SCIENTIFIC SUPPORT PLAN FOR A SUSTAINABLE DEVELOPMENT POLICY

Intermediary report - January 2003

SPATIAL ANALYSIS AND MODELLING BASED ON ACTIVITIES (SAMBA)
CP-41

FUNDP – RUG – UFSIA - UCL

SPSD II

PART 1

SUSTAINABLE PRODUCTION AND CONSUMPTION PATTERNS

- GENERAL ISSUES
- AGRO-FOOD
- ENERGY
- TRANSPORT

This research project is realised within the framework of the Scientific support plan for a sustainable development policy (SPSD II)

Part I "Sustainable production and consumption patterns"

The appendixes to this report are available at :
http://www.belspo.be (FEDRA)

Published in 2003 by the
Belgian Public Planning Service Science Policy
Rue de la Science 8 - Wetenschapsstraat
B-1000 Brussels
Belgium
Tel : 32/2/238.34.11 – Fax 32/2/230.59.12
http://www.belspo.be (FEDRA)

Contact person :
Mrs Aurore Delis (deli@belspo.be)
Tel : 02/238.37.61

Spatial Analysis and Modelling Based on Activities (SAMBA)

Scientific Report

Scientific support plan for a sustainable development policy (SPSD II)
Part I: “Sustainable consumption and production patterns“

1. Project title

Spatial **A**nalysis and **M**odelling **B**ased on **A**ctivities **(SAMBA)**

2. Introduction

2.1. Context and summary

This project, coordinated by TRG (Transportation Research Group, University of Namur) began on the 1st of November 2001 and will end on the 31st of October 2003. For this project, TRG collaborates with the Universities of Antwerp (UFSIA), Ghent (RUG) and Louvain-La-Neuve (UCL).

As the main purpose of this project is to integrate the spatial dimension in the description of the households' activity chains, a first useful and necessary step must be the *geocoding* of all trip data collected from the activity patterns described in the 1999 national survey MOBEL and other regional and local surveys (OVG 1999-2001). This description will allow the synthetic modelling of the population travels. The study of borders effects and fuzzy sets will help to understand the household decision making process in activity scheduling.

2.2. Objectives

The primary objective of this project is to assess the travel demand, for any reason, in Belgium with the help of the activity chains. Due to budget constraints, we won't be able to totally fulfil our purpose. Nevertheless, our project will go further in this direction by creating a synthetic population (see section 3). The activity patterns we are working with come from the 1999 national survey on the Belgian household's mobility (MOBEL) and from other regional (for the Flemish region) and local (for Antwerp, Ghent and Hasselt-Genk, Halle-Vilvoorde, Leuven, Mechelen and Aalst) surveys. To achieve our goal, we will examine the spatial aspects of the activity chains which until now have only been studied in their temporal components.

This project will be organised in three steps:

· *geocoding of all the information collected about activity locations*. The created databases will be treated by means of a Geographic Information System (GIS)

· *spatio-temporal modelling describing activity scheduling*. At this stage we will study if the models used for the temporal analysis of the activity chains (already adapted by TRG to Belgian workers and non-workers (Cirillo & Toint, 2001)) can be spread to a spatio-temporal study or if it turns out that other models are more suitable

· *creation of a synthetic population* i.e. an artificial one, based on variables previously defined. At this stage, starting from data collected for a Belgian household sample (and thus for individuals), we aim to create a population "close" to the real Belgian population.

The UCL is involved in the study of barrier effects on the travel behaviours whereas the RUG studies the contribution of the fuzzy sets techniques to the description of the decision making process around household activity scheduling. Finally UFSIA will work on the introduction of the spatial variables deduced from data files on e.g. land use and public transport networks.

2.3. Expected outcomes

This spatial analysis of trips will allow us to develop an - up to now - under investigated aspect of individual and household activity scheduling. It will contribute to a more realistic approach of activity-based travel patterns.

An achievement of this project would be an origin - destination matrix allowing us to deduce the trips involved in individual activity participating between every zone (e.g. the municipalities) and from there, the induced travel demand.

3. Detailed description of the scientific methodology

An important step is the specialisation of all the data collected from different surveys. The geocoded addresses will then be used to generate spatial variables which will be introduced in a model related to activity destination choice. Then, a synthetic population will be created to fill in a possible lack of exhaustive information in order to estimate travel demand of the Belgian population.

All SAMBA-project partners first studied literature about the *Activity Based Approach*.
Indeed in the present project the travel demand will be studied like as it is induced from the need for individuals and households to participate in different kinds of activities at different locations.

3.1 Literature study

In the field of travel demand research, various methods have been developed to predict travel decisions subject to certain constraints (e.g. costs, demographic composition of the population, external changes). In general, two dominant methods can be defined: the trip-based four-step modelling approach and the activity-based approach. The four-step modelling approach emerged in a number of early metropolitan transportation studies conducted in the U.S.A. in the mid to late 1950's (Pas, 1996). It was the primary tool for forecasting future demand and performance of transportation systems (McNally, 2000). Four-step models have a logical sequential basis: the four steps or stages in the model are *trip generation, trip distribution, modal split* and *assignment*.
The main criticism on the trip-based approaches is threefold: (i) structural absence of feedback; (ii) lack of the spatial and temporal interconnectivity in household travel behaviour; and (iii) no link with activity participation (McNally, 2000). The latter criticism is very crucial, because it is widely accepted that personal travel is derived from the need to participate in outdoor activities (Doherty, 2002). This realisation has lead to a shift from trip-based to activity-based approaches in the last two decades (Kitamura, 1988). This new approach in transport demand modelling, first denoted as the Transport Studies Unit (TSU) paradigm, was developed in the mid-70's at Oxford University (Jones, *et al.*, 1983). Today, it is better known as the activity-based modelling paradigm (Ettema & Timmermans, 1997).

The main idea is that you first need to understand activity behaviour to be able to 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 (Pas, 1996).

As previously stated, two definitions of an activity-based approach are used. A more general definition by Jones *et al.* (1990, p. 34) stating that *"the activity-based approach shares a common philosophical perspective whereby the conventional approach to the study of travel behaviour … is replaced by a richer, more holistic, framework in which travel is analysed as daily or multi-day patterns of behaviour, related to and derived from differences in lifestyles and activity participation among the population"*. A more elaborated definition by Ettema and Timmermans (1997, p. xv) states that *"activity-based approach describes which activities people pursue, at what locations, at what times and how these activities are scheduled, given the locations and attributes of potential destinations, the state of the transportation network, aspects of the institutional context, and their personal and household characteristics."*

Bates (2000) pointed out that spatial separation is the essence of travel demand. It is clear that there is a variation in transport demand over time, but there are also spatial implications. Not much research is conducted on the geographical dimension of travel demand. In a recent overview of space-time approaches to travel behaviour, Timmermans *et al.* (2002) stated that spatial variables have not been examined in studies on activity duration and time allocation. There is still no definitive conclusion whether spatial characteristics are influencing time use and activity patterns. Such a context leads to understand why our project aims to further investigate the spatial dimension of activity chains.

Travel demand analysis is intrinsically spatial. Bhat and Zhao (2002) highlight the need to accommodate spatial issues in travel modelling and advocate a spatial analysis of activity stop generation. When the spatial dimension is part of the modelling approach, the planning area consists of only some zones, each of which represents an aggregate spatial unit from where trips are produced and to where trips are attracted. One of the characteristics of the SAMBA-project is to examine the locations at a very detailed and disaggregated level allowing more accurate study of the spatial factors.

3.2. Data geocoding

The SAMBA project has a spatial focus and therefore the location of trips on a digital reference map is one of the most important steps (*geocoding process*). The result will allow us to highlight some spatial characteristics, especially in combination with other geographical layers, e.g. transportation networks or land use.

3.2.1 Geocoding process

***Geocoding** (also commonly known as **address matching**) is the process of creating geometric representations for descriptions of locations. A **geocoding service** defines a process for converting alphanumeric description of locations into geometric shapes*[1].

More commonly, the geocoding services allow matching between addresses collected from surveys to their corresponding coordinates on a digital reference map. Different geocoding levels exist: the coordinates can be attributed to zip code level, to street name level or to house number level.

The geocoding process is threefold: (i) the geocoding service standardizes the address; (ii) the geocoding service searches the geocoding reference data to find potential candidates and each potential candidate is assigned a score based on how closely it matches the address; and (iii) the address is matched to the candidate with the best score.

The **standardisation** consists on dividing the address into its components. For example, the address “12 Albertstraat 8370 Blankenberge” is standardized as the street number, “12”; the address name, “Albert”; the street type, ”st” (straat); and the zone code, “8370”. The municipality name is not considered. Unfortunately, the standardisation process provided by the commercial geocoding systems is based on the American address structure. The Flemish addresses can be easily geocoded (since their structure is close to the US one). On the other hand, we expected more difficulties with addresses in French due to their totally different structure (the street type is firstly indicated while for Flemish addresses it appears at the end of the street name: **boulevard** Ernest Mélot against Albert**straat**). TRG was particularly confronted with this problem. The TRG team tried to develop another methodology by

[1] *Using ArcMap*, Geocoding addresses p.411, ESRI, ArcMap, 2002

implementing a program able to manage character comparison. Unfortunately, the method appeared too complex and too much time would be needed to achieve it. Therefore, the TRG team has chosen to use manual geocoding when such a problem occurred.

An important geocoding service setting is the **spelling sensitivity**. It controls how much variation the geocoding service will allow when it searches for potential candidates in the reference data. The lower the sensibility, the more spelling errors are expected in the address name. Finally, the last interesting setting is the **parameter to specify the position** for a geocoded address along the reference feature (reference street arc). Since house numbers are available in the MOBEL travel data, addresses were geocoded at this more precise level by TRG. These geocoded addresses are then expected to be correctly positioned. The others teams geocoded addresses as located at the middle of the reference feature since OVG and local surveys don't have addresses at the house number level.

3.2.2 Data collection

The geocodable data were collected from the 1999 National Survey MOBEL and from regional (for the Flemish region) and local (for Antwerp, Ghent and Hasselt-Genk, Halle-Vilvoorde, Leuven, Mechelen and Aalst) surveys.

⇒ The 1999 Belgian National Travel Survey contains 21096 trips, which represent 26474 addresses to be geocoded (see table 3.1). The project was coordinated and supervised by a consortium led by TRG, who was also in charge of building and correcting the databases. The survey was conducted by Dimarso and funded by OSTC.

⇒ The Flemish Travel Behaviour Research-project (OVG 1999-2001) consists of eight data sets: seven at the level of urban districts and one at the Flemish level. The survey was conducted by Dimarso, while the Provincial College of Limburg (PHL) was responsible for the supervision and control on the data collection.
The data set of Antwerp contains 29 778 trips. For the region of Mechelen we have 37 956 trips. For the Ghent region, a data set of 35 878 trips is available. The Hasselt-Genk data set includes 37 976 trips. The Halle-Vilvoorde data set contains 34 910 trips. The data set for the urban region of Leuven holds 37 167 trips, and the one for the Aalst region 33 857 (see table 3.2).

Table 3.1 - Data from MOBEL

	Belgium
Year of survey	1999
Number of households	3063
Number of persons	7037
Number of trips	21 096

Table 3.2 - Data overview of OVG

URBAN DISTRICT	ANTWERPEN	GENT	HASSELT-GENK	LEUVEN	AALST	BRUSSELSE RAND	MECHELEN
Year of survey	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**

The collected data can be grouped into three broad categories (Toint & Cirillo, 2001):

1) background information on the household (e.g. household structure, household location, household resources, i.e. means of transport and their use)
2) background information on individuals (e.g. year of birth, gender, education level, driving licence)
3) description of each trip including information on associated activity (e.g. street name, zip code, hour of departure and arrival, distance, time, means of transport, purpose, etc.)

3.2.3 Data preparation

Before being suitable for automatic geocoding through a GIS software, the addresses required corrections to increase the matching score (or success). Indeed, addresses files presented a large amount of different and particular errors.

The spelling of street names or the zip codes were often incorrect or only a characterisation of the destination (*alias name*) was given instead of an address (for example when we found "CLINIQUE ST LUC" in the street name record and "BOUGE" in the municipality record, we had to search the hospital's address in the given locality). These errors could be fixed by using search engines on the Internet or by consulting street atlas maps (e.g. De Rouck). All teams were involved in such procedures.

TRG mainly worked on alias names correction and "abbreviation expansion" before treating addresses with GIS, while the UFSIA, RUG and UCL teams checked the complete structure of the address record before submitting it to the geocoding service. Most errors contained in the MOBEL databases were therefore corrected during the geocoding process by automatic or manual selection of the appropriate candidate.

Besides errors related to spatial dimension, UFSIA and RUG teams were confronted with some other errors: errors occurring in chaining trips, rounding of arrival and departure times and distances, etc. These errors could cause biased results in activity chains study. Hence, they also focused on how to correct these kinds of errors. We refer to the next section (section 4) for more details and results. For the MOBEL data set, these corrections have already been applied by TRG on activity chains data during previous studies on temporal activity scheduling (cf. Cirillo and Toint, 2001) and do not need to be treated during this project.

Data correction before or during GIS geocoding process is really time-consuming. But it allows a better overview and insight in the data. Moreover the geocoding step was the first important step to reach a spatial understanding of trips and it was essential to spend a lot of time to achieve it.

3.2.4 Geocoding result

At present, TRG has geocoded all addresses included in the MOBEL survey, while UFSIA has geocoded two activity-based data sets (Antwerpen and Mechelen) and RUG geocoded addresses from the Ghent and Hasselt-Genk database. The success rate for the geocoding of trips is shown in table 3.3. UCL has begun the Halle-Vilvoorde data geocoding (using a methodology similar to the one developed by UFSIA and RUG teams), but results are not yet available (May 2003).

Table 3.3 Geocoding result

	Antwerpen	Mechelen	Gent	Hasselt-Genk	Mobel-belgium
Number of trips	29 778	37 976	35 878	37 976	21 096
Trips successfully geocoded	25 338	33 088	29 851	31 147	15 804
Success (%)	85.1	87.2	83.2	82	75
Geocoded by	UFSIA	UFSIA	RUG	RUG	TRG

The difference in the success rate between MOBEL and OVG data come from the fact that a *MOBEL-trip* has been considered as well geocoded if and only if the origin and destination addresses were **both** successfully geocoded, while in the data of Antwerpen, Mechelen, Gent and Hasselt-Genk, a trip was considered as successfully geocoded when the destination was geocoded. You can find detailed results in section 4.

3.3. Modelling aspects

3.3.1. Discrete choice modelling (TRG)

One of our main objectives for this project is to build up tools in order to elaborate an integrated model based on activities. Such a model should describe the activity-travel patterns of travellers with emphasis on the effects of spatial, temporal and individual and household characteristics. We will base our methodology on Bhat & Singh's research (2000). This approach has already been used to create a temporal model of the travel-activity patterns for Belgian workers and non-workers applied to the 1999 national survey MOBEL (Cirillo & Toint, 2001). A spatial analysis in the study of activities scheduling is an **original process**. It seems to us that developing a destination choice model is a first step in the direction of spatial studies. Indeed, the expected spatial analysis finds its place in the study of the individual destination choice when he/she wants to participate in an activity. Moreover, destination choice models are less developed than mode choice models (Simma, Schlich & Axhausen, 2002). So it could be a real opportunity for us to make progress in that way. Finally, this model could be added later on to a more integrated model which describes the generation of activity-travel patterns.

Hence, our research turns towards the study of the literature related to destination choice modelling and therefore towards the study of discrete choice modelling. Our purpose at this stage of the project is to assess models describing destination choice for different activity types: according to discrete choice theory, the individual chooses among a set of alternatives (zones defined as activity locations) the one which is likely to offer him the highest utility (or satisfaction). Assumptions have to be made on the definition (or generation) of the choice set (definition, size and number of alternatives) but also on the selection of variables used in the model.

The model of Bhat and Singh already integrates a kind of destination choice model. Their model on the scheduling of activities is subdivided into two levels:

I. Modelling the pattern and tour level attributes (*tour, mode, number of stops*)
II. Modelling the stop-level attributes conditional on I. (in each tour for each period of time)
   - *Activity type, duration, travel time to stop* (1)

– *Location of stops* (destination choice model) conditional on (1) (2)

The destination choice model developed in (2) is a spatial disaggregated (individual – household level) destination choice model which predicts the individual choice of travel to aggregated zone (or spatial clusters) based on zonal attributes (travel impedance characteristics, attraction-end attributes, trips characteristics) and individual and household characteristics. It was developed by Bhat, Govindarajan & Pulugurta (1998). The term *attraction-end* stands for activity locations out of home. This model is the main reference in destination choice modelling and therefore will be the basis of our approach.

Based on reviewed literature the first step of our disaggregated destination choice model consists in defining the choice set. The alternatives would be the smallest geographical units for which spatial information could be collected. We already know that Belgian data are available at municipality level but we could consider working on statistic sectors[1] if it's possible. The number of alternatives could be determined by adding to the actually chosen alternative, a random sample of non-chosen alternatives (see Pozsgay and Bhat). But it could also be restricted to alternatives which are included in the household individual ***action space***. Indeed an original fact consists on examining activity destinations in a **spatio–temporal** context. To our knowledge, there is a lack of investigation in linking spatial and temporal constraints in destination choice models. According to Dijst & Vidakovic (1997), we plan to determine an ***action space*** which will be the area containing every reachable activity places, subjected to a set of temporal and spatial conditions (types, location of activity, available time interval, activity time consuming, travel speed, and travel time). As we will develop different models for some particular motives (shopping, leisure), we could restrict the choice set to locations more likely to accommodate such activity types.

The second main problem is the collection of spatial information. The more commonly used variables are impedance variables (in-vehicle travel time, out-of-vehicle travel time and travel cost), level of employment (retail and service employment), and area by land use type (airport, industrial, parking…) (Kockelman (1996), Simma, Schlich & Axhausen (2002) and Bhat, Govindarajan & Pulugurta (1998)). It seems important to us to also consider socio-demographic characteristics of individuals and households in our model. Indeed, we know from previous studies that socio-demographic characteristics play an important role in travel behaviour. This aspect was also shown in the paper *"Spatial Analysis and Modelling Based on Activities: A Pilot Study for Antwerp and Ghent"*. In this analysis, factors that determine activities have been examined. We can conclude that gender, age and income are three important socio-demographic variables determining activity patterns. Moreover, we could consider the interaction effects of these variables with zonal attributes and impedance variables (Bhat, Govindarajan & Pulugurta (1998)). Finally according to distance decay studies, a variable like the distance could be introduced in our model. The UFSIA, RUG and UCL teams will spend some time to generate variables to be used in the model which will be developed by the TRG team.

Previously developed destination choice models are trip-based while in our project travel is studied like induced from activity participating. Therefore we need to introduce activity-based variables like the location, duration, travel time and type of the previously scheduled activity in the original model we plan to develop in the SAMBA project.

[1] Smallest administrative areas considered in census data

3.3.2. Spatial variables generation (UCL, UFSIA and RUG)

The Universities of Antwerp, Ghent and Louvain-La-Neuve will introduce spatial variables to be used by TRG to create the model. They are mainly working on Antwerp and Ghent area.

(a) Descriptive analysis of OVG data (UFSIA, RUG)

This analysis of OVG activity chains is an initial step in the research of fundamental travel characteristics on which the spatial modelling can be built. Observations and conclusions are presented in the next section (section 4).

(b) Dynamics in urban regions (UFSIA, RUG)

At this moment, the UFSIA and RUG teams are working on the paper "Dynamics in urban regions". The aim of this paper is to get more insight in the spatial activity patterns of people living in different urban environments. Variables like "***zone of destination***" and "***zone of the household location***" are introduced. The first one indicates whether the trips end either in the city centre, the 19th century neighbourhoods, the suburbs or the urban fringe of Antwerp/Ghent or outside the Antwerp/Ghent metropolitan area. The second variable tells us in which metropolitan area the respondents live. The geocoded addresses are assigned to these defined zones. These variables will be used in further modelling. More information is available in section 4.

(c) Distance decay in activity chains and barrier effects (UFSIA, UCL)

By studying distance decay, the UFSIA and UCL teams are making a first exercise in modelling a spatial variable: ***distance.*** The constructed exponential model shows how the distance determines travel behaviour, particularly how sensitive respondents were in travelling longer distances for different trip purposes, modes, etc.

3.4. Synthetic population (TRG)

From actual Belgian census data, we can only derive an origin-destination matrix based on home-work/school trips. Moreover this kind of census is expected to disappear in our country. Therefore we have to find a method to estimate the Belgian population travel demand (for any kind of purpose). At this stage the only source of information would be data collected from surveys but on a sample of the population. We should be able to generate from these samples the information about the whole population. The method consists in creating *synthetic daily activity-travel patterns*. The method is threefold: the first step consists in building a synthetic baseline population (individuals belonging to defined household types synthetically close to the real population), then locating them on the transport network at a housing location. The second step is the assignment of an activity chain (from collected survey data) to each synthetic individual of the household. Finally locations will be attributed to each activity. Therefore, an origin-destination matrix could be built allowing the estimation of trips between zones of our study area and thus the induced travel demand.

Our aim in the SAMBA-project will be in first instance the creation of a baseline synthetic population (at Belgian or regional level). This research seems to be still an open issue as the studied literature indicates only one application: the *TRANSIMS (**TR**ansportation **AN**alysis **SIM**ulation **S**ystem) project* (Los Alamos Laboratory, USA).

A short description of the building process for a synthetic population in this case is presented here.

TRANSIMS is an integrated system of travel forecasting models which is designed to provide correct transportation planners, complete information on traffic impacts, congestion and pollution. So the TRANSIMS's microsimulator database contains every movement of individual travellers between activities at different points. Therefore a *synthetic population* is needed to represent all the households and persons in the studied area. The synthetic population will include the **demographics** and the **residential and activity locations** of the households and persons studied. All individuals participating in an activity at one location and taking part later on to another activity at an extra location are included in the TRANSIMS travel file. The activities and their locations are often driven by household and personal demographics. Thus not only individual households will be generated but also household and individual demographics will be estimated.

In the United States, the census data are distributed in geographical regions of different size (census block < census block group < census tract < census). Those concepts will be useful to define the different files used to create the synthetic population. These files are:

⇒ The ***Census Standard Tape File 3A*** **(STF-3A)** is a collection of summary tables of demographics, such as the number of persons per household, for census tract or census block group sized area. Most tables in STF-3A summarize one demographic characteristic.

⇒ The ***PUMS*** file consists of a 5% representative sample of almost complete census records (addresses and other unique identifiers are missing) from those contained in a collection of census tracts or other small geographic census areas, which collectively is called a Public Use Micro Area (PUMA). A PUMA is constructed so that it contains approximately 100 000 individuals. Each census block group (or census tracts) assembled in a PUMA must have the same correlation structure (to avoid inconsistencies).

The aim is to estimate a multiway table of demographics for a PUMA. As stated before, the PUMA is a collection of census tracts or block groups. Multiway summary tables are not available in STF-3A (containing only partial population information) and moreover marginal totals included in the multiway table for the PUMA have to match those included in STF-3A. Therefore the ***Iterative Proportional Fitting (IPF)*** technique needs to be used in order to estimate the expected table by means of PUMS data (since complete demographic data are available for each person and household in the sample) and STF-3A marginal totals. An example to illustrate the IPF method appears in appendix.

The demographics data from which the synthetic population will be created are selected from the STF-3A, and all the variables which describe travel behaviour are chosen. Therefore variables like age of householder, gender, vehicle ownership, etc. are indicated.

As soon as the multiway table is estimated, the synthetic population of households will be constructed with the selection of the whole households from the PUMS according to proportions given by the estimated table.

An Activity Location File is required here after which each household in the population is located at an activity location. This special file includes all the characteristics about the locations of the activities. By the way, all the activities are characterised by their land use. Thanks to these variables, it is possible to affect a weight to each activity location. This factor represents the relative likelihood of a housing unit being placed there, and then the synthetic household are located to residence location.

When the synthetic population is created, we can allocate each synthetic household to an activity pattern. At this moment, an activity list should be computed for every traveller from survey and network data. We will find in those lists the activity type, start time, stop time, travel time to the activity, the travel mode and a list of other participants. So an activity list will be built for each synthetic household individual with the help of TRANSIMS Activity Generator module. The main steps are the following: (i) from household activity survey, skeletal activity patterns are created by stripping locations from survey and organising via trips; (ii) the synthetic households are then matched to survey households on selected demographics; (iii) they are given an activity pattern of one of the survey households to which they were matched; (iv) the final step consists on locating the household on activity locations. Locations are selected by following two steps: firstly, a discrete choice model is used to select appropriate zones for all activities; and, secondly, land-use variables and employment data are used to find specific activity locations within zones for each activity.

From this, we can say that the "static" synthetic population is set to interact with other TRANSIMS modules to become "dynamic", but this new construction could be the subject of another research purpose.

3.5. Fuzzy logic (RUG)

Besides some other fuzzy set theories, the theory of rough sets seems to contain some appropriate possibilities for activity based modelling. The rough set approach is essentially a decision support tool from operations research, which tries to formulate decision rules of an 'if-then' nature. This approach is useful in cases of ambiguous data and when data is not convenient for traditional, ordinary statistical methods. The creation of models of the most representative objects for particular decision classes or reduction of all redundant objects and attributes in the information system so as to get the minimum subset of attributes ensuring a satisfactory approximation of objects' classification are just two of the possible areas of application[1].

4. Detailed description of the intermediary results, preliminary conclusions

4.1. Geocoding result

UFSIA, RUG and UCL teams worked together on a common methodology to correct and geocode OVG-data.

⇒ UFSIA

UFSIA teams used the MapInfo GIS software and the digital street map of Flanders delivered by OC-GIS Vlaanderen was used as reference map.

For 25% of the cases in Antwerpen, the addresses or purposes have been adjusted or corrected. Often the spelling of street names or the zip codes were incorrect or only a description of the destination was given. These errors could be adjusted by using search engines on the Internet. Inconsistencies appeared for example when the purpose 'going home'

[1] More explanations are available on the following website: http://www-idss.cs.put.poznan.pl/staff/slowinski/main.html.

was related to an address not corresponding with the home address. Other errors are blanks for street name, zip code or purpose. All addresses abroad (outside Belgium) are not included in this study. An overview of the frequencies of occurring errors is given in Table 4.1.

In sum, in both districts about 85% of all trips could be geocoded. This means that they can be used in our further spatial analysis. The remaining 15% that could not be located are not completely lost as they still contain a lot of useful information.

Table 4.1. Error codes Antwerpen and Mechelen

Code	Antwerpen			Mechelen		
	Freq	%		Freq	%	
Correct data	17 897	60,1%	Geocoded: 85,1%	26 381	69,5%	Geocoded: 87,2%
Corrected or adjusted data	7441	25,0%		6707	17,7%	
Description of the street (e.g. Supermarket 'X', Firm 'Y', etc.), but not locatable. Purpose and zip code all right	575	1,9%	Not geocoded: 14,9%	710	1,9%	Not geocoded: 12,8%
Zip code and/or street not locatable or missing. Purpose all right	1973	6,6%		2408	6,3%	
Purpose not correct, but zip code or street all right	87	0,3%		517	1,3%	
Locatable, but purpose missing	234	0,8%		143	0,4%	
Too much data missing	970	3,3%		783	2,1%	
Purpose 'going home' is inconsistent with indicated address	482	1,6%		159	0,4%	
Trips abroad	119	0,4%		148	0,4%	
Total	29 779	100,0%		37956	100,0%	

⇒ **RUG**

RUG used the ArcView GIS 3.2a software for geocoding. The digital street map of Flanders delivered by OC-GIS Vlaanderen was used as reference map.

In more than 30% of the cases the addresses or purposes could have been adjusted or corrected. An overview of the frequencies of occurring errors is given in Table 4.2.

Table 4.2: Different type of 'geocoding errors' for OVG Gent and OVG Hasselt-Genk

Code	Gent			Hasselt-Genk		
	Freq	%		Freq	%	
Correct data	17 325	48,3%	Geocoded: 83,20%	19536	51,4%	Geocoded: 82,00%
Corrected or adjusted data	12 526	34,9%		11611	30,6%	
Description of the street (e.g. Supermarket 'X', Firm 'Y', etc.), but not locatable. Purpose and zip code all right	3431	9,6%	Not geocoded: 16,80%	1487	3,9%	Not geocoded: 18,00%
Zip code and/or street not locatable or missing. Purpose all right	1210	3,4%		451	1,2%	
Purpose not correct, but zip code or street all right	135	0,4%		1	0,0%	
Locatable, but purpose missing	238	0,7%		480	1,3%	
Too much data missing	141	0,4%		3455	9,1%	
Purpose 'going home' is inconsistent with indicated address	499	1,4%		628	1,7%	
Abroad	373	1,0%		324	0,9%	
Total	35 878	100%		37976	100%	

⇒ **TRG**

Our geocoding service is provided by ArcGis software (2002). Téléatlas, multinet Belgique, (2001) was used as reference map.

Geocoding results of addresses from MOBEL data:

86% addresses were successfully geocoded. The un-geocoded addresses (14%) contain error of street name, zip code or an *alias* not easily to locate is given instead of a street name.

Table 4.3 Geocoded addresses

Provinces	***Addresses Mobel***	***Matching***	***% Matching***	***Ungeocoded address***
Antwerpen	2138	1865	87.2	273
Limburg	1641	1482	90.3	159
Oost-Vlaanderen	1738	1623	93.4	115
West-Vlaanderen	2014	1659	82.4	355
Vlaams-Brabant	1896	1698	89.6	198
Bruxelles	6441	5834	90.6	607
Namur	1911	1685	88	226
Hainaut	2608	2124	81.4	484
Liège	2724	2298	84.4	426
Luxembourg	1380	1118	81	262
Brabant-Wallon	1646	1331	80.9	315
Other	337	/	/	337
Total	**26474**	**22717**	**86**	**3757**

The '*other*' class contains addresses abroad (not considered in our study) and addresses with (not resolved) zip code error.

<u>*Geocoding results of trips from MOBEL data:*</u>
3176 trips have no geocoded origin and 3286 have no geocoded destination. According to table 4.4 we can see that 15 804 trips (75%) were entirely geocoded against 1170 trips (5%) with geocoding failure and that 4 122 trips (20%) are partially geocoded.

We are still trying to correct street name or zip code errors contained in un-geocoded addresses. At this stage, we are expecting 80 to 85% of exploitable trips.

Table 4.4 geocoded trips

Destination / Origin	Bruxelles geocoded	Wallonie geocoded	Flandre geocoded	Ungeocoded Destination	Total
Bruxelles geocoded	*3 802*	*231*	*390*	478	**4 901**
Wallonie geocoded	*242*	*5 401*	*0*	1 044	**6 687**
Flandre geocoded	*306*	*17*	*5 415*	594	**6 332**
Ungeocoded Origin	455	981	570	1 170	**3 176**
Total	**4 805**	**6 630**	**6 375**	**3 286**	21 096

We can point out that percentages of trips successfully geocoded are quite smaller than the percentages of successes in geocoding addresses. As mentioned above, a trip is only considered as well geocoded if and only if the origin and destination addresses are successfully geocoded. Moreover, in the MOBEL survey, some addresses were not indicated by the respondents although a trip was made. Consequently, the corresponding trips are lost for the geocoding process.

<u>Encountered problems during geocoding process:</u>

- Standardisation error for addresses in French

The automatic standardisation of the geocoding service does not recognise addresses in French structure. Hence we tried to modify the structure of our reference table as well as the one of our data to recover some geocoded addresses. But it was not sufficient. Then the un-geocoded addresses required interactive matching i.e. we carried out manual standardisation.

- Incomplete addresses

The geocoding services do not manage incomplete addresses (for example, the street name '*Boulevard Tirou*' cannot match because the complete street name is '*Boulevard Joseph Tirou*'). Once again, automatic geocoding is impossible; hence a research to the street atlas De Rouck allow us to introduce manually the appropriate standardisation.

- Postal code error

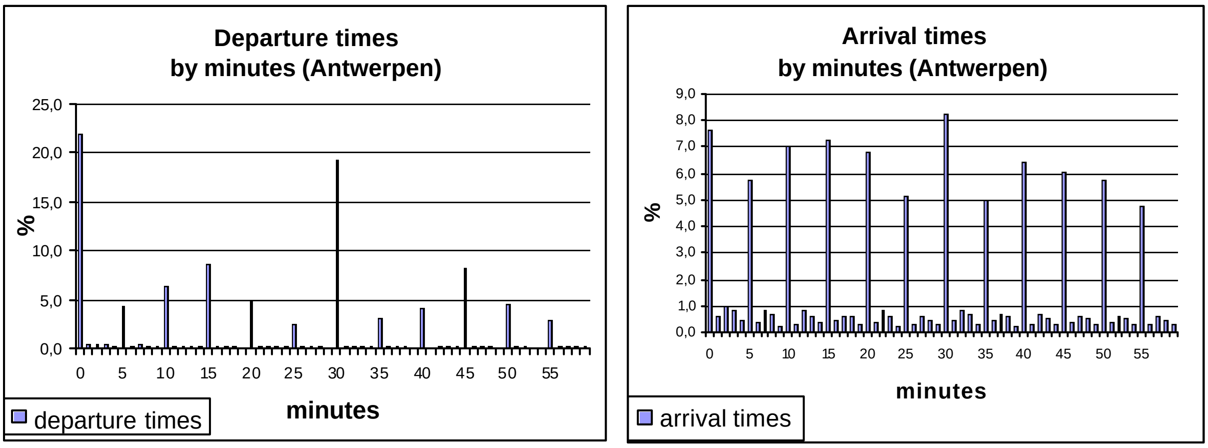
As postal code errors are not a priori known, they were only recognised during the geocoding process. When it occurs we used De Rouck atlas or internet search engine to find the exact address.

4.2. Rounding of arrival and departure times and distances (UFSIA, RUG and UCL)

Travel times are often used in research on travel behaviour. Here a potential problem may occur as pointed out by Rietveld (2001). Travel times are in most cases the outcome of subtracting reported departure times and arrival times. It is commonly known that people think and report in 'rounded' times, in fact in multiples of 5, 15 and 30 minutes. Since departure times appear to be rounded much more frequently than arrival times this will have some implications. In addition, the probability of rounding upward is much higher than rounding downward, which results in an overestimated travel time. This fact has a number of consequences because it affects the reliability of individual observations as well as the reliability of the districts' or regional averages, and can lead to erratic patterns in detecting or forecasting time bottlenecks in traffic. It is not easy to solve the rounding problem, but a useful method to counter the problem in travel behaviour analysis and modelling is computing the probability that the reported arrival time means the actual arrival time by means of a chi-square test.

Figure 4.1 depicts the frequencies of times in the survey of Antwerpen. Our findings are almost completely similar to those of the 1998 national transport survey in The Netherlands (Rietveld, 2001). For the urban district Antwerpen, about 90% of the departure times and about 75% of the arrival times are multiples of 5, while they only count for 20% of the possible clock times. Departure times are more rounded than arrival times. There is just one small, but remarkable anomaly in the reported arrival times in the Antwerpen data: the number of trips of 40 minutes is higher than the number of trips of 45 minutes, where we would expect it inversely.

Figure 4.1: Departure and arrival times by minutes (Antwerpen)



Not only travel times but also travel distances are rounded. However the rounding problem for distances is more complex than rounding of time. Figure 4.2 and 4.3 depict the frequencies of travel distances respectively for Ghent and Antwerpen.

Figure 4.2: Frequencies of the reported distances in % (Ghent)

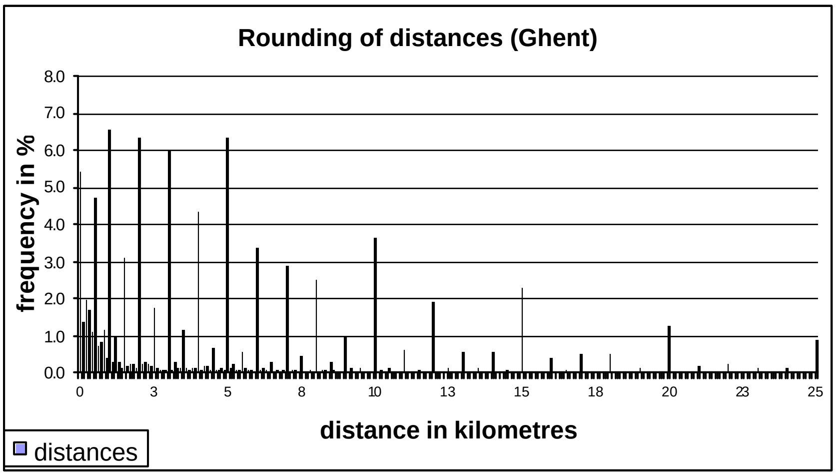


Figure 4.3: Frequencies of the reported distances in % (Antwerpen)

Unlike time, distance has no limited interval and since the accuracy of distance reporting depends on mode and/or distance itself, it shows a more diverse pattern. The higher the trip distance, the more rounded the distance reported. It is also obvious that reporting distances will be more accurate for e.g. car drivers, who can check the mileage counter, than for public transport users. In addition the accuracy depends on the repetition of trips. The probability to report exact distances is lower for once-only trips, except for special occasions, than for daily recurrent trips. Therefore exclusively basing on reported distances in the spatial dimension of travel modelling could lead to biased views.

Because the SAMBA-project works with geocoded data it is possible to measure the trip distance in a GIS. But we have no idea which path the respondents followed, which means we can only estimate the exact travelled distance under several assumptions and different methods are possible to measure the distance: crows flies distance, shortest path, fasted path, etc.

These calculated distances could be set out against the reported distances. Descriptive statistics can be presented to describe the characteristics of the distance data, including measures of location, dispersion and shape for the reported distances as well for the calculated distances. This can be done for different modes. In addition, reported distances can be set out against calculated distances in scatter plots and a regression and correlation analysis will be done.

Because time and distance are supposed to be closely related in travel, the results will be compared to time rounding errors. The impact of rounding will be investigated for distance and time, and their interwoven relationship analyzed. The outcome will allow us to trace the underlying mechanisms and influences involved in reporting distances in activity patterns.

Through this research, the main goals are first of all to get more insight in the accuracy of the travel behaviour data, and distance data in particular and in the second place to upgrade the reliability of the data. Therefore we try to determine how the rounding problem can be solved, and verify the possibility of cleaning the data by means of locating and/or repairing inconsistencies, errors and outliers in the data.

4.3. Tracing and eliminating errors and inconsistencies

Having already good insight in the data and some of the occurring errors and inconsistencies, UFSIA and RUG teams still have to trace more and complex inconsistencies and errors, i.e. missing data about the person, end time inconsistent with start time, first or last activity of the day is not equal to sleep, etc. Therefore they will try to co-operate with Eindhoven University of Technology to apply SYLVIA, a System for the Logical Verification and Inference of Activity Diaries, which is an interactive computer system to test the logical consistency and completeness of activity diaries (Arentze et al., 1999).

4.4. Spatial variables generation

- *Spatial analysis and modelling based on activities: a pilot study for Antwerpen and Ghent (UFSIA, RUG).*

This analysis provides an examination of the activity determinants. Socio-demographic variables like gender, age and income were revealed important activity patterns components. The different socio-demographic groups have a specific time-space pattern to participate in activities. Moreover the means of transport to reach the activity locations is very different for each group.

The most frequent trips that appear in the data set of Antwerpen are shopping, work, entertainment/sports/culture or school. The four activity groups mainly determine the activity pattern of most respondents. Often they are the central activities round which other activities are organised. During weekdays school and work are the main activities, but on Saturday and Sunday when shopping and leisure activities become more important no clear main activity can be indicated.

First shopping trips can be described. Most of the times they have a very short distance (shorter than 5 kilometres) and they do not have long trip duration. In cities as Antwerpen shops, supermarkets and hypermarkets can be found either in the city centre or in the suburbs, so this explains why 90% of the shopping trips cover less than 10 kilometres. In Antwerpen 30% of shopping trips have the city suburbs as destination. Here the big shopping complexes have their locations. Only 7% of the trips have its shopping destination outside the Antwerpen urban area. But although shopping trips seldom cover long distances, many respondents choose their car as transport mode probably because it is easy to carry shopping bags. Women and low-income earners are more likely to do shopping trips than men and high-income earners. It is striking that the elderly people are such devoted shoppers.

Further the second most frequent activity is work. If we examine the spatial part of the activity we can conclude that for work trips respondents are often willing to cover higher distances (about 13 kilometres on average). It explains that for Antwerpen 20% of the work trips has its activity place outside the Antwerpen urban area. The capital city Brussels for instance is located at approximately fifty kilometres from the metropolitan region of Antwerpen and it is an important work location of the survey respondents. In the Antwerpen trip data set almost 60% of the trips with a work purpose are made by car (as driver or passenger). Male respondents, heads of the household and high-income earners, have made most work trips.

Next the leisure activity including entertainment, sports and culture is treated. On average the trip distances are larger for leisure trips: there are many trips with a distance over 10 kilometres. For the Antwerpen data more than 40% of the leisure trips have its destination in the city suburbs or outside the urban area. We suggest that some sports cannot be practised in the city centre. There are also a significant number of tourist or holiday trips with an activity place outside the urban area. Since respondents often leave the city centre for their leisure activities the car (as a driver as well as a passenger) is a frequently used means of transport. Generally men, high-income earners, heads of the household, kids and young people do many trips with this purpose.

Finally for school trips we saw that respondents cover small distances. The average trip distance of a school trip is about 4 kilometres. In metropolitan regions as Antwerpen it is easy to see that many schools (primary and secondary schools as well as universities and colleges) can be found on a relative small area. For Antwerpen more than 65% of the school trips have its destination in the city centre. So respondents seldom have to leave the city to participate in these activities. Students most of times have no car on their own so they are obliged to use other means of transport, e.g. public transport (18%) or cars as a passenger (19%). But in most cases they count on their own strength using bicycle (28%) or they go on foot (25%).

- *Dynamics in urban regions. The intra-urban travel in Antwerpen and Ghent (paper in progress) (UFSIA, RUG).*

In the 70's H. Van der Haegen and M. Pattyn have defined the Belgian urban regions. A few years later they made some adjustments on the basis of the census of 1981 and taking into account the new administrative classification and the evolution at that moment. Seventeen urban regions were defined. From that moment on, the concept of 'urban region' received more and more recognition in scientific, administrative and economic environments. Also in different policies it is an often used geographical entity, like for example in urban and regional planning and the urban policy of the federal government.

According to H. Van Der Haegen et. al. (1996) the 'urban region' is the 'whole geographically extended structure within which the various basic activities of the urban community (particularly housing, work, education, shopping, culture and leisure activities) are principally situated'. There are intense relationships between the activities, so that a single functional entity comes into being. The urban region is a social, geographical and functional system.

The actual method of delimitation for the metropolitan areas is rather static/morphologic, where one sees morphology as the reflection of urban functions and relationships. By definition one relies especially on criteria like population density and type of housing. Only in the definition of 'urban fringe' and 'commuter residential areas' dynamic criteria, namely commuter travel and migration are taken into consideration. The dynamics are taken for granted without any form of empirical analysis. Dordregter already asked himself in 1971 whether we could still speak of a strong nodal attraction to the city centre or not. Not only the orientation to one central node has to be studied, but also the mutual relationships between every other urban area for different trip purposes. That is how the complementary functions of the different parts of the urban region can be discovered.

The data of the OVG of Antwerpen and Ghent permit to analyze the intra-urban dynamics by means of a study of the activity patterns as well as to study the attraction of the city centre. Flows inside or between metropolitan areas are an important part of our research. In the Belgian literature up to now not much is written on the link between travel behaviour and functions in the urban regions. Urban regions are at that time morphologically defined according to activities, but the question remains whether these activities actually follow this morphology or not.

The UFSIA and RUG teams have proceeded as follows: they went deeper on the subject of the origin of the concept 'urban region' and gave a description of the different metropolitan areas that later on will be used as research areas. Then, the possibilities and restrictions to perform this analysis on the level of the metropolitan areas have been considered and they study the intra- and inter-zonal trip behaviour for inhabitants of different urban areas with comparison between different transport modes, age groups, income groups to get better insight in the travel behaviour of households in an urban region.

❖ *Distance decay in activity chains analysis and barriers effects (paper in progress) (UFSIA, UCL).*

By the following preliminary analysis, we point out that distance and barrier effect are extremely linked. Indeed, the longer the distance between two places, the smaller the fluxes generated between these two points (Fotheringham & O'Kelly, 1989). The analysis of the distance decay functions in function of the number of trips will mainly deal with (1) the modelling of the decreasing number of trips with distance, and (2) the introduction of the differences between a trip and a chain of activities. This latter makes the exercise much more difficult. Let us remind here that - to our knowledge - there is up to now no or very little research conducted on distance decay, spatial barrier effects and modelling activity chains. We here first choose to simply concentrate on the estimation of the distance decay functions of the number of trips in the case of the modelling of activity chains (Taylor, 1975 et Grasland, 1999). The same analysis is conducted by the UFSIA team on regional level data set while UCL will work on MOBEL data (National level data set). We are using a common methodology. We do hope to end with a common paper and present it in two colloquiums in 2003.

Here is a summary of the methodology developed for this analysis:
The distance decay model is exponential: $y = be^{-ad}$,
where b and a are the parameters of the model, y the number of trips and d the distance.
The estimation of this model can be done in two different ways:

- directly, by using non-linear estimation techniques such as the maximum likelihood or the non-linear least squares;
- indirectly, by using a logarithmic transformation of the model leading to a classical linear model that is easily estimated by ordinary least-squares

In fact, we want to test if parameter *a* varies significantly with the type of trip maker, trip type, activity chain and trip purpose. We therefore used a maximum likelihood test. We then also analysed the graphical outputs (distance decay curves); this led to a discussion of the differences between the reported distances and the real ones and the over-estimations of the "5-distances". Reported distances are indeed usually rounded in multiples of five minutes/kilometres (see e.g. Rietveld et al. 1999). Graphs enable one to visualise the hierarchy of travels, from the most to the less sensitive to distance (see section 6.3).

In other word, is the friction of distance the same for everyone and for all types of trips? This question is very close to that of barrier effects. We can indeed expect that barriers (or some of them!) have a stronger or a different effect on some groups of individuals or trips. It is hence important to find which variables discriminate best the trips according to distance decay functions. These variables will hence be very useful for the estimation of barrier effects.
Variables are classified in three groups:

- Group 1: characteristics of the individual and of the household he/she belongs to (age, gender, socio-professional status, income, type of household (couple with or without children, family with one parent, etc.).
- Group 2: geographical characteristics (region, urbanisation level of the place of residence, etc.).
- Group 3: characteristics of the trip and the chain of activities (purpose, transportation mode, number of trips in the chain, etc.)

There is no problem for segmenting trips and chains for the two first groups of variables; this is not the case for group 3. Indeed, we here have to use trip chaining modelling techniques: what is the purpose of a trip when the chain of activities includes several different purposes? This question remains the same for all the other variables concerning the characteristics of the travel (transportation mode etc). We are here confronted to the well-known problem of the complexity of the chain of activities. This problem will be even more accurate for barrier effects modelling; we intend to here define an appropriate methodology. Some authors (see e.g. Hensher & Reyes 2000) suggest to simplify the chain of activities by – for instance – reducing the number of purposes or the number of transport modes and hence reducing the size of the chains (limiting the chains to 4 or 5 trips). In our case, this is not sufficient. We are still trying to find a solution to that problem, to find another definition of the chain, keeping in mind that they also have to be used in the modelling of the barrier effects.

In the present state of the research, we've already undertaken many evaluations of the MOBEL data. (These are impossible to include in the present report but we would be very happy to present them) and started to compared them to the UFSIA-team data. Most variables of the three groups defined previously were analysed; the model is validated. It is however still too early to conclude: there are still some statistical treatments to be done on the definition of some of the variables used; the control of each variable is not perfect. Moreover, in our distance decay analyses, space is supposed to be isotropic (no difference between directions). Example of application on distance decay for trip purpose is available in appendix.

5. Future prospects and future planning

Little work has still to be done on the data geocoding and data correction: the TRG team should geocode yet ungeocoded addresses from MOBEL data which contain many errors or are very incomplete. UCL will geocode the Halle-Vilvoorde and Leuven data sets with the help of the recently acquired GraphyMap. In the meantime, UFSIA and RUG teams will work on OVG-data inconsistencies.

Mainly, TRG will prepare a destination choice model based on discrete choice theory while UFSIA, RUG and UCL teams will progress in spatial variables analysis (e.g. introduction of land use variables, distance decay and barrier effects analysis, etc.).

The team of UFSIA already started to search for possible spatial variables that can be linked to travel data. At this moment, they are working on the introduction of some interesting spatial variables such as (1) the land use and land coverage characteristics of Flanders, (2) different density variables (e.g. population density,…) and (3) information on the public transport network (e.g. to detect the distance to public transport stops). Since it will be possible to combine these layers to the geocoded household locations and destinations, quite some spatial variables could be linked with the OVG-travel data in order to explain travel behaviour much more precisely afterwards. It is also possible to create 'mixed land use' variables. These variables explain the land use in a range of e.g. 1 kilometre from the destination or the household location (i.e. % of industrial area, % of housing area, etc.). This is interesting to know in order to study whether or not the travel behaviour in a very mixed land use area has other characteristics (such as distance, mode, etc.) than in a homogenous area.

When the analysis of the distance decay will be finished (January-February 2003), the UCL team will start with the barrier effect analysis. If this barrier effect can be seen as an extension of the distance decay, many other problems could occurred. Classical tools of geography will not enable us to measure the barrier effects in activity chains. These tools are indeed only valid for analysing trips (and not chains). They are mainly based on the analysis of the spatial distribution of the residuals of gravity models. Moreover, we here want to insist on the fact that the literature is lacking paper on activity chains models including spatial components. Two exceptions are the papers of Mc Nally (2000) and Dijst & Vidakovic (1997). They consider the "spatialisation" of the activities but they do not meet all our expectations. At this stage of our research, we intend to proceed by steps: the first step will identify barriers by type of trips at several spatial scales. We will therefore work on trips and not on chains. In the second step, we will introduce the detected border effects as explanatory variables in the activity chain models. We hence will be able (1) to test the effect on border on trip chaining and (2) measure their exact impact. At this stage of the analysis, we intend to use hierarchical logit models (Bowman & Ben-Akiva, 1997) that estimate the choice of the activity and the spatial destination.

Mid-February, a meeting will be organised in order to share the work already done in view of elaborating an integrated model around activities destination choice.

At the beginning of July, the TRG will start to work on the creation of a synthetic baseline population based on variables previously introduced by the others SAMBA project partners.

Finally, if the study of RUG concerning the possible introduction of fuzzy logic in activity based modelling turns out to be positive, some methods of fuzzy logic could be integrated to describe the decision making process in activity-travel scheduling.

6. Appendix

6.1. References

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.

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

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

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. & 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.

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

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

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

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

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.

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.

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 Applications*. Dordrecht : Kluwer Academic Publishers.

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

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

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.*

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

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

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., 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.

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

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. 35-52Nuyts, E. & Zwerts, E. (2001) *Onderzoek Verplaatsingsgedrag Vlaanderen en Stadsgewest Gent, januari 2000-januari 2001, Controle en begeleidingsopdracht*. Unpublished.

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

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

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

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.

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.

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

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

Van der Haegen, H., Van Hecke, E., Juchtmans, G., De Belgische stadsgewesten 1991, 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 Riet, C. (1992). Fundamenten, aard, omvang en ontwikkeling van het pendelverkeer met een toepassing op het Antwerpse stadsgewest. UFSIA, Antwerpen, 112 p.

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

6.2. Publications

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

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

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

- Paper in progress: Verhetsel A., Thomas I., Hammadou H. & Van Hofstraeten D. (2003), *'Distance decay in activity chains analysis. A Belgian case study'*.

6.3. Detailed results

➢ **An example to illustrate the IPF technique:**

Suppose that a PUMA consists of two census tracts.
Dem1 and Dem2 are household demographics.
The tables give the number of households corresponding to the demographic characteristics.

	Dem2=1	Dem2=2	Total
Dem1=1	?	?	1700
Dem1=2	?	?	1050
Total	1505	1245	2750

Table1 - Census tract 1 (STF-3A)

	Dem2=1	Dem2=2	Total
Dem1=1	?	?	1405
Dem1=2	?	?	905
Total	700	1610	2310

Table 2 - Census tract 2 (STF-3A)

First, the marginal tables for the two census tracts in the PUMA are added:

	Dem2=1	Dem2=2	Total
Dem1=1	?	?	3105
Dem1=2	?	?	1955
Total	2205	2855	5060

Table 3 - Census Tract 1 + Census Tract 2

The PUMS table (containing census tract 1 and 2 data) is the following:

	Dem2=1	Dem2=2	Total
Dem1=1	45	108	153
Dem1=2	63	37	100
Total	108	145	253

Table 4 -PUMS

Then a multiway table is estimated from PUMS by IPF against the marginal totals of table 3:

	Dem2=1	Dem2=2	Total
Dem1=1	949	2156	3105
Dem1=2	1256	699	1955
Total	2205	2855	5060

Estimated table for the PUMA

➢ **Example of application on distance decay for trip purpose**

First step: Estimation result on non-linear least squares for exponential model on distance decay.

	b	t value b	a (sensitivity)	t value a	r²
Bring	240,4	9,07	-0,130	-6,81	0,70
Home	1001,3	14,42	-0,141	-10,96	0,82
Work	229,3	14,41	-0,088	-10,65	0,78
School	196,0	8,87	-0,172	-6,85	0,75
Shopping	436,5	14,53	-0,157	-11,13	0,84
Prive	486,3	11,86	-0,177	-9,18	0,83
Leisure	271,6	9,81	-0,151	-7,44	0,73

Second step

1. Test of likelihood ratio

	Home	Work	School	Shopping	Prive	Leisure
Bring	0,88	17,13**	4,04*	4,39*	8,68**	1,23
Home		25,37**	2,04	1,38	4,72*	0,25
Work			20,17**	37,37**	39,03**	14,07**
School				1,14	0,08	0,99
Shopping					1,41	0,08
Prive						1,49

* significant 5%
** significant 1%

2. Graphical analysis