Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek
Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt

2011
2012

## FACULTY OF BUSINESS ECONOMICS

*Master of Management: Management Information Systems*

# Masterproef

*Identification of common characteristics in activity-travel behavior based on activity-travel diaries*

Promotor :
Prof.dr.ir Tom BELLEMANS

Michael Ceunen

*Master Thesis nominated to obtain the degree of Master of Management , specialization Management Information Systems*

2011
2012

# FACULTY OF BUSINESS ECONOMICS

*Master of Management: Management Information Systems*

# Masterproef

*Identification of common characteristics in activity-travel behavior based on activity-travel diaries*

Promotor :
Prof.dr.ir Tom BELLEMANS

Michael Ceunen

*Master Thesis nominated to obtain the degree of Master of Management , specialization Management Information Systems*

KNOWLEDGE IN ACTION

# Acknowledgements

This master thesis is the final step in my degree in Master of management: specialization Management Information Systems (MIS).  The realization of this thesis was an interesting but demanding job. Therefore I want to extend a word of gratitude to everybody who contributed to this thesis.

First, I want to thank my supervisor Tom Bellemans for his useful advice and professional opinion throughout the realization of this project. Furthermore, I would like to thank Feng Liu for her constructive feedback and close cooperation.

Finally, I want to thank my parents, friends and girlfriend for their help, support and distraction.

## Summary

The impact of traffic and transportation on the development and welfare of populations is increasing rapidly nowadays, making them important components of the economy. An effective policy concerning traffic and transportation issues is therefore needed. Traffic estimation and prediction models play an important role in developing this policy.

This master thesis aims at identifying different population segments in Flanders based on activity-travel behaviour. Furthermore, this study incorporates socio-demographic variables and the day of the week in the analysis. The ultimate goal of this research project is to predict an individual's activity-travel pattern based on his or her socio-demographic information and the day of the week.

In **chapter one** an introduction to the problem is given. The research objective of this master thesis is presented in **chapter two**. First, a general description of the problem is given. Next, the objective of this study is summarized using one central research question and several more specific research questions. The central research question of this study is: "Can different population segments be identified in Flanders based on activity-travel behaviour?".

A literature study is performed in **chapter three**. Different subjects are discussed in this chapter. First, the general concept of activity-based modeling is outlined. Second, scientific findings regarding the influence of socio-demographic variables on travel behaviour are summarized. Third, two statistical methods (classification and regression tree (CART) analysis and cluster analysis) are outlined. Fourth, the Transportation Analysis and Simulation System (TRANSIMS) is discussed.

In **chapter four** the analysis of this master thesis is presented. Before the actual analysis is performed, information is given concerning method, sample, measurements and software. Subsequently the analysis is outlined step by step. The different steps are based on TRANSIMS. In the first two steps, raw survey results are transformed into usable data. The third step encompasses the CART analysis. Multivariate regression tree (MRT) analysis is conducted to identify population segments based on travel behaviour. Socio-demographic variables and day of the week are also taken into account during the MRT

analysis. Various MRT analyses are performed using different settings to obtain the optimal tree.

**Chapter five** contains the discussion of this study. This chapter starts off with the interpretation and model performance. A comprehensive overview of the MRT analysis results is given here. Conclusions regarding the performance of the model in this study are also drawn in this section. Next, the constraints of this research project are discussed and recommendations for further research are given. In the final section of this chapter the general conclusion of this study is concisely presented.

# Contents

## List of figures

## List of tables

# 1. INTRODUCTION AND BACKGROUND

Traffic and transportation are very important components of the economy, impacting the development and welfare of populations. Various reports show that the importance of these two factors is increasing rapidly nowadays. This trend is fed by urbanization and globalization, which results in increased global trade and passenger traffic.

In Flanders, a significant increase in traffic has been observed as well. Since 1970 the total road traffic has more than tripled (Statistics Belgium, 2010). This expansion increased the impact of major problems around transportation. Issues vary from congestions, traffic-jams, poorly maintained roads, large infrastructure investments, overcrowding, accumulation of fine particles and increasing $CO_2$ (Proost et al., 2011).

This trend is likely to continue in the future (MORA, 2009). Until 2020 the number of inhabitants in Flanders will increase with 40.000 every year. This would result in an additional 15 million direct traffic kilometres a day. Furthermore, this population growth will indirectly stimulate freight traffic.

Traffic estimation and prediction models play an important role in the development of an effective policy concerning traffic and transportation issues. They have the potential of improving traffic conditions and reduce travel problems by facilitating better utilization of available capacity. The outcomes of these models can help achieve better short and long term-decisions concerning transportation and traffic.

Activity-based models are one type of models that can contribute to reasonable policy predictions and decisions regarding traffic and transportation. Activity-based travel analysis attempts to better understand the behavioural basis for individual decisions regarding participation in activities in certain places at given times (Bhat & Koppelman, 1999).

# 2. RESEARCH OBJECTIVE

As stated above, estimation and prediction models play an important role in policy predictions and decisions relating to traffic and transportation. Different types of models exist. Such models are based on data of current and past traffic. This research project is based on data collected by a survey.

During a large scale survey of the Flemish population, detailed activity-travel diaries were collected. A sample was taken from 2500 Flemish households and their family members. For seven consecutive days, a diary was filled in for each trip an individual in a household made. Besides trip variables, this survey also collected socio-demographic and other information about the participants and performed trips. The results from this study form the basis for this thesis.

This study aims at revealing different segments of the considered population based on activity-travel behaviour. Every identified segment will have a specific activity-travel pattern which describes travel behaviour within that segment. This study can be seen as the first step towards an operational estimation and prediction model.

Other variables will also be taken into account. Socio-demographic variables will be used to predict travel behaviour and describe segments. Additionally, this study will investigate which other variables play a significant role in travel behaviour of the inhabitant of Flanders.

The analysis will be preceded by a literature review. This review will first shed some light on activity-based models. Next, current findings on activity-travel behaviour will be reviewed. Third, the statistical methods for exploring and analyzing the data will be explained. Fourth, the Transportation Analysis and Simulation System (TRANSIMS) is discussed.

In general the objective of this research project can be summarized in the following central research question and specific research questions.

## 2.1 Central research question

Can different population segments be identified in Flanders based on activity-travel behaviour?

2.2 Specific research questions

- According to the literature, how do socio-demographic variables influence travel behaviour?
- Can we identify common characteristics in activity-travel behaviour in Flanders?
- What is the influence of socio-demographic variables on travel behaviour in Flanders?
- Which other variables have an influence on travel behaviour and what is the influence of these variables?

# 3. LITERATURE STUDY

## 3.1 Activity-based modelling

Activity-based transportation models are one type of models that can contribute to reasonable policy predictions and decisions. Since the late 1970s, extensive activity-travel research has been undertaken (Buliung & Kanaroglou, 2007). Before these methods became popular, trip-based methods were used as an approach to travel demand analysis. Clearly, there has been a distinct paradigm shift towards activity-based models in the past couple decades.

Activity-based travel analysis attempts to better understand the behavioural basis for individual decisions regarding participation in activities in certain places at given times (Bhat & Koppelman, 1999). The activity-based approach views travel as a derived demand. This demand is based on the need to pursue activities distributed in space (Recker, 1995). The basic travel unit in activity-based modelling is a tour. A tour can be seen as a sequence of trips starting and ending at home (Shiftan et al., 2003). The ultimate goal of travel behaviour analysis is to gain a full understanding of why people travel and to develop quantitative capabilities that facilitate the prediction of travel behaviour (Kitamura & Fujii, 1996).

### *3.1.1 Advantages of activity-based modelling*

Activity-based modelling has become a very important method for analyzing travel and transportation. It provides several distinct advantages in comparison to other models (Shiftan & Suhrbrier, 2002).

First, travel demand management measures and other types of transportation policies have a direct impact on travellers. Activity-based modelling is able to give a better understanding of this effect. This in turn can lead to a better prediction of travel and emissions impact.

Second, activity-based modelling is also able to consider secondary effects of travel demand management. For example, a commuter receives a subsidy for using the train as mode of travel. As a result he switches from driving alone to work to using the train. This switch is the primary effect. However, that person is no longer able to buy groceries on the way home. As a result he has to take the car when he gets home to do groceries. This is the secondary effect.

Third, activity-based models permit us to consider induced travel and the generation of new trips and trade-offs among different travel behaviour decisions as a result of transportation changes. This advantage is mainly realised in the disaggregate activity-based model.

## 3.2 Factors influencing travel behaviour

The how, where and why of travel behaviour is influenced by many factors. The most important personal factors are the needs, preferences, prejudices and habits of individuals and the households they live in. Other essential factors are the cultural and social norms of the community and the travel service characteristics of the surrounding environment (Bhat & Koppelman, 1999). These factors can be divided into four distinct categories (Curtis & Perkins, 2006). These are: urban form, socio-demographic variables, psycho-social variables and pricing.

Although different categories of factors affect travel behaviour of individuals, this research project mainly focuses on socio-demographic variables. A considerable amount of research on the link between socio-demographic variables and travel behaviour has already been conducted. Significant relationships between travel behaviour and variables such as age, gender, household composition, income, etc. have been identified. Even though these studies were conducted in other regions, the results can give a good indication of what to expect from the study results.

### *3.2.1 Age and travel behaviour*

In 2005 the influence of age on travel patterns was studied in Canada (Newbold, Scott, Spinney, Kanaroglou, & Páez, 2005). Data for this study was provided by the General Social Survey (GSS) of Canada. The researchers concluded that older Canadians make less daily trips than younger Canadians. They suggested that the main reason for these findings lies in the fact that participants were no longer employed and consequently did not have to make travel-to-work journeys. They also suggested health status as a second reason for their conclusion. According to their results, there was also a significant difference in transport mode between older and younger Canadians. The latter made more use of public transport as the principal travel mode. In other words, the reliance on the car was greater with the older Canadians.

### *3.2.2 Gender and travel behaviour*

A study in Germany attempted to determine whether there were gender related differences in car use and travel patterns for maintenance travel (Best & Lanzendorf, 2005). No significant differences were identified in distance travelled or in the total number of trips between men and women. Relating to the type or destination of trips, significant differences were determined. The researchers found that women made more journeys for maintenance activities and fewer journeys to work by car. These results were partially confirmed in a study of travel behaviour in southern California (Boarnet & Sarmiento, 1997). They found that women make fewer trips and trips were shorter than those of men. Studies with a somewhat different approach were carried out in Sweden (Polk, 2003), (Polk, 2004). The relationship between sustainable travel patterns and gender were investigated here. Polk concluded that women are more positive towards reducing the negative impact of transport on the environment. Women are more willing to reduce their use of the car than men, Polk concluded.

### *3.2.3 Household composition and travel behaviour*

The relationship between household composition and travel behaviour was investigated in 2005 in a study in Scotland (Ryley, 2006). Ryley used cluster analysis to categorize the sample into ten different segments, based largely on life stage. Results showed that households with children are highly dependent on cars as the primary source of travel mode. In addition, the results indicated that these households often own bicycles without using them. When used, these bicycles are mainly used for leisure and not for work. Riley also concluded that families with unemployed members, part-time workers without children and students primarily use non-motorised transport modes. In contrast, households consisting of retirees and members with a high income favour motorised transport modes. In conclusion, the results of this study indicate clearly that different stages of the household life cycle have an impact on travel behaviour. A research project in the Netherlands also studied the influence of household attributes on travel activities (Dieleman, Dijst, & Burghouwt, 2002). Results were similar to those of the study in Scotland. Their research showed that households on higher incomes own and use cars more often. Families with children also make more use of the car as means of transport.

### *3.2.4 Other socio-demographic variables and travel behaviour*

An empirical study in Sweden investigated the effects of socio-demographic variables on daily car use (Bergstad et al., 2011). The socio-demographic variables used were: sex, age, marital status, having children, education, employment, income, residential area and number of cars. Results showed that seven socio-demographic variables (sex, age, having children, percent employed, residential area and number of cars) account for 9% of the variance of weekly car trips. Furthermore, all the variables except 'age' and 'percent employed' account for 15% of the variance of the variable 'percent car use as driver'. Finally, they also concluded that sex, marital status, having children, employment and residential area account for 7% of the variance of the variable 'car use as passenger'.

In 1998 a study in the USA took a closer look at the relationships between socio-demographic variables, activity participation and travel behaviour (Lu & Pas, 1999). According to their results, employed people spend more time travelling. In addition, parents with more children generate more travel chains. On the other hand, people who work, make fewer chains than those who don't work. Households with more workers results in fewer chains.

## 3.3 Classification and regression trees (CART)

Classification and regression tree analysis is a common used classification method. This technique was developed in the 80s (Breiman, Friedman, Stone, & Olshen, 1984). The purpose of CART is to construct a so-called decision tree that is used to classify new data. In order to construct such a tree, historical data is used (Timofeev, 2004).

Decision trees are represented by a set of questions which splits the learning sample into smaller parts. The questions are exclusively yes/no questions. Possible questions are: "Is age greater than 18?" or "Does he/she own a car?". The CART algorithm is a technique for modelling a relationship between one or more dependent variables (response variables) and independent variables (explanatory variables). The algorithm considers all possible independent variables and their values in order to find the question that divides the data in two segments with maximum intrasegment homogeneity of the dependent variables. These two segments in turn can be split again using the same procedure and so on (Timofeev, 2004).

### *3.3.1 Classification and regression problems*

Tree-based models are used for both classification and regression problems. Classification trees use a categorical dependent variable. They are used when, for each observation of the learning sample, the class is known a priori. Regression trees use quantitative dependent variables. They do not have pre-assigned classes. In Figure 1 & 2 examples of classification and regression trees are given.

### *3.3.2 CART methodology*

CART methodology consists of three parts: construction of maximum tree, choice of the right tree size and classification of new data using the constructed tree (Timofeev, 2004).

In the first step the **maximum tree** is build. This is done by splitting the learning sample up to the last observations. This means that terminal nodes contain observations of one class. Different tree building procedures are used for classification and regression problems. This step is the most time consuming one.

Once the first step is completed, a very large and complex tree with hundreds of levels is obtained. Because such a tree is difficult to work with, it has to be optimized. This implies **choosing the right tree size** and cutting of insignificant nodes and even subtrees. Two algorithms are available in practice: optimization by number of points and by cross-validation.

In the third and final step the classification or regression tree can be used to **classify new data**. The output is an assigned class or response value to each of the new observations. This is done by the set of obtained questions in the tree.

### *3.3.3 Univariate and multivariate regression trees*

As stated earlier, regression trees use numeric dependent variables. Regression trees can be split up in univariate and multivariate regression trees.

Univariate regression trees (URT's) explain the variance of a single numeric dependent variable using explanatory variables that may be numeric and/or categorical (Breiman et al., 1984). Univariate regression trees are most common used. Extensive research had been done on this technique. In Figure 1 an example of a univariate regression trees is given.

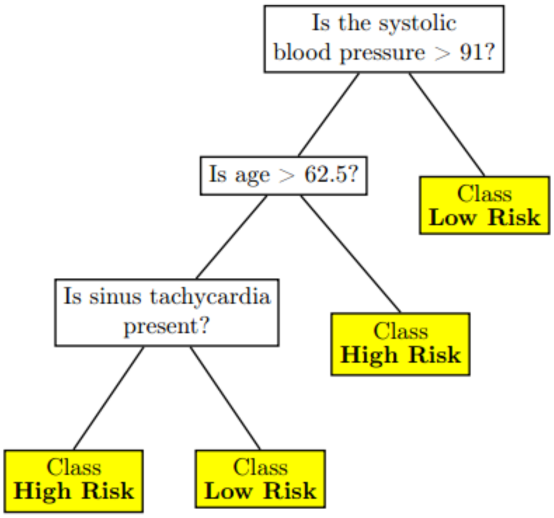

**Figure 1: Example of a univariate classification tree (Timofeev, 2004)**

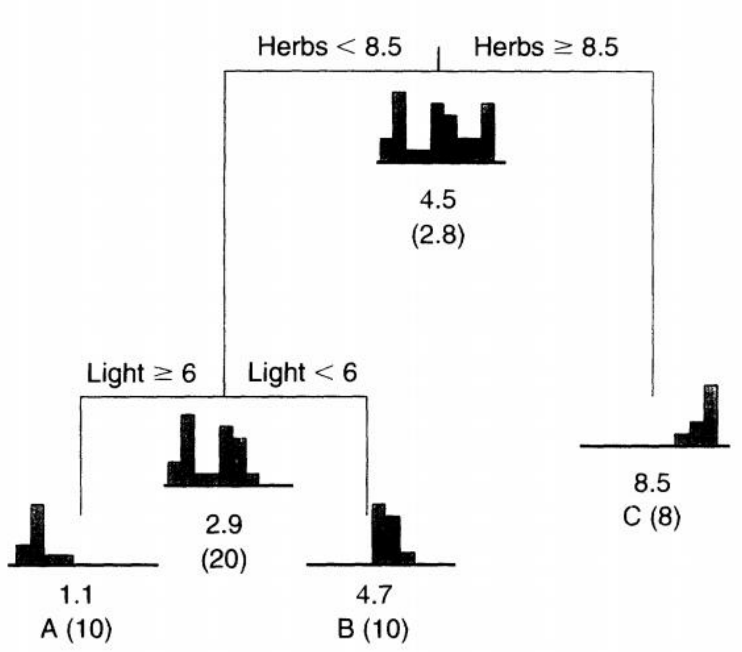

**Figure 2: Example of a univariate regression tree (De'ath, 2002)**

Multivariate regression trees (MRT's) are in fact a simple extension of URT's where the univariate response is replaced by a multivariate response. In addition, the impurity of a node has to be redefined by summing the univariate impurity measure over the multivariate response (De'ath, 2002). MRT analysis is not commonly used and very little research has been done on this method. Moreover, most established statistical software programs do not contain a MRT function. In Figure 3 an example of a multivariate regression tree is given.

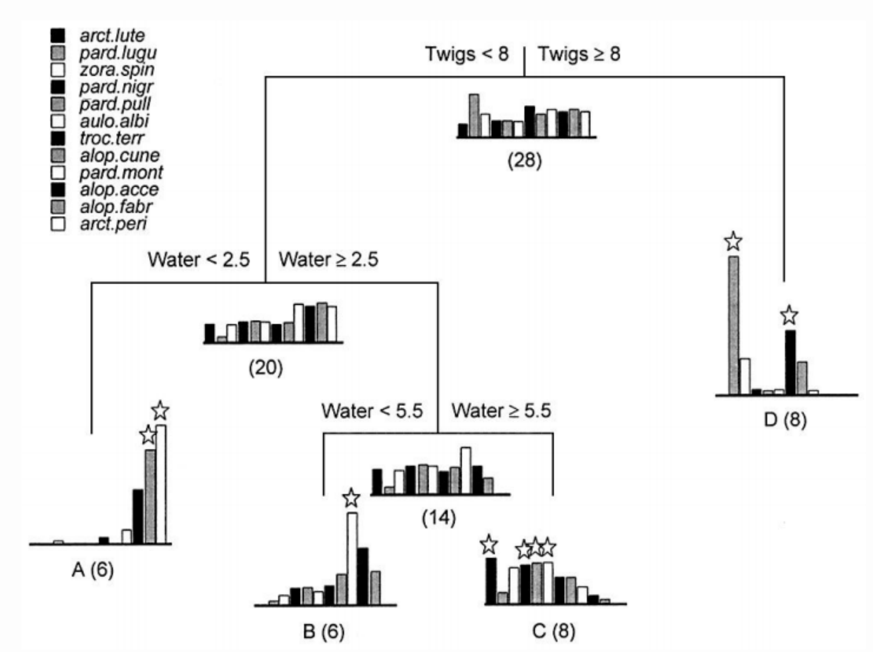

Figure 3: Example of a multivariate regression tree (De'ath, 2002)

### *3.3.4 Relative error and cross-validated relative error of trees*

Relative error and cross-validated relative error are two important characteristics of multivariate and univariate regression trees. Both metrics are outlined below (Borcard, Gillet, & Legendre, 2011).

The relative error (RE) is the sum of within-group sum of squares over all leaves divided by the overall sum of squares of the data [1]. In other words, this is the fraction of variance explained by the tree.

$$RE = \frac{SS\ over\ all\ leaves}{SS\ of\ data} \quad [1]$$

The relative error tends to give an over-optimistic estimate of how good a tree will predict new data. Using the relative error to describe trees would be an explanatory approach rather than a predictive approach. To assess the true predictive power of a tree a second metric is often used, namely: cross-validated relative error (CVRE). Cross-validation is an approach to estimate how well a model built from a training data set is going to perform on a new data set. To answer this question a subset of the objects (training set) is used to construct the tree. Next, the remaining objects (test set) are used to validate the result by allocating them to the constructed groups. A tree with a good prediction assigns the objects of the test set correctly. The performance of the tree is assessed by its predictive error which is measured by the CVRE. The function of the CVRE is presented in equation [2].

$$CVRE = \frac{\sum_{i=1}^{n}\sum_{j=1}^{p}\left(y_{ij(k)}-\hat{y}_{j(k)}\right)^2}{\sum_{i=1}^{n}\sum_{j=1}^{m}\left(y_{ij}-\bar{y}_j\right)^2} \quad [2]$$

Where $y_{ij(k)}$ is one observation of the test set k, $\hat{y}_{j(k)}$ is the predicted value of one observation in one leaf. The denominator represents the overall dispersion of the response data. In other words, the CVRE is the ratio between the dispersion unexplained by the tree (summed over the k test sets) divided by the overall dispersion of the response data. For perfect predictors the CVRE = 0. The CVRE is close to 1 for a poor set of predictors.

## 3.4 Cluster analysis

Cluster analysis groups objects or individuals into clusters so that objects in the same cluster are more similar to one another than they are to objects in other clusters. In other words, it aims at maximizing intragroup homogeneity and intergroup heterogeneity. If the classification is successful, objects within clusters will be close together when plotted geometrically. Different clusters will be far apart (F. Hair, Jr., C. Black, J. Babin, & E. Anderson, 2010).

Cluster analysis classifies objects on one or more selected characteristics. Univariate clustering uses one characteristic to group. Multivariate clustering deals with more characteristics. (F. Hair, Jr. et al., 2010).

### *3.4.1 Hierarchical clustering*

Cluster analysis is comprised of hierarchical cluster procedures and non-hierarchical cluster procedures. Hierarchical clustering groups data by creating a cluster tree or 'dendrogram'. The tree is not a single set of clusters, but rather a multilevel hierarchy, where clusters at one level are joined as clusters at the next level. This allows you to decide the level or scale of clustering that is most appropriate for your application. Hierarchical clustering can be subdivided into two types. The divisive methods start with all of the observations in one cluster and then proceeds to split them into smaller clusters. The agglomerative methods begin with each observation being considered as a separate cluster and then proceeds to combine them until all observations belong to one cluster (F. Hair, Jr. et al., 2010).

### *3.4.2 Non-hierarchical clustering*

Non-hierarchical clustering does not use a treelike construction. This method uses an iterative algorithm to assign objects to a pre-determined number of clusters. Starting from an initial classification, units are transferred from one group to another or swapped with units from other groups, until no further improvement can be made to the criterion value (F. Hair, Jr. et al., 2010).

### *3.4.3 The roles of cluster analysis*

Cluster analysis is commonly used for two purposes: data reduction and hypothesis generation (F. Hair, Jr. et al., 2010). Many data sets contain a large number of observations that are relative meaningless unless they are classified into manageable groups. Cluster analysis can be used to perform this data reduction objectively by reducing the information from an entire sample to information about specific groups. Cluster analysis can also be used to examine previously stated hypotheses. It also enables us to develop hypothesis relating to the nature of the data.

### *3.3.4 Measuring similarity*

The concept of similarity is fundamental to cluster analysis. Interobject similarity is an empirical measure of correspondence, or resemblance, between two objects. Different ways can be used to measure interobject similarity. The most commonly used approaches are correlation measures, distance measures, and association measures (F. Hair, Jr. et al., 2010).

1. <u>Correlation measures</u>

A correlation measure of similarity does not look at observed mean value or magnitude. Instead it looks at the patterns of movement seen as one traces the data for each case. Correlation represents patterns across the variables rather than the magnitudes. When a strong correspondence of patterns across characteristics is present, there is a high correlation. A lack of correspondence results in a low correlation.

2. Distance measures

Distance measures are the most commonly used measures of similarity. Here, similarity is seen as the proximity of observations to one another across the variables representing the objects or individuals. The proximity can be represented by various types of distance measures. To give some indication, the equation for two variables (u=(x1，y1) and v=(x2，y2)) is given for every distance measure. These equations are easily generalised to more than two variables.

**Euclidean distance** is the most commonly used distance measure. It is also referred to as straight-line distance. The Euclidean distance between u and v is given by the formula in equation [3].

$$\sqrt{(x1 - x2)^2 + (y1 - y2)^2} \quad [3]$$

**Squared Euclidean distance** is the same as the Euclidean distance, but no square root is taken. As a consequence, calculations are speeded up noticeably. The squared Euclidean distance between u and v is given in equation [4].

$$(x2 - x1)^2 + (y2 - y1)^2 \quad [4]$$

**City-block distance** uses the sum of the absolute differences of the variables. Although this distance measure is the easiest to calculate, it can lead to invalid clusters if the variables are highly correlated (Roger N., 1966). In equation [5] the city-block distance between u and v is given.

$$|x2 - x1| + |y2 - y1| \quad [5]$$

3. Association measures

When objects are measured in non metric terms, association measures of similarity are used to compare them. An association measure could assess the degree of agreement or matching between each pair of respondents. For example, an association measure would be the percentage of times respondents said ‘yes’ to a yes/no question.

### 3.5 Transportation Analysis and Simulation System (TRANSIMS)

TRANSIMS is an initiative of the Travel Mode Improvement Program (TMIP). This program is sponsored by the U.S. Department of Transportation, the Environmental Protection Agency (EPA) and the Department of Energy. TMIP aims at improving both analytical tools and the integration of these tools in the transportation planning process. The main goal of this program is to improve existing travel forecasting procedures in order to respond to emerging policy and technology issues. In addition, it concerns itself with redesigning the travel forecasting process and integrating forecasting techniques into the decision making process (‘Chapter 1: Transims overview,’ n.d.), (‘TRANSIMS Training Course at TRACC Part 1,’ 2009).

TRANSIMS develops technologies that can assist transportation planners in any urban environment. Increased policy sensitivity, detailed vehicle-emission estimates and improved analysis and visualization capabilities are offered by TRANSIMS. The philosophy underlying TRANSIMS is based on two key concepts: simulation and the activity-based approach. To study transportation, it tries to simulate travel in a study area based on individuals and their activity-travel pattern. Furthermore, census data and land use data are used to generate synthetic households and the people in it (‘Chapter 1: Transims overview,’ n.d.), (‘TRANSIMS Training Course at TRACC Part 1,’ 2009).

A series of modules is used by TRANSIMS to produce synthetic households, activities for household members, the choice of routes for movements among these activities, the microsimulation of these movements to create traffic dynamics on the network and consequently produced emissions (‘Chapter 1: Transims overview,’ n.d.), (‘TRANSIMS Training Course at TRACC Part 1,’ 2009). This framework is graphically presented in Figure 4.

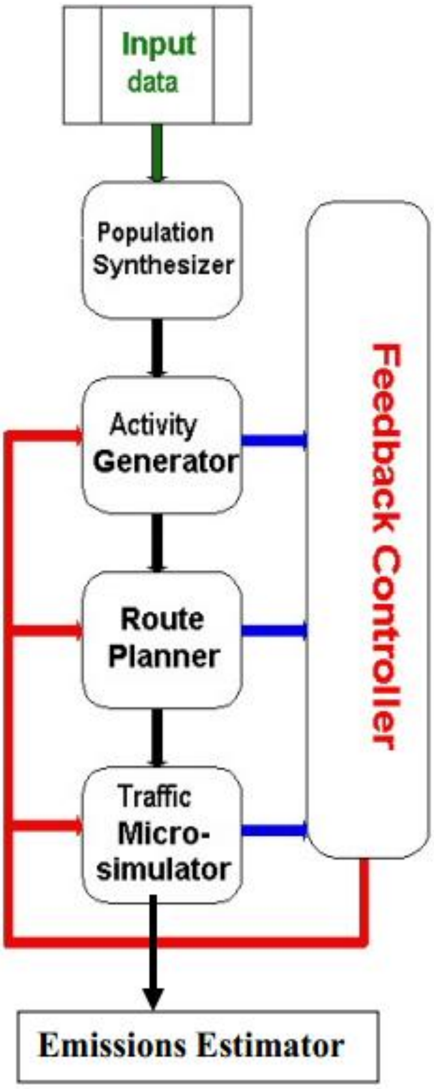

Figure 4: TRANSIMS framework ('Chapter 1: Transims overview,' n.d.)

The **population synthesizer** module generates all synthetic travellers using census data, land use data and network data. This module also estimates the number of synthetic households, the demographic characteristics of each household member and the locations of these households on the network. Thereafter the **activity generator** module creates a list of activities, activity times and activity locations for each synthetic traveller. Activities include work, shopping, school, etc.  These activity estimations are based on survey data. The next module is the **route planner**. Reading individual activities previously generated, this module calculates combined route and mode trip plans to accomplish the desired activities of each individual. Next, the **traffic microsimulation** module uses the paths developed in the route planner module to perform a regional microsimulation of vehicle interactions. The output can provide individual locations of all travellers and vehicles at all times over a 24-hour period. Finally, the **feedback controller** module manages the feedback of information among the different modules mentioned above.

### *3.5.1 The activity generator module*

The activity generator module is further outlined because the techniques used in this study are based on procedures of this module. The main task of this module is to generate activities for each member of a synthetic household over a 24-hour period. Different algorithms are used to perform this task. The algorithms involve the following five steps ('Chapter 4: Activity Generator,' n.d.):

1. Create skeletal activity patterns from the survey results
2. Use the CART (Classification and Regression Tree) algorithm to build a tree based on household demographic data
3. Match the given synthetic household with a survey household
4. Generate activity times and durations
5. Generate activity locations

With regard to the research objective of this master thesis, the first two steps are important and are further outlined. In the first step skeletal activity/travel patterns are created. This is done by organizing the activity lists of the members of each survey household and stripping of the locations. In addition, activities collected in the survey are grouped sequentially for each household member.

In the second step classification and regression tree analysis (CART) is used to build a tree based on household demographic data. The result of these two steps is a tree structure that partitions the data set into groups with similar values for the response variables. TRANSIMS uses the total time a household spends on different activity types as response variables. This approach will also be used in this research project. Furthermore, this analysis reveals the relevant explanatory variables. The explanatory variables in TRANSIMS are made up of various household demographic attributes. A simple example of such a tree structure is given in Figure 5.

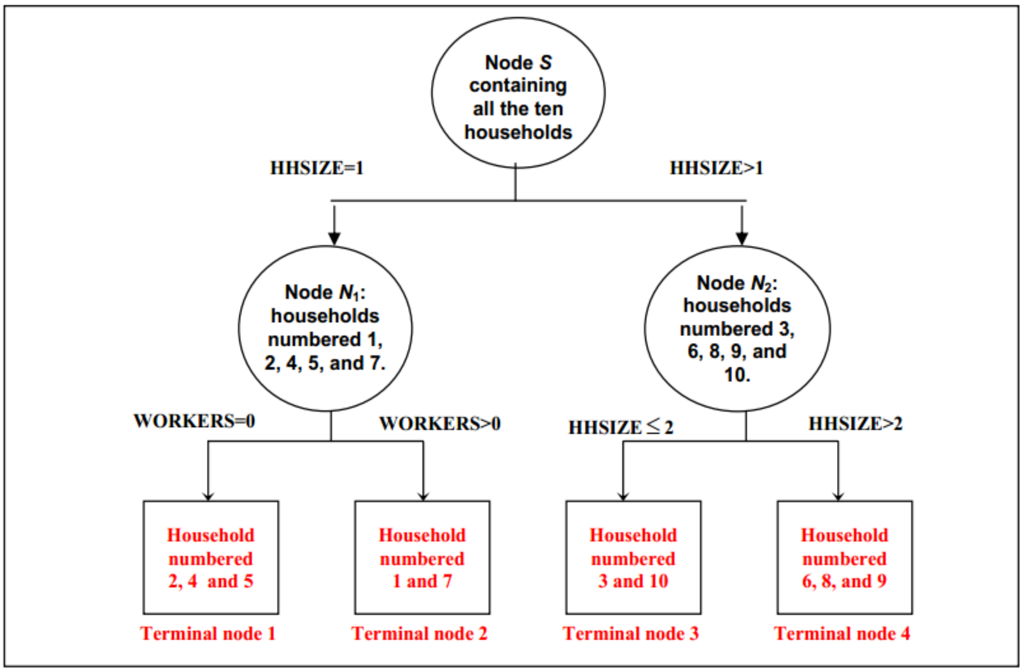
Node S containing all the ten households
HHSIZE=1
HHSIZE>1
Node N1: households numbered 1, 2, 4, 5, and 7.
Node N2: households numbered 3, 6, 8, 9, and 10.
WORKERS=0
WORKERS>0
HHSIZE ≤ 2
HHSIZE>2
Household numbered 2, 4 and 5
Household numbered 1 and 7
Household numbered 3 and 10
Household numbered 6, 8, and 9
Terminal node 1
Terminal node 2
Terminal node 3
Terminal node 4

**Figure 5: Example of a classification tree ('Chapter 4: Activity Generator,' n.d.)**

*3.5.2 TRANSIMS software*

TRANSIMS is an open source transportation modelling and simulation toolbox. To establish TRANSIMS as an ongoing public resource available to the transportation community, it is made available under the NASA Open Source Agreement Version 1.3. An important question that rose at the beginning of this project was to what extent TRANSIMS could be used in the analysis process. Although the techniques used in the analysis of this study are based on principles used in the activity generator module, it is not possible to use actual TRANSIMS software to conduct the analysis in this research project.

Two main reasons exist for not being able to use TRANSIMS software for the analysis in this research project. First, TRANSIMS makes use of household demographic attributes like household size, number of employed household members, etc. This study focuses on individual demographic variables instead of household demographic variables. Second, TRANSIMS consist of different integrated modules where the output of one module is the input of a next one. Taking out one module and transforming it to the specific needs of this analysis would be extremely complex, if not impossible. In summary, this study will not be making use of TRANSIMS software. Instead, other software will be used to perform analyses based on the techniques used in the activity generator module of TRANSIMS.

# 4. ANALYSIS

In this chapter the actual analysis of this research project will be performed. In the first section, a general explanation of the analysis will be given. Furthermore, information concerning the sample, measurements and software will be given here. In the second section, the analysis process will be outlined step by step. In the third and final section, results will be presented and discussed.

## 4.1 Method

The main goal of this research project is to identify different segments in Flanders based on activity-travel behaviour. Every identified segment will have a specific activity-travel pattern which describes travel behaviour of individuals within that segment.

Socio-demographic variables like age, gender, etc. Also play an important role in this analysis. They can be used to describe the different segments. The ultimate goal is to use these variables as predictors of activity-travel behaviour. Additionally, the study aims at taking other relevant variables into account.

The analysis will be performed using data collected in a large scale survey of the Flemish population where detailed activity-travel diaries were collected. This survey was taken within the framework of other projects. The method and design of this survey will not be outlined within this master thesis.

The way in which the actual analysis is performed is based on methods used by the Transportation Analysis and Simulation system (TRANSIMS). These methods and techniques have been outlined in the literature study. The analysis consists of three major steps that transform the raw survey data into usable results (Figure 6). First, skeletal activity patterns are created from the survey results. In the next step, the skeletal activity patterns are transformed into a usable data set. In the third step, CART analysis will be used to build a classification/regression tree. The result of these three steps is a tree structure that partitions the data set into groups with similar activity-travel behaviour.

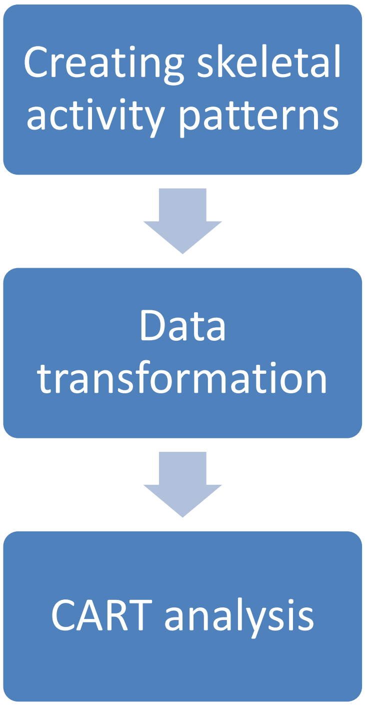

**Figure 6: Analysis steps**

### 4.2 Sample

During the survey of the Flemish population, detailed activity-travel diaries were collected for seven consecutive days (one week). The data were collected between 12/04/2006 and 17/04/2008. Information about a total of 11506 trips was gathered. These 11506 trips were performed by 699 individuals. This means that an individual in the sample performs an average of 16.5 trips per week. Data of 141 individuals was incorrect or incomplete. These were left out of the sample. Consequently, the final data set includes information about 558 people. This data set consists of 406 men and 152 women. The average age within the sample is 47.1 (standard deviation = 12.8).

### 4.3 Measurements

As stated earlier, the main goal of this project is to identify different segments based on activity-travel behaviour. Furthermore, this study aims to establish a link between travel behaviour, socio-demographic variables and other variables. Throughout this analysis, travel behaviour variables will be called 'response variables'. Socio-demographic variables and other variables will be called 'explain variables'.

### *4.3.1 Travel behaviour*

In this study, travel behaviour is described using daily and weekly travel time per activity type. This means the total daily or weekly time spent on travelling for a specific type of activity. 13 different activity types are used. The different activity types are further outlined in the actual analysis (Table 6). The first reason for using daily travel time and weekly travel time per activity type as measurement for travel behaviour is its correspondence with the survey. The survey was also concentrated around trip duration and travel time. Second, daily travel time and weekly travel time are understandable and tangible measurements. Furthermore, these measurements have the advantage of focusing exclusively on the travel/transportation part of an activity.

### *4.3.2 Socio-demographic variables*

The socio-demographic variables used in this study are age, gender, function in household, marital status, diploma, personal income, main profession, work situation, shift work, working hours and drivers license. These variables will be furthered outlined in the actual analysis (Table 6).

### *4.3.3 Other variables*

The survey data provides information of only one other variable that can be used as explain variable, namely: day of the week. In Table 1 an overview of the different variables and measurements used in this study is presented.

Table 1: Analysis variables and measurements

| *Variable* | **Travel behaviour** | **Socio-demographic variables and other variables** |
|---|---|---|
| *Measurement* | • daily travel time per activity type<br>• weekly travel time per activity type | • age, gender, function in household, marital status, diploma, personal income, main profession, work situation, shift work, working hours and drivers license<br>• day of the week |
| | = *response variables* | = *explain variables* |

## 4.4 Software

Three different computer programs have been used throughout the analysis: Microsoft Office Excel, Microsoft Office Access and R.

### *4.4.1 Microsoft Office Excel and Access*

Microsoft Office Excel is a well-known spreadsheet application used for calculation, graphing tools, pivot tables and a macro programming language. It is distributed by Microsoft. Within this master thesis Microsoft Excel is mainly used for data transformation. Different kinds of calculations have been made using Microsoft Excel. Also the pivot Table function of Microsoft Excel has been used throughout the analysis.

Microsoft Office Access is a database management system. It is also distributed by Microsoft and is like Microsoft Excel a member of the Microsoft Office suite of applications. In the analysis of this thesis Microsoft Access was employed to manage different databases. The survey results were collected in different databases. Microsoft Access was used to merge the different databases efficiently and effectively.

### *4.4.2 R*

R is a language and environment for statistical computing and graphics. It is based on the S language and environment which was developed at Bell

Laboratories (formerly AT&T, now Lucent Technologies). Some important differences exist between R and S language, but much S code runs unchanged under R.

R is a GNU project. This means it is a Unix-like operating system ultimately aiming to be wholly composed of free software. R runs on a wide variety of UNIX platforms and similar systems like Windows and MacOS. R can be seen as an integrated suite of software facilities for data manipulation, calculation and graphical display. R includes the following features ('The R Project for Statistical Computing,' n.d.):

- an effective data handling and storage facility
- a suite of operators for calculations on arrays, in particular matrices
- a large, coherent, integrated collection of intermediate tools for data analysis
- graphical facilities for data analysis and display either on-screen or on hardcopy
- a well-developed, simple and effective programming language which includes conditionals, loops, user-defined recursive functions and input and output facilities

R offers an array of different statistical and graphical techniques (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering). One of R's major advantages is its extensibility. R can be extended easily via user-created packages, which allow specialized statistical techniques, graphical devices, import/export capabilities, reporting tools, etc. A basic set of packages is included in the installation of R. Other additional packages are available at the Comprehensive R archive Network (CRAN). Two packages are used in this study, namely: mvpart package and MVPARTwrap package. ('The R Project for Statistical Computing,' n.d.).

**Mvpart** is an R package which allows you to run multivariate regression tree analysis in R. Mvpart is an extension of the rpart package. The rpart package allows the user to perform recursive partitioning and build regression trees for a univariate response. The mvpart package is maintained by Glenn De'ath (M. Therneau & Atkinson, 2012).

**MVPARTwrap** is a package of R that provides additional functionalities for the mvpart package. This packaged is maintained by Marie-Helene Ouelette (Ouelette & Legendre, 2011).

### 4.5 Step 1: creating skeletal activity patterns

In this step raw survey results are transformed into skeletal activity patterns. A skeletal activity pattern is organized around trips. Every trip is displayed on a new row. Every row consists of a person identification (who performs this trip?), a date (when is the trip performed?), an activity type identification (why is this trip performed?) and the duration of the trip (how long lasts this trip?). Locations are also stripped off. A sample of the skeletal activity patterns is shown in Table 2. The different variables are explained in Table 3.

**Table 2: Skeletal activity patterns**

| PersonID | VDate | U1 | … | U4 | M1 | … | M4 | WU1 | … | WU4 | WM1 | … | WM4 | Act |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| HH5GL11 | 12.04.06 | 0 | … | 0 | 5 | … | 0 | 0 | … | 0 | 0 | … | 0 | 16 |
| HH5GL11 | 13.04.06 | 1 | … | 0 | 30 | … | 0 | 0 | … | 0 | 10 | … | 0 | 26 |
| HH5GL11 | 13.04.06 | 1 | … | 0 | 30 | … | 0 | 0 | … | 0 | 10 | … | 0 | 16 |
| HH5GL11 | 13.04.06 | 0 | … | 0 | 5 | … | 0 | 0 | … | 0 | 0 | … | 0 | 18 |
| HH5GL11 | 15.04.06 | 0 | … | 0 | 25 | … | 0 | 0 | … | 0 | 10 | … | 0 | 16 |
| HH9GL56 | 24.04.06 | 1 | … | 0 | 5 | … | 0 | 0 | … | 0 | 0 | … | 0 | 22 |
| HH9GL56 | 24.04.06 | 0 | … | 0 | 5 | … | 0 | 0 | … | 0 | 0 | … | 0 | 19 |
| HH9GL56 | 24.04.06 | 1 | … | 0 | 15 | … | 0 | 0 | … | 0 | 0 | … | 0 | 16 |

**Table 3: Skeletal activity pattern variables**

| | |
|---|---|
| **PersonID** | This variable gives the unique ID for each individual in a household. This ID is made up of two parts. The first part is a code for the household (HH), while the second part is a unique code for the individual (GL). |
| **VDate** | This variable gives the date of the trip |
| **TDurationU1 (U1):** | This variable shows per trip how many hours one used the first transportation mode. This is only a specification of the hours (on hour level). |

| | |
|---|---|
| **TDurationU2 (U2),**<br>**TDurationU3 (U3),**<br>**TDurationU4 (U4)** | These variables show per trip how many hours (on hour level) one used the second, third and fourth transportation mode. |
| **TDurationM1 (M1)** | This variable shows per trip how many minutes one used the first transportation mode. |
| **TDurationM2 (M2),**<br>**TDurationM3 (M3),**<br>**TDurationM4 (M4)** | These variables show per trip how many minutes one used the second, third and fourth transportation mode |
| **WDurationU1 (WU1)** | This variable shows per trip how many hours (on hour level) one had to wait to use the first transportation mode. |
| **WDurationU2(WU2),**<br>**WDurationU3(WU3),**<br>**WDurationU4 (WU4)** | These variables show per trip how many hours (on hour level) one had to wait to use the second, third and fourth transportation mode. |
| **WDurationM1(WM1)** | This variable shows per trip how many minutes one had to wait to use the first transportation mode. |
| **WDurationM2(WM2),**<br>**WDurationM3 (WM3),**<br>**WDurationM4 (WM4)** | These variables show per trip how many minutes one had to wait to use the second, third and fourth transportation mode. |
| **Activity Type ID (Act)** | This variable indicates for which type of activity the trip was performed. This variable specifies for each trip the category number of the activity type (14 = activity at home; 15 = sleeping; 16 = working; 17 = services (e.g. going to a doctor); 18 = eating; 19 = daily shopping; 20 = shopping (non-daily goods); 21 = education; 22 = social activities; 23 = leisure activities; 24 = bring-get activities; 25 = touring (driving around for pleasure, walking around for pleasure…); 26 = other) |

## 4.6 Step 2: data transformation

Skeletal activity patterns are not directly usable for CART analysis. Different calculations and transformations have to be made in order to make the data usable for further analysis. Programs used for these operations are Microsoft Office Excel and Microsoft Office Access. The transformations are outlined in the following five steps.

1. Unique dates are transformed into days of the week.

| VDate | | Day of the week |
|---|---|---|
| 12.04.06 | | Wednesday |
| 13.04.06 | ➡ | Thursday |
| 15.04.06 | | Saturday |

2. Travel hours, travel minutes, waiting hours and waiting minutes are summed up to determine the total travel time for a particular activity type.

| U1 | U2 | U3 | U4 | M1 | M2 | M3 | M4 | WU1 | WU2 | WU3 | … |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | … |
| 1 | 0 | 0 | 0 | 30 | 0 | 0 | 0 | 0 | 0 | 0 | … |
| 1 | 1 | 0 | 0 | 30 | 5 | 0 | 0 | 0 | 0 | 0 | … |
| 0 | 0 | 0 | 0 | 5 | 5 | 0 | 0 | 0 | 0 | 0 | … |
| 0 | 0 | 0 | 0 | 25 | 10 | 0 | 0 | 0 | 0 | 0 | … |

| WU4 | WM1 | WM2 | WM3 | WM4 | | Total Time (Minutes) |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | | 5 |
| 0 | 10 | 0 | 0 | 0 | | 100 |
| 0 | 10 | 0 | 0 | 0 | ➡ | 165 |
| 0 | 0 | 0 | 0 | 0 | | 10 |
| 0 | 10 | 0 | 0 | 0 | | 45 |

3. The representation of the original survey results is build around trips. For every trip a new row is used in the survey data, irrespective of day and activity type. The data has to be transformed so it is concentrated around activity types and days of the week. The result of this transformation is a

Table in which the total travel time for different activity types is given for every particular person on a particular day of the week. Because of the fact the survey is taken during seven consecutive days, every day of the week will only occur once for every person.

| PersonID | Day of the week | 16 | … | 18 | … | 26 | Total |
|---|---|---|---|---|---|---|---|
| HH5GL11 | Wednesday | 5 | … | 0 | … | 0 | 5 |
| HH5GL11 | Thursday | 165 | … | 10 | … | 100 | 275 |
| HH5GL11 | Thursday | 10 | … | 0 | … | 0 | 10 |
| HH5GL11 | Saturday | 45 | … | 0 | … | 0 | 45 |

⬇

| PersonID | Day of the week | 16 | … | 18 | … | 26 | Total |
|---|---|---|---|---|---|---|---|
| HH5GL11 | Wednesday | 5 | … | 0 | … | 0 | 5 |
| HH5GL11 | Thursday | 175 | … | 10 | … | 100 | 285 |
| HH5GL11 | Saturday | 45 | … | 0 | … | 0 | 45 |

4. So far the data does not represent days of the week on which a particular person did not perform any travel activity. These days of the week have to be added.

| PersonID | Day of the week | 16 | … | 18 | … | 26 | Total |
|---|---|---|---|---|---|---|---|
| HH5GL11 | Wednesday | 5 | … | 0 | … | 0 | 5 |
| HH5GL11 | Thursday | 175 | … | 10 | … | 100 | 285 |
| HH5GL11 | Saturday | 45 | … | 0 | … | 0 | 45 |

⬇

| PersonID | Day of the week | 16 | … | 18 | … | 26 | Total |
|---|---|---|---|---|---|---|---|
| HH5GL11 | Monday | 0 | … | 0 | … | 0 | 0 |
| HH5GL11 | Tuesday | 0 | … | 0 | … | 0 | 0 |
| HH5GL11 | Wednesday | 5 | … | 0 | … | 0 | 5 |
| HH5GL11 | Thursday | 175 | … | 10 | … | 100 | 285 |
| HH5GL11 | Friday | 0 | … | 0 | … | 0 | 0 |
| HH5GL11 | Saturday | 45 | … | 0 | … | 0 | 45 |
| HH5GL11 | Sunday | 0 | … | 0 | … | 0 | 0 |

5. The travel behaviour results that have been transformed in the previous four steps and the socio-demographic information were collected in two different files. These two files are merged together to create a data set containing both travel behaviour information and socio-demographic information.

   Besides daily travel time, weekly travel time will also be used as a measure for travel behaviour. To obtain these values the daily travel times are summed for every individual. Because the survey was taken during seven consecutive days, weekly travel time is obtained this way.

   The result of these operations is a set of two similar data sets that can be used for CART analysis. Both data sets contain travel behaviour information (travel time) and socio-demographic information. The difference between the two data sets is the measurement used for travel behaviour. One is build around daily travel time per activity type. The second data set is concentrated around weekly travel time per activity type.

   A sample of both data sets in is given in Table 4 and Table 5. Table 6 outlines all the variables which were not outlined in Table 2.

Table 4: Sample final data set (daily travel time)

| PersonID | Day of the week | 16 | … | 18 | … | 26 | Total | Age | Gender | … |
|---|---|---|---|---|---|---|---|---|---|---|
| HH5GL11 | Monday | 0 | … | 0 | … | 0 | 0 | 39 | Male | … |
| HH5GL11 | Tuesday | 0 | … | 0 | … | 0 | 0 | 39 | Male | … |
| HH5GL11 | Wednesday | 5 | … | 0 | … | 0 | 5 | 39 | Male | … |
| HH5GL11 | Thursday | 175 | … | 10 | … | 100 | 285 | 39 | Male | … |
| HH5GL11 | Friday | 0 | … | 0 | … | 0 | 0 | 39 | Male | … |
| HH5GL11 | Saturday | 45 | … | 0 | … | 0 | 45 | 39 | Male | … |
| HH5GL11 | Sunday | 0 | … | 0 | … | 0 | 0 | 39 | Male | … |

Table 5: Sample final data set (weekly travel time)

| PersonID | 16 | … | 18 | … | 26 | Total | Age | Gender | … |
|---|---|---|---|---|---|---|---|---|---|
| HH5GL11 | 225 | … | 10 | … | 100 | 335 | 39 | Male | … |
| HH3GL23 | 120 | … | 0 | … | 20 | 140 | 27 | Male | … |
| HH1GL32 | 40 | … | 40 | … | 0 | 80 | 56 | Female | … |
| HH14GL4 | 65 | … | 10 | … | 100 | 175 | 22 | Male | … |
| HH6GL6 | 50 | … | 10 | … | 10 | 70 | 36 | Male | … |
| HH2GL11 | 85 | … | 0 | … | 0 | 85 | 27 | Female | … |
| HH8GL18 | 0 | … | 20 | … | 0 | 20 | 62 | Male | … |

Table 6: Variables final data set

**Day of the week** This variable specifies the day of the week

**Age** This variable represents the age of the person

**16,17…,26** These variables represent the daily/ weekly travel time per activity type 16, 17…26. (14 = activity at home; 15 = sleeping; 16 = working; 17 = services (e.g. going to a doctor); 18 = eating; 19 = daily shopping; 20 = shopping (non-daily goods); 21 = education; 22 = social activities; 23 = leisure activities; 24 = bring-get activities; 25 = touring (driving around for pleasure, walking around for pleasure…); 26 = other)

**Gender**

This Variable represents the sex of the individual and is either ‘male’ or ‘female’

**Function in household**

This variable represents the function of the person in its household and can have five values: Head of the family, Partner of the head of the family, Child of the head of the family, Other family member, Not defined

**Marital status**

This variable stands for the marital status of the individual. This variable can have six different values: Married, Unmarried, Not defined, Divorced, Cohabiting (or living together), Widow

**Diploma**

This variable represents the level of education of the individual, by specifying the highest diploma the individual attained. Possible values are: Primary education, lower secondary education: general, lower secondary education: technical or vocational, higher secondary education: general, higher secondary education: technical or vocational, Higher education (not university), Higher ( university education), NRC (Not defined)

**Personal income**

This variable specifies the personal monthly net income of the individual (including wages, salaries, compensations for unemployment and disability …). Possible values for this variable are: <750, 750-1250, 1250-1750, 1750-2250, 2250-2750, >2750, NRC (= Not defined).

**Main profession**

This variable gives the main profession of each individual. Possible values are: Student, exclusively active in own household), unemployed, Retired, Laborer, Public servants or public officials, disabled to work, clerk, no executive, executive clerk, liberal profession, Self-employed person, freelancer, independent worker, other (not professionally active), other, (professionally active), NRC (= not defined).

**Work situation** This variable represents the work situation, namely 'Full-time worker' or 'Part-time worker'

**Shift work** In this variable the participant indicates when he or she is working, during what time of the day. Possible values are: Exclusively during the day, Exclusively during the night, Working in shifts but not at night, Working in shifts including at night, Other, Not defined, Not applicable.

**Working hours** In this variable the participant indicates how stable the working hours are. Possible values are: fixed working hours on daily basis which are determined by employer, fixed working hours on daily basis which are determined by the worker self, regularly varying working hours which are determined by employer, regularly varying working hours which are determined by the worker self, Not defined, Not applicable.

**Driver's license** This variable indicates the possession of a driver's license for cars. 'NO' stands for no car driver's license, while 'YES' stands for the possession of a car driver's license.

## 4.7 Step 3: CART analysis

In this step the actual classification and regression tree analysis (CART) is performed. More specifically, a multivariate regression tree (MRT) analysis is conducted to identify segments based on travel behaviour (response variables) and socio-demographic variables (explain variables). The day of the week is also included as an explain variable. Background information concerning CART analysis is given in the literature review (section 3.3). The MRTs are conducted using the mvpart library in R. In this model the sum of squared Euclidean distances about the multivariate mean of samples is used as an impurity measure of each node. Each split in the tree is made to maximize the sum of squares between nodes and to minimize the sum of squares within nodes. Compared to cluster analysis, CART analysis is able to incorporate response variables and explain variables, where cluster analysis only uses one group of variables. In other words, a predictive approach is used by CART analysis. Cluster analysis makes use of an explanatory approach. A more detailed explanation of CART and cluster analysis is presented in the literature review. MRT's are strong candidates for community data modelling because of several reasons (De'ath, 2002):

- unlimited numbers of quantitative and categorical variables can be used as explanatory variables
- monotonic transformations of explanatory variables are allowed
- interactions between explanatory variables are automatically detected
- they are robust to the collinearity of explanatory variables and
- they are minimally affected by missing values

Overall fit of a tree is the fraction of variance not explained by that tree. More generally, fit is defined by the relative error (RE): the total impurity of the leaves divided by the impurity by the root node. RE is an easy-to-use measurement but tends to give over-optimistic estimate. This is why in CART analysis the cross-validated relative error (CVRE) is often used (De'ath, 2002). A more detailed explanation regarding the RE and CVRE can be found in the literature review.

The general mvpart model is given by (M. Therneau & Atkinson, 2012):

```
mvpart(form, data, minauto = TRUE, size, xv = c(‘1se’, ‘min’,
‘pick’, ‘none’), xval = 10, xvmult = 0, xvse = 1, snip = FALSE,
plot.add = TRUE, text.add = TRUE, digits = 3, margin = 0, uniform
= FALSE, which = 4, pretty = TRUE, use.n = TRUE, all.leaves =
FALSE, bars = TRUE, legend, bord = FALSE, xadj = 1, yadj = 1, prn
= FALSE, branch = 1, rsq = FALSE, big.pts = FALSE, pca = FALSE,
interact.pca = FALSE,  wgt.ave.pca = FALSE, keep.y = TRUE, ...)
```

The model consists of many arguments. Some arguments belong exclusively to the mvpart function. Other arguments originally belong to the rpart package and are transferred to mvpart. For most arguments, the default setting is kept throughout the study. The arguments for which this is not the case are outlined below (M. Therneau & Atkinson, 2012):

- **xv**: this argument defines the selection of tree by cross-validation: ‘1se’ - gives best tree within one standard error (SE) of the overall best, ‘min’ - the best tree, ‘pick’ - picks the tree size interactively, ‘none’ - no cross-validation.

- **cp**: cp stand for complexity parameter. Any split that does not decrease the overall lack of fit by a factor of cp is not attempted. For instance, with anova splitting, this means that the overall Rsquare must increase by cp at each step. The main role of this parameter is to save computing time by pruning off splits that are obviously not worthwhile. Essentially, the user informs the program that any split which does not improve the fit by cp will likely be pruned off by cross-validation, and that hence the program need not pursue it.

- **xvmult:** this argument defines the number of cross-validations. The value for this argument defines how many trees R creates. It will then pick the tree that is most consistently produced.

To extract more information out of the different MRT analyses a second package is employed: MVPARTwrap package. The MRT function of this package creates a modified output of the multivariate regression trees. This study is especially interested in the table with the total response variables variance portioned by response variables, by the tree and by the splits of the tree. This information is not provided by the mvpart package. The general MVPARTwrap model is given by (Ouelette & Legendre, 2011):

```
MRT(obj,percent,species=NULL,LABELS = FALSE,...)
```

Different multivariate tree analysis will be conducted in this study using two different response variables (daily travel time or weekly travel time) and complexity parameters. Different tree sizes per analysis will be investigated. In the final part of this step the use of another measurement (Bray-Curtis dissimilarity) will be explored.

### *4.7.1 Daily travel time per activity type (cp = 0.05)*

In this analysis a multivariate regression tree analysis will be performed using the travel time for the different activity types (16,17,…,26) as response variables (dependent variables). The explain variables (independent variables) used in this analysis are: day of the week, age, gender, function in household, marital status, diploma, personal income, main profession, work situation, shift work, working hours and driver's license. The complexity parameter is set on 0.05 and the parameter 'xv' is set on 'pick', which gives the opportunity to pick the tree size interactively. The number of cross-validations is set on 50 (xvmult = 50). An abstract of the design is given in the following summary.

| **Response variables** | Daily travel time per activity type |
|---|---|
| **Explain variables** | day of the week, age, gender, function in household, marital status, diploma, personal income, main profession, work situation, shift work, working hours and driver's license |
| **complexity parameter (cp)** | 0.05 |
| **Tree size (xv)** | 'pick' → Interactively |
| **multiple cross-validations (xvmult)** | 50 |

R first produces a graph where the tree size can be selected (Figure 7).

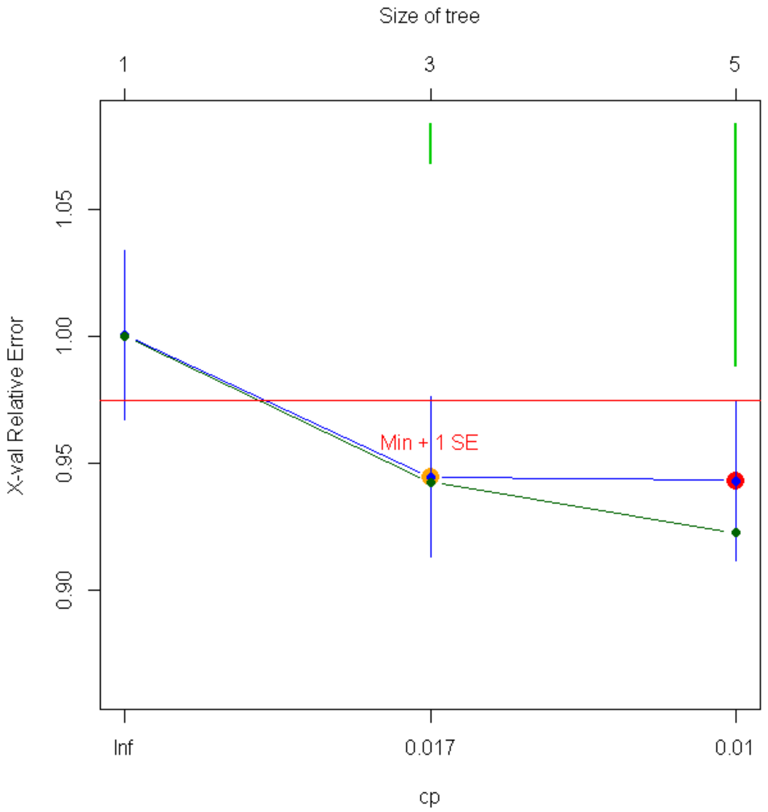

**Figure 7: Tree size selection 4.7.1**

The relative error is given by the green (lower) line. It decreases with tree size. The cross-validated relative error is given by the blue (upper) line. This error decreases to a minimum for a size of tree before flattening to a typical plateau. Normally, the smallest tree within one standard error (SE) of the best is selected (Breiman et al., 1984). This study takes a look at all the possible trees to get a comprehensive overview. First, the tree with size three is examined. The results are presented below.

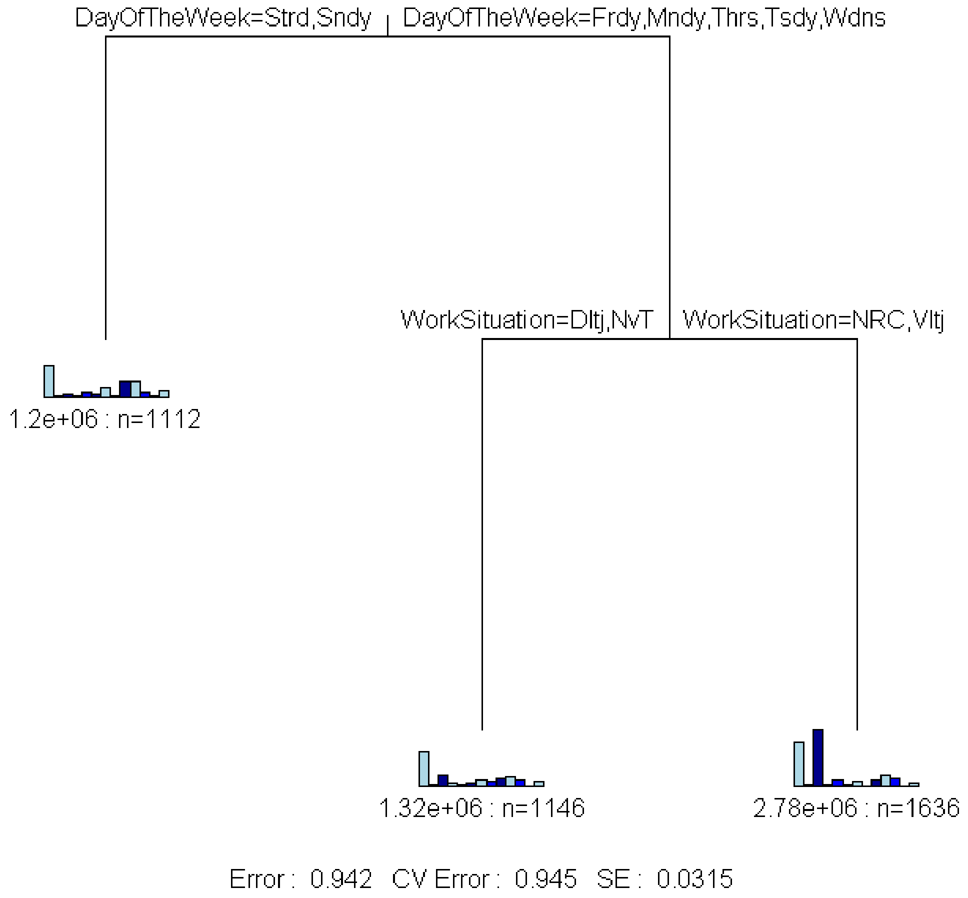

**Figure 8: Multivariate regression tree 4.7.1 (A)**

***Split 1 (Day of the week)*** *→ Strd = Saturday, Sndy = Sunday, Frdy = Friday, Mndy = Monday, Thrs = Thursday, Tsdy = Tuesday, Wdns = Wednesday*

***Split 2 (Work situation****) → Dltj = part time, NvT = not applicable, Vltj = full time, NRC = no value)*

Table 7: Travel behaviour in different segments 4.7.1 (A)

| | Activity type | Segment | | |
|---|---|---|---|---|
| | | 1 (n=1112) | 2 (n=1146) | 3 (n=1636) |
| 14 | activity at home | 10,88 | 12,27 | 15,64 |
| 15 | sleeping | 0,65 | 0,73 | 0,52 |
| 16 | working | 0,97 | 3,98 | 19,92 |
| 17 | services | 0,36 | 1,35 | 0,62 |
| 18 | eating | 1,87 | 0,93 | 2,15 |
| 19 | daily shopping | 0,98 | 1,23 | 0,74 |
| 20 | shopping | 3,04 | 2,24 | 1,70 |
| 21 | education | 0,30 | 1,81 | 0,48 |
| 22 | social activities | 5,40 | 2,91 | 2,38 |
| 23 | leisure activities | 5,34 | 3,58 | 3,89 |
| 24 | bring-get activities | 1,87 | 2,61 | 2,69 |
| 25 | touring | 0,64 | 0,35 | 0,30 |
| 26 | other | 2,19 | 1,67 | 1,26 |

The MRT analysis produces a three-leaf tree with splits based on 'day of the week' and 'work situation' (Figure 8). The tree only explains 5.8% of travel behaviour variance (RE = 0.942). The cross-validated relative error is slightly higher than the relative error (CVRE = 0.944). Despite the nature of multivariate community data this value is very low. Daily travel time for only one activity type varies strongly across the three segments: working. Segments one and two both have low values for working travel time. This is not illogical given the fact that segment one is characterized by Saturday or Sunday as day of the week. Segment two contains the individuals that work part time or do not work at all. Segment three is defined by high levels of travel time for working. Differences between travel times for other activity types are very small.

The contribution of each individual travel time category at each split and how well each category is explained by the tree can be quantified by tabulating the explained variance at each split.

Table 8: Tabulation of travel behaviour variance 4.7.1 (A)

| Act | Split 1 | Split 2 | Tree total | Travel behaviour total |
|---|---|---|---|---|
| 14 | 1.603202e-01 | 1.361601e-01 | 0.296480327 | 27.020737 |
| 15 | 2.465288e-05 | 5.608451e-04 | 0.000585498 | 1.212687 |
| 16 | 2.162465e+00 | 3.040154e+00 | 5.202619500 | 29.461298 |
| 17 | 4.460413e-03 | 6.414508e-03 | 0.010874921 | 1.594089 |
| 18 | 7.320983e-04 | 1.782768e-02 | 0.018559782 | 3.483115 |
| 19 | 1.988724e-05 | 2.866970e-03 | 0.002886857 | 1.250498 |
| 20 | 1.764282e-02 | 3.572586e-03 | 0.021215401 | 4.462151 |
| 21 | 7.415039e-03 | 2.129285e-02 | 0.028707893 | 2.389991 |
| 22 | 1.107104e-01 | 3.353002e-03 | 0.114063447 | 7.275304 |
| 23 | 3.502058e-02 | 1.171889e-03 | 0.036192465 | 10.235350 |
| 24 | 8.772576e-03 | 7.161182e-05 | 0.008844188 | 5.460018 |
| 25 | 1.431808e-03 | 3.881213e-05 | 0.001470620 | 1.507522 |
| 26 | 8.168420e-03 | 1.957568e-03 | 0.010125987 | 4.647239 |
| total | 2.517184e+00 | 3.235443e+00 | 5.752626886 | 100.000000 |

As can be seen from Table 7, the variance of travel time for activities at home comprises 27.02% of total travel behaviour variance. 29.46% of the total variance is comprised by travel time for working. It is notable that these two variables make up almost 60% of total travel behaviour variance. It was earlier noted that the total tree only explains 5.8% of total variance. 5.2% of this variance is comprised by daily travel time for working (16). Split one and two explain respectively 2.16% and 3.04%. None of the 13 activity types are well separated by the three segments, with the best explained, daily travel time for working, having 17.7% (5.2 of 29.5%) of its variance explained. According to the table, 2.5% of variance is explained by split one (day of the week) and 3.2% is explained by the second split (work situation).

Although cross-validated relative error flattens out, the multivariate regression tree with five segments is investigated to get an idea of the possible additional explain variables (Figure 7). The tree is presented in Figure 9.

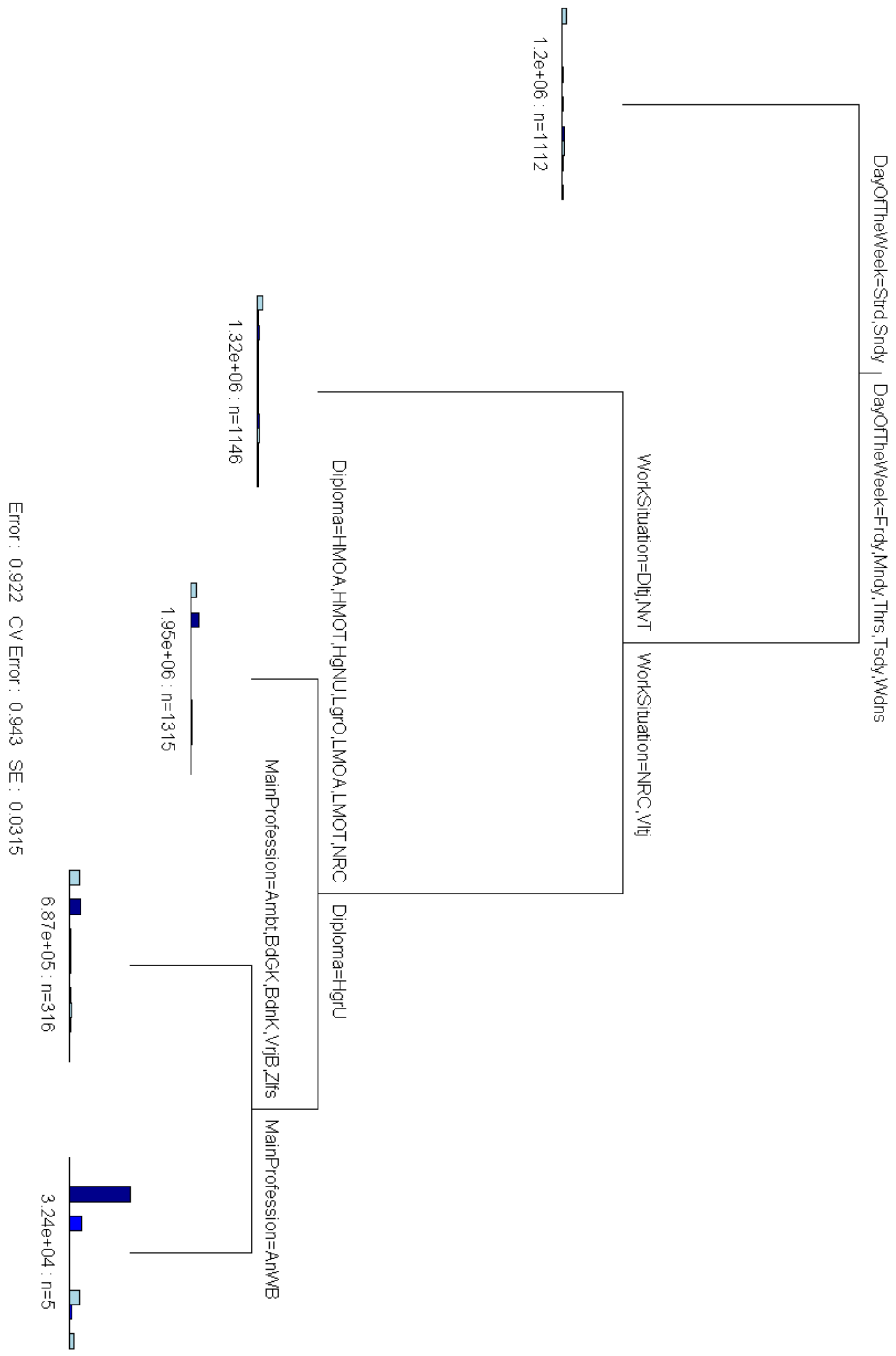

**Figure 9: Multivariate regression tree 4.7.1 (B)**

***Split 1 (Day of the week)*** → *Strd = Saturday, Sndy = Sunday, Frdy = Friday, Mndy = Monday, Thrs = Thursday, Tsdy = Tuesday, Wdns = Wednesday*

***Split 2 (Work situation)*** *→ Dltj = part time, NvT = not applicable, Vltj = full time, NRC = no value*

***Split 3 (Diploma)*** *→ HMOA = higher secondary education: general, HMOT = higher secondary education: technical or vocational, HgNU = higher education but not university, LMOA = lower secondary education: general, LMOT = lower secondary education: technical or vocational, NRC = no value, HgrU = higher university education*

***Split 4 (Main profession)*** *→ Ambt = public servants or public officials, BdGK = clerk but no executive, Bdnk = executive clerk, VrjB = liberal profession, Zlfs = self-employed person, freelancer, independent worker; AnWB = other (professionally)*

**Table 9: Travel behaviour in different segments 4.7.1 (B)**

| | Activity type | Segment | | | | |
|---|---|---|---|---|---|---|
| | | 1 (n=1112) | 2 (n=1146) | 3 (n=1315) | 4 (n=316) | 5 (n=5) |
| 14 | activity at home | 10,88 | 12,27 | 14,22 | 21,80 | 0,00 |
| 15 | sleeping | 0,65 | 0,73 | 0,56 | 0,32 | 1,00 |
| 16 | working | 0,97 | 3,98 | 17,95 | 26,17 | 141,00 |
| 17 | services | 0,36 | 1,35 | 0,61 | 0,66 | 0,00 |
| 18 | eating | 1,87 | 0,93 | 1,94 | 2,60 | 28,00 |
| 19 | daily shopping | 0,98 | 1,23 | 0,60 | 1,33 | 0,00 |
| 20 | shopping | 3,04 | 2,24 | 1,68 | 1,80 | 0,00 |
| 21 | education | 0,30 | 1,81 | 0,41 | 0,76 | 0,00 |
| 22 | social activities | 5,40 | 2,91 | 2,32 | 2,64 | 1,00 |
| 23 | leisure activities | 5,34 | 3,58 | 3,77 | 4,11 | 22,00 |
| 24 | bring-get activities | 1,87 | 2,61 | 2,52 | 3,32 | 6,00 |
| 25 | touring | 0,64 | 0,35 | 0,27 | 0,41 | 0,00 |
| 26 | other | 2,19 | 1,67 | 1,25 | 1,16 | 11,00 |

A five-leaf tree is produced by the MRT analysis. Four splits are made. The splits are based on the ‘day of the week’, ‘work situation’, ‘diploma’ and ‘main profession’. These parameters make up the explain variables of this model. Although two additional splits and two extra segments are formed in comparison to the previous tree, this five-leaf tree only explains 7.8% of travel behaviour variance (RE = 0.922). This is 2% more than the three-leaf tree but still low, even when the nature of multivariate community data is considered. The cross-validated relative error dropped almost nothing in comparison to the three-leaf tree (CVRE = 0.943). ‘Working’ is again the activity type for which daily travel time varies most between the different segments (Table 8). Differences in daily travel time for activities at home seem slightly larger for these five segments than in the previous tree. Travel time for all the other categories barely varies between the segments. The two additional segments are formed by splitting segment three of the three-leaf tree. Two splits were made based on 'diploma' and 'main profession'. According to the tree, people with a higher university diploma tend to spend more time travelling for work.

### Table 10: Tabulation of travel behaviour variance 4.7.1 (B)

| Act | Split 1 | Split 2 | Split 3 | Split 4 |
|---|---|---|---|---|
| 14 | 1.603202e-01 | 1.361601e-01 | 2.404117e-01 | 4.155377e-02 |
| 15 | 2.465288e-05 | 5.608451e-04 | 2.544020e-04 | 4.083941e-05 |
| 16 | 2.162465e+00 | 3.040154e+00 | 4.586558e-01 | 1.152524e+00 |
| 17 | 4.460413e-03 | 6.414508e-03 | 8.097146e-06 | 3.860206e-05 |
| 18 | 7.320983e-04 | 1.782768e-02 | 5.065851e-03 | 5.641397e-02 |
| 19 | 1.988724e-05 | 2.866970e-03 | 2.319178e-03 | 1.544083e-04 |
| 20 | 1.764282e-02 | 3.572586e-03 | 4.481144e-05 | 2.843948e-04 |
| 21 | 7.415039e-03 | 2.129285e-02 | 5.204099e-04 | 5.041902e-05 |
| 22 | 1.107104e-01 | 3.353002e-03 | 3.950372e-04 | 2.357798e-04 |
| 23 | 3.502058e-02 | 1.171889e-03 | 1.786711e-03 | 2.796252e-02 |
| 24 | 8.772576e-03 | 7.161182e-05 | 3.260523e-03 | 6.264879e-04 |
| 25 | 1.431808e-03 | 3.881213e-05 | 8.353212e-05 | 1.479308e-05 |
| 26 | 8.168420e-03 | 1.957568e-03 | 1.512804e-05 | 8.471727e-03 |
| Total | 2.517184e+00 | 3.235443e+00 | 7.128212e-01 | 1.288372e+00 |

| Act | Tree total | Travel behaviour total |
|---|---|---|
| 14 | 0.5784457448 | 27.020737 |
| 15 | 0.0008807394 | 1.212687 |
| 16 | 6.8137997494 | 29.461298 |
| 17 | 0.0109216199 | 1.594089 |
| 18 | 0.0800396045 | 3.483115 |
| 19 | 0.0053604432 | 1.250498 |
| 20 | 0.0215446076 | 4.462151 |
| 21 | 0.0292787217 | 2.389991 |
| 22 | 0.1146942640 | 7.275304 |
| 23 | 0.0659416918 | 10.235350 |
| 24 | 0.0127311995 | 5.460018 |
| 25 | 0.0015689451 | 1.507522 |
| 26 | 0.0186128425 | 4.647239 |
| total | 7.7538201733 | 100.000000 |

Table 9 displays the explained variance at each split of daily travel time for each activity type for the five-leaf tree. The five-leaf tree explains 2% more variance than the three-leaf tree. This model explains 23.1% (6.8 of 29.5%) of the variance of the travel time the activity type 'working'. The variance of travel time for the other activity types is barely explained by this tree.

*4.7.2 Daily travel time per activity type (cp=0.005)*

In this section the same response variables (dependent variables) are used as in section 4.7.1, namely: daily travel time per activity type (16, 17… 26). Likewise, the same explain variables (dependent variables) are used. These are: day of the week, age, gender, function in household, marital status, diploma, personal income, main profession, work situation, shift work, working hours and driver's license. In comparison to analysis 4.7.1, the complexity parameter is set on a lower value (0.005). This approach aims at revealing more segments and relevant explain variables by allowing more splits to be performed. The parameter 'xv' is set on 'pick' so the tree size can be picked interactively. The number of cross-validations is set on 50 so R can pick the tree that is most consistently produced. An overview of the design of this MRT analysis is given in the following overview.

| | |
|---|---|
| **Response variables** | Daily travel time per activity type |
| **Explain variables** | day of the week, age, gender, function in household, marital status, diploma, personal income, main profession, work situation, shift work, working hours and driver's license |
| **complexity parameter (cp)** | 0.005 |
| **Tree size (xv)** | 'pick' → Interactively |
| **multiple cross-validations (xvmult)** | 50 |

The tree size selection graph produced by the mvpart package in R is presented in Figure 10.

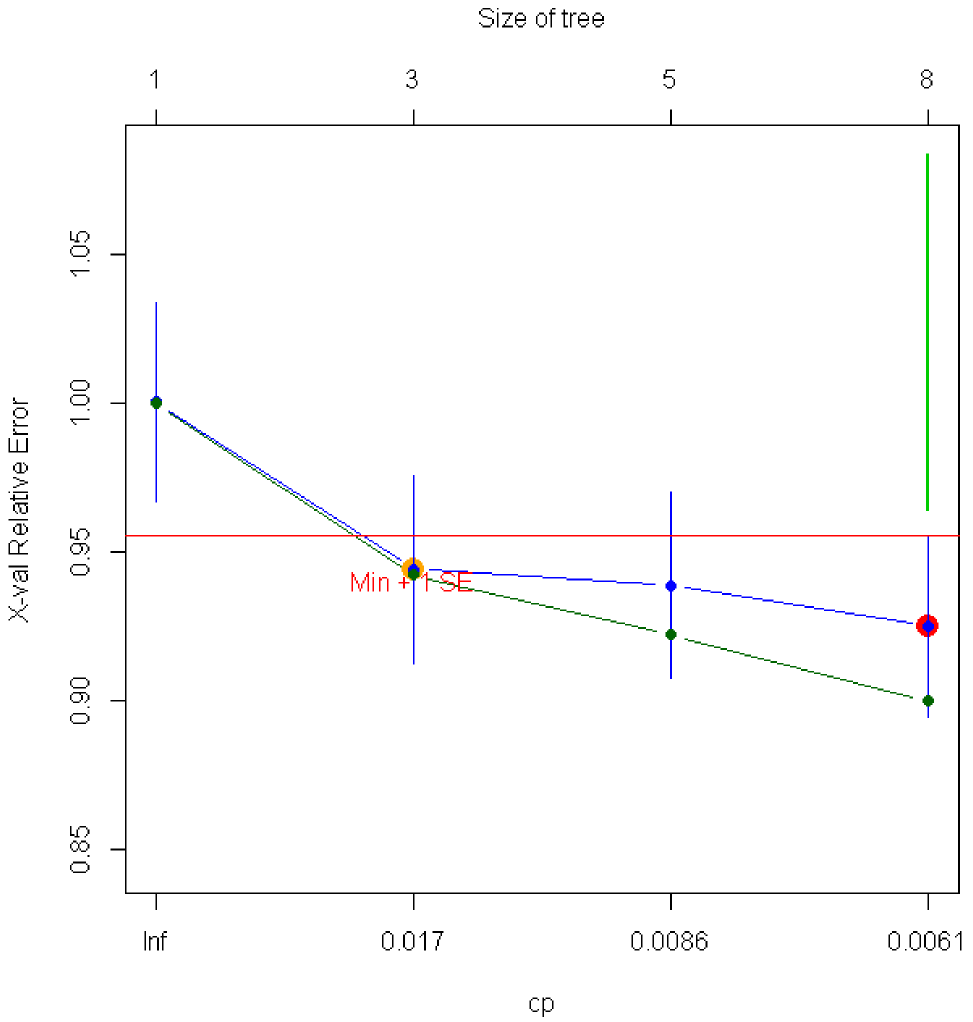

**Figure 10: Tree size selection 4.7.2**

As it can be observed from Figure 10, an additional tree size is brought up by lowering the complexity parameter. Both relative error (lower line) and cross-validated relative error (upper line) continue to drop with tree size. The first two trees (three-leaf and five-leaf) are identical to the ones build in analysis 4.7.1 and are consequently not included in this section. The eight-leaf tree is presented in Figure 11.

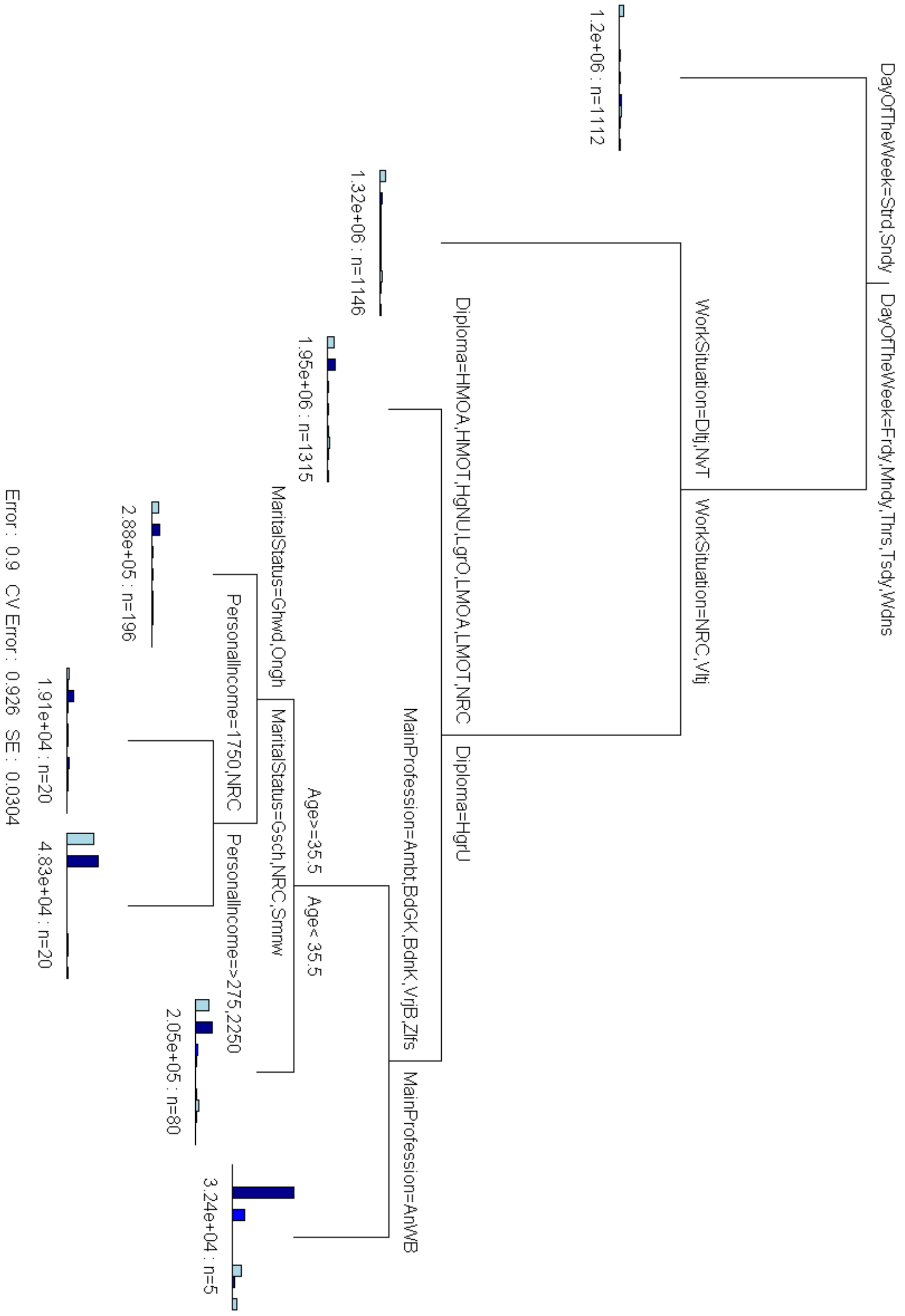

**Figure 11: Multivariate regression tree 4.7.2**

***Split 1 (Day of the week)*** *→ Strd = Saturday, Sndy = Sunday, Frdy = Friday, Mndy = Monday, Thrs = Thursday, Tsdy = Tuesday, Wdns = Wednesday*

***Split 2 (Work situation****) → Dltj = part time, NvT = not applicable, Vltj = full time, NRC = no value*

***Split 3 (Diploma)*** *→ HMOA = higher secondary education: general, HMOT = higher secondary education: technical or vocational, HgNU = higher education but not university, LMOA = lower secondary education: general, LMOT = lower secondary education: technical or vocational, NRC = no value, HgrU = higher university education*

***Split 4 (Main profession)*** *→ Ambt = public servants or public officials, BdGK = clerk but no executive, Bdnk = executive clerk, VrjB = liberal profession, Zlfs = self-employed person, freelancer, independent worker; AnWB = other (professionally)*

***Split 5 (Age)*** *→ Age >= 35.5, Age < 35.5*

***Split 6 (Marital staus)*** *→ Ghwd = married , Ongh = unmarried, Gesch = divorced ,NRC = not defined , Smnw = cohabitating*

***Split 7 (Personal income)*** *→ 1750 = €1750-€2250, NRC = not defined, >275 = > €2750, 2250 = €2250-2750*.

**Table 11: Travel behaviour in different segments 4.7.2**

| | **Activity type** | **Segment** | | | | |
|---|---|---|---|---|---|---|
| | | **1 (n=1112)** | **2 (n=1146)** | **3 (n=1315)** | **4<br>(n=196)** | **5 (n=20)** |
| 14 | activity at home | 10,88 | 12,27 | 14,22 | 15,48 | 4,50 |
| 15 | sleeping | 0,65 | 0,73 | 0,56 | 0,23 | 1,50 |
| 16 | working | 0,97 | 3,98 | 17,95 | 17,63 | 16,00 |
| 17 | services | 0,36 | 1,35 | 0,61 | 0,82 | 2,25 |
| 18 | eating | 1,87 | 0,93 | 1,94 | 2,45 | 0,00 |
| 19 | daily shopping | 0,98 | 1,23 | 0,60 | 1,22 | 3,50 |
| 20 | shopping | 3,04 | 2,24 | 1,68 | 2,42 | 2,25 |
| 21 | education | 0,30 | 1,81 | 0,41 | 1,12 | 0,00 |
| 22 | social activities | 5,40 | 2,91 | 2,32 | 3,16 | 6,00 |
| 23 | leisure activities | 5,34 | 3,58 | 3,77 | 3,24 | 2,75 |
| 24 | bring-get activities | 1,87 | 2,61 | 2,52 | 4,06 | 2,50 |
| 25 | touring | 0,64 | 0,35 | 0,27 | 0,59 | 0,00 |
| 26 | other | 2,19 | 1,67 | 1,25 | 1,20 | 1,25 |

|  | Activity type | Segment |  |  |
|---|---|---|---|---|
|  |  | 6 | 7 | 8 |
|  |  | (n=20) | (n=80) | (n=5) |
| 14 | activity at home | 62,00 | 31,56 | 0,00 |
| 15 | sleeping | 0,00 | 0,31 | 1,00 |
| 16 | working | 71,50 | 38,31 | 141,00 |
| 17 | services | 0,00 | 0,06 | 0,00 |
| 18 | eating | 0,00 | 4,25 | 28,00 |
| 19 | daily shopping | 0,00 | 1,38 | 0,00 |
| 20 | shopping | 1,00 | 0,38 | 0,00 |
| 21 | education | 0,00 | 0,25 | 0,00 |
| 22 | social activities | 0,00 | 1,19 | 1,00 |
| 23 | leisure activities | 2,25 | 7,06 | 22,00 |
| 24 | bring-get activities | 1,50 | 2,19 | 6,00 |
| 25 | touring | 0,00 | 0,19 | 0,00 |
| 26 | other | 3,00 | 0,56 | 11,00 |

In comparison to the five-leaf tree, this tree contains three additional segments and splits. The three extra splits are based on ‘age’, ‘marital status ‘and ‘income’. The tree explains 10% (Error = 0.9) of travel behaviour variance, which is 2.2% more than the five-leaf tree. The cross-validated relative error is 0.926. The variance explained by the total tree remains rather low despite the size of the MRT. Again, travel behaviour varies most for the ‘working’ category and the ‘activity at home’ category (Table 10). Differences regarding other activity types are small between the eight segments. Segment six, seven and eight are all defined by high levels of travel time for working.

Table 12: Tabulation of travel behaviour variance 4.7.2

| Act. | Split 5 | Split 6 | Split 7 | Tree total | Travel behaviour total |
|---|---|---|---|---|---|
| 14 | 1.811521e-01 | 1.861845e-01 | 0.5871238780 | 1.532906176 | 27.020737 |
| 15 | 2.976494e-08 | 1.597667e-04 | 0.0003995550 | 0.001440091 | 1.212687 |
| 16 | 2.804219e-01 | 4.025556e-01 | 0.5469907978 | 8.043768086 | 29.461298 |
| 17 | 6.895000e-04 | 5.620782e-05 | 0.0008989988 | 0.012566326 | 1.594089 |
| 18 | 5.210607e-03 | 3.538086e-03 | 0.0000000000 | 0.088788298 | 3.483115 |
| 19 | 4.005170e-06 | 1.629148e-04 | 0.0021753550 | 0.007702718 | 1.250498 |
| 20 | 3.883296e-03 | 3.761098e-04 | 0.0002774688 | 0.026081482 | 4.462151 |
| 21 | 4.937837e-04 | 7.432439e-04 | 0.0000000000 | 0.030515749 | 2.389991 |
| 22 | 4.026507e-03 | 1.572483e-05 | 0.0063928800 | 0.125129376 | 7.275304 |
| 23 | 1.653803e-02 | 3.228657e-04 | 0.0000443950 | 0.082846982 | 10.235350 |
| 24 | 2.451708e-03 | 2.493998e-03 | 0.0001775800 | 0.017854485 | 5.460018 |
| 25 | 9.535376e-05 | 2.030868e-04 | 0.0000000000 | 0.001867386 | 1.507522 |
| 26 | 6.679264e-04 | 5.058704e-04 | 0.0005438388 | 0.020330478 | 4.647239 |
| total | 4.956347e-01 | 5.973180e-01 | 1.1450247471 | 9.991797634 | 100.000000 |

Table 11 presents the explained variance for the additional splits (5,6 & 7) for each activity type. The previous splits are not displayed because their values are already displayed in Table 9. The total tree explains 10% of total travel behaviour variance. The total variance is almost fully comprised by activity type ‘activity at home’ (1.53%) and ‘working’ (8.04%). As can be seen from Table 11, the category for which most variance is explained is again ‘working’. 27.3% (8.04 of 29.46%) of this activity type is explained by this model. Variance explained of the other variables remains low.

### *4.7.3 Weekly travel time per activity type (cp=0.05)*

In this section, daily travel time for the different activity types is no longer used as measurement for travel behaviour. Instead, weekly travel time will be used. Weekly travel time is a broader measurement which makes no distinction between days of the week and weekends. The explain variables are: age, gender, function in household, marital status, diploma, personal income, main profession, work situation, shift work, working hours and driver’s license. The complexity parameter is set on 0.05. The tree size is picked interactively (xv = ‘pick’). The number of cross-validations is set on 50 so R can pick the tree that is most consistently produced.

| | |
|---|---|
| **Response variables** | Weekly travel time per activity type |
| **Explain variables** | day of the week, age, gender, function in household, marital status, diploma, personal income, main profession, work situation, shift work, working hours and driver’s license |
| **complexity parameter (cp)** | 0.05 |
| **Tree size (xv)** | ‘pick’ → Interactively |
| **multiple cross-validations (xvmult)** | 50 |

The tree size selection graph is presented in Figure 12.

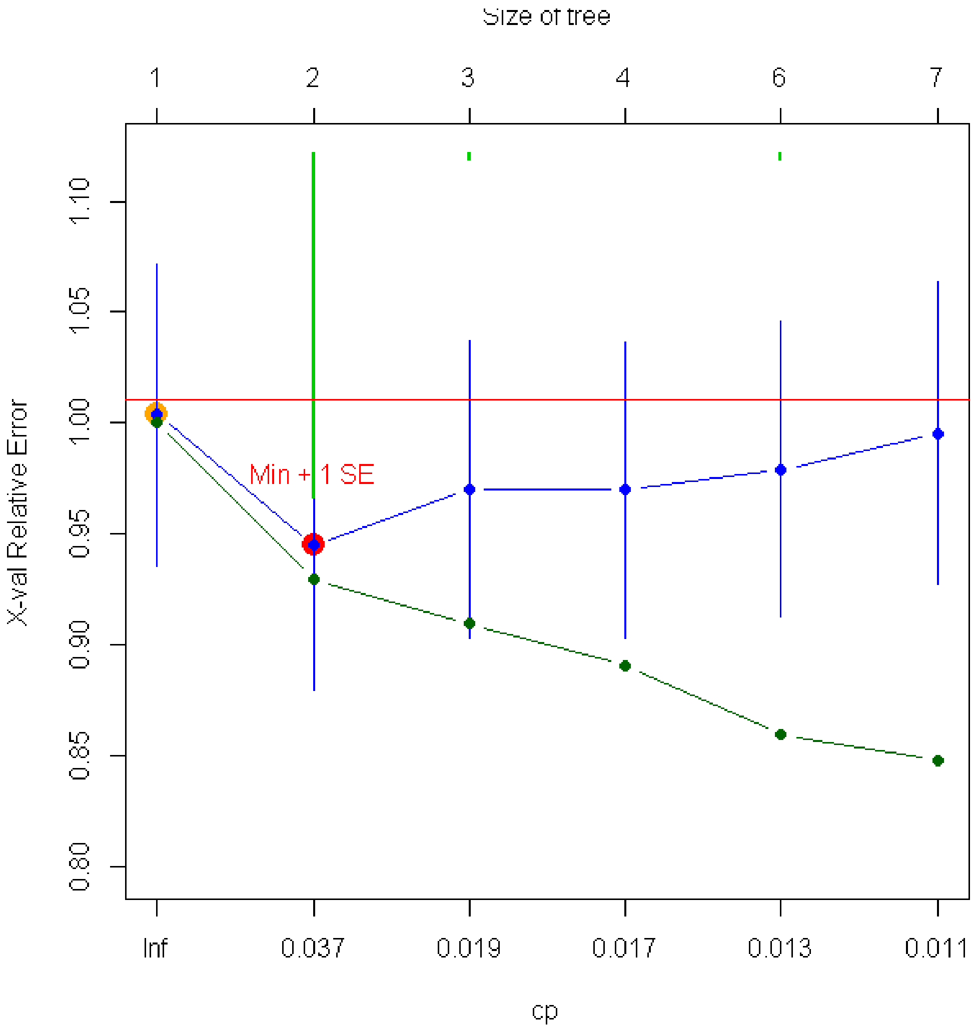

**Figure 12: Tree size selection 4.7.3**

This study was initially interested in daily activity patterns (daily travel time). The use of weekly travel time as measurement has the disadvantage of not making a distinction between days of the week and the weekend. This results in a less accurate description of activity-travel behaviour of the different segments.

When looked at relative error, all trees built in this analysis explain the variance in travel behaviour slightly better compared to the trees build in section 4.7.1 and 4.7.2. However, the cross-validated relative error increases when the tree size is higher than two. This was not the case in previous MRT analyses where daily travel time was used. The CVRE tends to increase to a value of one, which implies that the tree is a poor predictor. Despite the high CVRE, this study will take a look at the seven-leaf model for two reasons. First, the explained variance by the tree is the highest so far. Second, the seven-leaf tree does contain information about relevant socio-demographic (explain) variables. The explain variables found by this tree, could form the starting point for future research. The tree is presented in Figure 13.

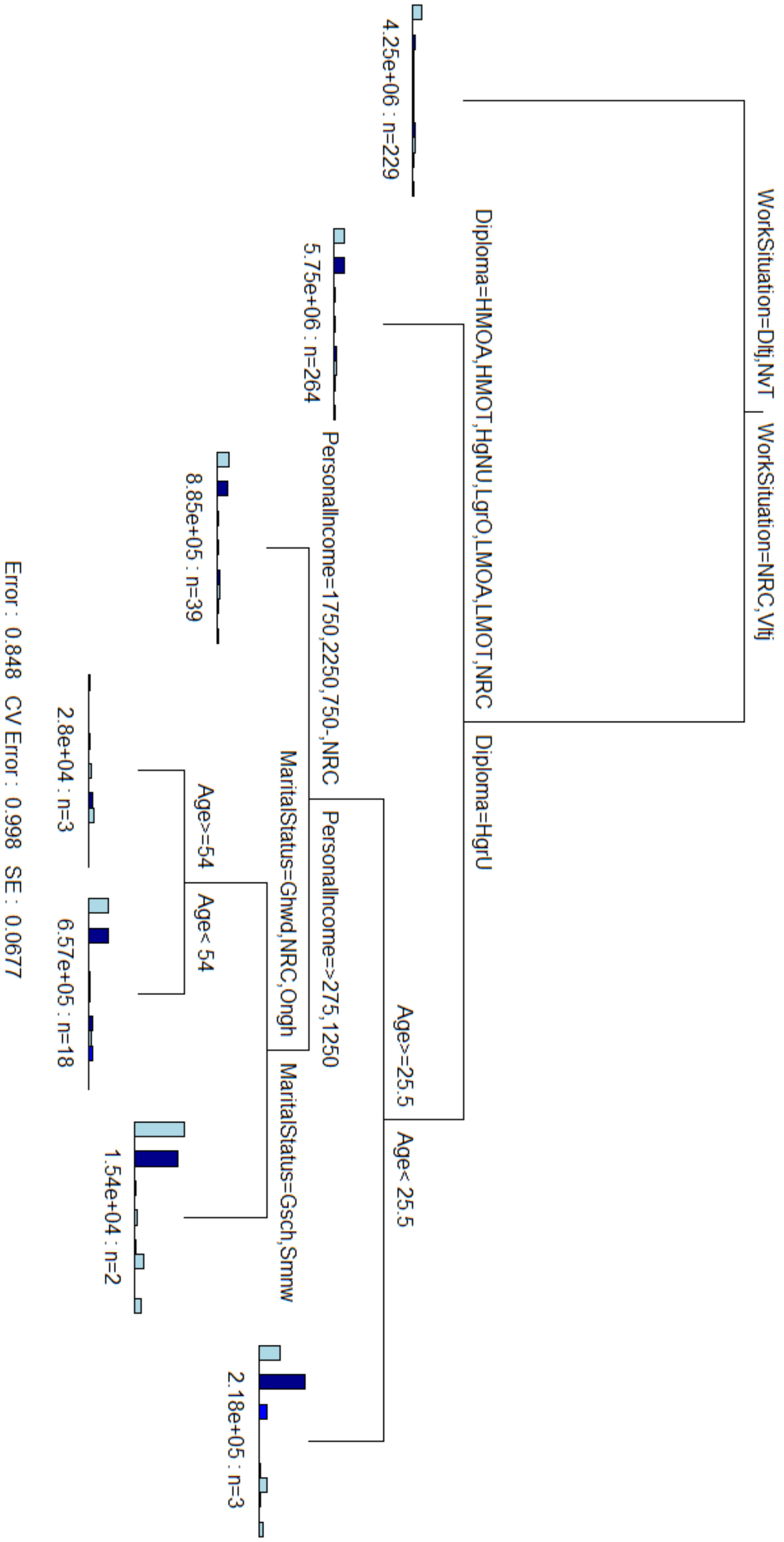

Figure 13: Multivariate regression tree 4.7.3

***Split 1 (Work situation)*** *→ Dltj = part time, NvT = not applicable, Vltj = full time, NRC = no value*

***Split 2 (Diploma)*** *→ HMOA = higher secondary education: general, HMOT = higher secondary education: technical or vocational, HgNU = higher education but not university, LMOA = lower secondary education: general, LMOT = lower secondary education: technical or vocational, NRC = no value, HgrU = higher university education*

***Split 3 (Age)*** *→ Age >= 25.5, Age < 25.5*

***Split 4 (Personal income)*** *→ 1750* = €1750-€2250, NRC = not defined, >275 = > €2750, 2250 = €2250-€2750

***Split 5 (Marital staus)*** *→ Ghwd = married , Ongh = unmarried, Gesch = divorced ,NRC = not defined , Smnw = cohabitating*

***Split 6 (Age)*** *→ Age >= 54, Age < 54*

**Table 13: Travel behaviour in different segments 4.7.3**

| | Activity type | Segment | | | | |
|---|---|---|---|---|---|---|
| | | 1 (n=229) | 2 (n=264) | 3 (n=39) | 4 (n=3) | 5 (n=2) |
| 14 | activity at home | 79,82 | 92,92 | 115,30 | 193,30 | 460,00 |
| 15 | sleeping | 4,67 | 4,15 | 5,39 | 1,67 | 0,00 |
| 16 | working | 20,92 | 92,20 | 103,70 | 416,70 | 392,50 |
| 17 | services | 7,56 | 3,66 | 4,23 | 0,00 | 0,00 |
| 18 | eating | 8,14 | 13,48 | 20,13 | 73,33 | 12,50 |
| 19 | daily shopping | 7,95 | 5,15 | 7,05 | 0,00 | 5,00 |
| 20 | shopping | 15,94 | 14,81 | 19,87 | 8,33 | 27,50 |
| 21 | education | 9,35 | 2,97 | 4,74 | 0,00 | 0,00 |
| 22 | social activities | 23,30 | 22,92 | 26,28 | 15,00 | 12,50 |
| 23 | leisure activities | 26,38 | 30,81 | 29,23 | 75,00 | 84,80 |
| 24 | bring-get activities | 16,09 | 15,95 | 20,13 | 20,00 | 0,00 |
| 25 | touring | 3,21 | 2,63 | 3,33 | 0,00 | 0,00 |
| 26 | other | 11,79 | 11,10 | 10,26 | 43,33 | 62,50 |

| | Activity type | Segment | |
|---|---|---|---|
| | | 6 (n=3) | 7 (n=18) |
| 14 | activity at home | 13,33 | 179,40 |
| 15 | sleeping | 5,00 | 0,56 |
| 16 | working | 0,00 | 180,80 |
| 17 | services | 0,00 | 5,83 |
| 18 | eating | 13,33 | 8,61 |
| 19 | daily shopping | 0,00 | 12,78 |
| 20 | shopping | 28,33 | 13,06 |
| 21 | education | 0,00 | 4,44 |
| 22 | social activities | 35,00 | 40,28 |
| 23 | leisure activities | 50,00 | 29,72 |
| 24 | bring-get activities | 0,00 | 41,67 |
| 25 | touring | 10,00 | 0,56 |
| 26 | other | 3,33 | 6,39 |

The seven-leaf tree produced by the MRT analysis using weekly travel time per activity type has splits based on work situation, diploma, age, personal income, marital status and (again) age. When compared to the eight-leaf tree in section 4.7.2, different splits are made. In this tree 'day of the week' and 'main profession' were not used as basis for splits (day of the week was not used as possible explain variable). A different split value was used for age as well (35.4 vs. 54). According to the relative error (RE = 0.848), the tree explains 15.2% of travel behaviour variance. On the contrary, the cross-validated relative error is almost one for this tree. Note that the segment sizes are smaller than in previous trees. This can be attributed to the fact that the size of the data set was divided by seven when daily travel times were summed to become weekly travel times. 'Activity at home' and 'working' remain the activity types for which travel times vary most across the segments.

**Table 14: Tabulation of travel behaviour variance 4.7.3**

| Act | Split 1 | Split 2 | Split 3 | Split 4 | Split 5 |
|---|---|---|---|---|---|
| 14 | 0.5083216354 | 9.224122e-01 | 5.828293e-02 | 4.653063e-01 | 1.214348e+00 |
| 15 | 0.0003664683 | 7.769686e-05 | 9.268866e-05 | 1.919214e-03 | 1.858753e-05 |
| 16 | 6.4315611802 | 9.904156e-01 | 1.684186e+00 | 5.376883e-01 | 7.397882e-01 |
| 17 | 0.0140088559 | 9.310769e-05 | 3.897646e-04 | 1.162297e-05 | 3.278840e-04 |
| 18 | 0.0397202077 | 1.076766e-02 | 6.706413e-02 | 1.159399e-02 | 1.355031e-04 |
| 19 | 0.0049028158 | 2.877592e-03 | 1.418043e-03 | 1.189575e-03 | 4.646881e-04 |
| 20 | 0.0002261715 | 3.996656e-03 | 2.144567e-03 | 1.322436e-03 | 1.971951e-03 |
| 21 | 0.0367196655 | 4.561169e-04 | 3.754626e-04 | 1.663667e-04 | 1.903363e-04 |
| 22 | 0.0008626677 | 1.680984e-02 | 4.825276e-03 | 1.232718e-02 | 9.577966e-03 |
| 23 | 0.0253598136 | 4.372313e-03 | 3.754626e-02 | 6.919851e-03 | 3.950240e-02 |
| 24 | 0.0023378485 | 2.765141e-02 | 4.652849e-04 | 1.618538e-02 | 1.672877e-02 |
| 25 | 0.0003267022 | 1.107107e-07 | 1.545158e-04 | 2.640868e-04 | 4.758407e-05 |
| 26 | 0.0002559997 | 3.044571e-04 | 2.217761e-02 | 3.906609e-05 | 4.193811e-02 |
| Total | 7.0649700320 | 1.980235e+00 | 1.879123e+00 | 1.054933e+00 | 2.065040e+00 |

| Act | Split 6 | Tree total | Travel behaviour total |
|---|---|---|---|
| 14 | 0.5096013112 | 3.678272679 | 34.4905165 |
| 15 | 0.0003648112 | 0.002839466 | 0.6895604 |
| 16 | 0.6039349260 | 10.987574227 | 33.8493759 |
| 17 | 0.0006284443 | 0.015459679 | 1.0462348 |
| 18 | 0.0004118376 | 0.129693325 | 2.2228842 |
| 19 | 0.0030153924 | 0.013868107 | 0.8471947 |
| 20 | 0.0043107571 | 0.013972538 | 2.3143716 |
| 21 | 0.0003648112 | 0.038272759 | 2.4880163 |
| 22 | 0.0005144408 | 0.044917365 | 4.7245789 |
| 23 | 0.0075940576 | 0.121294689 | 7.4037400 |
| 24 | 0.0320634822 | 0.095432172 | 6.2912798 |
| 25 | 0.0016473505 | 0.002440350 | 0.7183095 |
| 26 | 0.0001724303 | 0.064887673 | 2.9139374 |
| Total | 1.1646.240.521 | 15.208.925.029 | 100.0000000 |

The variances of activity types 'activity at home' and 'working' again comprise the majority of total travel behaviour variance with respectively 34.5% and 33.8%. These values are even higher than in the analyses in section 4.7.1 and 4.7.2. Almost half (7.1 of 15.2%) of total travel behaviour variance is explained by the first split which is based on work situation. The best explained activity type is again 'working' having 32.5% (10.99 of 33.85%) of its variance explained.

### *4.7.4 Weekly travel time per activity type (cp=0.005)*

In this analysis, weekly travel times per activity type are again used as response variables. The explain variables are the same as in the previous analyses (4.7.3): age, gender, function in household, marital status, diploma, personal income, main profession, work situation, shift work, working hours and driver's license. The complexity parameter is set on 0.005, so more segments and relevant explain variables can be identified. The tree size is picked interactively (xv = 'pick'). The number of cross-validations is set on 50 so R can pick the tree that is most consistently produced

| | |
|---|---|
| **Response variables** | Weekly travel time per activity type |
| **Explain variables** | day of the week, age, gender, function in household, marital status, diploma, personal income, main profession, work situation, shift work, working hours and driver's license |
| **complexity parameter (cp)** | 0.005 |
| **Tree size (xv)** | 'pick' → Interactively |
| **multiple cross-validations (xvmult)** | 50 |

The tree size selection graph is presented in Figure 14.

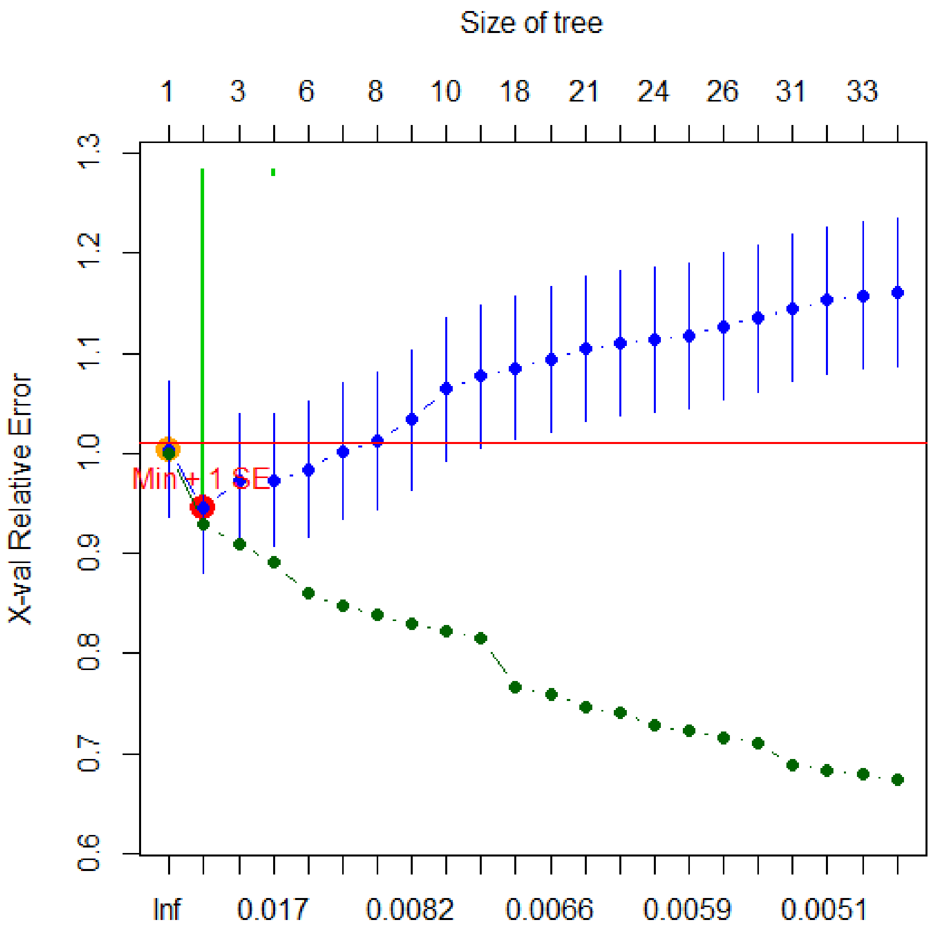

**Figure 14: Tree size selection 4.7.4**

Figure 14 confirms the trends noticed in the tree size selection graph of section 4.7.3 (Figure 12). The relative error continues to decrease and the cross-validated relative error continues to increase with tree size. When tree size is eight, the CVRE reaches a value of one. From then on the CVRE is always higher than one. For this reason no trees will be further investigated.

*4.7.5 CART analysis using Bray–Curtis dissimilarity*

None of the multivariate regression trees produced in the previous analyses (4.7.1 - 4.7.4) was able to explain a considerable amount of travel behaviour variance. For this reason, this study also takes a look at the use of Bray-Curtis dissimilarity for building multivariate regression trees. This statistic has proven to be an effective measure with non-normal data. Bray-Curtis is a modified Manhattan measurement, where the summed differences between the variables are standardized by the summed variables of the objects. (Faith et al. 1987, De'ath 1999). The MRT analyses based on the Bray Curtis measurement are also performed using the mvpart package, but a different model is added, namely ‘gdist’. This model has to be incorporated in the mvpart model to build trees based on Bray-Curtis dissimilarity. The mvpart model was explained earlier in the introduction of section 4.7. The same principles apply regarding the arguments. The gdist model is given by (M. Therneau & Atkinson, 2012).

```
gdist(x, method=‘bray’, keepdiag=FALSE, full=FALSE, sq=FALSE)
```

To compare the two measures (Euclidean distance measure & bray-Curtis dissimilarity), exactly the same design for the corresponding analyses is used as in section 4.7.1, 4.7.2, 4.7.3 and 4.7.4 regarding response variables, explain variables, complexity parameter (cp), tree size selection (xv) and number of cross-validations (xvmult). Full explanation of the different designs is therefore not repeated here. This can be found in the corresponding analysis in section 4.7.1, 4.7.2, 4.7.3 or 4.7.4. Furthermore, only the results that are noteworthy are included in this report.

#### ***A) Daily travel time per activity type (cp = 0.05)***

The same design as in section 4.7.1 is used here, except for the use of Bray-Curtis dissimilarity instead of Euclidian distance measures.

The tree size selection graph is presented in Figure 15.

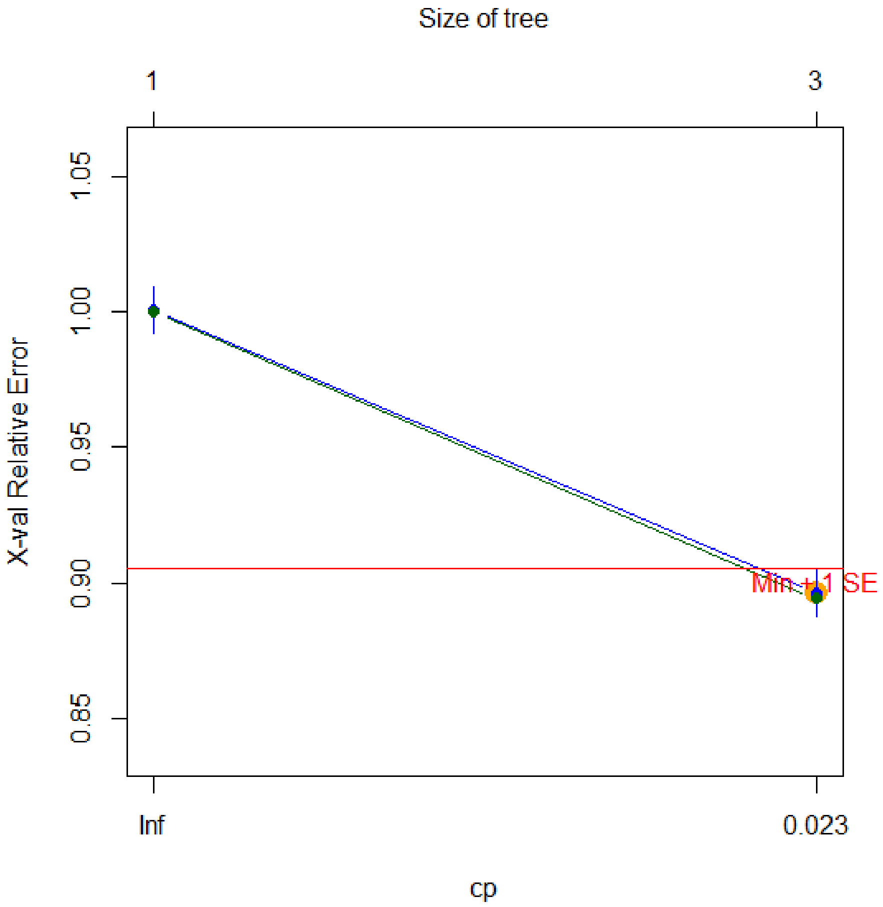

**Figure 15: Tree size selection 4.7.5 (A)**

As can be seen from Figure 15, the relative error and the cross-validated relative error practically fall together. Both errors are lower compared to the corresponding three-leaf tree based on Euclidean distance measures build in section 4.7.1. The multivariate regression tree is presented in Figure 16.

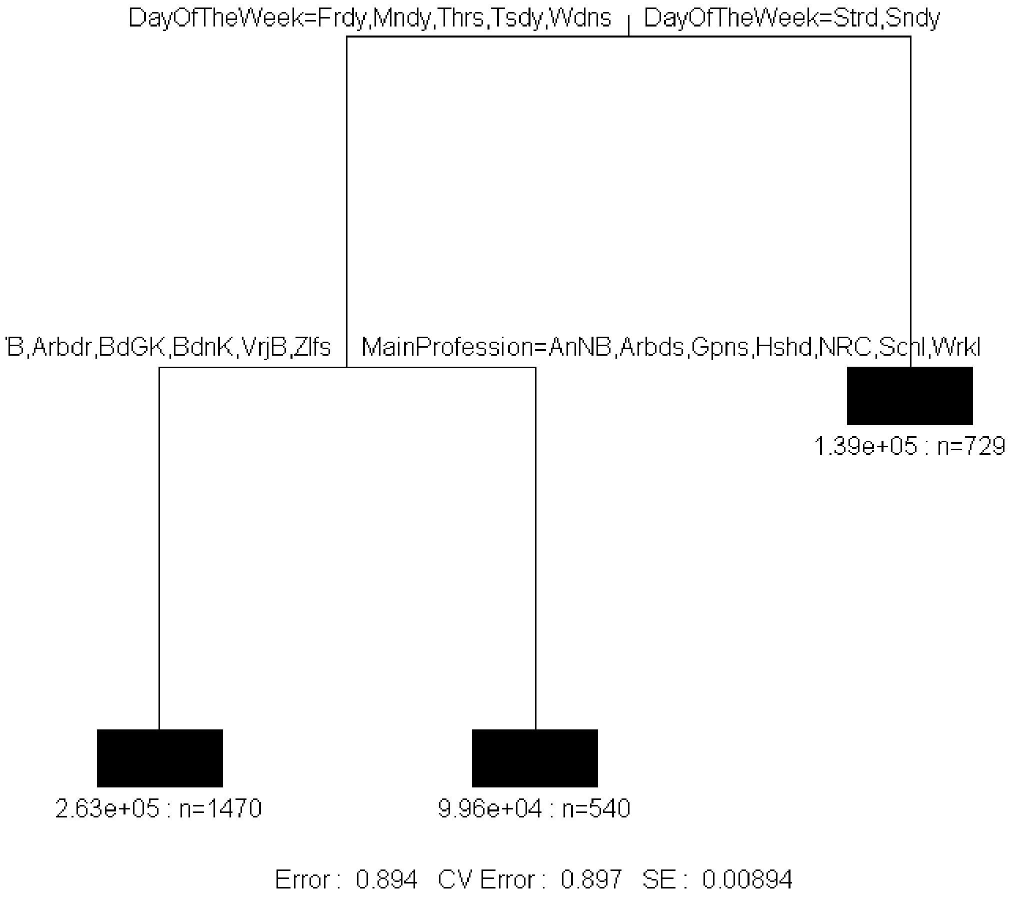
DayOfTheWeek=Frdy,Mndy,Thrs,Tsdy,Wdns
DayOfTheWeek=Strd,Sndy
'B,Arbdr,BdGK,BdnK,VrjB,Zlfs
MainProfession=AnNB,Arbds,Gpns,Hshd,NRC,Schl,Wrkl
1.39e+05 : n=729
2.63e+05 : n=1470
9.96e+04 : n=540
Error : 0.894 CV Error : 0.897 SE : 0.00894

**Figure 16: Multivariate regression tree 4.7.5 (Bray-Curtis)**

***Split 1 (Day of the week)*** *→ Strd = Saturday, Sndy = Sunday, Frdy = Friday, Mndy = Monday, Thrs = Thursday, Tsdy = Tuesday, Wdns = Wednesday*

***Split 2 (Main profession)*** *→ Ambt = public servants or public officials, BdGK = clerk but no executive, Bdnk = executive clerk, VrjB = liberal profession, Zlfs = self-employed person, freelancer, independent worker, AnWB = other (professionally)*

The three-leaf tree produced by the MRT analysis has splits based on 'day of the week' and 'main profession'. The splits of the three-leaf tree in section 4.7.1 were based on 'day of the week' and 'work situation'. The tree based on Bray-Curtis dissimilarity explains 10.6% of the travel behaviour variance (RE = 0.894). The cross-validated relative error is only slightly higher than the relative error (CVRE = 0.897). Both errors are lower than in section 4.7.1, but remain high.

As can be observed from Figure 16, a total of 2739 observations are divided into three segments. 1155 Observations were deleted during the analysis due to missingness. These observations are the days where no activity was performed and subsequently the travel times for all activities were zero for a particular individual. These observations are deleted when using Bray-Curtis dissimilarity.

When Bray-Curtis dissimilarity is used for building MRT's, R only produces limited information. No travel time averages for the different segments are shown and the table with the variance explained at each split for every activity type is not retrievable. Future R packages could provide a solution here.

***B) Daily travel time per activity type (cp = 0.005)***

The same design as in section 4.7.2 is used here, except for the use of Bray-Curtis dissimilarity instead of Euclidian distance measures. However, lowering the complexity parameter did not lead to more possible trees. The same output as in Figure 15 was produced.

***C) Weekly travel time per activity type (cp = 0.05)***

The same design as in section 4.7.3 is used here, except for the use of Bray-Curtis dissimilarity instead of Euclidian distance measures.

The tree size selection graph is presented in Figure 17.

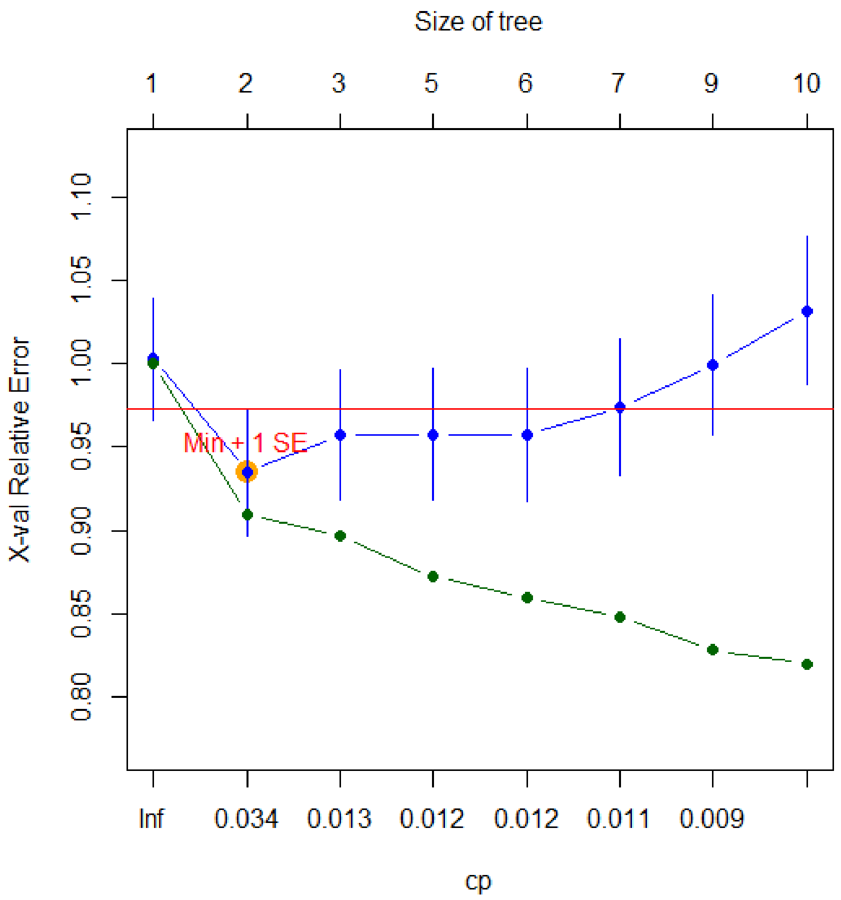

Figure 17: Tree size selection 4.7.5 (C)

Compared to the results in section 4.7.3, the trees built using Bray-Curtis dissimilarity are not able to explain more travel behaviour variance. Both relative error and cross-validated relative error are similar. Subsequently the different trees will not be analyzed in greater detail.

***D) Weekly travel time (cp = 0.005)***

The same design as in section 4.7.4 is used here, except for the use of Bray-Curtis dissimilarity instead of Euclidian distance measures.

The tree size selection graph is presented in Figure 18.

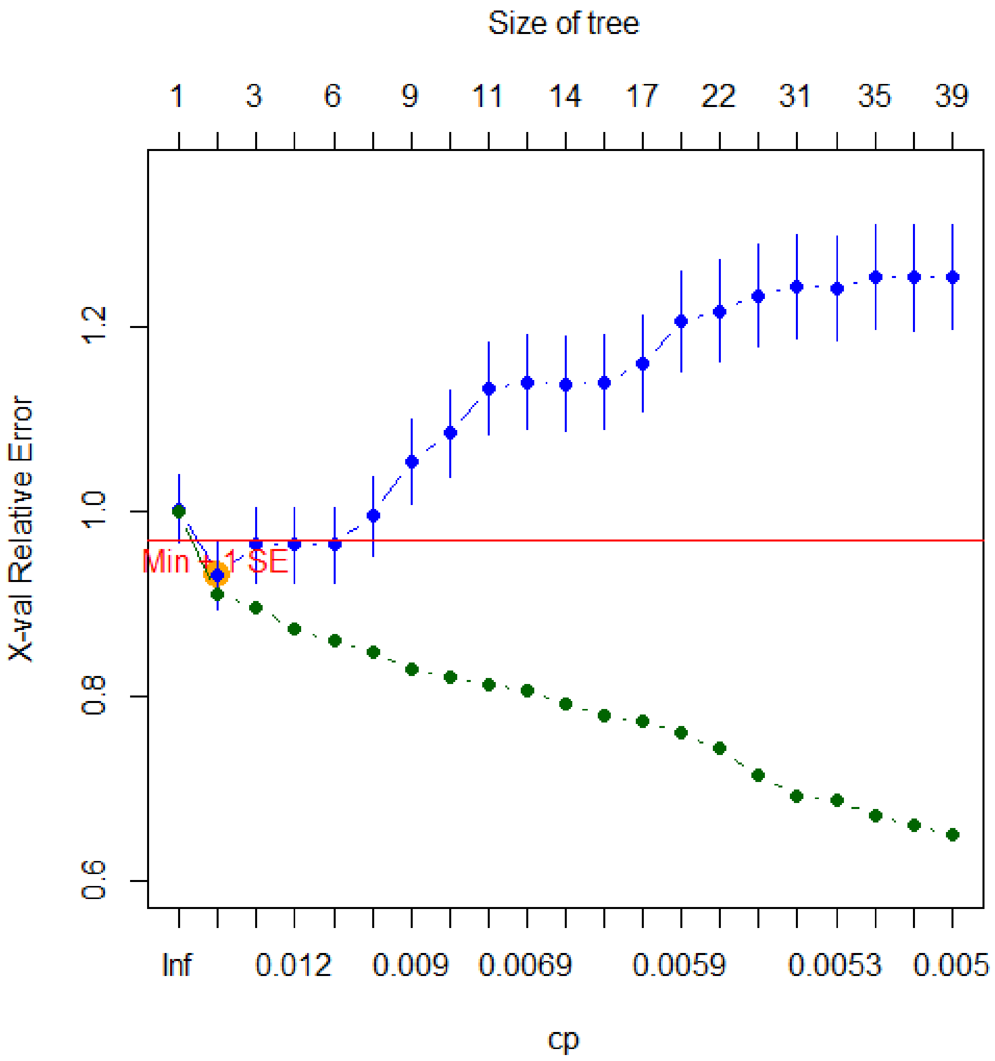

**Figure 18: Tree size selection 4.7.5 (D)**

Lowering the complexity parameter did lead to more possible trees. The trees, however, are not able to explain more travel behaviour variance than the corresponding trees in section 4.7.4. Both relative error and cross-validated error show similar trends compared to the values in section 4.7.4. No trees are further investigated.

# 5. DISCUSSION

## 5.1 Interpretation and model performance

The main question of this thesis was if common characteristics in activity-travel behaviour for different population segments in Flanders could be identified. Furthermore this study aimed at taken socio-demographic variables and day of the week into account. Based on TRANSMIMS, multivariate regression tree analysis was chosen as statistical method.

Various multivariate regression trees (MRT's) were built using different measurements for travel behaviour (daily travel time vs. weekly travel time), complexity parameters (0.05 vs. 0.005) and distance measures (Euclidean distance vs. Bray-Curtis dissimilarity) in order to find the optimal tree.

In general, explained travel behaviour variance was low in all MRT's. Three MRT's were produced using daily travel time as measurement for travel behaviour. The three-, five- and eight-leaf trees respectively explained 5.8%, 7.8% and 10% of travel behaviour variance. The MRT's built using weekly travel time tended to explain a little more variance of travel behaviour, but suffered from a high cross-validated relative error. Furthermore, weekly travel time was less favoured as measurement for travel behaviour because it makes no distinction between weekends and days of the week. Only one tree was further investigated here. The seven-leaf tree explained 15.4% of total travel behaviour variance, but had a cross-validated relative error of almost one. In the final section of the analysis, Bray-Curtis dissimilarity was used as distance measure. Again, this study found no tree explaining a decent amount of travel behaviour variance.  In addition, limited information about the trees was provided by R when Bray-Curtis dissimilarity was used.

Twelve activity types were used in the analysis to describe travel behaviour. The contribution of activity types and how well travel time for each activity type was explained by the tree was tabulated for every tree. According to these results, travel time for the activity types 'working' and 'activity at home' comprised the major part of total travel behaviour variance. When daily travel time was used as measurement, 'working'' and 'activity at home' respectively comprised 27.0% and 29.5% of total travel behaviour variance. These values were even higher when weekly travel time was used as measurement, with respectively 34.5% and 33.8%.

The different MRT's were not able to separate daily and weekly travel time for the different activity types, besides 'working' and to a lesser extent 'activities at home'. The three-, five- and eight-leaf trees using daily travel time respectively explained 17.7%, 23.1% and 27.3% of the variance in daily travel time for working and 1.1%, 2.1% and 5.6% of the variance in daily travel time for activities at home. The seven-leaf tree using weekly travel time explained 32.5% of the variance in weekly travel time for working and 10.7% of the variance in weekly travel time for activities at home.

Different variables (socio-demographic and day of the week) were used as basis for splits.  Generally speaking, the different multivariate regression trees only managed to explain a significant amount of variance in daily and weekly travel time for the activity type 'working'. That is why only cautious conclusions regarding this activity type are drawn. The dominant split variables in the MRT analysis using daily travel time were 'day of the week' and 'work situation', explaining respectively 2.5% and 3.2% of travel behaviour variance. According to those MRT's, daily travel time for working was less on Saturday and Sunday. Furthermore, the trees suggested that people who work full time spend more time travelling for work than people who work part time or do not work at all. This confirms the results found by Lu & Pas (1999). Other split variables that were used in these analyses were: diploma, main profession, age, marital status and personal income. However, these splits did not manage to explain a significant amount of variance. The dominant split variable in the seven-leaf MRT using weekly travel time was 'work situation', explaining 7.1% of travel behaviour variance. The seven-leaf MRT also suggested that people who work full time spend more time travelling for work than people working part time or not working at all. The other (weaker) split variables were: diploma, age, personal income, marital status and (again) age. They explained between 1.1% and 2.1% of travel behaviour variance. The split variables for the three-leaf MRT using Bray-Curtis similarity were 'day of the week' and 'main profession'.

## 5.2 Constraints and recommendations

In this research project, multivariate regression tree analysis was used because this technique allows to simultaneously incorporate travel behaviour, socio-demographic variables and day of the week. As noted earlier, the trees built in the analyses did not explain the hoped-for variance in travel behaviour. A possible explanation could be an absence of segments having distinct travel behaviour patterns. Future research could be conducted focusing on revealing segments based on travel behaviour without incorporating other variables, i.e. socio-demographic variables and day of the week. If different segments can be identified, the emphasis of research could be switched back to socio-demographic, day of the week or other (see next paragraph) variables.

A total of twelve possible explain variables were used in the MRT analyses (day of the week, age, gender, function in household, marital status, diploma, personal income, main profession, work situation, shift work, working hours and driver's license). The selection of these specific variables was based on the survey. These variables failed to explain a decent amount of variance in travel behaviour. The use of additional variables could lead to better results. Two variable categories come to mind: household composition and demographics. The existence of a relationship between household composition and travel behaviour was proven earlier by Ryley (2006) and Dieleman et al. (2002). It is also not illogical to presume that demographic variables influence travel behaviour. For example, home-work distance and environment (city, rural area...) could have an influence on travel behaviour. Future research including all these types of variables (socio-demographic, day of the week, household composition and demographics) could definitely be interesting.

This study used (daily and weekly) travel time for different activity types as a measure for travel behaviour. Several reasons for this choice were mentioned in section 4.3 of this thesis (tangibility, understandability, correspondence to survey and focus on travel part). Travel time however does not fully encompass travel behaviour. Two individuals can have exactly the same daily travel time values for all the activity types, but the sequence in which those activities are performed could be different. Likewise, two individuals spending the same weekly travel time on the same activity types could have a different travel behaviour pattern. Activities could be performed during different days or using a different sequence. The use of a more complex measure for travel behaviour could be adopted in future research.

Besides the use of a more complex measure, a simple alternative could also be used, namely: activity time, i.e. total time an individual spends on a specific activity (type). In comparison to travel time, this measure also takes the duration of the actual activity into account. Unfortunately, the survey did not provide enough information to use this measure.

The activity types were also based on the survey (activity at home, sleeping, working, services (e.g. going to a doctor, eating, daily shopping, shopping (non-daily goods), education, social activities, leisure activities, bring-get activities, touring (driving around for pleasure, walking around for pleasure),other). According to the results, travel behaviour variance was largely comprised by travel time for working and activities at home. The use of different or less categories could be recommended for future research. Various activity types, each explaining a little bit of travel behaviour variance, could be merged into fewer activity types. Factor analysis could provide a solution here.

Travel information of every individual was collected during seven consecutive days (one week). Travel information regarding a longer period could lead to better results. For some activity types (services, social activities, shopping (non daily goods)) the use of a longer observation period is recommended because these activities are often not performed every week.

According to the results of this study, variance in travel time for working and activities at home comprised the majority of total travel behaviour variance. Accordingly, future research exclusively concentrated on these two activity types could be interesting.

## 5.3 General conclusion

This thesis aimed at revealing different population segments in Flanders based on travel behaviour, socio-demographic variables and the day of the week. Multivariate regression analysis was chosen as statistical method for this study.

Travel time for different activity types was used as a measure for travel behaviour. A number of multivariate regression trees (MRT's), using different explain variables and distance measures, were build throughout this research. Overall, these MRT's were characterized by high errors and failed to form well separated population segments.  Cautious conclusions could only be drawn for the activity type 'working'.

The multivariate regression trees did not produced the hoped-for results. However, multivariate regression tree seems an interesting method for analyzing travel behaviour. The incorporation of explain variables and response variables provides an interesting approach. The incorporation of different or more variables could lead to interesting results. Furthermore, indications brought up by the different MRT analyses in this master thesis could form the starting point of future research.

# References

Bergstad, C. J., Gamble, A., Hagman, O., Polk, M., Gärling, T., & Olsson, L. E. (2011). Affective–symbolic and instrumental–independence psychological motives mediating effects of socio-demographic variables on daily car use. *Journal of Transport Geography*, *19*(1), 33–38.

Best, H., & Lanzendorf, M. (2005). Division of labour and gender differences in metropolitan car use: An empirical study in Cologne, Germany. *Journal of Transport Geography*, *13*(2), 109–121.

Bhat, C. R., & Koppelman, F. S. (1999). A retrospective and prospective survey of time-use research. *Transportation, 26*(2)*,* 119-139.

Boarnet, M. G., & Sarmiento, S. (1997). Can land-use policy really affect travel behaviour? A study of the link between non-work travel and land-use characteristics. *Urban Stud June, 35*(7)*,* 1155-1169.

Borcard, D., Gillet, F., & Legendre, P. (2011). *Numerical Ecology With R*. New York, USA: Springer

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees* (1st ed.). Belmont, USA: Wadsworth International Group.

Buliung, R. N., & Kanaroglou, P. S. (2007). Activity–travel behaviour research: conceptual issues, state of the art, and emerging perspectives on behavioural analysis and simulation modelling. *Transport Reviews*, *27*(2), 151–187.

Chapter 1: Transims overview. (n.d.). Retrieved November 18, 2011, from http://tmip.fhwa.dot.gov/resources/clearinghouse/docs/transims_fundamentals/ch1.pdf

Chapter 4: Activity Generator. (n.d.). Retrieved November 15, 2011, from http://tmip.fhwa.dot.gov/resources/clearinghouse/docs/transims_fundamentals/ch4.pdf

De'ath, G. (2002). Multivariate trees: a new technique for modelling species-environment relationships. *Ecology, 83*(4)*,* 1105-1117.

Dieleman, F. M., Dijst, M., & Burghouwt, G. (2002). Urban form and travel behaviour: micro-level household attributes and residential context. *Urban Studies (Routledge)*, *39*(3), 507–527.

F. Hair, Jr., J., C. Black, W., J. Babin, B., & E. Anderson, R. (2010). *Multivariate Data Analysis: a global perspective*. New Jersey, USA: Pearson Education Inc.

Kitamura, R., & Fujii, S. (1996). Two computational process models of activity-travel behaviour. *Theoretical Foundations of Travel Choice Modeling,* 251-279

Lu, X., & Pas, E. I. (1999). Socio-demographics, activity participation and travel behavior. *Transportation Research Part A: Policy and Practice*, *33*(1), 1–18.

M. Therneau, T., & Atkinson, B. (2012). Package "mvpart." Retrieved from http://cran.r-project.org/web/packages/mvpart/mvpart.pdf

MORA. (2009). Mobiliteitsrapport van Vlaanderen 2009. Retrieved 4, 2012, from http://www.serv.be/uitgaven/1556.pdf

Newbold, K. B., Scott, D. M., Spinney, J. E. L., Kanaroglou, P., & Páez, A. (2005). Travel behavior within Canada's older population: a cohort analysis. *Journal of Transport Geography*, *13*(4), 340–351.

Ouelette, M.-H., & Legendre, P. (2011). Package "MVPARTwrap." Retrieved May 10, 2012, from http://cran.r-project.org/web/packages/MVPARTwrap/MVPARTwrap.pdf

Polk, M. (2003). Are women potentially more accommodating than men to a sustainable transportation system in Sweden? *Transportation Research Part D: Transport and Environment*, *8*(2), 75–95.

Polk, M. (2004). The influence of gender on daily car use and on willingness to reduce car use in Sweden. *Journal of Transport Geography*, *12*(3), 185–195.

Proost, S., Driesen, J., Loeckx, A., Nemery, B., Steenbergen, T., Tampère, C., Van den Bulck, E., et al. (2011). Traffic mobility in Flanders. Retrieved February 6, 2012, from http://www.kuleuven.be/metaforum/docs/pdf/wg_7_e.pdf

Recker, W. W. (1995). The household activity pattern problem: General formulation and solution. *Transportation Research Part B: Methodological*, *29*(1), 61–77.

Roger N., S. (1966). Metric structures in ordinal data. *Journal of Mathematical Psychology*, *3*(2), 287–315.

Ryley, T. (2006). Use of non-motorised modes and life stage in Edinburgh. *Journal of Transport Geography*, *14*(5), 367–375.

Shiftan, Y., Ben-Akiva, M., Proussaloglou, K., de Jong, G., Popuri, Y., Kasturirangan, K., & Bekhor, S. (2003). Activity-based modelling as a tool for better understanding travel behaviour. Paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne.

Shiftan, Y., & Suhrbrier, J. (2002). The analysis of travel and emission impacts of travel demand management strategies using activity-based models. *Transportation, 29*(2), 145-168.

Statistics Belgium. (2010). Afgelegde afstanden in het verkeer. Retrieved 4, 2012, from http://statbel.fgov.be/nl/statistieken/cijfers/verkeer_vervoer/verkeer/afstand/

The R Project for Statistical Computing. (n.d.). Retrieved May 2, 2012, from http://www.r-project.org/

Timofeev, R. (2004). Classification and Regression Trees (CART) Theory and Applications. Retrieved November 29, 2011, from http://edoc.hu-berlin.de/master/timofeev-roman-2004-12-20/PDF/timofeev.pdf

TRANSIMS Training Course at TRACC Part 1. (2009). Retrieved December 28, 2011, from http://transims.googlecode.com/svn/v4/trunk/documentation/training/TRANSIMS%20-%204%20-%20Activities%20(final).pdf

# Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
**Identification of common characteristics in activity-travel behavior based on activity-travel diaries**

Richting: **Master of Management-Management Information Systems**
Jaar: **2012**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

**Ceunen, Michaël**

Datum: **8/06/2012**