and entertainment, sports, cultural activities while less earning people participate to walk, drive and cycle, visit someone, shopping and bring/get someone. Age is also a determinant in activity scheduling: young people are more often involved in school and leisure activities, the 18 to 60 years old visit more work and shopping locations while older people, most often non-workers, prefer shopping, then leisure and visiting someone.

Through this descriptive analysis we have highlighted some chain characteristics which seem interesting to introduce in the further modelling of OVG travel data.

Finally, in the framework of this project, UG undertook two extra activity-based analyses. The first one aimed at discussing the potential modal shift from car to bicycle in the urban region of Ghent (see *"Analysing bicycle travel behaviour in the Ghent region: An activity-based approach"*, Witlox F. and Tindemans H. (2003)). Based on OVG-travel data set the relationship between mode, activity, distance, location and socio-demographic background were explored. In addition, specific attention was paid to urban level of the origin and destination of each trip and to the influence of distance and speed. The second analysis was an exploration of the current mobility of children among different socio-demographic groups in Ghent (see *"Children's space-time activity behaviour: A case study for Gent"*, Tindemans H. and Witlox F. (2004)). It aimed at detecting major patterns in children's activity behaviour by incorporating a spatial analysis and identifying important socio-demographic and spatial factors determining the space-time behaviour of children.

6. Spatial activity chains modelling

One of our main concerns in the SAMBA project has been the introduction of the spatial dimension in studying travel demand. All analyses which have been undertaken aimed therefore to highlight spatial factors that can explain the spatial distribution of trips (in particular activity chains). At the end, destination choice models have been built to conclude our spatial analysis.

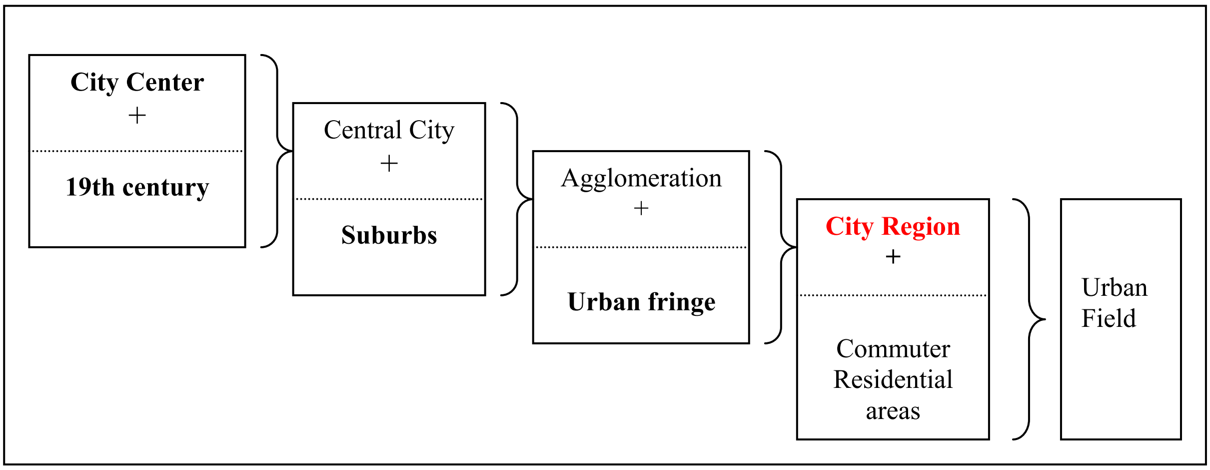
6.1. Preliminary analysis

The analysis undertaken in this phase was related to the spatial components of destination choice involving *distance* and *land use* patterns. A first analysis has been realised to study if the spatial distribution of the activities fitted the morphologically delimited metropolitan areas of city regions like Ghent and Antwerp or not. It has allowed a better understanding of how people travel in these city regions. After an analysis of the spatial distribution of chains in city regions, the most commonly used spatial component of trips, *distance,* has been studied through *distance decay.* The distance decay study aimed at measuring and testing the impact of distance in trips and activity chains.

6.1.1. Dynamics in city region

Antwerp and Ghent are two of the so-called city regions of Belgium. According to the functional typology of Van der Haegen (1996), those city regions are built up by a city center, a 19th-century area, suburbs and an urban fringe (see Figure 6.1.1).

Figure 6.1.1.1: Schematic spatial structure of the city regions as part of the urban field (Source: Van der Haegen et al., 1982)



The different zones are defined as a social, geographical and functional system in a wide geographical environment, containing various interrelated basic urban activities.

The **city centre** is the small area in the heart of the city with only a small residential population. It is the decision-making and activity centre with the largest concentration of retail and services having a regional function. A frame of densely built-up areas, the **19th**

century area, encloses the city centre. This area is characterised by its residential developments along with small to large-scale activities. The **suburbs** have a prior residential function but secondary shopping and service centres have been developed as well as green spaces have been retained. This zone takes up a large part of the city region and contains also important manufacturing areas and space for traffic. The **urban fringe** is considered as the current, mainly residential, urban growth zone. Finally, the commuter residential area gathers residential zones from which commute patterns to the core of the city region are generated.

Nevertheless, the usual delimitation methods were mainly based on the morphological and demographic variables and did not hold on activity patterns data. Within the framework of this study the definitions of the metropolitan areas have been tested by analysing the activity destinations for different kinds of activity purposes within the city region relative to the home locations of the individuals (see for example, Figures 6.1.1.2 and 6.1.1.3).

It has been found that most people living in the city region do not very often leave their city region and most of all they do even not leave their own metropolitan area. Indeed the hypothesis already found in '*SAMBA: Spatial Analysis and Modelling Based on Activities: A pilot study for Antwerp and Gent'* , Verhetsel *and al.* (2002), saying that most activity locations are close to the household residence has also been established. As many urban residents live in the suburbs and urban fringe, it can be stated that, in terms of number of services and facilities, the suburbs attract most urban activities so that the city centre is not the most attractive metropolitan area. The main activities realised close to the household residence are shopping and service related activities. Cultural and leisure activities are found to a large extent in the city centre as well as in the suburbs, while trips with a visiting purpose are mainly outward oriented to the suburbs, urban fringe or even beyond the city region borders. Work trips are spread over the city region and even beyond. The city centre seems to stay attractive for education trips even if most of them are performed within the home location metropolitan area.

Figure 6.1.1.2: Activity range for 19th century area inhabitants of Antwerp and Ghent

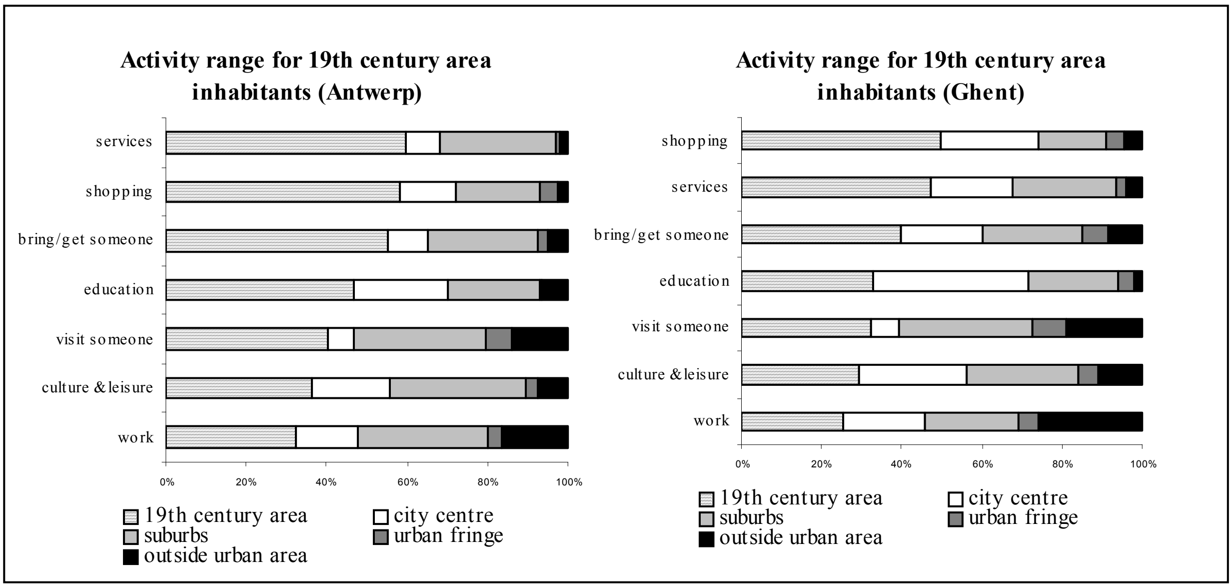
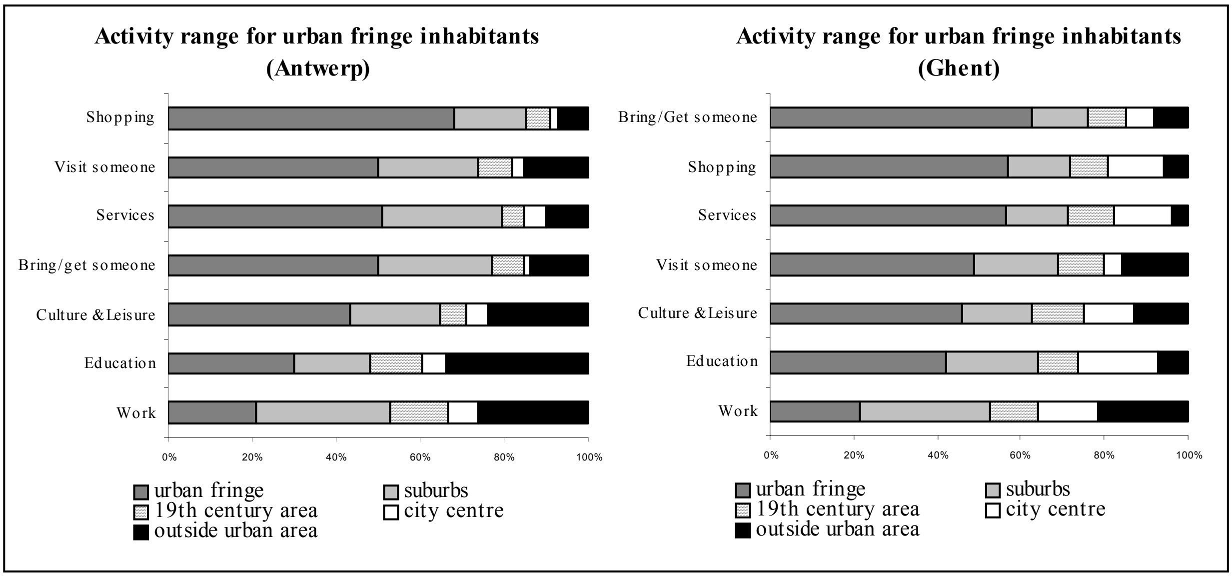


Figure 6.1.1.3: Activity range for urban fringe inhabitants of Antwerp and Ghent



This analysis proves that the city centre does not play (anymore) its expected role of nodal attraction. Moreover the different metropolitan zones seem containing various kinds of facilities that land use patterns analysis can detect. Therefore it appears interesting to introduce land use variables in further analysis.

Complete results are available in « *Dynamics in city regions. The intra-urban travel patterns in Antwerp and Ghent*», Verhetsel *and al*. (2002).

6.1.2. Distance decay

The nature of the relationship between distance and number of trips and chains that individuals performed has been explored. The analysis has been realised at the national level (MOBEL data survey) and also at regional one (OVG-Antwerp data survey). Until now, distance decay has only been studied on trips but, in this study, the activity based approach has been investigated and compared to the trip version. The difference between both approaches is that the activity-based approach considers for the cumulative structure in chain distance which is not possible in a trip based approach.

Among others methodologies, the exponential model appears to be the most convenient approach for distance decay. It shows how sensitive respondents were in travelling longer distances for different trips and chains in regards of different modes, travel purposes, composition of household and urban levels of the destination. The exponential model is given as follows:

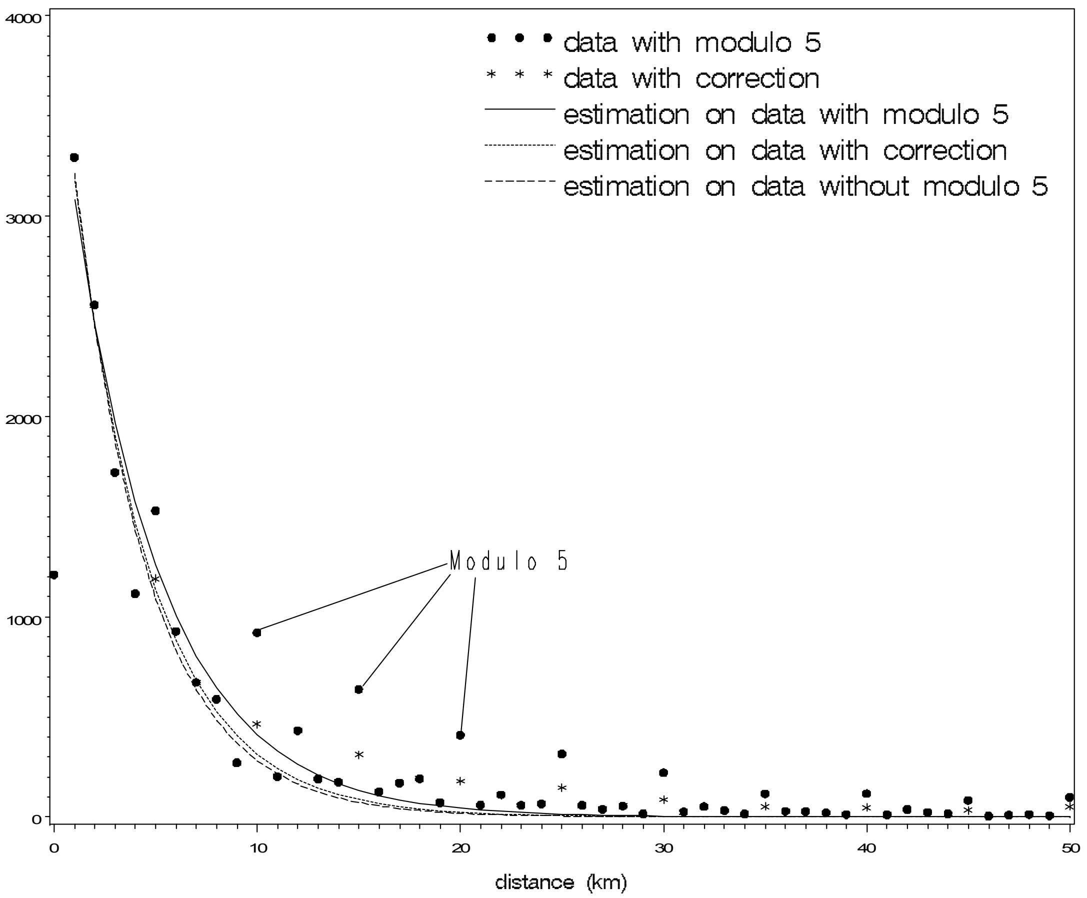
$$I_{ij} = ae^{-bD_{ij}}$$

where I_{ij} is the interaction between origin *i* and destination *j*, D_{ij} is the distance, *a* and *b* are the constants to be estimated and *e* the Neper constant. The most interesting parameter is the constant *b* because it represents the sensitivity to distance. Most of time this parameter is estimated to be negative, meaning that increasing distance decreases interaction. In addition a highly negative parameter indicates that distance is a strong deterrent to interaction whilst larger estimate indicates a weak deterrent to interaction. A maximum likelihood test is used for parameters estimation.

Because each tour is supposed to have its own sensitivity to distance, a tour-based approach is preferred to an activity-based one. Each home-based tour was considered and simplified in order to avoid too complex schemes. Less complex tours were obtained by limiting the number of trips to maximum three trips and defining one single purpose when the tour includes several different purposes. A tour containing only work trips (whatever the number of trips in the tour) receives the single purpose "work". When work and shopping trips are combined in a tour, this tour receives the single purpose "work-shopping" (whatever the number of work and shopping trips, and whatever the temporal scheduling).

It has been found in MOBEL as well as in OVG travel data sets that respondents often reported distances which are multiple of 5. This is a phenomenon that occurs very often when treating data resulting from revealed preferences and it can not be ignored (Rietveld et al., 1999). Rietveld gave different ways to correct those distances using different definition of distance such as network distance or distance as the crow flies. But, since no information about the travel path is available in our travel data set, those correction methods can not be applied. Therefore UCL has measured the importance of the bias of estimation on the distance decay models by first comparing the model without modulo 5 (where all observations with rounded distances are deleted) to the model containing modulo 5. Finally a test was included to check whether or not there is a significant difference between the parameters estimated by these models. It has been found that the model with modulo 5 tended to underestimate the friction of distance while the model without modulo 5 overestimated it. A temporary correction has been introduced. A model with a reduced bias of estimation of distance decay is obtained with this correction. Results, with or without corrections, are presented in Figure 6.1.2.1. Another source of bias in estimation is missing information in activity chains. For example in the original travel data sets more than 10% of chains did not end at home. By analysing the information reported in chains we can fill in blanks.

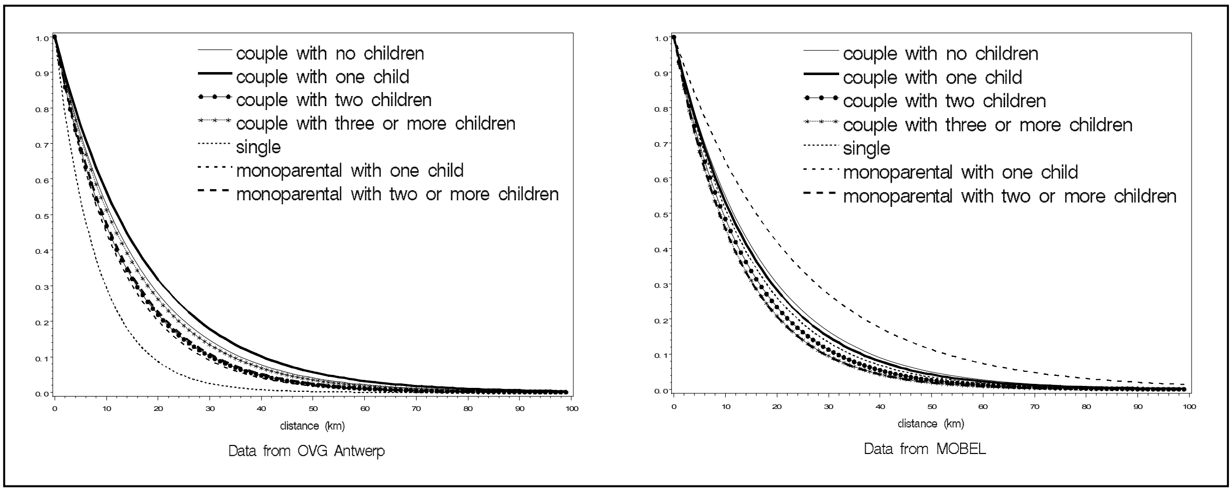
Figure 6.1.2.1 Distance decay graph of data with or without corrections or modulo 5



Surprisingly, since the beginning it has been noticed that the model gives similar results for trips and chains whatever the criterion. It is an important finding because it casts doubt over the need for trip chain analysis (with the limitations implied by its definition in this context), as trips are easier to use and they give comparable estimates of model parameters.

The distance decay is first analysed with regard to the *household composition* (see Figure 6.1.2.2). We found that single and mono-parental with children households were more sensitive to distance in Antwerp city region. At the national level, we found that the higher the number of children the higher the sensitivity to distance. Nevertheless household composition was not very discriminatory at the national level.

Figure 6.1.2.2: Distance decay for the household typology – Chain approach



The distance decay was then studied from the point of view the *urban level of the home location* (see Figure 6.1.2.3) and then, with regard to the *destination urban level* (see Figure 6.1.2.4). For the OVG-Antwerp data the urban levels were those defined previously for the delimitation of a city region. For MOBEL data different urban levels were introduced to account for the national level (urban and non-urban zones). We finally found that individuals living in the central city of Antwerp were more sensitive to distance than people living in outer central urban areas. It could be explained by the fact that the centre provides more opportunities to participate in various activities and so implies shorter trips. Longer trips are therefore seen as more constraining. For the national level, we could conclude that the individuals were mostly less sensitive to distance than individuals living in the city region Antwerp. Nevertheless, we also found that people living in agglomerations were more sensitive to distance than people living in the urban fringe possibly because inhabitants of the urban fringe are richer and own more cars allowing them to cover more easily longer distance. Looking at the urban level of destination in OVG data, tours performed in the suburbs and in the urban fringe had a similar friction to distance. However we found a higher sensitivity to distance for the 19th century city than for the city centre although the contrary was expected. Once more land use analysis could maybe explain it: has the 19th century area more mixed land use than the city centre so that higher distances have to be travelled between different kinds of activities? In MOBEL, the group of individuals moving in the agglomeration of Brussels and outside the urban area was more sensitive to distance than the group of people moving in all the other urban levels of destination.

Figure 6.1.2.3: Distance decay for the urban level of household location – Chain approach

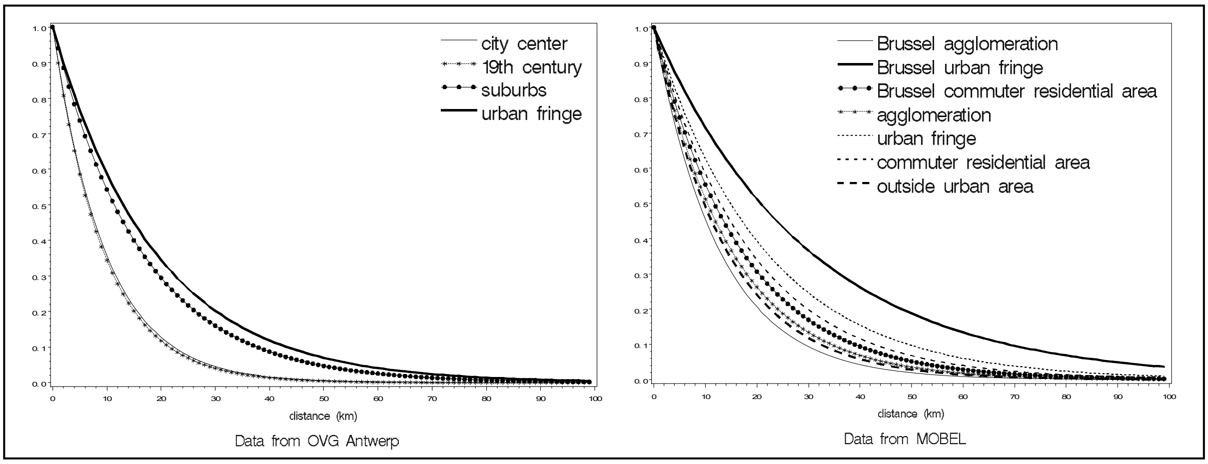
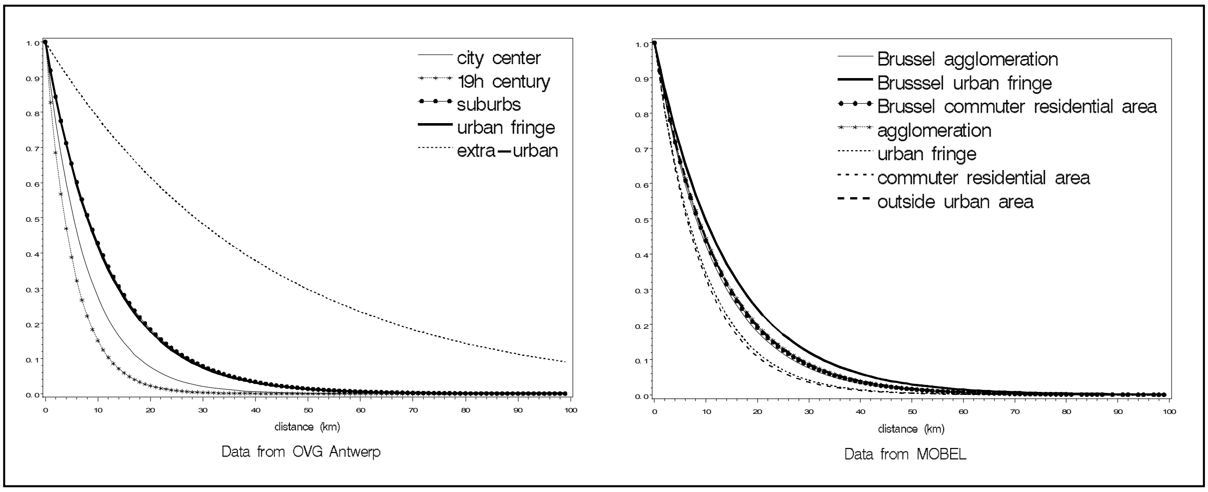
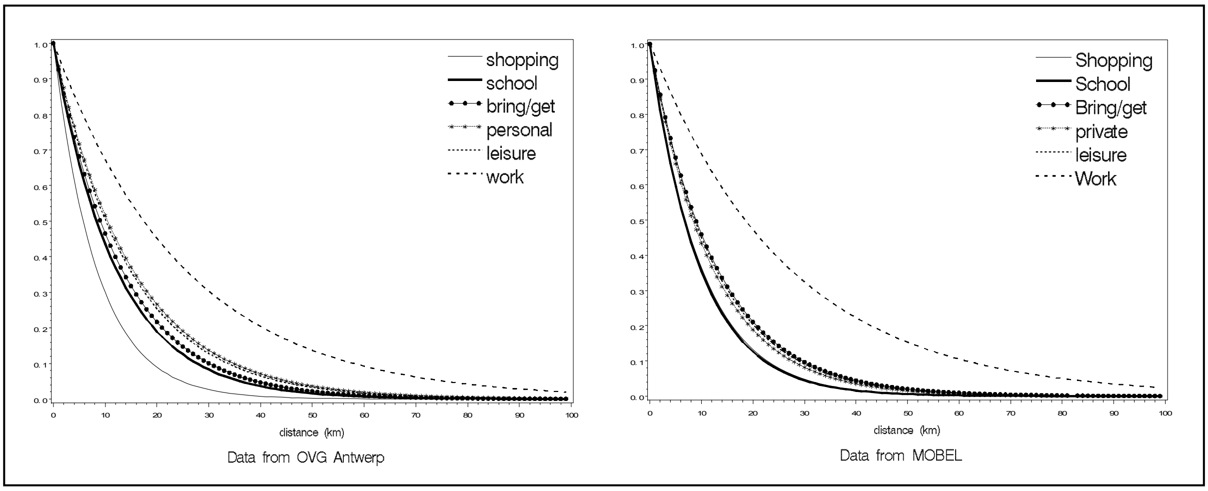


Figure 6.1.2.4: Distance decay for the destination – Chain approach



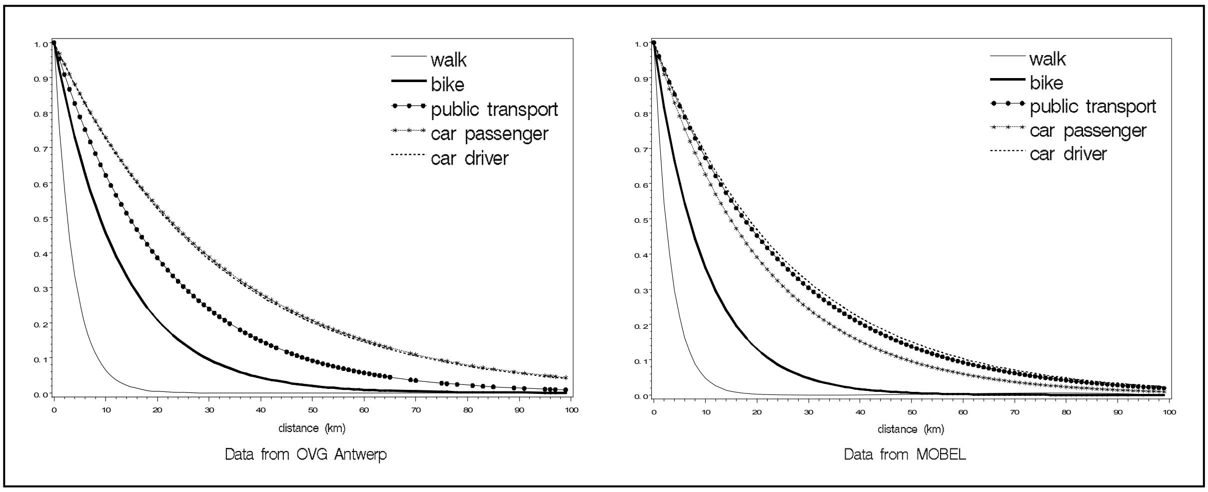
The next criterion was the *trip purpose* (Figure 6.1.2.5). The trend was more or less the same for both surveys. The work activity was the least sensitive to distance. Indeed, people seemed ready to travel more for working while others purposes (bring/get someone, personal, leisure) and particularly shopping and school involved shorter trips.

Figure 6.1.2.5: Distance decay for the chain purposes – Chain approach



Finally, distance decay was studied regarding *transport mode* (Figure 6.1.2.6). We were not surprised to find that walking and biking had more friction to distance. Travelling by faster modes (car and public transport) allows increasing the travel distance and thus leads to less sensitivity. At the national level, we found that the respondents who choose to travel by public transport were less sensitive to distance than in the city region Antwerp. In addition, in the Antwerp city region, we noticed a similarity between car passengers and car drivers while at the national level, we found that car passengers were more sensitive to distance than car drivers.

Figure 6.1.2.6: Distance decay for the transport mode – Chain approach



We may conclude that the distance decay effect for different household compositions and urban levels of the household location is less strong than for the mode of transport, the purpose and the urban level of the destination. In addition, the different geographical scales, namely the national and regional level, often lead to different results for the sensitivity parameters.

Finally a clustering analysis aiming at comparing simultaneously the distance decay parameter *b* for most variables (see Table 6.1.2.1) was achieved. The different clusters that result from that analysis were ordered by increasing sensitivity to distance. Cluster 1 gathered the variables with the weakest friction to distance while cluster 5 had the strongest sensitivity to distance.

Complete results have been published in *"Distance Decay in activity chains analysis. A Belgian case study"*, Hammadou H. *and al.* (2003).

Table 6.1.2.1 Cluster analysis on the distance decay sensitivity parameter (chain approach)

Cluster	OVG Antwerp	Parameter *b*
1	Car driver Car passenger Destination outside Antwerp urban area	*Max: -0,024* *Mean: -0,029* *Min: -0,033*
2	Leisure Personal Residence Antwerp suburbs Couple with no children Couple with 1 child Couple with 3 or more children Residence Antwerp urban fringe Work Public Transport	*Max: -0,040* *Mean: -0,058* *Min: -0,068*
3	Destination Antwerp suburbs Bicycle School Destination Antwerp urban fringe Couple with 2 children Bring or collect someone Mono-parental with 2 children or more Mono-parental with 1 child	*Max: -0,074* *Mean: -0,081* *Min: -0,087*
4	Residence Antwerp city centre Residence Antwerp 19th century area Shopping Single person Destination Antwerp city centre	*Max: -0,104* *Mean: -0,117* *Min: -0,129*
5	Destination Antwerp 19th century area Foot	*-0.189* *-0.220*

Cluster	MOBEL data	Parameter *b*
1	Mono-parental with 1 child Residence outside urban area Work Public Transport Residence urban fringe Car passenger, Car driver	*Max: -0,033* *Mean: -0,041* *Min: -0,046*
2	Residence Brussels urban fringe Single person Couple with no children Residence agglomeration Couple with 1 child, with 2 children Residence Brussels commuter residential area Residence commuter residential area	*Max: -0,054* *Mean: -0,064* *Min: -0,073*
3	Couple with 3 children or more Mono-parental with 2 children or more Leisure Destination agglomeration Residence Brussels agglomeration Destination outside urban area Destination Brussels commuter residential area Personal Bring or collect someone Destination Brussels agglomeration	*Max: -0,078* *Mean: -0,081* *Min: -0,086*
4	Shopping Bicycle School Destination urban fringe Destination commuter residential area	*Max: -0,101* *Mean: -0,105* *Min: -0,111*
5	Foot	*-0,303*

6.2. Destination choice models

All teams were involved in the destination choice modelling although they were working with different methodologies and different travel data sets. GRT has built a first model concerning trips made in the city region Antwerp and another one on the Flemish region from MOBEL data set while UG, UA and UCL have built a model for the four following city regions from OVG travel data set: Antwerp, Ghent, Mechelen and Hal-Vilvoorde. All models were based on the discrete choice theory but different assumptions had led to different formulations. In this section we first present the discrete choice theory on which the models are based and then the models developed by each team.

6.2.1. Discrete choice theory

The discrete choice theory is based on the fact that an individual facing a set of mutually exclusive alternatives will choose the one that maximises its satisfaction. This satisfaction can be built for each alternative and is described by a utility function. However because the choice made by a human being involves some unobserved factors, its choice will be properly expressed as a probability of making that choice.

The utility function is composed of a deterministic part which is a function of the attributes of the alternative and of the characteristics of the individual, and a stochastic part (or error term) representing the factors that the modeller does not or can not observe. Assuming an individual n faces a set of J alternatives, the utility which the individual n assigns to an alternative j is therefore given by

$$U_{nj} = V_{nj} + \varepsilon_{nj}$$

where V_{nj} represents observed factors and ε_{nj} is the error term. The individual is then supposed to choose the alternative that maximises his/her utility. The probability that individual n chooses alternative i is

$$\begin{aligned} P_{ni} &= P(U_{ni} > U_{nj} \quad \forall j \neq i) \\ &= P(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \quad \forall j \neq i) \\ &= P(\varepsilon_{ni} - \varepsilon_{nj} > V_{nj} - V_{ni} \quad \forall j \neq i) \end{aligned}$$

Depending on how the modeller represents the choice, he/she specifies a density function to the error term ε_{nj} which can therefore leads to different discrete choice models.

If each error term is supposed independent and identically distributed (IID), meaning that the unobserved factors are uncorrelated over alternatives and that they have the same variance for all alternatives, with GUMBEL distribution given by

$$f(\varepsilon_{nj}) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}}$$

then, the probability that individual *n* chooses alternative *i* is

$$P_{ni} = \frac{\exp(V_{ni})}{\sum_{j}^{J} \exp(V_{nj})}.$$

This is the **logit** choice probability leading to multinomial **logit models** (MNL).

The deterministic part of the utility can be written as a linear function of parameters such as $V_{nj} = \beta X_{nj}$, where β is a vector of parameters and X_{nj} is a vector of attributes related to the alternative and to the individual, therefore the choice probability becomes

$$P_{ni} = \frac{\exp(\beta X_{ni})}{\sum_{j}^{J} \exp(\beta X_{nj})}.$$

Assume that a sample of N individuals is available for the estimation and that each individual has expressed his choice concerning any kind of action (for example, choosing a travel mode among the set of alternatives including bus, car and walking). The probability of individual *n* choosing the alternative he has actually chosen is given by

$$\prod_{i} (P_{ni})^{y_{ni}}$$

where $y_{ni}=1$ if individual *n* chose alternative *i* and zero otherwise. If we assume that each individual's choice is independent from the others, the probability of each individual in the sample choosing the alternative that he was actually observed to choose is

$$L(\beta) = \prod_{n=1}^{N} \prod_{i} (P_{ni})^{y_{ni}} .$$

The vector of parameters, β, being considered as fixed across individuals and alternatives, is then estimated by maximising the log-likelihood function given by

$$LL(\beta) = \sum_{n=1}^{N} \sum_{i} y_{ni} \ln P_{ni} .$$

Many statistical packages are available for the estimation of such model (Gauss, Alogit, etc.)

This logit specification is however limited in three important ways. It can represent systematic taste variation but only when it relates to observed factors. To illustrate that point assume that an individual has to choose a travel mode and that the utility function is given by

$$U_{nj} = \alpha_n C_{nj} + \beta T_{nj} + \varepsilon_{nj}$$

where C_{nj} is the cost and T_{nj} is the travel time. Each individual is assumed to perceive travel time in the same way. However the cost variable can be differently perceived depending on the individual income, I_n. Therefore a new parameter, γ, is introduced such that $\alpha_n = \frac{\gamma}{I_n}$. The utility then becomes

$$U_{nj} = \gamma \frac{C_{nj}}{I_n} + \beta T_{nj} + \varepsilon_{nj}$$

and γ is therefore estimated accounting of cost compared to income. However the logit model can not represent random taste variation. Indeed, in actual decision making process, it is not always possible to measure all factors influencing the choice, especially those related to the human nature and implying that different people make different choices.

The next limitation is that logit specification exhibits the Independence from Irrelevant Alternatives property (IIA) such that the ratio of probabilities is independent of the attributes or existence of all other alternatives. It means that if an alternative is removed or added, the probability of all other alternatives increases or decreases in the same proportion.

Finally, the logit model cannot handle situations where unobserved factors are correlated over time.

More flexible models had therefore to be developed to introduce taste variation or correlations due to unobserved factors. We present here two models which overcome some logit limitations.

A first generalisation of the logit model which is able to capture correlation among alternatives is the **nested logit model** (NMNL). Its particularity is that the modeller assumes that the set of alternatives can be partitioned into subsets, called nests, so that the decision making process is hierarchical. The alternatives are grouped in a nest so that IIA assumption is satisfied. The utility that individual *n* chooses alternative *j* in nest *l* is

$$U_{nj} = W_{nl} + V_{nj} + \varepsilon_{nj}$$

where W_{nl} depends only on variables that describe nest *l*, these variables differ over nests but not over alternatives within each nest; V_{nj} depends on variables that describe alternative *j,* these variables vary over alternatives within nest *l.* The vector of error terms $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nJ})$ has cumulative distribution

$$\exp\left(-\sum_{l}^{L} \left(\sum_{j \in l} e^{-\varepsilon_{nj} / \lambda_l} \right)^{\lambda_l} \right).$$

While each ε_{nj} was independent in the logit model, in the nested logit the ε_{nj} 's are correlated within nests. For any alternatives in different nest, the error part of utility is still uncorrelated. The parameter λ_l measures the degree of independence in unobserved utility among the alternatives in nest *l.*

The choice probability has the particularity of resulting from the product of two standard logit probabilities: the probability P_{nl} of choosing an alternative in nest *l* and the probability $P_{nj|l}$ that alternative *j* is chosen given that an alternative in nest *l* is chosen. These probabilities are expressed as

$$P_{nl} = \frac{\exp(W_{nl} + \lambda_l \ln(\sum_{i \in l} \exp(V_{ni} / \lambda_l))}{\sum_{k=1}^{L} \exp(W_{nk} + \lambda_k \ln(\sum_{i \in k} \exp(V_{ni} / \lambda_k)}, \text{ and}$$

$$P_{nj|l} = \frac{\exp(V_{nj} / \lambda_l)}{\sum_{i \in l} \exp(V_{ni} / \lambda_l)}.$$

In the previous expression, the quantity $\lambda_k \ln(\sum_{i \in k} \exp(V_{ni} / \lambda_k)$ is called the inclusive value (or inclusive factor) and represents the expected utility that decision-maker, *i*, receives from the choice among the alternatives in nest *k*.

Then the probability that individual *n* chooses alternative *j* is given by

$$P_{nj} = \frac{\exp(V_{nj} / \lambda_l)\left(\sum_{i \in l} \exp(V_{ni} / \lambda_l)\right)^{\lambda_l - 1}}{\sum_{l=1}^{L}\left(\sum_{i \in l} (\exp(V_{ni} / \lambda_l)\right)^{\lambda_l}}$$

The parameters of a nested logit model can be estimated by standard maximum likelihood techniques.

The next model obviates the three limitations of the standard logit while the nested model only overcomes the second limitation. **Mixed logit models** are characterised by their probability choice which is an integral of standard logit probabilities over a density of parameters (as we will see later). These models can be derived under a variety of different behavioural specifications, and each derivation provides a particular interpretation. In this section we deduce mixed logit models from the hypothesis of taste variation and correlation. The first one leads to a random coefficients formulation: the β parameters are no more fixed across individuals but are supposed to vary in the population with a given density.
The β parameters are now written β_n to represent the taste of individual *n* and they are supposed to follow the density $f(\beta_n|\theta^*)$ where θ^* are the true parameters of the distribution. According to the IID (independent and identically distributed) extreme value distribution of the error term, the probability that individual *n* chooses alternative *j* is now given by

$$P_{nj}(\beta_n) = \int L_{nj}(\beta_n) f(\beta_n \mid \theta^*) \, d\beta_n$$

where $L_{nj} = \frac{\exp(\beta_n X_{nj})}{\sum_{k}^{J} \exp(\beta_n X_{nk})}$ is the logit probability that individual *n* chooses alternative *j* for a given β_n.

We point that a **mixed nested logit model** (MXNMNL) can also be deduced by introducing random coefficients in the nested logit model as well as random coefficients in a logit model leads to a mixed logit model.

The second mixed logit formulation introduces error components that create correlations among the utilities for different alternatives. The utility is then specified as

$$U_{nj} = \alpha X_{nj} + \gamma_n Z_{nj} + \varepsilon_{nj}$$

where X_{nj} and Z_{nj} are vectors of observed variables relating to alternative j, α is a vector of fixed parameters while γ_n is a vector of random parameters with zero mean and finally ε_{nj} are IID Gumbel distributed error terms. The stochastic part of the utility is here given by

$$\gamma_n Z_{nj} + \varepsilon_{nj}$$

which can be correlated among alternatives depending on the specification of Z_{nj}.

With mixed logit the probability choice is not in a closed form so that exact maximum likelihood estimation is impossible. The probability choice is therefore approximated through simulation and the simulated log-likelihood function is maximized.

In addition of choosing the more suitable model formulation, other important tasks of modelling are the definition of a set of alternatives among which each individual is supposed to make a choice, and the selection of a set of observed variables which the individual based his choice on. The choice set and the way that variables will enter the model also determine which model formulation is the more convenient.

Let us now describe the different models developed by each team.

6.2.2. GRT destination choice model over the city region of Antwerp

The model has been developed on MOBEL travel data set relative to trips made in the city region Antwerp. Data represent a set of 351 geocoded daily activity chains. First, each chain reported in the data set has been split into home-based or work-based tour(s) and commute(s) legs between home and work respectively to the framework presented in Section 2. Next, for workers, in each tour, at most two out-of-home stops and, in each commute, at most one out-of-work stop have been kept for modelling while, for non-workers, three out-of-home stops are considered in the main tour and two in the before and evening tour. Each stop induces a choice of destination that has been modelled. Home and work locations are supposed fixed so that the choice of their destination is not modelled. The first step is now to build for each stop a set of alternatives.

- **Description of the set of alternatives: action space**

Destination choice models have the particularity that their choice sets which include all places where activities could take place can become huge. An alternative is in general defined as an aggregation of elemental locations (Daly, 1982). We can for example assume that a statistical sector is a good aggregation of a set of elemental activity locations. Moreover it seems suitable to estimate that an individual chooses among all these statistical sectors where his/her activities can take place. But even at this level of aggregation, the set of statistical sectors is still huge. That's why in order to avoid problems in the estimation process, modellers often decide to work on a restricted subset of alternatives or to aggregate alternatives which are similar in view of decreasing the number of alternatives. The random sampling is the most commonly used technique (Bhat et al., 1998; Srour et al, 2000). It consists of randomly selecting a fixed number of alternatives from the entire choice set to which the chosen alternative is added. Thanks to its formulation, in particular the IIA property, the multinomial logit is the only model which allows consistent estimation of the parameters on a subset of alternatives (see Train, 2002 for more details). Sampling of alternatives could nevertheless be used with other models like mixed logit to avoid complexity of estimation. The choice set must always include the alternative that the individual has really chosen, since this choice has to be confronted to the other alternatives. As we will see, in the different models developed during this project, all teams have worked on different ways of building the set of alternatives.

In the GRT model, an alternative is represented by a statistical sector. Then in order to restrict the set of sectors among which the individual has to make a choice, GRT has used the notion of action space introduced by Hägerstrand (1970). The action space defines a restricted area where the individual is supposed to achieve all out-of-home or out-of-work activities contained in his/her daily activity chain. It is built based on the assumptions that the individual organised his/her activity chains around fixed bases as home and work subjected to some temporal and spatial characteristics: time-budget (sum of travel time and activity duration), travel speed and distance between home and work. Every time an individual leaves home or work to undertake one or several activities, the temporal and spatial characteristics induced by those activities, are used to build an action space.

Dijst and Vidakovic (1997) gave the following definition which allows (using properties of an ellipse) implicit building of an action space for each home-based tour, work-based tour and commute legs between home and work:

$$\frac{(x-\frac{l}{2})^2}{(\frac{\tau TV}{2}+\frac{L}{2}(1-\tau))^2}+\frac{y^2}{(\frac{\tau TV}{2}+\frac{L}{2}(1-\tau))^2-(\frac{L}{2})^2}\leq 1 \tag{1}$$

where T is the time-budget, V is the travel speed, L is the distance between bases, τ is the travel time ratio (ratio between the travel time and time budget) and x, y are the coordinates of points belonging to the action-space. Therefore an action space is a set of statistical sectors whose centroids satisfy this property.

The action space is an ellipse when activities are made during a commute between home and work, it degenerates into a circle (because of a vanishing L) when activities are performed in a home-based or work-based tour.

In practice, the action space is built in the following way: we compute the longer principal axis from (1), we use the bifocal definition of the ellipse in order to reformulate the action space as the set of locations *S* satisfying the following inequality:

$$dist(B_1,S)+dist(B_2,S)\leq LA \tag{2}$$

where B_1 is the first base, B_2 is the second base, and *LA* is the measure of the longer principal axis for the action-space ellipse.

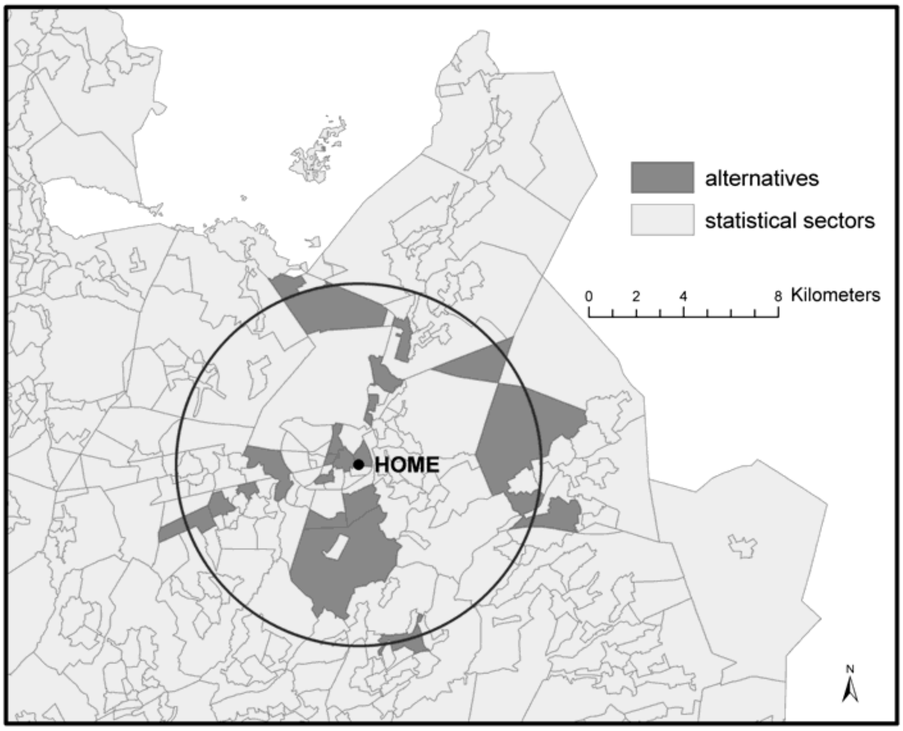
In the case of non-base stops included in a home-based or work-based tour, (2) reduces to:

$$dist(B,S)\leq R$$

where *B* is the base and *R* is the radius of the circle defining the action space.

Therefore accounting of information reported in the travel data set, an action space is first built for each tour and commute. We can assume that the action-space related to a tour or a commute is built to represent the alternatives that the individual is really confronted to. Then for each modelled stop, nine alternatives have been randomly selected from the respective action space to which we add the destination really chosen by the individual. By this way, we get a choice set of ten alternatives for each observation. Figure 6.2.2.1 is an example of all alternatives faced by a non-worker who wants to achieve an activity during his *before tour*. The black sectors are those selected to belong to the choice set.

Figure 6.2.2.1 Action space



Now the alternatives have to be characterised by a set of selected variables.

- **Variables**

In destination choice models, the most commonly used variables are *land use* characteristics of the destination zone and variables measuring the *attractiveness* and *accessibility* of the destination zone (Kockelman (1996); Simma, Schlich and Axhausen (2002) and Bhat, Govindarajan and Pulugurta, (1998)). Badoe and Miller (2000) and Stead (2001) also indicated that land use, density and accessibility are three important groups of spatial variables that explain travel choice behaviour. The attractiveness is measured by means of variables that estimate the size of the destination zone to give an indication of the importance of the destination in terms of facilities (leisure and cultural activity locations, shopping centres, school, etc.) and in terms of employment and population. The expected effect on travel behaviour is obvious: the larger the place, the more destinations within the activity range, and hence, the more trips and multipurpose trips (Van Wee, 2002). These variables are also known as *size variables*. On the other hand, accessibility is defined here in terms of travel time and travel distance.

Land use variables

The land use data have been built by the teams of the Universities of Antwerp and Ghent from a digital land use map for the Flemish region. This map has first been provided in 1996 by the OC-GIS Flanders and then updated in 2001. The data set is based on satellite images, soil information and road network. By using an automatic classification procedure, satellite information was converted into 19 categories of land use (see Table 6.2.2.1, from OC-GIS

Vlaanderen, 2001). Other land use characteristics as administrative borders, road network and specific land use information such as 'built-up area" was provided in TeleAtlas (MultiNet data, 2001).

The surface occupied by each type of land use (in square meters and percentage of the total surface) was computed for each of the statistical sectors.

Table 6.2.2.1 Overview of the 19 categories of land use

1	**Agriculture/meadowland**	Agriculture and open space
2		Meadowland
3		Alluvial meadowland
4		Orchards
5	**Forests and parks**	Coniferous forests
6		Broad-leaved forests
7		Mixed forests
8		Municipal parks
9	**Development and industry**	Densely-built housing
10		Housing and other development
11		Industrial and commercial area
12	**Infrastructure**	Highways
13		District roads
14		Airport infrastructure
15		Port infrastructure
16		Other infrastructure (railway…)
17	**Heath land and dunes**	Heath land
18		Dunes
19	**Water**	

Source: Land cover and land use data set – OC-GIS Vlaanderen (2002)

Attractiveness variables

The following commercial variables have been introduced in the model to estimate the size or attractiveness of destination zones.

- shopping;
- financial (banks);
- hotel / restaurant / café;
- cinemas;
- sport activities;
- cultural, recreational and leisure activities (museum, library, school of music, zoo, nature reserve, theatres, casino and so on);
- car retail;
- personal service (beauty center and so on).

As proxy, the number of employments in each of these sectors was used for measuring the attractiveness.

These variables were provided by the Statistical Services (SES), the Ministry of Walloon Region, and calculated using data from EURO DB and the National Office for Social Security (N.O.S.S.).

Accessibility variables

The accessibility of destination zones has been computed by means of cost and in-vehicle travel time (IVTT) variables. For each origin-destination pair, the distance has been used to calculate the COST and IVTT variables for the chosen travel mode. Four categories of travel mode were considered: *walk and bike*, *train*, *bus* and finally *car*. For the first one, we assign a cost equal to zero. The cost of train is established by means of the Belgian railways tariff distances. A fixed cost chosen to be 1.1 euros is assigned to the bus mode. When the car is used, we assumed that it costs 0.17 euros per kilometer. IVTT is obtained by the ratio of the crow-flight travel distance (between base and non-base stop) and the mean travel speed (determined, from the MOBEL data, for each tour and commute). The accessibility variable is then expressed as

$$Access = IVTT + \alpha\, COST\ ,$$

where α is a parameter allowing the conversion of COST into time value. The value of time was chosen to be 5 euros/hour as in a European study comparing the different values of time across the European Union (Trace, 1999); it means that each euro of travel cost is equivalent to 12 minutes of travel time which is the value employed for α.

- **Model formulation**

Two models have been estimated. The first one is a standard logit model while the second is a mixed logit where parameters related to accessibility are assumed random.

Let I be the population size and $A(i)$ the set of available destinations for individual i, $i = 1,..., I$. For each individual i, each alternative A_z, $z = 1, ..., |A(i)|$, has an associated utility, depending on the individual characteristics and on the relative attractiveness of the alternative. In this destination choice model, the utility can be specified as:

$$U_{iz} = \lambda^T \ln y_z + \ln M_z + \sum_j \gamma\, x_{izj} + \varepsilon_{iz}$$

where:

y_z is a vector of observed **land use zone-specific variables** and λ is a column vector of coefficients fixed across all zones; in order to specify a destination choice that is not sensitive to zonal aggregation, we represent utility with parameters inside a log operation (Train, 2002);

$M_z = \sum_{k=1}^{s} e^{\beta_k} M_{zk}$ is the measure of the size of the alternative z, where M_{zk} is the kth **size variable** for zone z, β_k is the corresponding coefficient;

x_{izj} are the components of a column vector of exogenous **accessibility variables** for individual i in zone z and γ is a vector of parameters;

ε_{iz} is the error term, which is assumed to be independently and identically Gumbel distributed across zones alternatives and individuals.

In the logit formulation the vector of parameters γ is considered fixed across individuals. However to measure perception of accessibility between three geographical levels, we have introduced a variable (URBAVAR) which measures the density of each statistical sector. This variable can take the three following values: 0 for a zone which lies in rural area (outside built-up area or in a built-up area including less than 7 000 inhabitants), 1 for a built-up area with more than 7 000 inhabitants included in the set of the less dense sectors and finally, 2 for a built-up area with more than 7 000 inhabitants and included in the set of the most dense sectors.

Mixed logit assumes that individual parameters vectors $\gamma(i)$, $i = 1,..., I$, are realizations of a random vector $\gamma^{'}$. We assume that $\gamma^{'}$ is itself derived from a random vector δ and a parameters vector θ, which we express $\gamma^{'} = \gamma^{'}(\delta,\theta)$. For example, if $\gamma^{'}$ is a J-dimensional normally distributed random vector, we may choose $\delta = (\delta_1, \delta_2, ..., \delta_J)$ with $\delta_j = N(0,1)$, and let θ specify the means and standard deviations of the components of $\gamma^{'}$. The probability choice is then given by (see section 6.2.1):

$$P_{iz}(\theta) = E_P[L_{iz}(\delta,\theta)] = \int L_{iz}(\delta,\theta)P(d\delta) = \int L_{iz}(\delta,\theta)f(\delta)d\delta$$

where P is the probability measure associated with δ and $f(\bullet)$ is its distribution function, and L_{iz} is the logit formula given by

$$L_{iz} = \frac{e^{V_{iz}}}{\sum_{l=1}^{|A(i)|} e^{V_{il}}}.$$

The vector of parameters θ is then estimated by maximizing the log-likelihood function, i.e. by solving

$$\max_{\theta} LL(\theta) = \max_{\theta} \frac{1}{I} \sum_{i=1}^{I} \ln P_{iz_i}(\theta)$$

where z_i is the alternative choice made by the individual i.

This involves the computation of $P_{iz_i}(\theta)$ for each individual i, $i = 1, ..., I$, which is impractical since it requires the evaluation of one multidimensional integral per individual. The value of $P_{iz_i}(\theta)$ is therefore replaced by an estimate obtained by sampling over δ (*Halton random draws* (Bhat, 1999)), and given by

$$SP_{iz_i}^{R} = \frac{1}{R}\sum_{r=1}^{R} L_{iz_i}(\delta_r, \theta)$$

where R is the number of random draws δ_r, taken from the distribution function of δ. As a result, θ is now computed as the solution of the simulated log-likelihood problem

$$\max_{\theta} SLL^{R}(\theta) = \max_{\theta} \frac{1}{I} \sum_{i=1}^{I} \ln SP_{in_i}^{R}(\theta)$$

We will denote by θ_R^* a solution of this last approximation (often called Sample Average Approximation, or SAA), while θ^* denote the solution of the true problem.

McFadden (1973) shows that the log-likelihood function is globally concave for linear-in parameters utility. Many statistical packages are available for estimation of these models. However when parameters enter the utility function non-linearly, a more adapted estimation code has to be written. Due to the size variables which are specific to destination choice models, the utility function is no more linear in its parameters. It implies that standard maximum likelihood can not be used. Therefore the special logit estimation procedure developed by Daly (1982) within ALOGIT has been used to estimate the multinomial logit model while the following two steps calibration was developed to estimate the mixed logit model:

Step 1 Estimate the "size variables" and calculate the attraction utility part, lnM_z as follows

$$lnMz = \ln(M_{z0} + \exp(\beta_1) * M_{z1} + \exp(\beta_2) * M_{z2} + ... + \exp(\beta_s) * M_{zs})$$

(note that not all size variable coefficients can be identified, so it is necessary to impose a scaling restriction such as $\beta_0 = 1$)

Step 2 Reinstate the fixed attraction utility part (computed in step 1) in the prime utility function (U_{iz}) and estimate the mixed logit model with random parameters on impedance measure.

- **Results**

In the MNL (first column of Table 6.2.2.2), the accessibility on the three spatial levels identified by the variable URBAVAR has been found negative and significant. The trend of those variables does not increase with density; the largest negative value is found when URBAVAR is equal to one where we expected mainly residential locations, the smallest value is associated with URBAVAR equal to two, where most of the shopping activities are expected. Four land use variables affect the destination choice behaviour; all except green area have been found positive and significant. Agricultural area is found attractive since meadowland is convenient for leisure and recreational activities. Built up area and densely built area increased activities facilities and are therefore found chosen as destination trips. Four size variables fit into the model, of which three could be estimated. The density of employment has been found strongly negative since destination of work has not been modelled. Moreover areas attracting work trips do not offer great opportunity for shopping and leisure activities. Surprisingly cinema and restaurants have been found not significant and are not shown in the table. The size variables that affected the destination choice model are the number of shops and the number of recreation activities defined as the sum of the number of recreation, sport and cultural activities reported in the statistical sector.

In the second column of Table 6.2.2.2, we report results of the mixed logit models; ten variables are reported to be significant. The MXL model allows the accessibility of destination locations to vary with normal distribution across the observations. Both mean and standard deviation of the accessibility parameter have been found significant; the mean is negative. Three land use parameters (all positive) have been included into the model. Housing area is in this case estimated and its parameter sign means that such zones are chosen as destination especially for visit to family trips. Six size variables are reported to be significant. The area of industry and shopping activities is the reference, the density of employment is the only variable found to be negative (as in the MNL) and four variables (recreation locations, cinema, restaurants, and shops) were found positive and significant.

The improvements in the value of the log-likelihood at convergence and in the number of estimated variables, comfort us in the use of a mixed logit formulation, moreover it allows analyses not possible under classical MNL assumptions.

Table 6.2.2.2 GRT Destination choice model on Antwerp: results

Variable	Multinomial logit (MNL)		Mixed logit (MXL)	
	Parameter	*t*-Stat.	Parameter	*t*-Stat.
Accessibility Variables				
Accessibility (Mean)			-0.4067	2.6
Accessibility (S.D.)			0.1353	2.9
Accessibility (URBAVAR = 0)	-0.0079	5.0	-	-
Accessibility (URBAVAR = 1)	-0.0089	8.2	-	-
Accessibility (URBAVAR = 2)	-0.0068	6.6	-	-
Land use variables				
Ln (agricultural area)	0.1139	1.7	0.2569	2.3
Ln (green area)	-0.1264	3.2	-	-
Ln (built area)	0.4204	11.7	0.1176	5.2
Ln (densely built area)	0.2156	3.7	-	-
Ln (housing area)	-	-	0.458	1.1
Size Variables				
Industry/Shopping area	1	-	1	-
Employment	-8.075	3.5	-3.084	4.0
Recreation	1.137	1.1	2.939	2.8
Cinema	-	-	7.927	8.3
Restaurants	-	-	3.375	3.4
Shopping	3.529	4.4	2.469	2.2
Number of observations	448		448	
Log-likelihood at convergence	-865.58		-852.92	
Degrees of Freedom	11		10	
Adjusted ρ	0.3766		0.3858	

Complete results are available in « *Combining Spatial and Temporal Dimensions in Destination Choice Models* », C. Cirillo, E. Cornélis, L. Legrain and Ph. Toint (2003).

6.2.3. GRT destination choice model over Flanders

This model is an improvement of what has already been achieved in the previous formulation on Antwerp travel data. The model includes now all trips from MOBEL travel data which have both origin and destination in Flanders. Socio-demographics and correlations between stops made at different time of the day are as well estimated. A logit formulation has been compared to a mixed logit with error components to hold on correlation among destinations. We point here that such a model could not be developed on the entire Belgium (as expected), because at that time data were not all available (e.g. land use characteristics).

- **Description of the set of alternatives**

In the previous model, the individual choice set was determined by randomly selecting nine alternatives from the set of statistical sectors in the action-space to which the effectively chosen destination was added (see Section 6.2.2.). In order to estimate a mixed logit model with error components and to allow alternatives to be correlated over unobserved factors, all the actually chosen destinations corresponding to observed stops in the individual daily activity chain have also been added to the set of alternatives leading to a set of at most 19 alternatives. Those destinations are classified into three groups:

Group I:	Group II:	Group III
▪ Home ▪ Work (workers) ▪ Main activity stop (non-workers)	▪ Main stop in the morning commute/ before tour ▪ Main stop in the evening commute/ evening tour	▪ 1 Secondary stop in the morning commute/ before tour ▪ 1 Secondary stop in the evening commute/ evening tour ▪ 2 Secondary stops in principal tour

- **Variables**

Because we thought that choices are sensitive to characteristics of the individual (as already seen in MOBEL and "SAMBA: Spatial Analysis and Modeling Based on Activities: A pilot study for Antwerp and Gent", Verhetsel *and al*. (2002)), the following characteristics has been introduced: age, gender, number of children (0, 1 or more), car ownership (0, 1, 2 or more), number of workers (0, 1, 2 or more) and marital status. But since these characteristics are always fixed whatever the alternatives, taste variation among individuals due to these observed factors is only measured when interacting for example with accessibility. The estimated parameters take account of the current socio-demographic characteristics as explained in Section 6.2.1 about the limitation of the logit formulation. Land use variables, size variables and accessibility variables are those previously described in the Antwerp destination choice model.

- **Model formulation**

The utility function is mainly the same as in the previous logit model. The mixed logit formulation is quite different since error components are introduced to handle correlation due to unobserved factors. The aim has been to bring out the correlation structure among chosen destinations corresponding to the following stops in the individual daily activity chain:

- Main stop in the before tour,
- Main stop in the evening tour,
- Secondary stop in the before tour,
- Secondary stop in the evening tour,
- Stop during the morning commute to work / first intermediary stop in the principal tour and,
- Stop during the evening commute from work / second intermediary stop in the principal tour.

To achieve this goal different kinds of error structure in the utility function can be tested.

The error structure used in this model is the one provided by Walker (2001). She introduced flexible error structures which can capture interdependencies among alternatives.

It is assumed that individual n faced a set of J_n alternatives described by K observed factors and that the vector of utilities associated to each alternative is expressed as follow:

$$U_n = X_n\beta + F_n\xi_n + \upsilon_n$$

where

X_n is a $(J_n \times K)$ matrix of explanatory variables;
β is a $(K \times 1)$ vector of unknown parameters fixed across individuals and alternatives;
F_n is a $(J_n \times M)$ matrix in which the structure reflects the unobserved factors entering the error (F_n includes fixed and/or unknown parameters);
ξ_n is a $(M\times 1)$ vector of M multivariate distributed unobserved factors and
υ_n is a $(J_n \times 1)$ vector of IID Gumbel random variables with zero location parameter and scale parameter equal to $\mu > 0$. The variance is g/μ^2, where g is the variance of a standard Gumbel ($\pi^2/6$).

For estimation, it is desirable to specify the factors such that they are independent, so ξ_n is decomposed as follows

$$\xi_n = T\zeta_n$$

where:

T is a $(M \times M)$ lower triangular matrix of unknown parameters corresponding to the Cholesky factorisation of the covariance matrix of ξ_n (i.e. $TT' = \mathrm{cov}(\xi_n)$;
ζ_n is a $(M \times 1)$ vector of IID random variables with zero mean and unit variance. The vector of utilities then becomes

$$U_n = X_n\beta + F_nT\zeta_n + \upsilon_n\,.$$

The expressions of F_n and T yield two different kinds of error structure such as heteroscedastic mixed logit, mixed logit with nested or cross-nested structure, mixed logit with random parameters or else error components specification. This last structure is in fact a generalization that includes the heteroscedastic, nested and cross-nested structures.

The unknown parameters are μ, β, those in F_n, and those in T.

In order to bring out some kinds of correlation structure, the model has been estimated with different error structures until a significant one has been found. Hence, among various frameworks the following structures have been explored: (1) heteroscedasticity (with no correlation), (2) two-nests and three nests structure and finally, (3) cross-nested error structure. The heteroscadastic specification is expressed as follows:

F_n is an identity matrix (6×6) and the T matrix is diagonal with standard deviation of each factors:

$$T = \begin{pmatrix} \sigma_1 & & & & & \\ 0 & \sigma_2 & & & & \\ 0 & 0 & \sigma_3 & & & \\ 0 & 0 & 0 & \sigma_4 & & \\ 0 & 0 & 0 & 0 & \sigma_5 & \\ 0 & 0 & 0 & 0 & 0 & \sigma_6 \end{pmatrix}.$$

This formulation supposed that there is no correlation between the six stops and that the variance due to unobserved factors is different for each stop.

If we assume now that the first two alternatives belong to one nest and the last three alternatives belong to a second nest, we get a nesting error component structure. Two error components are therefore defined for each nest. These assumptions lead to the following F_n and T matrices:

$$F_n = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad T = \begin{pmatrix} \sigma_1 & \\ 0 & \sigma_2 \end{pmatrix}.$$

A three-nests structure which is defined by those matrices has also been experienced:

$$F_n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \; T = \begin{pmatrix} \sigma_1 & & \\ 0 & \sigma_2 & \\ 0 & 0 & \sigma_3 \end{pmatrix}.$$

Finally, a cross-nested structure which allows the second alternative belonging to two nests is tested.

$$F_n = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

In the present model it turns out that the three-nested structure has allowed to emphasize correlation among the main stop in the evening tour and the secondary stop performed in the same tour, the main stop in the before tour choice and the destination of the first intermediary stop in the principal tour and finally choice destinations of the main stop in the evening tour with the two intermediary stops of the principal tour.

- **Results**

With the MNL formulation (first column of Table 6.2.3.1), we have estimated one accessibility variable and eight interaction variables between accessibility and socio-demographic characteristics, seven land use variables and four size variables. All are reported significant and signs are coherent to what we expected. The interaction variables have a significant impact on the utility function. In particular we note that accessibility decreases with age; in particular the less negative value corresponds to people over 55 years old while the more negative value is associated to the 18-25 years old; old people often have a less constrained schedule than young people and are therefore less sensitive to travel time. Within the estimated land use variables just two of them have negative impact on the attraction of the destination zones: the population for which a higher density is synonymous of less activity opportunities and the percentage of land dedicated to agricultural activities. Infrastructure, housing area, built area, densely built area and surface have a positive effect on individual utility function.
The composite size measure of destination zone is represented by the industry/shopping area, which is normalized to one, the number of employed in the shopping sectors, in recreation and spectacle services, and in restaurants. The model tells us that 13 km^2 of industrial/shopping area are equivalent, for destination attractivity, to 0.8 individuals employed in the recreational sector, to 5 in the spectacle sector, to 13 in the catering and 20 in shops.

Within the Mixed Logit formulation (second column of Table 6.2.3.1), we estimated the accessibility variable as random distributed (normal), to capture random taste variation on this measure of travel time and cost. Both mean and standard deviation are found to be significant; we do not report the percentage of the population with positive accessibility values. Land use variables stay significant and with the same sign as in the logit model, except for the infrastructure variable. We were able to estimate again four size variables, but the employed in recreational activities are replaced by those working in sport activities. Three error component terms are estimated: (1) induces correlation among the evening main destination and the stop performed in the same tour, (2) accounts for spatial interactions across the morning main destination choice and the destination of the first stop in the principal activity, (3) is the term that correlates choice destinations of the evening tour with the stops in the principal tour. The error component (1) is negative, the other two are positive. A positive sign means that the destinations are chosen in the same destination zone while a negative sign supposes that the destinations are more spread in space.

When comparing goodness of fit we find that the mixed logit model performs much better than the logit model.

Table 6.2.3.1 GRT destination choice model on Flemish Region: results

Variable	Multinomial logit (MNL)		Mixed logit (ML)	
	Parameter	*t*-Stat.	Parameter	*t*-Stat.
LOS Variables				
Impedance (Mean)			-0.1157	-5.4
Impedance (s.d.)			-0.0671	-4.8
Impedance	-0.0307	-3.1		
Socio-demographic interaction with impedance				
Female	0.0082	2.6	-	-
Number of HHLD actives = 1	-0.0172	-3.4	-	-
Number of HHLD actives = 2	-0.0182	-3.5	-	-
Age 18-25	-0.0199	-3.0	-	-
Age 25-55	-0.0182	-2.2	-	-
Age > 55	-0.0103	-2.1	-	-
Number of cars 0-1	0.0152	2.1	-	-
Number of cars =2	0.0251	2.7	-	-
Land use variables				
Ln (population)	-0.4951	-9.7	-0.3567	-5.9
Ln (agricultural area)	-0.2817	-8.3	-0.3933	-6.8
Ln (infrastructure)	0.0645	2.1	-0.0514	1.0
Ln (housing)	0.1117	2.8	0.1669	2.4
Ln (built area)	0.0691	3.5	0.1164	3.5
Ln (densely built area)	0.1492	3.8	0.2003	2.8
Ln (surface)	0.6884	14.8	1.0740	12.4
Size variables				
Industry/Shopping area	1	-	1	-
Recreation	7.077	9.7	4.035	5.3
Spectacles	5.205	5.6	-	-
Restaurants	4.493	5.2	1.753	2.2
Shopping	4.519	9.0	1.912	5.8
Sports	-	-	2.851	2.9
Error components				
Evening Main + Stop Evening	-	-	-0.4072	-3.4
Morning Prin. + Stop1 Prin.	-	-	0.2559	3.6
Evening Main + Stop1 Prin. + Stop2 Prin.	-	-	0.4016	3.9
Number of observations	1582		1582	
Log-likelihood at zero	-3424.85		-3424.85	
Log-likelihood at convergence	-2961.68		-2676.33	
Degrees of Freedom	21		16	
Adjusted ρ	0.129		0.214	

Complete results appear in « *Introducing spatial correlation in destination choice models* », C. Cirillo, L. Legrain et Ph. L. Toint (2004).

6.2.4. UG, UA, UCL destination choice models over Antwerp

In this model the destination of the main stop in each home-based tour has been considered. Work is assumed to be a mandatory activity at a fixed place so that only non-work stops are modelled. The four possible tour structures are:

- Home-Main-Home
- Home-Intermediate-Main-Home
- Home-Main-Intermediate-Home
- Home-Intermediate-Main-Intermediate-Home

- **Description of the alternatives: spatial zoning algorithm**

An algorithm has been developed by UA (see "*De Introductie van Ruimtelijke Variabelen en Zoneringstechnieken in de Analyse van het Verplaatsingsgedrag"*, Van Hofstraeten, D. and Verhetsel, A. (2004)) in order to aggregate statistical sectors in a city region into more manageable zones. The algorithm is based on a clustering or classification of sectors by their spatial characteristics: land use, accessibility and attractiveness.

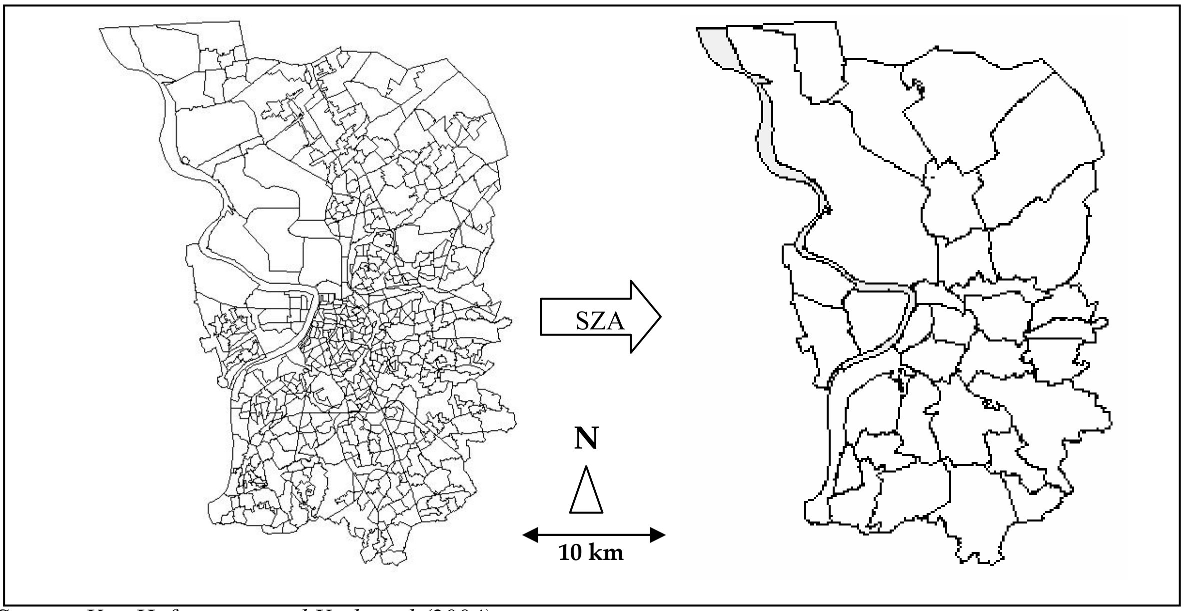
First, the correlated spatial characteristics are clustered through a principal component analysis with Varimax rotation. It results into a certain number of (uncorrelated) components each containing variables with the same patterns. Then a score on each of the defined components is computed for each statistical sector. A clustering method is then applied on the sectors to group sectors that are similar in terms of scores (see Table 6.2.4.1). The last step consists of defining a reasonable number of homogenous destination zones by grouping contiguous statistical sectors based on the clustering results. The gathering follows the assumptions that neighbouring sectors with similar land use are grouped together, and that the geographical delimitation of a destination zone takes into account physical barriers (river) and major infrastructures (highways, ring way, as first borderlines and main roads, main railways and administrative borders as secondly borderlines) (see Figure 6.2.4.1).

In the estimation process, each individual faces the entire set of the so-aggregated destination alternatives.

Table 6.2.4.1 Principal component analysis and clustering for the city region Antwerp

Components		Clusters	
1	Densely-built area, employment	1	Infrastructure
2	Residential built-up area	2	Densely-built area, large employment
3	Industry, port	3	Industry, port
4	Green area	4	Agriculture
5	Parks	5	Residential built-up area
6	Infrastructure	6	Green area
		7	Parks
		8	Green residential area
		9	Open area with scattered housing

Figure 6.2.4.1: From 608 sectors to 33 Destination Choice Zones for the city region Antwerp



Source: Van Hofstraeten and Verhetsel (2004)

- **Variables**

Land use Variables

The land use variables estimated in this model are the same as those used in GRT models. See Section 6.2.2.

Attractiveness variables

The destination size is here expressed in terms of

- number of **inhabitants** provided by the National Institute of Statistics (NIS, 2001);
- concentration of **employment** provided by the Regional Development Agency for 2001;
- the **number of schools** available in each statistical sector of the city region which were provided by the Department of Education of the Flemish Government, and consisted in a list of addresses of all the Flemish schools (primary and secondary schools, colleges, and universities); unfortunately, the size of each school is not known;
- the **number of shopping centres**. Shopping data on all commercial activities is not easy to obtain. However, this is a crucial variable in this analysis since shopping trips are frequent in our travel data set. An internet data source, called *SCOOT (www.scoot.be)*, gave the opportunity to find the addresses of most shopping alternatives, and some additional information on them. The computation resulted in approximately 6,000 stores (great and small) in the studied city region. Geo-coding these addresses made it possible to assign this information to each statistical sector. Due to a lack of data, no reference is made to the relative importance of stores (surface, etc.), the presence/absence of stores is only retained.

Accessibility variables

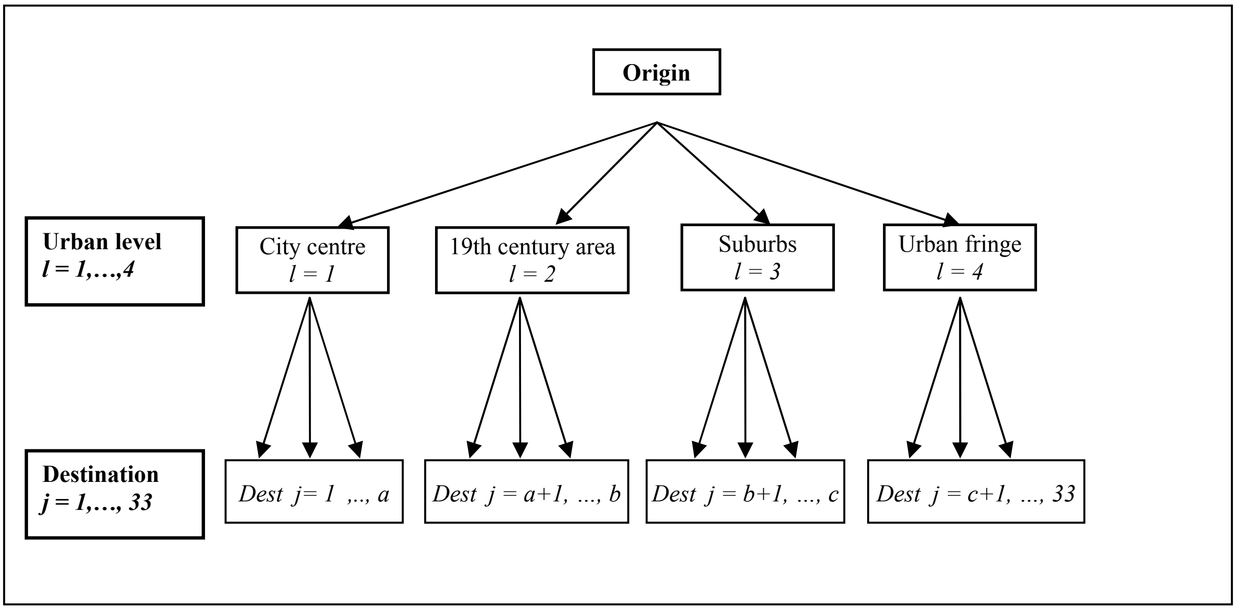
The shortest path network distance and time were computed between each centroid of the 608 statistical sectors by means of the *StreetNet* 2001 network and *ArcView Network Analyst*. The *StreetNet* road network includes information on traffic regulations, such as closed streets, one-way streets, underpass and overpass and travel surplus. However, waiting time at traffic lights or extra time to take turns have not been taken account. In order to partly compensate, all network segments have been assigned a slightly lower speed level than the actual maximum authorized speed. For the shortest route between each centroid, *ArcView Network Analyst* implements a modified Dijkstra algorithm (ESRI, 2003; Sherlock et al., 2002; Wise, 2002). The *StreetNet* software enables one to compute two variables: a network distance and a travel time by car. For the other transport modes (foot and bike) network distances are used to estimate travel time. The assumptions made on speed factors are 4 km/h for walking and 15 km/h for biking. Hence, the variable time refers to the shortest time between two centroids of the studied area, whereas the way of computing this time depends upon the transportation mode(s) used.

Socio-demographics

The most significant estimated parameters are those associated to the following characteristics: age, income (< 500 €) and household type (mono parental with two children and more; couple with two children and more).

- **Model formulation**

The destination choice process is supposed hierarchical so that a nested logit formulation is adopted for estimation. Each individual is assumed to first choose an urban level and then, within that broad spatial zone, to choose a precise destination. We have already seen in Section 6.1.2 that the urban level of the destination induces more friction to distance so that it makes sense to assume that the individual first chooses an urban level before an elemental alternative within the urban zone. This leads to the nested-logit formulation with two levels of decision for destination (see Section 6.2.1 for the formulation). The decision-making structure is the following:



A mixed logit formulation has also been estimated. The utility function associated to a destination j for an individual i can be considered as follows:

$$U_{ij} = \gamma T_{ij} + \beta \boldsymbol{X}_{ij} + \varepsilon_{ij}$$

For each individual i and for each destination j, T_{ij} denotes the travel time, $\boldsymbol{X}_{ij}$ is a vector of observable attributes and β are the parameters to be estimated. The residuals ε_{ij} are assumed following a Gumbel distribution. The parameter γ measures the travel time effect. If γ is assumed fixed across individuals, we get the standard logit formulation and the probability choice is

$$P_{nj} = \frac{\exp(\gamma T_{jn} + X_{jn}\beta)}{\sum_{j=1}^{J} \exp(\gamma T_{jn} + X_{jn}\beta)}$$

The mixed logit formulation is obtained by assuming γ random and normally distributed.

The density function $f(\gamma)$ is

$$f(\gamma) = \frac{\exp\left[-\frac{1}{2}\left(\frac{\gamma-\omega}{\sigma}\right)^2\right]}{\sigma\sqrt{2\pi}} = N(\omega, \sigma^2)$$

where $N(\omega,\sigma^2)$ is the density function of a normal distribution with mean ω and variance σ^2.

The mixed logit choice probability for choosing a destination j is then

$$P_n(j) = \int_0^\infty P(j \mid \gamma).N(\omega, \sigma^2)d\gamma$$

where $P(j|\gamma)$ is the standard logit formulation given by

$$P_n(j \mid \gamma) = \frac{\exp(\gamma T_{jn} + X_{jn}\beta)}{\sum_{j=1}^{J} \exp(\gamma T_{jn} + X_{jn}\beta)}$$

These different models have been estimated on the city region Antwerp. First the nested logit model has been compared to a standard logit model. Next the coefficient associated to the travel time variable is assumed random so that a mixed logit and a mixed nested logit formulation are compared. Finally, in order to reduce heterogeneity and misspecification of the global estimators, a mixed logit and a nested logit have been estimated only for shopping and leisure activities. Models have been estimated using the *Biogeme* software.

- **Results**

The MNL and nested MNL have first been estimated on the city region Antwerp (Table 6.2.4.1). The log-likelihood slightly drops from –6743 in the MNL case to –6697 on the nested logit and the coefficients of the inclusive factors for each branch of the nested logit are significantly different from 1. Hence, we can confirm that there is a hierarchical structure of stop choices in the studied area. In Antwerp, an individual will first choose a large sub-region (i.e. suburb, urban fringe, city centre, 19th century centre) and then within that sub-region, he/she will choose a more precise sector of destination. The nested logit enables one – in a certain way – to consider the nested effect of spatial scale as well as spatial correlation between destination choices. Let us also mention here that, as the coefficients of the inclusive variables are all significantly higher that 1.0, the IID property does not hold anymore. The MNL is no longer robust for explaining the destination choices. This gives us a good reason for preferring the nested logit formulation.

Table 6.2.4.1 Multinomial logit versus nested logit on Antwerp

	Multinomial Logit		Nested Logit	
	Value	**t-test**	**Value**	**t-test**
Accessibility variables				
Time	-0,43	-45,52	-0,40	-33,31
Land Use variables				
Agriculture	0,024	6,62	0,013	3,76
Industry	-0,017	-4,93	-0,019	-5,91
Housing	0,012	12,12	0,010	11,04
Parks	0,027	5,45	0,017	3,55
Green area	0,086	5,75	0,097	7,51
Size variables				
Employment	0,00002	3,60	0,00002	3,77
Number of shopping	0,00256	18,02	0,00191	13,26
Socio demographic variables				
Age	0,007*	1,43	0,012	3,74
Income				
Income <500 euro	0,73*	1,06	0,77*	1,13
Household type				

Monoparental with 2 Children and more	1,00*	1,76	0,95*	1,74
Couple with 2 children and more	1,01*	1,78	1,12	2,04
Household location				
Suburb	-1,45	-4,56	0	Fixed
Characteristic of chains				
Purpose : Service	1,02	2,00	1,04	2,10
Mode : bike	0,37*	0,81	0,31*	0,71
inclusive value 1			1	Fixed
inclusive value 2			1,49	7,54
inclusive value 3			1,06	2,10
inclusive value 4			1,07	0,94
Number of estimated parameters:	17		19	
Sample size:	3497		3497	
Null log-likelihood:	-12227		-12227	
Final log-likelihood:	-6743		-6697	
Likelihood ratio test:	10968		11061	
Rho-square:	0,45		0,45	

(* non significant at 5% level)

Next the mixed logit and the mixed nested logit are compared (see Table 6.2.4.2). A better fit and a better quality on the estimators are obtained comparing to Table 6.2.4.1: the log-likelihood ratio is much better and the rho-square much higher. As already shown in Table 6.2.4.1, the choice structure is once again nested. The inclusive variables are significantly higher than 1. Given that the model includes a random coefficient, the mixed nested logit model is chosen as the best estimation of the spatial choice process in the case of the destination choice process in the city region of Antwerp. This confirms other results obtained recently and independently by Suarez et al. (2004) on the best fit of nested logit in selecting shopping centres.

As expected, travel time appears to be one of the most important variables in the decision making process. Hence, destinations located further away from the place of residence or intermediate stop location, will be less likely selected as destination for an activity.

The choice of a destination is additionally and significantly influenced by other spatial and socio-economic variables. Spatial variables include attractiveness as well as land use information. The influence on the number of shopping and employment alternatives is positive: they increase the attractiveness of a destination and hence its choice as a destination. Moreover, the higher the percentage of surface affected to housing, parks or green area, the higher the probability of choosing this zone as destination. In contrast, industrial land use has a negative effect on the attractiveness of the destination. These results correspond to our expectations. Note that "agriculture" also has a positive influence on the probability to choose a zone: open "agricultural spaces" are attractive in terms of recreation and leisure, especially in an urban agglomeration.

Most socio-demographic variables as well as the characteristics of the tours themselves are less or not important in the explanation of the destination, with the exception of the location (suburbs) and the type (two children and more) of household. This means that personal characteristics are not very important in the modelling process, at least in Antwerp and for trip purposes. In other words, an individual chooses the destination that optimizes his/her utility under the constraint of the characteristics of the destination without reference to its personal characteristics with the exception of the transportation mode. This is a quite important result in terms of town planning: planning new attractive shopping malls or other shopping/service alternatives will have a drastic effect on transportation fluxes within the city. Let us also mention the particular importance of the suburbs.

Table 6.2.4.2 Mixed logit versus mixed nested logit: results

	Mixed Multinomial Logit		Mixed Nested Logit	
	Value	**t-test**	**Value**	**t-test**
Accessibility variables				
Time : mean	-0,75	-34,41	-0,64	-29,14
variance	0,51	12,30	0,38	11,82
Land Use variables				
Agriculture	0,014	3,63	0,011	3,18
Industry	-0,021	-5,81	-0,013	-4,59
Housing	0,003	2,30	0,009	10,02
Parks	0,010*	1,79	0,013	3,01
Green area	0,103	6,12	0,092	6,81
Size variables				
Employment	0,00002	3,12	0,00002	3,23
Number of shops	0,00305	22,02	0,00173	12,37
Socio demographic variables				
Age	0,005*	0,83	0,006*	1,30
Income				
Income <500 euro	0,75*	0,96	0,71*	0,99
Household type				
Monoparental with 2 Children and more	1,25	2,09	1,10	2,02
Couple with 2 children and more	0,88*	1,48	0,90*	1,71
Household location				
suburb	-1,46	-4,21	0	Fixed
Characteristic of chains				
Purpose Service	1,11	2,03	1,06	2,07
Mode bike	0,40*	0,80	0,31*	0,70
inclusive value 1			1,00	Fixed
inclusive value 2			1,62	7,50
inclusive value 3			1,22	5,43
inclusive value 4			1,58	3,48
Number of estimated parameters:	18		21	
Sample size:	3497		3497	
Null log-likelihood:	-12227		-12227	
Final log-likelihood:	-6554		-6498	
Likelihood ratio test:	11347		11458	
Rho-square:	0,46		0,47	

(* non significant at 5% level)

Finally a mixed nested logit model has been estimated on shopping and leisure trips to reduce heterogeneity in the data and thus avoid misspecification of the global estimators.

Results reported in Table 6.2.4.3 are close to those presented in Table 6.2.4.2: socio-economical characteristics of the individuals have only a small impact on the destination choice in Antwerp; travel time and spatial characteristics remain important explanatory factors in destination choice. However, the analysis on shopping and leisure trips give a better fit, as far as shopping is concerned; the pseudo rho-square increases from 0.47 to 0.53. This is not the case for leisure trips where rho-square decreases significantly from 0.47 to 0.37; this is probably due to the vague definition of the leisure activity which includes sport, cinema as well as simply "walk". Moreover, the location of this kind of activities is often not accurate. Even if our data set is quite large, it is not large enough to consider separate modelling of smaller subgroups of leisure activities. Leisure activities tend to occur at random. "Shopping" corresponds to a more homogeneous definition; that's why the results are much better in terms of rho-square. Compared to the results presented in Table 6.2.4.2, we observe that (1) the number of jobs does not affect the destination choice for shopping purposes (at the level of analysis, jobs are indeed concentrated in other places than shops), and (2) age enters positively and significantly the equation. Shopping seems – in our case – to affect more adults than youngsters.

Table 6.2.4.3 Mixed nested logit for shopping and leisure trips: results

	Shopping		Leisure	
	Value	**t-test**	**Value**	**t-test**
Accessibility variables				
Time : mean	-0,67	-20,72	-0,52	-15,43
variance	0,32	7,27	0,33	6,65
Land Use variables				
Agriculture	0,01	2,93	0,01*	1,65
Industry	-0,02	-3,60	-0,01*	-1,30
Housing	0,01	6,50	0,01	5,95
Parks	0,02	2,67	0,02	2,36
Green area	0,11	6,01	0,05	2,27
Size variables				
Employment	0,00001*	1,14	0,00002	2,63
Number of shopping	0,00183	8,91	0,00177	7,49
Socio demographic variables				
Age	0,02	2,86	-0,01*	-1,47
Income : Income <500 euro	1,41*	1,44	1,65*	0,00
Household type				
Monoparental with 2 Children and more	0,40*	0,52	1,88*	1,82
Couple with 2 children and more	0,12*	0,17	1,61*	1,78
Characteristic of chains				
Mode : bike	0,11*	0,18	-0,04*	-0,05
inclusive value 1	1,00	0,00	1,00	0,00
inclusive value 2	1,77	14,28	1,59	10,18
inclusive value 3	1,23	21,07	1,25	17,48
inclusive value 4	1,09	8,56	2,44	4,81
Number of estimated parameters:	17		17	
Sample size:	1881		1090	
Null log-likelihood:	-6577		-3811	
Final log-likelihood:	-3068		-2397	
Likelihood ratio test:	7018		2828	
Rho-square:	0,53		0,37	

(* non significant at 5% level)

The complete results are presented in « *How to incorporate the spatial dimension in destination choice models?* », H. Hammadou, I. Thomas, H. Tindemans, D. Van Hofstraeten, A. Verhetsel and F. Witlox (2004).

6.2.5. UA, UCL and UG model to compare Antwerp, Ghent, Mechelen and Aalst modelling

Based on the same methodology (in terms of choice set and variables definitions) developed for the city region Antwerp and described in Section 6.2.4, a logit model has been estimated on the Ghent and Mechelen city regions and on Aalst which is not a city region as such. Before undertaking this modelling, some explorative analyses on the different travel data sets have been realised (not reported here). Among them an analysis has tried to emphasize if a model can be developed on the Leuven and Halle-Vilvoorde travel data set. But unfortunately the data do not allow such modelling. Those analyses are described in several unpublished notes[1].

Only shopping trips have been taken into account in the modelling. The aim of this comparison is to see how the urban level can modify the results of the destination choice modelling. Moreover the aggregation technique of statistical sectors (based on the spatial zoning algorithm) has been compared to the random sampling technique. The results are presented in Tables 6.2.5.1 and 6.2.5.2. It has been found that the model with random sampling of alternatives performed less good looking at Rho value and significance of parameters. The parameters often take the sign which was expected. All attractiveness variables enter the utility positively which is quite expected for shopping activities and we found that industrial areas are not attractive for shopping trips. For Ghent and Antwerp city regions, shopping trip purpose is rarely achieved outside the residential urban level. Two groups of behaviours can be emphasized. The first one gathers Antwerp and Ghent, and the second one contains Aalst and Mechelen. In each group the cities seem to have common results in terms of rho square and average interpretation of parameters. We also point that Aalst has some results quite different from the others probably due to its size and the fact that it is not a city region.

The paper related to this modelling part is still in progress: «*Does the size of the city influence spatial choice behaviour: a comparison of four Flemish cities* », H. Hammadou, I. Thomas and A.Verhetsel.

[1] « Analyse descriptive des données de déplacement de Aalst », F. Riguelle and I. Thomas (2004), « Aalst : description de la procédure et des résultats du zonage » (UCL), « Samba : Analyse exploratoire des données de déplacement de la périphérie bruxelloise » (UCL)

Table 6.2.5.1 Logit model with aggregated sectors with the spatial zoning technique

	Ghent		**Antwerp**		**Aalst**		**Mechelen**	
	Value	t-test	Value	t-test	Value	t-test	Value	t-test
Accessibility variables								
Time	-0,34	-35,85	-0,64	-40,51	-0,50	-35,37	-0,44	-24,90
Land Use variables								
Industry	-0,046	-2,59	-0,020	-4,46	-0,013	-3,22		
Housing	0,075	9,83	-0,002	-0,73*	0,015	4,53	0,013	4,85
Built up area	0,012	3,13	0,006	3,02	0,003	1,40*	0,021	6,56
Size variables								
Number of supermarket	0,225	2,20	0,160	4,11	0,214	6,69	0,412	12,02
Number of shopping	0,003	2,54	0,003	6,31	0,004	3,36	0,007	2,30
Socio demographic variables								
Age	-0,01	-2,97	-0,03	-5,88	0,00	-0,85*	-0,01	-4,30
Income								
Income <500 euro	-0,464	-1,87*	-0,18	-0,55*	0,61	2,16	0,03	0,12*
Household type								
Couple with 1 child	0,234	1,02*	-0,30	-0,89*	0,39	2,61	-0,21	-0,86*
Household location								
stay in the same urban level	1,16	18,45	0,47	5,82				
Characteristic of chains								
Mode								
car	0,52	3,34	0,60	2,04	0,30	2,10	0,65	3,69
Nombre d'alternatives	31		33		29		31	
Sample size:	1856		1881		1538		1014	
Null log-likelihood:	-6373		-6577		-5179		-3482	
Final log-likelihood:	-3047		-2961		-3529		-1888	
Likelihood ratio test:	6652		7232		3300		3188	
Rho-square:	0,522		0,550		0,319		0,458	

Table 6.2.5.2 Logit model with random sampling of alternatives

	Ghent		Antwerp		Aalst		Mechelen	
	Value	t-test	Value	t-test	Value	t-test	Value	t-test
Accessibility variables								
Time	-0,37	-57,52	-0,67	-34,84	-0,49	-28,79	-0,43	-22,81
Land Use variables								
Industry	-0,012	-8,73	0,000	-0,16*	0,001	0,54*		
Housing	0,009	6,72	-0,011	-12,44	0,005	2,73	0,002	1,48*
Built up area	0,002	1,62*	0,001	1,66*	-0,001	-1,04*	-0,004	-1,44*
Size variables								
Number of supermarket	0,203	6,80	0,025	2,03	0,486	4,00	0,141	18,69
Number of shopping	0,022	13,15	0,030	21,82	0,025	17,24	0,006	2,72
Socio demographic variables								
Age	9,35	9,40	0,01	0,64*	0,00	0,00*	-0,05	-11,65
Income								
Income <1250 euro	0,007	0,00*	1,11	0,85*	0,00	0,00*	-0,15	-0,37*
Household type								
Couple with 1 child	0,023	0,00*	0,23	0,41*	0,00	0,00*	-0,10	-0,22*
Household location								
stay in the same urban level	0,75	0,00*	-0,54	-3,10				
Characteristic of chains								
Mode								
car	0,12	0,00*	-1,08	-1,42*	0,00	0,00*	0,23	0,75*
Sample size:	1849		1878		1538		974	
Null log-likelihood:	-5623		-5713		-4669		-2956	
Final log-likelihood:	-2941		-2934		-3090		-1482	
Likelihood ratio test:	5365		5558		3158		2946	
Rho-square:	0,477		0,486		0,338		0,498	

6.3. Border effects

A border is usually defined as a limit between two states. But there exist other kinds of borders that have an impact on travel behaviour such as in Belgium, the linguistic border (Dujardin, 2001). Although border effects have been already studied at a national level, they have never been observed at smaller geographical level such as agglomeration or city. This study shows that for example a river or a highway can constrain trips emission at city region level.

The most commonly used model to measure border effects is the gravity model. Nevertheless discrete choice theory can also be used to emphasize border effects by means of destination choice models. In this study the two approaches have been estimated on the OVG-Antwerp travel data set and compared. Unfortunately, MOBEL data were not sufficient for this kind of spatial data analysis. Hence, the effect of the linguistic border could not be analysed.

Border effects can be detected by two different approaches both applicable to the gravity and destination choice models. These two approaches are the "deductive approach" and the "explicative approach". The first one is based on analysing residual flows resulting from the model estimation while the second one concerns the estimation of explicative variables related to the *expected border*. An "indicative variable" is then usually introduced. This variable takes the value "1" if a trip between an origin and a destination implies crossing a suspected border like a river or a highway, and value "0" otherwise. Let us first describe the gravity model.

6.3.1. Gravity model

The deductive approach uses the following modelling structure:

$$F_{ij} = a_i O_i b_j D_j f(d_{ij})$$

$$s.c. \quad a_i = \frac{1}{\sum_j (b_j D_j f(d_{ij}))}$$

$$b_j = \frac{1}{\sum_i (a_i O_i f(d_{ij}))}$$

where F_{ij} is the flow between origin zone i and destination zone j, O_i is an emission factor from zone i, D_j is the attraction factor to zone j, $f(d_{ij})$ is the distance impedance function, and finally a_i, b_j, the parameters.

Generally, the deductive approach allows detecting border effects by analysing the residual flows on a map. Three kinds of residuals can be defined:

Rough residual flows: $F_{ij} - F_{ij}^*$

Relative residual flows: $(F_{ij} - F_{ij}^*) / F_{ij}^*$

Migrating residuals: $sign(F_{ij} - F_{ij}^*)\left[(F_{ij} - F_{ij}^*)^2 / F_{ij}^*\right]$

where F_{ij}^{*} is the estimated flow and F_{ij} is the observed flow.

If the model overestimates the flows, it means that in reality an obstacle puts a curb on trips spreading.

The explicative approach consists of a gravity model under its logarithmic form as proposed by McCallum (1995):

$$\ln(X_{ij}) = \alpha + \beta_1 \ln(Y_i) + \beta_2 \ln(Y_j) + \beta_3 \ln(D_{ij}) + \beta_4 front_{ij} + \varepsilon_{ij}$$

where X_{ij} is the flow between zone i and j, Y_i and Y_j are explicative variables related to the respective zone, D_{ij} is the distance between zone i and j. The variable, $front_{ij}$, is introduced to measure the border effect. Its value is 1 if the expected border is crossed when travelling from zone i to zone j and zero otherwise.

6.3.2. Destination choice model

A multinomial logit model has been estimated (see Section 6.2.2). Then an origin-destination matrix is built based on the aggregated predicted choices of the individuals from which flows between zones can be deduced. As for the gravity model, the first approach to emphasize border effects is based on residuals computed from estimated flows as follow:

Rough residual flows: $N_{ij} - N_{ij}^{*}$

Relative residual flows: $(N_{ij} - N_{ij}^{*}) / N_{ij}^{*}$

Migrating residuals: $sign(N_{ij} - N_{ij}^{*})\left[(N_{ij} - N_{ij}^{*})^2 / N_{ij}^{*}\right]$

where N_{ij}^{*} is the estimated flow computed from aggregated predicted choices of the individuals given by $N_{ij}^{*} = \sum_{n_i=1}^{N_i} P(j \mid x_{n_i})$ where $P(j \mid x_{n_{1i}})$ is the probability that an individual located in zone i chooses zone j as destination, and N_{ij} is the observed flow.

The second approach is based on the estimation of a variable measuring a border effect. As in the gravity model, an additional variable taking value 1 when the border is crossed and zero otherwise is estimated.

6.3.3. Data

Models have been estimated on OVG-Antwerp travel data sets. The travel data set consists of home-based tours containing at most four trips. These tours were described in terms of a main activity and intermediary stops such as in Section 6.2.4.

The variables to be estimated in the two models are the *travel time*, computed by means of a GIS-software, it gives the travel time corresponding to the shortest path between the centres of each statistical sector; the *population*, estimated only in the gravity model, it measures the flows emission capacity from one zone to all the other ones; the *number of employments* and *number of shops* are estimated in the gravity model, those variables quantify the size of a destination zone; finally, the *number of shops* and *land use variables* are estimated in the destination choice model as spatial characteristics of the destination.

6.3.4. Results

By comparing the results obtained by the models, we found that the gravity model deductive approach does not allow concluding on a border effect. However by introducing an indicative variable, we found that the coefficients related to the Schelde river and the highway are significantly different from zero and negative meaning that those elements induce border effects. The deductive approach of the destination choice model is more convincing: it shows that the model overestimates flows between zones located on both sides of the river and highway. According to the "gravity model", the coefficients of the indicative variables are found significantly different from zero and negative, meaning that the probability to choose a destination for which the trip requires crossing the river or the highway decreases.

The paper related to border effects is still in progress: « *Effets de frontière et choix de la destination* », H. Hammadou and I. Thomas.

7. Synthetic population

In Belgium, the available origin-destination matrices have until now been built from the census data of the Belgian population. But the information collected during these censuses, on the mobility behaviours of the Belgian residents relate only to their daily home-work or home-school trips. However, the Belgian mobility survey, MOBEL, reveals that these kinds of trips do not make the majority of those carried out by the population (Hubert and Toint, 2001). Indeed, people also travel for leisure, shopping, visits to family etc. Moreover, an increasing percentage of the population does not form part of the active population (and thus generates other trips but no "home-work" ones). Finally, no trips made apart from working days are taken in account in the collected information. In addition, based on the information collected in the MOBEL survey, it is not possible to build origin-destination matrices since the obtained sample is not representative on the trips level or, more exactly, on the level of each zone considered as origin or destination (unless an aggregation at the province or even region level). Hence we see that both sources of available information present some disadvantages and do not allow, by their own, estimating origin-destination matrices by trip purpose and/or mode of transport with an acceptable level of space desegregation. However, it would be interesting to obtain such matrices for the whole Belgian population. That would allow us to estimate the travel demand for all purposes and the whole population residing in Belgium.

The **Synthetic population** methodology is a tool which should allow us to fill this objective. The aim is that starting from detailed information - but partial because relative to a sample of the population (for example, contained in a mobility survey) - and aggregated information - but exhaustive on the real population (drawn from the census) - it is possible to build a population which is statistically close to the "true population" i.e., in our case, a population synthesis of Belgian population which satisfies the distributions of the true population and satisfies the statistics relative to travel behaviours observed in the survey. This method was already used, in particular in the TRANSIMS project in Los Alamos or in FAMOS ("*A Multimode Activity-Based travel demand Modelling System for Florida*" University of Florida, Department of Civil and Environmental Engineering) or even by Frick and Axhausen (2004). However, unlike these other case studies, we have approached, in this project, these techniques of synthetic population through a new methodology which seemed more adapted to our data.

The approach used for estimating the travel demand can be summarized by the following steps. First a synthetic population is built for each of the 589 Belgian municipalities. Then, an activity chain (drawn among chains collected, for example, in the MOBEL survey) is assigned to each individual of this *baseline* synthetic population and finally, a destination is chosen for each activity which the synthetic individual carries out (home, work, school, shopping centre, etc.).
As, in the framework of this project, we aim at developing tools which can be used to estimate the travel demand, we have therefore concentrated our efforts on creating the *baseline* synthetic population and then on developing an explorative method which allows assigning activity patterns to the population.

7.1. Data

The *baseline* synthetic population (which is defined as the first stage of a synthetic population), will consist of a set of individuals only characterized by a given number of socio-demographic variables. Therefore our first step is to define the characteristics which will describe the synthetic individuals. It seems better to focus on characteristics which appeared, at the time of MOBEL analysis, relevant for the description of travel behaviours.

The following variables were thus selected in order to define our various classes of individuals:

- Gender (male; female)
- Age (6-17; 18-39; 40-59; + from 60 years)
- Type of household (isolated with children; isolated without child; couples with children; couples without child)
- Professional status (student; active; inactive)
- Level of education or *diploma* (primary; secondary; superior; no diploma)
- Availability of a driving licence (yes; no)

In this way, we can define a little more than 700 categories of individuals (results of all possible crossings between these various variables). In order to reduce this number, and finally to get a set of 536 classes of individuals, we made the following non-restrictive assumptions:

- No individuals having the driving licence in the class of ages 6-17 years.
- No active or inactive in the age group of the 6-17 years
- No student in the age group over 60 years
- No superior diploma in the age group 6-17 years.

The information used to build the baseline population must, according to the method specificities, be related to the real Belgian population (the goal being of course to build a synthetic population close to this "true" population). The necessary data must provide exhaustive information but rather aggregate on the overall population. This information can be collected via a census or extracted from administrative bases like, for example, INS databases or National Register (RN). For our concern, we called upon various sources and we obtained the following marginal totals:

INS 2001

- Municipality **x** age **x** gender

SPF mobility, data 2004

- Municipality **x** age **x** gender **x** driving licence

Data obtained from the GEDAP (Louvain-La-Neuve) (in December 2004)

- Municipality **x** type of household (extracted from 2001 census data)
- Municipality **x** professional status (extracted from 1991 census data)
- Municipality **x** level of education (extracted from 1991 census data)
- District **x** type of household **x** age (extracted from 2001 data)
- District **x** gender **x** professional status (extracted from 1991 census data)
- District **x** age **x** professional status (extracted from 1991 census data)
- District **x** gender **x** level of education (extracted from 1991 census data)
- District **x** professional status **x** level of education (extracted from 1991 census data)
- District **x** age **x** level of education (extracted from 1991 census data)

As it can be noted, these data were available with various levels of geographical desegregation: mainly the municipality or the district was used as spatial entity. But considering that the problem of creating a synthetic population can be formulated in order to benefit from all available information, using together data at the municipality level and at the district one can be envisaged as an advantage. However it is necessary to point the lack of homogeneity in our data: the fact that they come from different sources and years will be awkward, and it has been necessary to hold account of it in the formulation of our problem.

These aggregated but exhaustive data sets will constitute the *constraints* of our problem and we also often refer them as *margins or marginal totals*. The built synthetic population must indeed satisfy as good as possible these totals in order to be as close as possible to the true population.

7.2. Building the baseline synthetic population

The most popular method to create a synthetic population is an iterative process called *IPF*: Iterative Process Fitting. This algorithm was presented by Deming and Stephan (see Little and Wu, 1991). It consists in updating, in all its dimensions, a multidimensional table, obtained from crossings between categories of the socio-demographic variables (for example gender **x** age **x** level education), in order to satisfy the margins on each dimension; each cell thus corresponding to a number of individuals belonging to some category of individuals.
It is worthwhile mentioning that, since it acts as an update, this method requires an initial matrix (or a multidimensional table). Such a table can generally be given by a representative sample of the population.

With this method, a multidimensional table would be first estimated for each district by taking in account the known margins for the current district, with the proviso of having a sample of the population for the district, which is not the case for MOBEL (the investigated sample is not statistically significant at district level). It would be then possible to estimate a table for each municipality of the considered district taking now into account the known margins for the municipalities. Once the various tables have been estimated for each municipality, the population can be randomly drawn according to the estimated distributions.

However in our case, as we have already noted, our sample, i.e. the MOBEL survey sample, is only significant when drawn at the level of the whole country or of the three Belgian regions. Consequently, we are not able to apply the iterative process presented in the previous paragraph. Moreover, this method only works successfully if the margins taken in account are coherent between them. This last condition was not met for our dataset, since different data sources were sometimes inconsistent. For example, we present in Table 7.2.1 different levels of inconsistencies. The column *prop* is the ratio between the 2001 population (from INS data source) and the total population given by summing the different crossings collected to build the synthetic population. Let us point that the 2001 population is equal to the margins given by the constraint age x gender x municipality by summing on all municipalities.

We see that the data sets for Charleroi are more consistent than the ones regarding Nivelles or Sint-Niklaas (which exhibits the largest inconsistencies).

Table 7.2.1 Example of constraints inconsistencies

	Charleroi		Nivelles		Sint-Niklaas	
Real population (INS, 2001)	**405 491**		**342 028**		**173 474**	
	Margins	Prop	Margins	Prop	Margins	prop
Household type x municipality	**380 653**	0.94	**321 777**	0.94	**208962**	1.20
Diploma x municipality	**426 372**	1.05	**321 144**	0.94	**215903**	1.25
Professional status x municipality	**396 594**	0.97	**299 145**	0.87	**206197**	1.19
Age x household type x district	**357 884**	0.88	**302 240**	0.88	**196961**	1.14
Age x diploma x district	**398 582**	0.98	**299 443**	0.88	**202728**	1.17
Age x professional status x district	**385 024**	0.95	**289 989**	0.85	**200457**	1.16
Profess. status x diploma x district	**396 594**	0.98	**299 145**	0.88	**206197**	1.19
Gender x diploma x district	**426 079**	1.05	**320 922**	0.94	**215876**	1.24
Gender x profess. status x district	**396 363**	0.98	**299 007**	0.87	**206174**	1.19

Therefore, given the current state of our data, the IPF does not represent, independently from other reasons, the best method for generating the synthetic population in our case study.

Consequently, we have thought of another method better suited to our data. In order to understand our approach, let us recall the required goal: generating a population which satisfies some constraints (given by the margins known for the "true population") while remaining close to the observed (through a sample) population. This problem can be formulated as a least squares problem.
Moreover, in this new approach, we build a population for each municipality by considering each district (and its municipalities) separately. We give up building the whole Belgian population in one go because creating simultaneously synthetic populations for all districts at once leads to a problem which is too big to be practically manageable. Consequently, since we do not have a representative observed sample for each district, the least square problem will not reveal information of this kind. This least squares problem can therefore be formulated as follows.

But, first, to simplify the exposition, let us assume that the population is built by considering only three socio-demographic variables: age, gender and professional status (the method can be generalised without difficulty to more dimensions).
Here are the available data and their notations:

- at the district level, we have the following crossins: (age **x** gender) and (gender **x** professional status)**.**
- at the municipality level, we have margins for all categories in each of the three variables but no crossing.

Let us define the two following matrices (for each district)

- a matrix specific to a district *a* (age **x** gender **x** professional status) whose elements are noted α_{ijk}, , with *i* for the age index, *j* for the gender index and *k*, for the professional status index.
- a matrix (age **x** gender **x** professional status **x** commune) whose elements are noted c_{ijkc}, $c = 1,...,\#c$, with *i,j* and *k* defined as previously, *c* being the index of municipality and #*c*, the number of municipalities in the district.

Following these definitions, each cell contains a number of individuals (belonging to a class of individuals of age i, of gender j and of professional status k - in the municipality c when it is necessary to specify it).

Our objective is to estimate the four dimensional (age **x** gender **x** professional status **x** commune) matrix, e_{ijkc}, given the following totals:

- Known margins at the district a level :

$$\alpha_{ij.} = \sum_k \alpha_{ijk}$$
$$\alpha_{.jk} = \sum_i \alpha_{ijk}$$

- Known margins at the municipality c level:

$$T_{i..c} = \sum_{j,k} c_{ijkc},$$
$$T_{.j.c} = \sum_{i,k} c_{ijkc},$$
$$T_{..kc} = \sum_{i,j} c_{ijkc}$$

The associated least square problems (one per district) are then expressed as:

$$\min \quad \sum_{i,j,k} \left(\sum_c e_{ijkc} \right)^2$$

$$s.c. \quad \boxed{\begin{array}{ll} \sum_{j,k} e_{ijkc} = T_{i..c} & \forall c \in a \\ \sum_{i,k} e_{ijkc} = T_{.j.c} & \forall c \in a \\ \sum_{i,j} e_{ijkc} = T_{..kc} & \forall c \in a \end{array}}$$

Municipality margins constraints

$$\boxed{\begin{array}{l} \sum_{c \in a} (e_{i..c} + e_{.j.c}) = \alpha_{ij.} \\ \sum_{c \in a} (e_{.j.c} + e_{..kc}) = \alpha_{.jk} \end{array}}$$

District margins constraints

If crossings (e.g. age **x** gender) are known at the municipality level, we also get the following margins:

$$\gamma_{ij.c} = \sum_k c_{ijkc}$$

The constraints block related to these additional margins will then contain this new constraints set:

$$(e_{i.c} + e_{.j.c}) = \gamma_{ij.c}.$$

This problem has been encoded according to SIF (Standard Input Format) language in order to be able to use LANCELOT package for its resolution (Conn, Gould and Toint, 1992). Forty-three data SIF files (each one corresponding to a Belgian district) have been created. The size of each problem can be found in Table 7.2.2. As it can be seen most of the problems are of large dimension.

However, this first formulation has led to problems due to the presence of linear dependences in the constraints. These dependences arise from the fact that some constraints could be found to be linear combinations of other constraints. Moreover, the suppression of the redundant constraints revealed the fact that the problem was inadmissible because of the inconsistency of our constraints. Consequently the least squares problems have to be modified, on one hand, for preventing the appearance of linear dependences, and on the other hand, so that the inconsistency of our data could be taken care of.

We have thus considered a weighted least squares formulation. For each district *a* including *c* municipalities, we thus solve the following problem:

$$\begin{aligned}
\min \quad & \sum_{i,j,k,c} \left(w_v e_{ijkc}\right)^2 + \sum_{c \in a} \left(w_{c_1} \left(\sum_{j,k} e_{ijkc} - T_{i..c} \right) \right)^2 + \sum_{c \in a} \left(w_{c_2} \left(\sum_{i,k} e_{ijkc} - T_{.j.c} \right) \right)^2 + \sum_{c \in a} \left(w_{c_3} \left(\sum_{i,j} e_{ijkc} - T_{..kc} \right) \right)^2 \\
& + \left(w_{a_1} \left(\sum_{c} (e_{i..c} + e_{.j.c}) - \alpha_{ij.a} \right) \right)^2 + \left(w_{a_2} \left(\sum_{c} (e_{.j.c} + e_{..kc}) - \alpha_{.jka} \right) \right)^2 \\
s.c \quad & e_{ijkc} \geq 0
\end{aligned}$$

Weights related to each kind of constraint are respectively written $w_v, w_{c_1}, w_{c_2}, w_{c_3}, w_{a_1}\ et\ w_{a_2}$.

This formulation has the advantage of enabling us to keep all the constraints since in this formulation the linear dependences are no longer a concern.

Table 7.2.2 Least squares problems size

District	Number of Municipalities	Number of Constraints	Cells to be estimated
Number of Variables	**536**		
Aalst	10	316	5360
Anvers	30	816	16080
Arlon	5	191	2680
Ath	8	266	4288
Bastogne	8	266	4288
Brugge	10	316	5360
Bruxelles	19	535	10184
Charleroi	14	416	7504
Dendermonde	10	310	5360
Diksmude	5	191	2680
Dinant	15	441	8040
Eeklo	6	216	3216
Gent	21	585	11256
Halle Vil	35	941	18760
Hasselt	18	516	9648
Huy	17	491	9112
Ieper	8	266	4288
Kortrijk	12	366	6432
Leuven	30	810	16080
Liège	24	660	12864
Maaseik	13	391	6968
Marche	9	291	4824
Mechelen	13	389	6968
Mons	13	385	6968
Mouscron	2	116	1072
Namur	16	466	8576
Neufchateau	12	366	6432
Nivelles	27	741	14472
Oostende	7	241	3752
Oudenaarde	11	341	5896
Philippeville	7	241	3752
Roeselare	8	266	4288
SintNiklaas	7	241	3752
Soignies	8	266	4288
Thuin	14	416	7504
Tielt	9	291	4824
Tongeren	13	385	6968
Tournai	10	316	5360
Turnhout	27	741	14472
Verviers	29	791	15544
Veurne	5	191	2680
Virton	10	316	5360
Waremme	14	416	7504

7.3. Results

We have thus estimated a synthetic population for each municipality in the 43 districts of Belgium.
Let us focus on results for the district Antwerp. Antwerp district is composed of 30 municipalities. For each municipality, we estimated a number of individuals for each particular individual class. Table 7.3.1. provides an extract of our results file for the Antwerp municipality (NIS 11002). The first column contains some of the individual classes. The second gives the municipality index and finally, in the last column, we find the estimated number of individuals created for the corresponding individual class. For example, the first line suggests that 2296 individuals, male students aged between 18 and 39, without diploma and without driving licence live in a household constituted of a couple with children.

In Table 7.3.2, we then present the size of the synthetic population for each district compared to the real size of this district population. The accuracy of the results closely follows, as expected, the (in)consistency of the problem data: when coherent marginal sums are available, these marginal sums are well reproduced (with accuracy around 1%) while incoherent sums lead to proportionally poorer results. We present in Table 7.3.3, the error measured from the constraints violation. For each constraint a penalty term is computed which represents the number of individuals created in excess or in default. By comparing this penalty to the real number of individuals which have to be created (given by the INS 2001 data source), we notice that the penalties reflect the inconsistency observed in the data source. For example we observe for Sint-Niklaas, which has great inconsistencies (cf. Table 7.2.1), that, on average, the constraints exceed the real population with about 20% - this excess appears in the penalties (first column of Table 7.3.3) and we similarly find that the synthetic population exceeds with 19% the true size of the population (Table 7.3.2). The conclusions are similar for the district of Charleroi.

The inconsistencies in the data source therefore represent an unavoidable limitation on the quality (consistency) of the baseline population, irrespective of the computational method.

Table 7.3.1 Example of synthetic individuals

Individual Class	Municipality Number	Estimated Count
Antwerp Municipality		
FHEOO2[2]	2	2296
FHESO2	2	0
IHEOP2	2	1029
IFEOP2	2	92
CHEOP2	2	1432
CHEPP2	2	0
CFEOP2	2	495
IHEOP3	2	1180
IFEOP3	2	149
IFEPP3	2	0
IFEUP3	2	0
NHEOP3	2	244
FHEOP3	2	2652
IHAPP2	2	121
IHASP2	2	2461
IHAUP2	2	944
IHIOP2	2	370
IHIPP2	2	0
IHISP2	2	243
IHIUP2	2	0
IFAOP2	2	1328
IHAOO2	2	2111
IHAPO2	2	38
IHASO2	2	2250
NHASO2	2	1354
FFAOO3	2	1315
FFAPO3	2	0
FFASO3	2	1051
FFAUO3	2	0
FFIOO3	2	1424
FFIPO3	2	0
FFISO3	2	895
IHAOP4	2	2168
IHAPP4	2	0
IHASP4	2	1436
IFIOO4	2	3935

[2] * Position 1: household type (N: isolated with children; I: isolated without child; F: couples with children; C: couples without child)
Position 2: gender (H: male; F: female)
Position 3: professional status (A: active; E: student; I: inactive)
Position 4: diploma (P: primary; S: secondary; U: superior; O: no diploma)
Position 5: availability of driving licence (O: no; P: yes)
Position 6: age (1: 6-17; 2: 18-39; 3: 40-59; 4: +from 60 years)

Table 7.3.2 Synthetic Population by district

District	Synthetic population	Real population (INS 2001)	Matching
Aalst	248546	255355	97%
Anvers	877514	905233	97%
Arlon	46920	50541	93%
Ath	72891	76825	95%
Bastogne	36290	39648	91%
Brugge	251839	264957	95%
Bruxelles	864229	930628	93%
Charleroi	399680	405491	99%
Dendermonde	175434	181538	97%
Diksmude	45303	46732	97%
Dinant	89011	97030	92%
Eeklo	75910	77271	98%
Gent	462728	481754	96%
Halle Vil	510684	545239	94%
Hasselt	350148	375723	93%
Huy	89475	97449	92%
Ieper	99349	101469	98%
Kortrijk	264278	270485	98%
Leuven	412662	444720	93%
Liège	550694	611131	90%
Maaseik	194967	215787	90%
Marche	44333	49053	90%
Mechelen	283701	297932	95%
Mons	234509	240430	97%
Mouscron	68322	67580	101%
Namur	255922	274757	93%
Neufchâteau	51146	53986	95%
Nivelles	304365	342028	89%
Oostende	132324	140196	94%
Oudenaarde	107539	111062	97%
Philippeville	56356	59883	94%
Roeselare	132475	136820	97%
Sint-Niklaas	206668	173474	119%
Soignies	159849	168375	95%
Thuin	135856	141708	96%
Tielt	82747	85482	97%
Tongeren	174219	186056	94%
Tournai	133295	136093	98%
Turnhout	368897	398518	93%
Verviers	239963	258376	93%
Veurne	51335	55762	92%
Virton	43827	46876	93%
Waremme	60781	66800	91%
Entire Population	9446981	9966253	94,79%

Table 7.3.3 Constraints violation in the synthetic population

	Sint-Niklaas				Charleroi		
	Penalty	True Value of the Constraint	Error		Penalty	True Value of the Constraint	Error
age x gender x municipality	33036	173474	1,19		-5966	405491	0,99
household type x municipality	-2438	208962	1,19		18896	380653	0,99
diploma x municipality	-9380	215903	1,19		-26848	426372	0,99
professional status x municipality	6065	206197	1,22		2929	396594	0,99
age x household type x district	9562	196961	1,19		41654	357884	0,99
age x diploma x district	3792	202728	1,19		961	398582	0,99
age x professional status x district	330	200457	1,16		14510	385024	0,99
professional status x diploma x district	321	206197	1,19		2938	396594	0,99
gender x diploma x district	-9360	215876	1,19		-26548	426079	0,99
gender x professional status x district	346	206174	1,19		3174	396363	0,99

This part of the research, of a very exploratory nature, turned out to be very difficult. The main problem was to handle the data inconsistencies in an acceptable way, and we tried a number of formulations before settling on that presented here. There is no doubt that it can be further improved if new information (data itself or data accuracy estimates) become available.

7.4. Assigning activity chain to the baseline population

Unfortunately, time limitations have prevented us to complete our research in computing an origin-destination matrix associated with our synthetic population. The methodology was however investigated and we describe how in this section.
The next step would be allocating an activity chain to each individual of our synthetic population. This process would be based on mobility behaviour data collected in MOBEL. For our study, we have more precisely decided to consider in MOBEL the observations relative to individuals having carried out a maximum of six trips during the day. It is not such a constraining assumption since all these chains up to six trips represent more than 80% in the whole set of MOBEL reported chains. If the individual has made a trip, his/her chain is required imperatively to start and finish at home (H (home) - RH (return to home)). Each trip is characterized by a departure time (categorized at morning peak (M), evening peak (S) or out-of-peak (O)), a trip purpose (the categories selected are: work (T), school (E), shopping

(S), pick up/drop off (A) or other (O)) and finally a transport mode (the modes are gathered into "slow" mode (with foot or bicycle) (D), car driver (V), car passenger (P) or public transport (C)). According to the number of trips it contains, the chain will be known of length 0 in the case of no travel, of length *j*, (*j=2..., 6)* for a number of *j* carried out trips. Let us point that a trip of length 1 does not exist insofar as the built chains require a return (thus a displacement) to home. In case of travel, a chain thus has a minimal length of two. Each selected observed individual will thus be characterized by his/her daily activity chain and his/her socio-demographic category determined by the crossing of the variables being used in the building of the synthetic population (see Section 7.2). After elimination of the incomplete chains, we obtain a set of 5290 individuals from which 1581 have made no trips. These individuals define 285 classes of individual type and have realised 3177 distinct chains. We then assign to each considered socio-demographic category all chains observed for the individuals belonging to this category. For each class, we thus know the various chains which could be assigned to a synthetic individual who belongs to it. Moreover, in each class, we have associated to each chain its observed probability (i.e. the proportion of observations of this chain in the whole set of chains collected for this class).

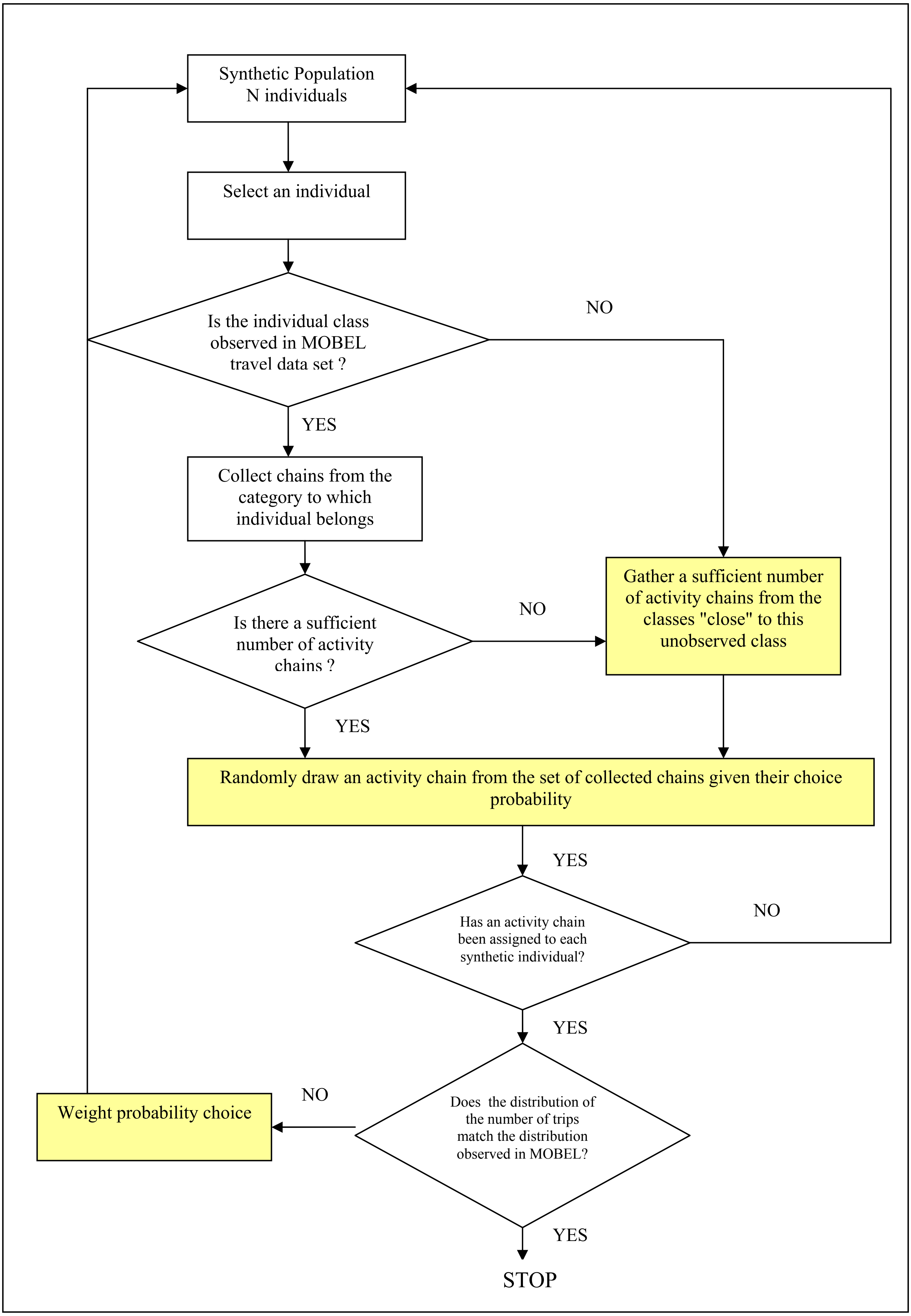
Let us take an example: we consider individuals belonging to the class (CFIOP3) which is the one of the women living in couple, inactive, without diploma, having a driving licence and being between 40 and 59 years old. For them, we have collected the 20 following chains:

Length	Chain	Number of individuals having realized this chain	Proportion
0	**000**	13	0.65
2	**OCS SCRH**	1	0.05
2	**OVA MVRH**	1	0.05
2	**OVO SVRH**	1	0.05
3	**MPS SPA SPRH**	1	0.05
3	**MVO MVA OVRH**	1	0.05
3	**OPO SPO SPRH**	1	0.05
5	**OVS OVO OVS OVS SVRH**	1	0.05

The assignment of an activity chain to a synthetic individual consist thus in associating to this synthetic individual one of the chains collected for the category to which he/she belongs. Hence, we have to randomly select a chain respectively to the observed probabilities of each chain. However as we have already explained previously, the class of a synthetic individual can have not been observed in the MOBEL sample. In this case a direct matching is impossible. A process has consequently to be developed to gather a sufficient number of activity chains from the classes "close" to this unobserved class on the assumption that this class of (synthetic) individuals, not observed in MOBEL, will behave in manner "similar" to those of its neighbourhood. This same process can also be implemented if, for a category, the number of MOBEL observations is considered to be insufficient (i.e. below a given threshold).

Once an activity chain has been assigned to each synthetic individual, it is necessary to take care of satisfying, in this whole set of "synthetic" chains, some statistics observed in MOBEL. Indeed, the goal of our synthetic population is to rebuild, at the level of the whole Belgian population, travel behaviours observed in the national mobility survey on a sample of this population. At a first step, we limit ourselves to build a synthetic population which satisfies the distribution of the number of trips. Therefore, if, after the assignment of the activity chains to the synthetic individuals, we find across the synthetic population the same distribution for the number of trips such as the one observed in MOBEL, we estimate to be satisfied by our population. If not, we will have to restart an assignment process of the chains but now with constraining this assignment so that it reproduces the desired distribution for the number of displacements, i.e. by weighting the proportions used during the random selection of the chain. This process continues until no further improvement can be made. A detailed algorithm for this process appears on the following page.

We have started implementing this algorithm. But work has still to be done in order to get a module which successfully assigns a chain. In particular the selection of a chain among the set of activity chains, available for a class of individuals, has to be improved. Moreover problems encountered with synthetic individuals not observed in the MOBEL survey have also to be solved. Once such an activity chain "allocator" would be obtained, the next step should consist of locating synthetic individuals' travels to activity destinations. This module should use the destination choice models previously developed. But we could also imagine another method which would be based on action space theory and data collected to characterised destination zones and therefore would allow computing a probability of choosing a zone depending on the trip purpose. We hope to find the opportunity to apply this methodology to its full extent in future researches.


Synthetic Population
N individuals
Select an individual
Is the individual class observed in MOBEL travel data set ?
NO
YES
Collect chains from the category to which individual belongs
Is there a sufficient number of activity chains ?
NO
Gather a sufficient number of activity chains from the classes "close" to this unobserved class
YES
Randomly draw an activity chain from the set of collected chains given their choice probability
YES
Has an activity chain been assigned to each synthetic individual?
NO
YES
Does the distribution of the number of trips match the distribution observed in MOBEL?
NO
Weight probability choice
YES
STOP

8. Rough sets: exploratory analysis

We conclude our research description by discussing how the UG investigates in a new technique to emphasize rules entering the decision-making process relative to the activity chains scheduling. This new approach is the **rough sets technique**. Until now, this technique was successfully applied in a series of research domains but has never been used to model daily activity chains. This technique was introduced to manage enormous databases. According to Pawlak (1981), "*the rough sets technique is a mathematical tool to search large, complex databases for meaningful decisions rules*". Indeed, the main problem faced by the researcher working with big data sets is the lack of interpretation. Unlike other techniques, rough sets provide results which are expressed in a more or less natural language, allowing an easier interpretation.

In activity-based transport modelling, we face larger and larger travel data sets, and it could therefore be interesting to study if rough sets can help us in our understanding of activity chains scheduling. This study was conducted on the MOBEL and OVG travel data.

First, let us describe briefly rough sets theory and later the results. More details are available in the paper: "*The Application of Rough Sets Analysis in Activity-Based Modelling; Opportunities and Constraints*", Witlox F. and Tindemans H. (2003).

The rough sets theory was introduced in 1981 by Pawlak. The theory is a mathematical framework which deals with vagueness and uncertainty, and can be viewed as an alternative to probability theory. The quintessence of the rough sets theory is discovering relationships in data – a set of objects - only by using their structure and information (no external parameters are used). The information about objects can be represented in the form of a decision table: rows contain condition attributes while columns contain decision attributes.

The basic concepts of rough sets are the *indiscernibility*, the *reduct and core* and finally the *decision rule*. The *indiscernibility* is the central notion of rough sets; it means that all objects characterized by the same information are indiscernible and define an "elementary set". However, all sets of objects can not be related to an elementary set without ambiguity. Some of them belong certainly to the set; they form the "upper approximation" while others possibly belong to the set, in this case those objects form the "lower approximation". A rough set is then defined through its lower and upper approximation. The *reduct* is the minimal attributes subset still providing the same object classification as with the full set of attributes. The *core* is the intersection of all reducts. Finally, the *decision rule*, presented as an "if-then-else" structure, reflects the relationship between a set of conditions and a conclusion or a decision, or more precisely, it combines the reducts with the values of the data. An example of a decision rule could be:

if gender(female) *and* age(35-45) *and* purpose(shopping) *then* mode(car) or mode(bike)

where *gender*, *age*, *purpose* are reducts and *mode* is a decision variable.

The rough sets modelling process can be describe in the five following stages:

- Data selection
- Pre-processing stage: removing obvious outliers, replacing or deleting blanks, discretization and finally, splitting the selected data set in a training data set and a test data set used in the final stage.

- Generation of the reducts
- Generation of the decision rules
- Evaluation: to assess the performance of the output, the revealed patterns can be evaluated according to robustness (rate of correct prediction on the training set), predictive ability (rate of correct prediction on the test data set), usefulness and simplicity.

All this process is supported by a software package called ROSETTA (Ohrn, 1999).

Two travel decisions were examined by means of rough sets technique. The first one concerns the destination choice and the second is relative to the mode choice. The condition and decision variables are presented in Table 8.1 and Table 8.2.

Table 8.1

	Variable	Classes
condition variables	gender	female; male
	living situation	head of the family; partner; child; other
	income	< 500 euro/month; 500-1250 euro/month; 1250-2500 euro/month; >2500 euro/month
	profession	student; worker; employee; executive; retired; other, with job; other, no job
	age	<16; 16-24; 25-34; 35-64; 65+
	car use	daily; >= 1 a week; not weekly, never
	mode	bicycle; car; foot; moped/motor; public transport; train
	purpose	bring/get someone; education; leisure; shopping; visit; work
	home location	Central Business District (CBD); 19th century rim, suburbs; fringe
decision variable	trip destination	Central Business District (CBD); 19th century rim, suburbs; fringe; outside city region

Table 8.2

	Variable	classes
	gender	female; male
condition variables	living situation	head of the family; partner; child; other
	income	< 500 euro/month; 500-1250 euro/month; 1250-2500 euro/month; >2500 euro/month
	age	<16; 16-24; 25-34; 35-64; 65+
	car use	daily; >= 1 a week; not weekly, never
	moped use	daily; >= 1 a week; not weekly, never
	motor use	daily; >= 1 a week; not weekly, never
	bus use	daily; >= 1 a week; not weekly, never
	tram use	daily; >= 1 a week; not weekly, never
	train use	daily; >= 1 a week; not weekly, never
	bike use	daily; >= 1 a week; not weekly, never
	distance to bus	0-249m; 250-499m; 500-999m; 1km-1,999km; 2km-5km; >5km
	distance to tram	0-249m; 250-499m; 500-999m; 1km-1,999km; 2km-5km; >5km
	distance to train	0-249m; 250-499m; 500-999m; 1km-1,999km; 2km-5km; >5km
	purpose	bring/get someone; education; leisure; shopping; visit; work
	trip distance	<= 1km; 1,1-2km; 2,1-5km; 5,1-10km; 10,1-15km; 15,1-25km; 25,1-40km; >40km
decision variable	mode	bicycle; car; foot; moped/motor; public transport; train

Following the reducts generation, all variables appear necessary to classify the information of the data set. For both cases the ROSETTA-system does quite well in classifying new objects, but only by producing a lot of detailed rules (see Table 8.3). Most produced rules were supported by just one or two objects. The high amount of rules makes the whole interpretation very hard although each rule is individually interpretable. Several attempts to reduce the number of rules were considered but even after completion (removing blanks in data set), or redefining the variables classes or else reducing the number of condition variables, the number of rules remains too high for human interpretation though it was reduced and classifying performance still good.

Table 8.3: Rough sets technique results

	Destination choice	**Mode choice**
Objects	8500	11750
Rules	>4000	8000
Robustness	0.75	0.95
Predictive ability (Naive Bayes method)	0.47	0.75

The main problem with activity chains travel data set is its complexity. Decisions made by a human being remain hard to understand especially since a lot of human and environmental factors are entering the decision making process. Rough sets technique needs to be more investigated in order to improve its capability of finding easily interpretable patterns in travel behaviour. Indeed, rough sets technique is quite new and still in progress. Some researchers are trying to base indiscernibility on a more human and sensitive method that accounts for decision-makers preferences by means of preference relations. More recently, hybrid techniques appear to fill gaps or combine performances (for example, neural networks or decision trees techniques are combined to rough sets). Research is therefore far from complete or conclusive in this domain.

9. Conclusions and prospects

The research program undertaken in the framework of this project, aimed at having a spatial insight of travel demand through distance decay, border effects and destination choice modelling. Distance decay shows the influence of distance and border effects emphasize how trips emission can be constrained by natural or infrastructural borders. Those elements have contributed to the development of some destination choice models: to know which variables have to be introduced in the model and to help in defining the choice set.

Much was learned by estimating different models for the same basic question: destination choice. Indeed, while being broadly comparable on many aspects, the estimated models differ on some significant aspects. Differences can be noted in the choice set definition: GRT used action space theory to account for chaining of activities while UA, UCL and UG developed a spatial zoning algorithm to aggregate similar statistical sectors based on land use characteristics. All models estimate zone size variables as well as land use variables. Socio-demographic characteristics were estimated differently in the methodology developed in GRT modelling and for UA, UCL and UG. GRT chose to analyse interaction between socio-demographics and accessibility, while socio-demographics were invariant in UA, UCL and UG models. Zone size variables were also estimated differently: they enter in a logarithmic form in GRT models and not in models developed by UA, UCL and UG. Finally, different model formulations have been used depending on the assumptions made for each used methodology and depending on what everyone expected to emphasize in its own modelling. GRT was interested in studying taste variation among individuals, measuring difference of accessibility perception and spatial correlations between destination choices in the daily activity pattern, while UA, UCL and UG expected to emphasize a hierarchical decision making process concerning destination choice. Despite the methodological differences, all results point out the fact that mixed logit or nested mixed logit formulations are clearly more suitable than techniques where the distribution of some important factors, like accessibility, is not parameterised. Estimated parameters mostly vary in the same direction in all models (except for agriculture and green areas). We may therefore conclude that, although broadly coherent, the obtained results differ between models, and that these differences may often be a result of different model formulations. Comparing statistics thus remains a somewhat perilous exercise which indicates that further work is probably needed on assessing model formulation (rather than on model estimation).

The other important task of the project was to produce tools in order to allow a further travel demand estimation. We used the synthetic population technique to create a set of individuals which reproduces as close as possible the margins observed in the 2001 Belgian population. This part of the research successfully brings a spatial component (the residence location) to a number of socio-demographics factors (such as education level, household type, or professional status), as well as to sub-classes of individuals obtained by crossing these factors. The research in this direction is however far of being concluded, as further exploitation of these results in activity chain modelling and travel demand estimation (the original long term goal of the project) remain, for now, in the future.

10. Publications and Presentations

- *« SAMBA: Spatial Analysis and Modelling Based on Activities: A pilot study for Antwerp and Gent*», Tindemans H., Van Hofstraeten D., Verhetsel A. and Witlox F. (2002). Paper presented at 5th Symposium of IUPEA, Oxford, 23-26 September 2002.

- « *Dynamics in city regions. The intra-urban travel patterns in Antwerp and Ghent*», Verhetsel A., Witlox F., Tindemans H., Van Hofstraeten D. (2002), in "The Land", Volume 6.2, p. 107-128

- «*Dynamiek binnen stadsgewesten. De intrastedelijke verplaatsingen in Antwerpen en Gent*», Verhetsel A., Witlox F., Tindemans H. & Van Hofstraeten D. (2003*)*, Forthcoming.

- «*Onderzoek verplaatsingsgedrag in het Gentse stadsgewest: het fietsgedrag nader geanalyseerd*», Tindemans H. & Witlox F. (2003). Paper presented at 'Eerste Belgische Geografendag', Liège, 12 March 2003.

- «*A spatial approach to the analysis of household activity surveys in Belgium*», paper presented at the EC-Workshop "Behavioural Responses to ITS", April 1-3 2003; Eindhoven

- «*Distance Decay in activity chains analysis. A Belgian case study*», Hammadou H., Thomas I., Van Hofstraeten D. & Verhetsel A. (2003); in DULLAERT W., JOURQUIN B. & POLAK J. (eds), Across the border. Building upon a quarter century of transport research in the Benelux, Antwerpen, De Boeck, pp 1-26.

- «*Analysing Bicycle Travel Behaviour in The Ghent Region: An Activity-Based Approach*», Witlox, F. & Tindemans, H. (2003). Presented at 1st Belgian Geographers Day, Liège. To be published in Belgeo (2nd review).

- « *Combining Spatial and Temporal Dimensions in Destination Choice Models* », C. Cirillo, E. Cornélis, L. Legrain and Ph. Toint (2003). FUNDP, département de mathématique, Rapport 2003/13

- « *Combining Spatial and Temporal Dimensions in Destination Choice Models* », C. Cirillo, E. Cornélis, L. Legrain and Ph. Toint (2003). Proceedings of European Transport Conference, October 2003

- «*The Application of Rough Sets Analysis in Activity-Based Modelling; Opportunities and Constraints*», Witlox, F. & Tindemans, H. (2003). In: Colloquium Vervoersplanologisch Speurwerk 2003: No pay, no queue? Oplossingen voor bereikbaarheidsproblemen in steden, Delft, CVS. Submitted to European Journal of Transport and Infrastructure Research (EJTIR)

- «*The Application of Rough Sets Analysis in Activity-Based Modelling; Opportunities and Constraints*», Witlox, F. & Tindemans, H. (2004). In: Expert systems with applications, 27, pp 585-592

- «*Evaluating bicycle-car transport mode competitiveness in an urban environment*», Witlox, F. & Tindemans, H. (2004). Journal of World Transport Policy and Practice. York, University of York, vol. 10 (4), pp 32-42

- « *Introducing spatial correlation in destination choice models* », C. Cirillo, L. Legrain et Ph. L. Toint (2004), paper presented at « workshop on progress in activity based modelling », Maastricht, May 2004.

- «*How to incorporate the spatial dimension in destination choice models? The case of Antwerpen*», Hammadou H., Thomas I., Tindemans H., Witlox F., Van Hofstraeten D. and Verhetsel A. (2004). In: Papers presented at Research, Istanbul, Turkey, July 4-8 2004

- «*How to incorporate the spatial dimension in destination choice models? The case of Antwerpen*», Hammadou H., Thomas I., Tindemans H., Witlox F., Van Hofstraeten D. and Verhetsel A. (2004). In: Regions and Fiscal Federalism: proceedings of the 44th Congress of the European Regional Science Association, Porto, 25-29 August 2004

- «*Children's space-time activity behaviour: a case study for Gent*», Witlox F. and Tindemans H. (2004). In: Child in the city: Second European Conference, 20, 21 and 22 october 2004, London, Child in the city Foundation, pp 66-81

- «*De Introductie van Ruimtelijke Variabelen en Zoneringstechnieken in de Analyse van het Verplaatsingsgedrag*» Van Hofstraeten, D. and Verhetsel, A. (2004). Antwerpen, Universiteit Antwerpen, Research paper.

- « *Estimation de la demande de mobilité par la création d'une population synthétique*», Cornélis, E., Legrain, L. and Toint, Ph. (2004). FUNDP, département de mathématique, Rapport 2004/14, présenté à l'ATEC 2005, Issy-les-Moulineaux, janvier 2005.

- « *Does the size of the city influence spatial choice behaviour : a comparison of four Flemish cities* », Hammadou, H., Thomas, I. and Verhetsel, A. (paper in progress)

- « *Effets de frontière et modèle du choix de la destination* », Hammadou, H. and Thomas, I. (paper in progress)

- « *Ruimtelijke analyse van de bestemmingskeuze van personen: een gevalstudie voor het Antwerpse stadsgewest*», Verhetsel A., Van Hofstraeten D, & Kandil I. (2005), Tijdschrift voor Vervoerswetenschappen, nr 3 (aanvaard voor publikatie).

11. References

Algers S, Bergstrom P., Dahlberg M and Dillen J. (1998). *Mixed logit estimation of travel time*. Working paper. dept of infrastructure and planning, Royal Institute of technology. Sweden

Ananthanarayana, V.S., Narasimha Murty, M., Subramanian, D.K. (2003). *Tree structure for efficient data mining using rough sets*. *Pattern Recognition Letters*. Vol. **24**, pp. 851-862.

Anderberg, M., (1973). *Cluster Analysis for Applications*. New York, Academic Press, 359 p.

Aptech Systems Inc (2002) GAUSS 5.0 for Windows.

Arentze, T. A., Hofman, F., van Mourik, H., Timmermans, H. J. P. & Wets, G. (2000). *Using decision tree induction systems for modelling space-time behaviour*. *Geographical Analysis*. Vol. **32** (4), pp. 330-350.

Arentze, T. & Timmermans, H.J.P. (Eds.) (2000). *ALBATROSS: A Learning-Based Transportation Oriented Simulation System*. Eindhoven, European Institute of Retailing and Services Studies (EIRASS), Technische Universteit Eindhoven.

Axhausen, K. and T. Garling (1992). *Activity-Based approaches to travel analysis: conceptual frameworks, models and research problems*. Transport Reviews, 12, 324-341.

Badoe, D.A. and Miller, E.J. (2000). *Transportation-land-use interaction: Empirical Findings in North America and their implications for modelling*. Transportation Research D, Vol. 5, pp. 235-263.

Bates, J. (2000). *History of demand modelling*. In: Handbook of Transport Modelling, ed. D.A. Hensher and K.J. Button, Pergamon, Oxford, pp. 11-33

Barette P., Cirillo C., Cornelis E., Dessy A. and Toint P.L (2001). *Enquête Nationale sur la mobilité des ménages*, Rapport final Publications of Mathematics Department N.2001/11 at FUNDP – GRT.

Bastin F., Cirillo C. and Toint P.L. (2003). *Numerical experiments with AMLET, a new Monte-Carlo algorithm for estimating mixed logit model.* In Proceedings of IATBR International Association for Travel Behaviour Research 2003 Lucerne Switzerland.

Beckman R.J., Baggerly K.A. and McKay *M.D.* (1996). *Creating synthetic baseline populations.* Transportation Research part A 30, p.415-429.

Ben Akiva M. and Bowman J. (1995). *Activity based travel demand model system*. In: Activity-based approaches to travel analysis, ed. D. Ettema, and H. Timmermans, Pergamon, Oxford

Ben Akiva M. Gunn H. and Silman L. (1984). *Disaggregate trip distribution models*. Proceedings Japanese society of civil engineers.

Ben Akiva M., Lerman S.R., (1985). *Discrete choice analysis: theory and application to travel demand*. Cambridge, MA: MIT Press.

Bhat, C.R. (2000). *A multi-level cross-classified model for discrete response variables*. Transportation Research B, Vol. 34, pp. 567-582.

Bhat, C. R. (2001). *Quasi-random maximum simulated likelihood estimation of the mixed multinomial logit model*, Transportation Research, **35B** (7) 677-693.

Bhat C.R. and Guo J. (2004). *A mixed spatially correlated logit model: formulation an duplication to residential choice modeling*, Transportation Research part B, vol.38, nb2, pp 147-168.

Bhat, C. & Koppelman, F. (1999). *Activity-based modelling of travel demand*. In: Handbook of Transportation Science, ed. R.W. Hall, Kluwer, The Netherlands, pp. 35-61

Bhat C.R. and Singh S.K. (2000). *A comprehensive daily activity-travel generation model system for workers* – Transportation Research part A 34, pp.1-22.

Bhat, C. & H. Zhao (2002). *The spatial analysis of activity stop generation*. Transportation Research Part B: Methodological 36, pp. 557-575

Bhat, C., Govindarajan, A. and Pulugarta, V. (1998). *Disaggregate attraction-end choice modeling (formulation and empirical analysis)* - Transportation Research Record 1645, pp.60-68

Boarnet, M.G. and Sarmiento S. (1996). *Can Land use policy really affect travel behavior? A study of the link between non-work travel and land use characteristics*. Paper prepared for 1996 Lincoln Land Institute TRED Conference, The university of California at Berkeley.

Bodpa A. & Grasland C. (1994). *Migration régionalisations et régionalismes Cameroun*. *Espace, Populations, Sociétés*, N°1, p. 109-129.

Bouckhuyt, A. (2002). *Génération d'une population synthétique.* FUNDP (promoteur du mémoire Ph. Toint)

Bours, A. (1971). *Van stad en gewest naar stadsgewest*. In: In: Stad en Stadsgewest in de ruimtelijke orde, Van Gorcum & Comp., Assen, p.309-317

Bowman J. (2004). *Logit Kernel (or Mixed Logit) models for large multidimensional choice problems: identification and estimation*.

Bowman, J. & M. Ben-Akiva (1997). *Activity-based travel forecasting*. In: Activity-based travel forecasting conference. Washington, DC: U.S. Department of Transportation, Report DOT-97-17.

Bowman J.L. , Ben-Akiva M.E. (2001). *Activity-based disaggregate travel demand model system with activity schedules*. Transportation Research part A 35, p.1-28.

Bradley M., Daly A., Hofman, Jong, Vellay C. (2001) *A model for time of day and mode choice using error components logit.* 2001 ETC proceeding

Bruinsma, F., Nijkamp, P. & Vreeker, R. (2002). *A comparative industrial profile analysis of urban regions in western europe: an application of rough set classification*. Tijdschrift voor Economische en Sociale Geografie. Vol. **95** (4), pp. 454-465.

Cabus, P. (1989). *De stedelijke ontwikkeling in Vlaanderen en Brussel tussen 1947 en 1988. Feitelijke situatie en houding hiertegenover van het (ruimtelijk) beleid*. In: Stadsvlucht maakt vrij, verlag van het colloqium stadsvlucht, ed. P. Cabus, C. Kesteloot and H. Van Der Haegen, Brussel, p.1-34

Cadwallader, M. (1985). *Analytical Urban Geography*, Prentice-Hall, New Jersey.

Cascetta, E. (2001). *Transportation Systems Engineering: Theory and Methods*. Applied Optimization Vol.49, Dordrecht, Kluwer.

Cervero, R. (1996). *Mixed land-use and commuting: Evidence from the American housing survey.* Transportation Research A, Vol. 30(5), pp. 361-377.

Cervero, R. and Knockelman, K. (1997). *Travel demand and the 3D's: Density, diversity and design*. Transportation Research D, Vol. 2(3), pp. 199-219.

Cirillo C. and Axhausen K.W. (2002). *Modal choice of complex tours*. In Proceedings of European Transport Conference Cambridge UK

Cirillo, C. & Ph. Toint (2001). *An activity-based approach to the Belgian National Travel Survey*. Groupe de Recherche sur les Transport, FUNDP, 24 p.

Clark, W.A.V. (2000). *Mono-centric to Polycentric: New urban forms and old paradigms*. In: BRIDGE, G & S. WATSON (Eds) *A Companion to the City*, Blackwell Publishers, Oxford, pp. 141-154.

Cramer, J.S. (1991). *The Logit Model: An Introduction for Economists*. London, Edward Arnold.

Cromley, R.G., (2002). *Spatial interaction/Spatial diffusion*. Course syllabus, University of Connecticut (downloadable from website http://www.geog.uconn.edu/courses/spring2003/g104.pdf).

Daly A. (1982). *Estimating choice models containing attraction variables*. Transportation Research, 16**B** (7) 5-15.

De Brabander, G. (1988). *Het Antwerpse stadsgewest: een economisch-geografisch portret*. SESO, Antwerpen, 47 p.

De Bruyn, F. (2001). *Mobiliteitsforum Kernstad*. Antwerpen Mobiel, on-line beschikbaar op: http://www.antwerpenmobiel.be/Studies/Mobiliteitsplan/Verslagen/Mobiliteitsforum15_06_2001.htm

Deming W.E. and Stephan F.F. (1940). *On a least squares adjustment of a sampled frequency table when the expected marginal tables are known*. The annals of mathematical statistics, 11, 427-444.

Devogelaer D. (2002). *Stedelijke woondynamiek van de Belgische bevolking en haar gezinnen*. Working paper, Federaal Planbureau, Brussel, 128 pp.

Dijst M. & Vidakovic (1997). *Individual Action-Space in the city*. In *Activity-based approaches to travel analysis*, D.F. Ettema et H.J.P Timmermans, Pergamon, p.117-134.

Doherty, S. *et. al.* (2002). *Conceptual frameworks for understanding activities and travel.* Social Sciences and Human Research Council of Canada, available on: http://www.wlu.ca/~sdoherty/mcri/detailed.html

Domencich, T. and McFadden, D. (1975). *Urban Travel Demand: A Behavioural Analysis*. Amsterdam, North-Holland Publishing Company.

Dordregter, P. Ph. (1971). *Het stadsgewest als verplaatsingssysteem: operationalisering en de theorie in het praktijkonderzoek*. In: Stad en Stadsgewest in de ruimtelijke orde, Van Gorcum & Comp., Assen, p.292-308

Dujardin C. (2001). Effet de frontière et interaction spatiale : Les migrations alternantes et la frontière linguistique en Belgique. L'Espace géographique, n°4, p. 307-320.

Düntsch, I. & Gediga, G. (1995). *Rough set dependency analysis in evaluation studies: an application in the study of repeated heart attacks*. Informatics Research Reports. Vol. **10**, pp. 25-30.

ESRI (2003), *Technical Article, No. 18856*. Downloadable from website: http://support.esri.com

Ettema, D., & H. Timmermans (1997). *Theories and models of activity patterns.* In: Activity-based approaches to travel analysis, ed. D. Ettema, and H. Timmermans, Pergamon, Oxford, pp. 1-36.

Evans Carolyn (2001). *Border Effects and the Availability of Domestic Products Abroad.* Federal Reserve Bank of New York Staff Report 127.

Fotheringham, A.S. (1981). *Spatial Structure and Distance-Decay parameters*. In: Annals of the Association of American Geographers, vol. 7, n° 3, p. 425-436

Fotheringham A.S., & O'Kelly M.E. (1989). *Spatial Interaction Models: Formulations an Application*s. Dordrecht : Kluwer Academic Publishers.

Frick M. and Axhausen K. (2004). *Generating synthetic populations using IPF and Monte Carlo techniques: some new results*. Technical report, conference paper STRC 2004.

Gärling, T., M.-P. Kwan & R. Golledge (1994). *Computational process modelling of household activity scheduling*. Transportation Research B: Methodological (28), pp. 355-364

Glenn, P., Thorsen, I. and Uboe, J. (2000). *Some theoretical aspects of distance deterrence functions in modelling journeys to work*. Paper presented at 3rd KFB Research Conference, Stockholm, Sweden, on-line beschikbaar op: http://www.kfb.se/junikonf/upps/Uboe.pdf

Grasland C. (1999). *Interaction spatiale et effets de barrière. 1ère partie : des modèles d'interaction spatiale aux modèles d'interaction territoriale. CNRS-UMR géographie-Cités mimeo.*

Greco, S., Matarazzo, B. & Slowinski, R. (1999). *Rough approximation of a preference relation by dominance relation.* European Journal of Operational Research. Vol. 117, pp. 63-83.

Greco, S., Matarazzo, B. & Slowinski, R. (2001). *Rough sets theory for multicriteria decision analysis*. European Journal of Operational Research. Vol. 129, pp. 1-47.

Greco, S., Matarazzo, B. & Slowinski, R. (2002). *Rough sets methodology for sorting problems in presence of multiple attributes and criteria*. European Journal of Operational Research. Vol. 138, pp. 247-259.

Grimmeau J.P. (1995). *Les migrations entre la Flandre et la Wallonie*. *Bulletin du Crédit Communal*, n°192, p.23-41.

Head, Keith and Thierry Mayer (2002). *Illusory Border Effects : How internal Geography Affects External Trade Volumes*. CEPII working paper n°1.

Head, Keith and Thierry Mayer (2001). *Effet frontière, integration économique et "Forteresse Europe"*. CEPII working paper n°6.

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

Hanou, M. (1995). *Stadsgewest en pendel. Geografie*. jg. 4, nr. 2, p. 31

Harms, L. en R. Van der Wouden (2002). Stadsgrenzen in beweging: mobiliteitspatronen en ruimtelijke schaalvergroting in de Randstad. Sociaal en Cultureel Planbureau, on-line beschikbaar op: http://plandag.org/pdf/Stadsgrenzen.pdf

Haynes, R.M. (1974). *Application of exponential distance decay to human and animal activities*. In: Geografiska Annaler, n°56 B, p. 90-104

Henscher, D.A. & Button, K.J. (eds.) (2000). *Handbook of Transportation Modelling*. Amsterdam, Pergamon.

Hensher D.A. and Greene, W. (2002). *Specification and estimation of the nested logit model: Alternative normalisations*. Transportation Research B, 36, 1-17.

Hensher, D.A. and Johnson, L.W. (1981). *Applied Discrete-Choice Modelling.* London, Croon Helm.

Hensher D.A., & Reyes A.J. (2000). *Trip chaining as a barrier to the propensity to use public transport*. Transportation, n°27, p.341-361.

Horowitz, A.J. (1998). *Through trip tables for small urban areas: A method for quick response travel forecasting*. on-line beschikbaar op: http://my.execpc.com/~ajh/Through.pdf

Johnson, D.S. (1974). *Approximation algorithms for combinatorial problems*. Journal of Computer and System Sciences. Vol. **9**, pp. 256-278.

Johnson, N.L. and Kotz, S. (1970). *Continuous Univariate Distributions-1*, New York, Houghton Mifflin Company.

Jones, P.M., F. Koppelman, & J.P. Orfeuil (1990). *Activity analysis: state-of-the-art and future decisions*. In Developments in Dynamic and Activity-based Approaches to Travel Analysis, Gower, Aldershot, UK

Jones, P.M., M.C. Dix, M.I. Clarke, & I.G. Heggie (1983). *Understanding travel behaviour.* Aldershot, Gower

Kitamura, R. (1988). *An evaluation of activity-based travel analysis*. Transportation 15, pp. 9-34

Kitamura R. (2004). *Spatial Processes*. In: Hensher D., Button K., Haynes K., Stopher P. (eds.) Handbook of Transport Geography and Spatial Systems. Oxford, Elsevier, pp. 513-531.

Kitamura R., Nishii, K. and Goulias, K. (1990). *Trip chaining behaviour by central city commuters: a causal analysis of time-space constraints* –- Developments in dynamic and activity-based approaches to travel analysis – ed. Peter Jones- Oxford studies in transport.

Klaassen L.H., Wagenaar S., & Van der Weg A. (1972). *Measuring psychological distance between the Flemings and the Walloons*. Papers of the Regional Science Association, 29, 45-62

Kockelman, K. (1996). *Travel behavior as function of accessibility, land use mixing, and land use balance: Evidence from San Fransico bay area* , Transportation Research Record 1607.

Krasowski, H. (1988). *Aircraft Pilot Performance Evaluation using Rough Sets*. Rzeszow, University of Rzeszow, Ph.D. thesis.

Kumar, A. & D.M.. Levinson (1995). *Chained trips in Montgomery County, Maryland*. ITE (Institute of Traffic Engineers) Journal, pp. 27-32

Kirby, A & D. Lambert (1984). *Space and Society 3: Urban systems*, Longman, Essex, 80 pp.

Leemans, S., Pattyn M., Rousseau S. & Van Der Haegen, H. (1989). *De Belgische stadsgewesten 1981*, In: Statistische Studiën 1989, Leuven, pp. 5-25.

Li, P. & Wang Z. (2003). *Mining classification rules using rough sets and neural networks*. European Journal of Operational Research (forthcoming; available on-line).

Liao, S. (2003). *Knowledge management technologies and applications – literature review from 1995-2002*. *Expert Systems with Applications*. Vol. **25**(2), pp. 155-164.

Little R.J.A. and Wu M (1991). *Models for contingency tables with known margins when target and sampled populations differ*. Journal of the American Statistical Association, March 1991, vol.86, n°413, Theory and Methods.

Mak, B. & Munakata, T. (2002). *Rule extraction from expert heuristics: a comparative study of rough sets with neural networks and ID3*. European Journal of Operational Research. Vol. **36**, pp. 212-229.

McFadden, D. (1973). *Conditional logit analysis of qualitative choice behaviour*. In: P.C. Zarembka (Ed.) *Frontiers in Econometrics*, New York, Academic Press, pp. 105-142.

McFadden, D. (1978). *Modelling the choice of residential location*. In: A. Karlqvist, L. Lundqvist, F. Snickars and J.W. Weibull (eds.) *Spatial Interaction Theory and Planning Models*, Amsterdam, North-Holland Publishing Company, pp. 75-96.

Maddala, G.S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge, Cambridge University Press.

Mayhew, S., (1997). *A Dictionary of Geography*. 2nd ed., Oxford, Oxford University Press.

McNally, M. & W.W. Recker (1986). *On the formation of household travel/activity patterns*. Institute of Transportation Studies, University of Canada, Irvine, CA, USDOT Final Report

McNally, M. (2000). *The activity-based approach*. In: Handbook of Transport Modelling, ed. D.A. Hensher and K.J. Button, Pergamon, Oxford, pp. 53-69

McNally, M. (2000). *The four-step model*. In: Handbook of Transport Modelling, ed. D.A. Hensher and K.J. Button, Pergamon, Oxford, pp. 35-52

Nuyts, E. & Zwerts, E. (2001) *Onderzoek Verplaatsingsgedrag Vlaanderen en Stadsgewest Gent, januari 2000-januari 2001, Controle en begeleidingsopdracht*. Unpublished.

Miller, H.J. (1999). *Potential contributions of spatial analysis to Geographic Information Systems for Transportation (GIS-T)*. *Geographical Analysis*, Vol. 31, pp. 373-399.

Miller E. (2004). *Intergrated land use/ tansport model requirements*. In: Hensher D., Button K., Haynes K., Stopher P. (eds.) Handbook of Transport Geography and Spatial Systems. Oxford, Elsevier, pp. 147-165.

Nerella, S. and Bhat, C. (2003). *A numerical analysis of the effect of sampling of alternatives in discrete choice models.* For presentation TRB 2004.

NIS, (2001). *Bevolking van het Rijksregister op 1 januari 2001*. Databank NIS

OC-GIS Vlaanderen (2002). *Bodembedekkings- en Bodemgebruiksbestand*. CD-ROM.

Openshaw, S. and Taylor, P.J. (1979). *A million or so correlated coefficients: Three Experiments on the modifiable areal unit problem*. In: N. Wrigley and R.J. Bennet (eds.), *Statistical Applications in the Spatial Sciences*, London, Pion, pp. 124-144.

Øhrn, A. (1999). *Discernibility and Rough sets in Medicine: Tools and Applications*. Trondheim, Norwegian University of Science and Technology, Ph. D. Thesis.

Ortuzar, J.D. and Willumsen, L.G. (2001). *Modelling Transport*. Chichester, John Wiley and Sons.

Papola, A. (2004). *Some developments on the cross-nested logit model*. Transportation Research B. Vol. 38(9), pp. 833-851.

Pas, E. (1996). *Recent advances in Activity-based travel demand modelling*. In Conference on Activity-based Travel Forecasting June 2-5, Texas Transportation Institute, Arlington, Texas, 1996

Pattyn, M. & Van Der Haegen, H. (1979). *De Belgische stadsgewesten. I*n: Statistische Studiën 1979, Leuven, nr. 56, pp. 3-18.

Pawlak, Z. (1981). *Information systems – Theoretical foundations*. Information Systems. Vol. **6**(3), pp.205-218.

Pawlak, Z. (1991). *Rough sets. Theoretical Aspects of Reasoning about Data - Theory and Decision Library, Series D: System Theory, Knowledge Engineering and Problem Solving (Vol. 9)*. Dordrecht, Kluwer Academic Publishers.

Pawlak, Z., Slowinski, K. & Slowinski, R (1986). *Rough classification of patients after highly selective vagotomy for duodenal ulcer*. International Journal of Man-Machine Studies. Vol. **24**, pp. 413-433.

Poulain M. & Foulon M. (1998). *Frontières linguistique, migrations et distribution spatiale des noms de famille en Belgique* . L'Espace géographique, Vol. 27, n°2, p. 53-62.

Pozsgay, M.A. and Bhat, C.R. (2001). *Modelling attraction-end choice for urban recreational trips: implications for transportation, air quality and land-use planning* — transportation research record.

Recker, W.W. (2001). *A bridge between travel demand modelling and activity-based travel analysis.* Transportation Research Part B: Methodological 35, pp. 481-506

Reilly, W.J., (1931). *The Law of Retail Gravitation.* New York, The Knickerbocker Press

Reinhart, A., Stawski, B. & Weber, T. (1989). *Application of rough sets to study the water outflow from the river basin*. Bulletin of the Polish Academy of Sciences, Technical Sciences. Vol. **37** (1), pp. 97-104.

Revelt, D. and Train K. (1998) *Mixed logit with repeated choices: household's choices of appliance efficiency level*. The review of Economics and Statistics. Vol. LXXX, No.2

Revelt, D. and Train, K. (1999). *Customer-specific taste parameters and mixed logit*. Working paper, Department of Economics, University of California, Berkeley. Paper available on: http://elsa.berkeley.edu/users/train/index.html

Rietveld & Janssen (1990). *Telephone calls and communications barriers : the case of Netherlands* . The Annals of Regional Science, vol 24, n°4, p.307-318.

Rietveld, P., Zwart, B. *et. al.* (1999). *On the relationship between travel time and travel distance of commuters*. In: The Annals of Regional Science, vol. 33, p. 269-287

Rosen, K.T. & Resnick, M., (1980). The size distribution of cities: an examination of *the Pareto Law and primacy.* Journal of Urban Economic, vol. 8, pp.165-186.

Shafer, G. (1976). *A Mathematical Theory of Evidence.* Princeton, Princeton University Press.

Sherlock, R., Mooney, P., Winstanley, A. and Husdal, J. (2002). *Shortest Path Computation: A Comparative Analysis*, Paper presented at the GISRUK 2002 Conference, Sheffield, University of Sheffield.

Schönfelder, S. and Axhausen, K.W. (2003). *Measures of social exclusion.* Submitted to Transportation Policy.

Schwanen, T., Dieleman, F.M. en M. Dijst (2002). *The impact of metropolitan structure on commute behaviour in the Netherlands*. Paper presented for the 42nd ERSA Conference

Simma, A., Schlich, R. and Axhausen, K.W. (2002). *Destination choice modelling for different leisure activities*. Institut für Verkehrsplanung und Transporttechnik, Strasses und Eisenbahnbau, ETH Zürich, Zürich.

Slavin H. (2004). *The role of GIS in land use and transport planning*. In: Hensher D., Button K., Haynes K., Stopher P. (eds.) *Handbook of Transport Geography and Spatial Syste*ms. Oxford, Elsevier, pp. 329-356.

Slowinski, R. (ed.) (1992). *Intelligent Decision Support: Handbook of Applications and Advances of the Rough Sets Theory*. Dordrecht, Kluwer Academic Publishers.

Slowinski, K., Slowinski, R. & Stefanowski, J. (1989). *Rough sets approach to analysis of data from peritoneal lavage in acute pancreatitis*. *Medical Informatics*. Vol. **13**, pp. 143-159.

Srour, I.S., Kockelman, K and T.P. Dunn. *Accessibility indices: A connection to residential land prices and locations choices*. Paper presented at the 81st Annual Meeting of the transportation research board. Forthcoming in Transportation Research Record.

Stad Antwerpen (2002), *Nieuw Zuid: De Nieuwe Place To Be,* Antwerpen. Stad Antwerpen, Unpublished paper.

Stead, D. (2001). *Relationships between land use, socio-economic factors and travel patterns in Britain*. Environment and Planning B, Vol. 28(4), pp. 499-529.

Suarez A., Rodriguez del Bosque I., Rodrigues-Poo J. and Moral I. (2004). *Accounting for heterogeneity in shopping, centre choice models*. *Retailing and Consumer Services*, Vol.11, pp. 119-129.

Taylor, P. J. (1975). *Distance Decay Models in Spatial Interactions*. In: Catmog (Concepts and Techniques in Modern Geography), n° 2, 35 pp.

Tieleman, H. (1998). *Modern mobiel*. In: Cultuur en Mobiliteit, ed. H. Achterhuis en B. Elzen. SDU, Den Haag

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

Train, K. (1986). *Qualitative Choice Analysis, Theory, Econometrics, and an Application to Automobile Demand*. Cambridge (Mass.) and London, MIT Press.

Train K.E. (2002). *Discrete Choice Methods with Simulation.* Cambridge University Press.

Train K. and McFadden D. (2000). *Mixed MNL models for discrete response*. *Journal of Applied Econometrics*, Vol.15, pp. 447-470.

TRANSIMS – Los Alamos national laboratory – http://transims.tsasa.lanl.gov

Ubbels, B. & Nijkamp, P. (2002). *Unconventional funding of urban transport*. Transportation Research D. Vol. **7**, pp. 317-329.

Van der Haegen, H. ed., (1982). West-European Settlement Systems. Acta Geographica Lovaniensia, vol. 22, 370 p.

Van der Haegen, H., Van Hecke, E., Juchtmans, G. (1991). *De Belgische stadsgewesten*. In: Statistische Studiën, nr. 104, 1996, p.5-42

Van der Laan, L. (1995). *Forensisme en megatrends op de arbeidsmarkt*. Geografie, jg 4, nr. 2, p.10-12.

Van der Vegt, C. en W.J.J. Manshanden, Steden en stadsgewesten: economische ontwikkelingen 1970-2015, SEO, Den Haag, 1996, 190 p.

Van der Wouden, R. en E. de Bruijne (2001). *De stad in de omtrek; problemen en perspectieven van de vier grootstedelijke gebieden in de Randstad*. Sociaal en Cultureel Planbureau, Den Haag

Van Dyck, B. (2003). *Brownfields Redevelopment: Governance and Partnerships*. Rotterdam, Erasmus University Rotterdom/EURICUR.

Van Riet, C. (1992). *Fundamenten, aard, omvang en ontwikkeling van het pendelverkeer met een toepassing op het Antwerpse stadsgewest*. UFSIA, Antwerpen, 112 p.

Van Wee, B. (2002). *Land use and transport: Research and policy challenges*. Journal of Transport Geography, Vol. 10, pp. 259-271.

Verbeeck, J. (2001). *Duurzame mobiliteit voor het stadsgewest Antwerpen*. Antwerpen Mobiel, on-line beschikbaar op: http://www.antwerpenmobiel.be/Nieuwsbrieven/NB3_02.htm

Verhetsel, A. (1998). *The impact of spatial versus economic measures in an urban transportation plan.* Computers, Environment and Urban Systems 22 (6), pp. 541-555

Verhetsel, A. (2001). *The impact of planning and infrastructure measures on rush hour congestion in Antwerp, Belgium.* Journal of Transport Geography 9, pp. 111-123

Vinterbo, S. & Øhrn, A. (2000). *Minimal approximate hitting sets and rule templates*. *International Journal of Approximate Reasoning*. Vol. **25** (2), pp. 123–143.

Walker, J. (2001). *Extended discrete choice models: integrated framework, flexible error structures and latent variables*. Submitted to the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the Degree of Philosophy in Transportation sytems.

Walker, J. (2002). *The mixed logit (or logit kernel) model: dispelling misconceptions of identification*. Transportation Research Record, 1805, 86-98.

Wilson, A.G. (1967). A statistical theory of spatial distribution models, Transportation Research 1, 253-269.

Wise, S. (2002). *GIS Basics*. London, Taylor and Francis.

Witlox, F. (2003). *Matisse: A relational expert system for industrial site selection*. Expert Systems with Applications. Vol. **24** (1), pp. 133-144.

Zadeh , L.A.(1965). *Fuzzy sets*. Information and Control. Vol. **8**, pp. 338-353.

Zipf, G.K. (1946). *The P_1P_2/D Hypothesis: On the Intercity Movement of Persons*. American Sociological Review, vol. II, pp.677-686.

Zwerts, E., and Nuyts E. (2001). *Onderzoek Verplaatsingsgedrag Stadsgewest Antwerpen (april 1999-april 2000, Deel 1: Methodologische analyse).* Provinciale Hogeschool Limburg, Departement Architectuur, 21 p.