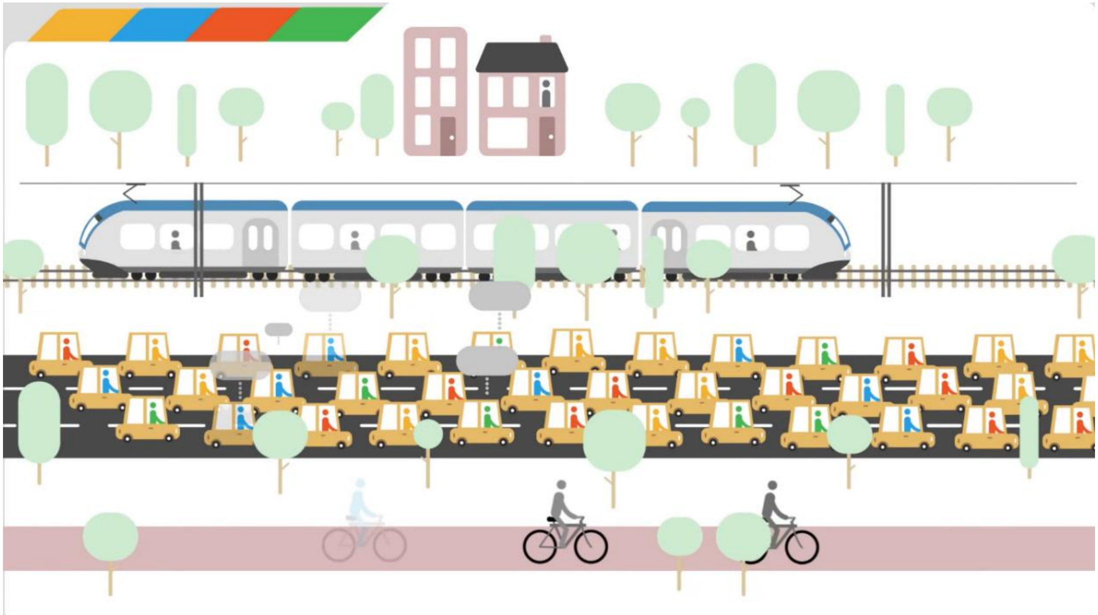


Developing a Population Synthesis Method based on Lifestyles towards Mobility for Travel Demand Modelling

Fieke Beemster

November 2016

Title page image from TNO Fountain promotional video.

https://youtu.be/-A-Ba8m2Ieo

Developing a Population Synthesis Method based on Lifestyles towards Mobility for Travel Demand Modelling

Master thesis

Fieke Isabelle Beemster
Student number 1547097

Master's programme: Transport, Infrastructure & Logistics
Faculty of Civil Engineering and Geosciences
Delft University of Technology

Graduation Committee

Chair:	Prof. Dr. ir. Bart van Arem Section Transport and Planning
1st supervisor:	Dr. Dimitris Milakis Section Transport and Planning
2nd supervisor:	Dr. ir. Sander van Cranenburgh Section Transport and Logistics
TNO supervisor:	Prof. Dr. ir. Erik de Romph TNO, Section Transport and Planning

Preface

This document represents the result of my Master Thesis Project of the Master program Transport, Infrastructure and Logistics at the Delft University of Technology. The thesis has been conducted in collaboration with the company TNO, department Smart Urban Mobility and Safety in Delft. The research focuses on a first step to improve travel demand modelling with which effect from new emerging transport technologies can be estimated. A model named Fountain which combines agent based modelling, discrete choice modelling and behavioral aspects in one simulation tool is improved. In this project the population synthesis for this simulation tool has been improved. Firstly I would like to thank my professors from TU Delft, namely Bart van Arem, Dimitris Milakis and Sander van Cranenburgh for their help, cooperation and advises during the entire project. Of course I would like to thank my colleagues and fellow-interns at TNO but I want to give a special thank to Erik de Romph, my supervisor at TNO who helped me through all the processes of my thesis. I am looking forward to work together at TNO. I would like to thank Selmar Smit from TNO for all his hours of help with debugging Fountain. I learned a lot in those hours. Then I would like to thank Anton Hagens from the municipality of Amsterdam for giving me permission to use the VMA and I would like to thank Luuk Brederoode from Goudappel Coffeng for his expertise on population synthesis. Of course I would like to thank my family for supporting me during my study in Delft. I could not have made it without them. At last a very very special thank you to Jurriaan who has been supporting me every step of the road.

Fieke Beemster

September, 2016

Contents

List of Figures

List of Tables

Summary

Introduction

Innovations in the field of technology have seen an exponential growth in the last decades. The same can be said about technologies in mobility. All modalities became faster, lighter, cleaner and safer and are still being improved. Nowadays, a more complicated system of transport is arising. Concepts as for example ‘mobility as a service’ and ‘automated vehicles’ have the potential to completely change the transport system as we know it. In the past, we have seen what mayor influence these kind of innovations had on the transport system. What the impact of the upcoming innovations exactly is, is the most interesting question.

This research first provides an overview of the current methods to address impacts of mobility: the main findings show that the current methods do not meet the needs for modelling new mobility concepts. However, a model named Fountain is being developed. This model uses lifestyles towards mobility as predictor of travel behavior instead of generic socio-demographic characteristics. In this freshly developed model, multiple simplifications are currently being used, that in their turn currently limit the usability of the model. In this research the focus is on eliminating one of the simplifications: The method of population synthesis. Population synthesis is the procedure of using microdata from a reference sample, combined with aggregate data to estimate the population of agents for a certain geographical area. This simplification was chosen due to the following reason: the population synthesis phase creates the basis of the model. When the population is not a usable representation of the actual population, an error is introduced that will impact every step of the model. Fountain currently assumes that a “lifestyle towards mobility” can better predict transportation choices than socio-demographic characteristics, which are commonly used in normal traffic models. Fountain currently bases the distribution of lifestyles upon one study. In this particular study, seven lifestyles are defined and the distribution of those lifestyles for a sample of the Dutch city of Utrecht is given. Currently, Fountain uses this distribution for every simulation run towards any city or geographical area. It is however expected that this distribution in reality is different for a city center than for a suburb area. Following this hypothesis, the main research question is:

How can we improve population synthesis, taking into account the lifestyles towards mobility that are used in Fountain as predictor and thereby improve travel demand modelling?

To answer this question, several sub questions have been used.

1. *Which population synthesis algorithm is most suitable to apply as a basis in the population synthesis tool?*
2. *Which studies are available in the field of lifestyles towards mobility and are there patterns that justify the use of this data?*
3. *How can a population synthesis tool be constructed that is able to construct the amount of people traveling between zones with a certain lifestyle based on the socio-demographic characteristics age and gender?*
4. *Is the population estimated by the population synthesis tool an improvement compared to the original population in Fountain?*

Literature study on population synthesis algorithms

A literature study towards different population synthesis algorithms is presented. Several algorithms have been found from which the Iterative Proportional Fitting (IPF) and Iterative Proportional Updating (IPU) are seen as the most promising. Based on several criteria, the IPF algorithm is chosen as a basis for the new population synthesis phase.

Meta-analysis on lifestyles towards mobility

To get from the socio-demographic data (the data that is available for all zones in the Netherlands), to the distribution over lifestyles for different zones, a tested method of convergence has to be established. Currently, the lifestyles in Fountain are based upon a study from Anable (2011). In this study the relation between different socio-demographic characteristics and lifestyles is given, but the relation is not strong enough to use without any further notice. For this reason a meta-analysis of lifestyles towards mobility was undertaken. This analysis resulted in seven relevant studies. In all studies several lifestyles have been identified. The underlying characteristics of the lifestyles in the different studies have been compared and resulted in the categorization in four overall groups in which all different lifestyles from the studies are classified. It was also expected that there is a certain relation between socio-demographic characteristics and lifestyles. Such a relation is given in some of the studies and have been compared. In all studies, similar patterns were found with respect to the age and gender within the different categories of lifestyles. Even though four more socio-demographic characteristics have been analyzed, the lack of data on these characteristics resulted in their exclusion for the tool as constructed in this thesis. However, the fact that the studies had no real contradictions is seen as promising and gives no reason for not using the data from the original study from Anable. (2011). As a result the data from Anable (2011) is used as input for the population synthesis tool. It forms the basis to estimate a distribution of lifestyles based on age and gender.

The population synthesis tool

An extensive description of the constructed population synthesis tool is given. Three steps have been distinguished. The first step is the creation of a population with IPF Multizone: a population is estimated which presents the amount of people living in a zone and the distribution of those people over gender and age classes. In the second step, the lifestyles are assigned to the population based upon the relation between lifestyles and the socio-demographic characteristics of gender and age as presented by Anable (2011). The third step in the population synthesis tool is assigning the workplace (destination) to each person in de database. This is done by using an existing OD matrix, in this case extracted from the traffic models available for the case studies. Then multiply this matrix by the distribution of lifestyles that was computed in the lifestyle assignment. The output of the population synthesis tool are seven OD matrices. The matrices together represent a detailed population for the whole geographical area with living places, working places (origin and destination) and lifestyle per person.

Case studies to measure the performance of the population synthesis tool

To measure the performance of the population synthesis tool, two case studies have been performed, ‘Amsterdam VMA’ and ‘Utrecht rush hour avoidance’. The first case study, ‘Amsterdam VMA’, was chosen based on the available actual data on the population of Amsterdam, this is used to compare the performance of the IPF Multizone algorithm used in the population synthesis tool. The second case study, ‘Utrecht rush hour avoidance’, was chosen based on the earlier implementation in the model Fountain and the available actual data of the project rush hour avoidance. The output of the IPF Multizone, the lifestyle assignment and the destination assignment, has been analyzed.

Case study Amsterdam VMA

The first case study is ‘Amsterdam VMA’. In Amsterdam VMA the IPF Multizone method and the lifestyle assignment have been analyzed. We found that the IPF Multizone method in the population synthesis tool performs as expected by literature and that the tool performs less diverse than comparable studies, which results in a more robust prediction of the population. Furthermore a mean absolute percentage error was calculated based on the comparison between the distribution over lifestyles towards mobility in the synthetic population and the average distribution of Anable (2011). When this value is zero, there is no difference between the overall distribution of the city and the distribution of the separate zones (such as the city center). The mean absolute percentage error is found for every lifestyle and they range between 2,05% and 9,38%. This indicates that the distribution over lifestyles towards mobility in the synthetic population is, as expected, different between zones.

Case study Utrecht rush hour avoidance

The second case study is ‘Utrecht rush hour avoidance’, in this case study the lifestyle assignment and the destination assignment in the population synthesis tool have been evaluated. The analysis of the lifestyle assignment resulted in a range of the mean absolute percentage error between 1,6% and 9,0%. Again, this indicates that the distribution over lifestyles towards mobility in the synthetic population is, as expected, different between zones. A more extensive analysis was performed on the average lifestyle distribution of different areas in the study area. Based on the synthetic population from the population synthesis tool, the conclusion was drawn that explainable distributions towards mobility were captured with the population synthesis tool for typical areas. When we used the population created by the population synthesis tool in Fountain and compared the output with the actual data of the project rush hour avoidance and to the output of Fountain with the original synthetic population in Fountain, we saw that the overall amount of avoidances is overestimated by the model, but that the type of avoidance is better predicted by a synthetic population and even slightly more better predicted with a lower persons per agent rate. The modality choice per zone is better predicted with the synthetic population of the tool when compared to the expectations of the analyzed zones.

Conclusions

Based on the results, the conclusion is made that the population synthesis tool performs better compared to the original method that Fountain used to estimate a population. The estimated population by using the population synthesis tool is more accurate, has a more deviated distribution of lifestyles per zone compared to the average and performs better in the case ‘rush hour avoidance’. Population synthesis while taking into account the lifestyles towards mobility is improved. As a final result, a new population synthesis method has been constructed, which can be used as a basis in travel demand modelling.

1. Introduction

Innovations in the field of technology have seen an exponential growth in the last decades. With this growth, the accessibility has seen a growth, which resulted in more trips and activities (e.g. work, school, leisure) per person, further from home (Annema, Brink, & Walta, 2012), (CBS, 2015). Nowadays new mobility concepts are emerging such as the automated vehicle or “mobility as a service”. In this thesis, a first step is done to improve methods to assess impacts from new mobility concepts.

While only a century ago, cars, trains and buses were seen as new and maybe dangerous modalities, nowadays a worldwide acceptance is created for these modes. In the meantime, new emerging transport technologies have taken over this place of the unknown. For example, automated vehicles are seen as the newest improvement in the field of mobility, or the new concept “mobility as a service” (Transportation Research Board, 2016), where mobility is introduced as a door to door intermodal service instead of separate services. These emerging transport technologies and new mobility concepts, will have impact on how we see the future of mobility. In the past we have seen which immense impacts these innovations had on the transport system. For this reason new impacts are to be expected.

When we look for example at automated driving, several studies have been done in the field of assessing impacts, see Schultz van Haegen (2014), Department of Transport (2015a), Arnaout & Bowling (2011), van Arem, van Driel & Visser (2006), Shladover, Dongyan & Lu (2012), Litman (2014), Snelder, van Arem, Hoogendoorn & van Nes (2015), Corwin et al. (2015) and Fagnant & Kockelman (2015). Most studies are positive, Corwin et al. (2015) gives the following advantages of automated vehicles:

- The first advantage comes from a safety perspective. Automated vehicles hardly ever crash, autonomous operation removes the cause of almost all accidents: human error (NHTSA, 2014).
- The second advantage is in terms of congestion. Traffic jams will be rarities, thanks to sensors allowing for less space between vehicles and guidance systems with real-time awareness of congestion.
- Energy demand drops, since smaller weight allows cars to be propelled by more compact, efficient, and environmentally friendly powertrains.
- Trip costs decrease because of higher utilization rates of the vehicle and less fuel is needed.
- Infrastructure is funded by charges for actual usage, since connected-car technology allows systems to precisely calculate personal road use.
- Parking lots disappear, as the rise of autonomous-drive and carsharing models diminish parking needs.
- Law enforcement ceases to concern itself with traffic, since autonomous vehicles are programmed not to violate traffic laws.
- Speed of deliveries quickens and costs decrease through the rise of fully autonomous networks of longhaul trucks that can operate for more extended periods of time and cover longer distances with lower labor costs.

A more critical image is given by Puyleart (2016), this research presents an opposing force. The simulations of Puyleart show that the personal benefits enabled by automated vehicles, such as the ability to do something else while driving and a reduced fuel consumption lead to an increased use of the car, which will lead to more emissions, noise pollution and has a negative influence on safety. The extra car traffic leads to a decrease for all other modes, since public transport users will make the switch to car use. Extra trips by car will also result to higher CO2 or other emissions. The same applies to the vehicle loss hours, the vehicle loss hours are a way to measure the total number of loss hours as a consequence of traffic jam. The vehicle loss hours increases a little. Under 40% penetration of automated vehicles on the road there are no advantages, from there on the advantages will rise. The same penetration rate of 40% is mentioned by Milakis, Arem & Wee (2015), a 40% penetration rate of Coöperative Adaptive Cruise Control (CACC) appears to be a critical threshold for realizing significant benefits on capacity.

Conclusions of these studies do not agree. What is definite about this matter, is that the technologies are improving and that manufacturers start producing. The transport system has to cope with the circumstances, while impacts for the network as a whole could be both positive and negative. For policy makers to prepare for these circumstances, it is very important to understand the impacts of emerging transport technologies. In this chapter we will give an overview of multiple methods to assess impact of transport in section 1.1, a method is chosen and will be explained in section 1.2 and 1.3. In section 1.4 the objective and the research questions of this master thesis are given, in section 1.5 the scope of the research is given and in section 1.6 an overview of the entire report is given.

1.1 Methods to model impacts of transport

To understand which impacts these new mobility concepts can have on the transport system, different methods are used to model impacts. In Figure 1 an overview of methods is given.

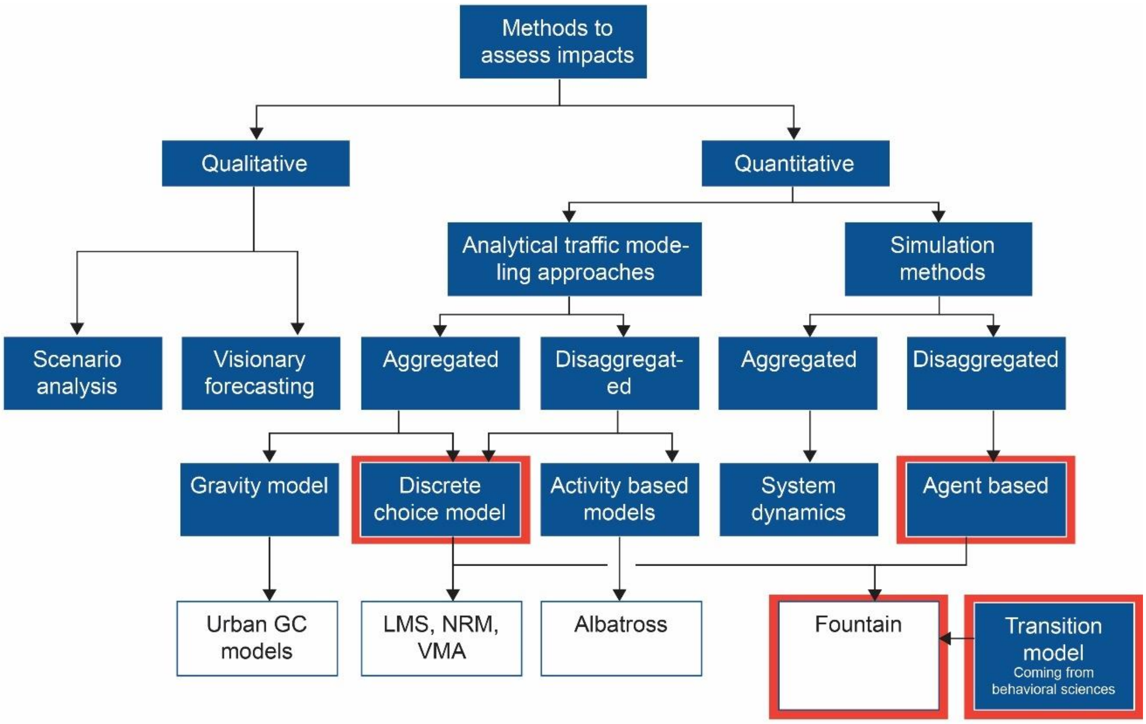


Figure 1 State of the art travel demand modelling tools

Figure 1 shows a decision tree with methods. For every method a short explanation is given. Different qualitative methods to assess impacts have been identified. First of all the scenario analysis which is used to outline possible futures in a certain field. Scenario analysis is the process of estimating the expected future after a given period of time, most important unknowns are diversified and give possible future scenarios. Expected is, that the future lies in one of the scenarios or between the boundaries of the scenarios (Kahn, 1962). Another qualitative method to assess impacts from new technologies is visionary forecasting or expert estimations (Chambers, Mullick, & Smith, 1971). In this method, experts estimate the likelihood that a certain event will occur and at what time horizon. There are different opinions about this method. According to Armstrong (1985) experts estimations can be reliable up to a certain point. When the expected growth is linear, experts are able to grasp the impact. However, growth in transportation is exponential, exponential growth is hard to handle even by experts. Also according to Wagenaar & Sagaria (1975), forecasts of exponential growth are highly conservative. The qualitative research methods give a first insight in impacts. The studies realized in this field are (KiM, 2015), (Milakis, Snelder, Arem, Wee, & Correira, 2015), (Gruel & Stanford, 2015) and (Zmud, Tooley, Baker, & Wagner, 2015). The studies mentioned give insights in possible futures of automated driving and mobility as a service. Still a very broad spectrum of futures is left. For this reason a more quantitative method is required.

In the right side of the tree, quantitative methods can be found. Different methods are identified, starting with analytical transport models. In the field of analytical transport models a distinction can be made between aggregated models and disaggregated models. To understand the different analytical transport models, first an explanation is needed on the difference between aggregated models and disaggregated models. In aggregated models, the data needed is on the zonal scale. In disaggregated models the data needed is on the individual scale. In practice disaggregated methods give more specific results. Disadvantage is the amount of data necessary for this methods compared to the aggregated models. An aggregated model is for example the gravity model, the gravity model is the prevailing framework with which to predict population movement. It illustrates the macroscopic relationships between places (for example home and work). Distance, time and cost between two locations have a negative effect on movement between the two places while the amount of activities have a positive effect on movement between the two places (Isard, 1956). An example of a disaggregated method is activity based modelling (for example Albatross). Activity based models predict for individuals where and when specific activities (e.g. work, leisure, shopping) are conducted. Travel demand is derived from activities that people wish to perform (McNally, 2000). Another method are discrete choice models (LMS, NRM & VMA), they can be both aggregated and disaggregated. Discrete choice models describe and predict choices between two or more discrete alternatives. In transport modelling this is used to describe and explain for example the mode choice (McFadden, 1975).

Besides the analytical transport models, simulation models can be used. System Dynamics is a method that uses causal relations, feedback loops, information flows and accumulations to assess behaviour of a system based upon differential equations (Forrester, 1971). Another simulation method is agent based modelling. An agent-based model (ABM) is one of a class of computational models for simulating the actions and interactions of autonomous agents (both individual or collective entities such as organizations or groups) with a view to assessing their effects on the system as a whole (Grimm, 2005). Some of these methods have been used to assess impacts from automated driving. For example system dynamics by Puyleart (2016) and agent based modelling by CPB (2015), Hwan Kim et al. (2015) and Fagnant & Kockelman (2015).

A challenge with these studies is that impacts and effects of emerging transport technologies are largely based upon behaviour of people. To predict behaviour and choices, travel demand modelling tools use population data with socio-demographic characteristics to predict which choices the population in the models will make with respect to, for example, modalities and their departure time. A full dataset of the to be replicated population is almost never available. For this reason, “population synthesis” methods have been constructed to estimate a synthetic population based on aggregate data to obtain disaggregate data. Based on the socio-demographic characteristics on the disaggregate level, travel behavior is predicted. According to Hunecke et al. (2011), lifestyles towards mobility have a higher predictive power than socio-demographic characteristics. However datasets containing lifestyles towards mobility are almost never available and there is no population synthesis method that can estimate lifestyles towards mobility based on the available socio-demographic data.

In order to improve general travel demand modelling and to make a first step towards obtaining insights in impacts from new mobility concepts, a new population synthesis method was constructed in this thesis. This population synthesis method uses the same input as the other methods, however its output is an distribution over lifestyles towards mobility instead of generic socio-demographic characteristics.

In the overview with travel demand modelling methods (see Figure 1), one of the listed methods is the simulation method “Fountain”. Fountain is a combination between ABM, discrete choice models and a transition model coming from behavioural sciences (transition study is a fairly new field of research. A transition can be defined as a specific type of social change that is characterized by non-linearity, a long time frame and a structural transformation (Bosch, 2010)). As an exception on most travel demand models, Fountain already uses lifestyles towards mobility as a predictor of travel behavior. For this reason Fountain has been chosen as the specific method to be directly improved. Furthermore, Fountain is used to test the constructed population synthesis tool during the case studies. In the following two sections, more information about Fountain can be found.

1.2 Fountain

Fountain is a simulation method that combines ABM, discrete choice modelling and a transition model. Fountain creates virtual agents that have a certain lifestyle towards mobility. With these lifestyles a distinction is made between several characteristics of people. The agents in Fountain all start with a default choice (modality and travel time). All default choices of the agents in the model together are the input for a traffic model “DTA Lite” that does the assignment to the network. The output of DTA Lite are the travel time, travel cost and a comfort grade for every agent for one day. After the first cycle, a new day starts and a utility value (based upon the travel time, travel costs, comfort grade, impact from a social network (ABM)) and a personal value (based on the lifestyle towards mobility of the agent) decides which choice the agent will make the next day; the default choice or another choice such as choosing another modality or departure time.

An intervention can be simulated in Fountain, this can have as a results that the agents have an extra modality choice or get a reward for a certain travel choice. The main goal of Fountain is to give insight in the consequences of these interventions. An example of an intervention is the project ‘rush hour avoidance’, where a financial reward is given to people that avoid the rush hour by travelling before rush hour, after rush hour, decide to work at home or use another modality than the car. Effects of such an intervention on both the short and long term can be modelled with Fountain. An extensive description of Fountain is given in appendix A.

In this new developed method, some simplifications are identified which are currently limiting the model. This list of simplifications can be found in appendix A. One of the simplifications is the population synthesis phase, which is the focus of this thesis. In the next section a more extensive explanation of the population synthesis phase of Fountain is given.

1.3 Population synthesis phase

Goal of this research is to improve the population synthesis phase. To understand why the population synthesis phase was chosen, first the simplified points of this part are addressed. An overview of the population synthesis phase is constructed in Figure 2.

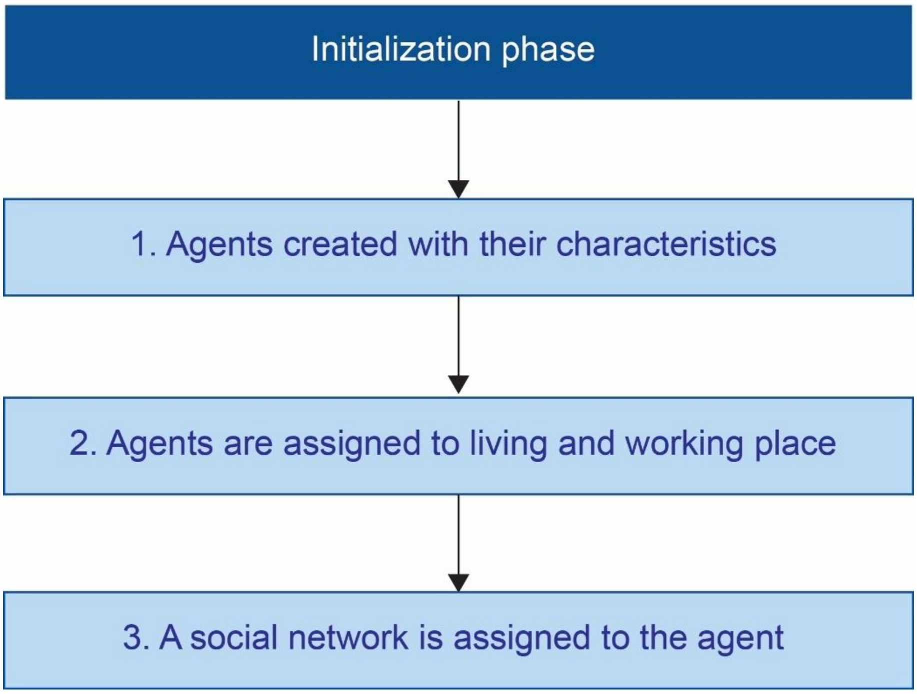


Figure 2 Overview processes in population synthesis phase

1.3.1 Creation of agents and lifestyle assignment

In the population synthesis phase of Fountain the following processes take place: First, the agents are created and a lifestyle is assigned to the agent. In Fountain it is assumed that a lifestyle towards mobility can predict choices better than socio-demographic characteristics, which are commonly used in normal traffic models. This assumption is supported by Hunecke et al. (2010). The distribution of people over lifestyles is based upon one study (Anable, 2011). In this study 7 lifestyles towards mobility are defined and the distribution of those lifestyles for a sample of Utrecht is available. The assumption is made that this distribution can be used for all situations. Because of this assumption, all zones have the same distribution of lifestyles. Expected is that the distribution is different in a city center than for example a suburb. This is the first aspect that can be improved.

1.3.2 Assignment of agents to living and working place

The second step in the population synthesis phase is the assignment of agents to living and working place. In Fountain all zones have a similar amount of agents living in the zone and similar amount of agents working in every zone. While in reality there are zones with more living places and zones with less living places. The same statement can be applied for working places. Since the data about the number of people living and working in zones is known, this can be improved too. The improvement on the amount of people living and working in a zone is the second aspect that can be improved in the population synthesis phase.

1.3.3 Assignment of social network to the agent

The third step is the assignment of the social network. Every agent has a social network of four neighbors, four colleagues and four friends. The social network is created with the following algorithm: pick 100 random agents and determine which four agents have characteristics that are closest to the agent in terms of living place for neighbors, working place for colleagues and type of lifestyle for friends.

The population synthesis phase was chosen for the following reasons: first of all, the population synthesis phase creates the basis of the model. When the modelled population is not representative for the empirical population, an error is introduced. This error has influence in every step of the model. Besides this, the population synthesis phase can be improved on more than one aspect. It can be improved on the distribution over agents on lifestyles and the amount of people living and working in a zone can be made more accurate.

1.4 Objective & research questions

The initial objective of this research was to improve a method to assess impacts of new mobility concepts such as 'mobility as a service' or 'automated vehicles'. This was done by exploring the usability of Fountain which is a simulation tool that combines agent based modelling, discrete choice modelling and behavioral aspects. In Fountain several simplifications have been found. The population synthesis phase is chosen to improve. Once improved, the population synthesis phase can be used in other travel demand models too and thereby a new population synthesis method is constructed. Therefore the main objective of this research is to improve population synthesis, taking into account lifestyles towards mobility. The main research question is as follows:

How can we improve population synthesis, taking into account the lifestyles towards mobility that are used in Fountain as predictor and thereby improve travel demand modelling?

To answer the main research question, several sub questions are posed. Behind the questions, the chapter numbers can be found where the specific question are answered.

1. *Which population synthesis algorithm is most suitable to apply as a basis in the population synthesis tool?*
2. *Which studies are available in the field of lifestyles towards mobility and are there patterns that justify the use of this data?*
3. *How can a population synthesis tool be constructed that is able to construct the amount of people traveling between zones with a certain lifestyle based on the socio-demographic characteristics age and gender?*
4. *Is the population estimated by the population synthesis tool an improvement compared to the original population in Fountain?*

1.5 Scope of the research

In this research the focus is on the population synthesis phase of Fountain. This definition applies to the creation of agents, the assignment of lifestyles to the agents and the assignment of living and working place of the agent. The other identified simplified points in Fountain are not in the scope of this master thesis.

1.6 Overview of the report

In this report, the following structure will be used. A literature study is conducted to obtain insight in population synthesis. This can be found in chapter 2. With some more knowledge about population synthesis, the approach for the rest of the research is given in chapter 3. In chapter 4, a meta-analysis on studies about lifestyles towards mobility is reported. After this, a population synthesis tool is created and described in chapter 5. To test Fountain as a tool for assessing impacts from emerging transport technologies, two case studies have been conducted on which is reported in chapter 6 (Amsterdam (VMA) and 'Utrecht rush hour avoidance'. In these case studies the estimated population is analyzed, used in Fountain and compared to other populations. In chapter 7 conclusions & recommendations are given, in chapter 8 the scientific and practical contribution are provided and finally in chapter 9 reflections upon the process are given.

2. Literature study on Population Synthesis

As described in the previous chapter, the focus of this research was on the population synthesis for transport models. Population synthesis can be described as the process of estimating characteristics of a population for a geographical area based on available data sources. The goal of the population synthesis was to estimate a population that resembles the actual population as much as possible. The main goal of this chapter is to answer the following research question:

'Which population synthesis algorithm is most suitable to apply in the population synthesis tool?'

To achieve this, section 2.1 gives a general explanation of population synthesis and the main expressions that are used in population synthesis, section 2.2 will present an overview of the available algorithms for population synthesis in literature and section 2.3 discusses which of the previously described algorithms is most applicable for the population synthesis tool.

2.1 Population synthesis

In this section a general explanation is given on population synthesis. In subsection 2.1.1 the main goal of population synthesis is explained. In subsection 2.1.2 an explanation is given on population synthesis is given.

2.1.1 Why population synthesis

The main goal of population synthesis is the creation of a population that can be used as input for another model. Agent- based simulation models for land use or transportation, simulate the behavior of agents over time. Agents often represent the individuals living in a geographical area. The models simulate decisions of agents, in order to predict future states of the system (Müller & Axhausen, 2010). Agent-based models allow for more detailed, accurate simulation and prediction of land pricing and travel demand than traditional aggregate models according to (Müller & Axhausen, 2012). The first process in an agent-based model is the definition of agents and their relationships. However, almost never a full database of the actual population is available, therefore algorithms are introduced to estimate a synthetic population.

2.1.2 What is population synthesis

In the process of population synthesis, microdata from a reference sample is combined with aggregate data to estimate the population of agents. As mentioned, almost never a full database of the actual population is available. Therefore many countries collect the required disaggregate data on a regular basis, but for only a small part of the population. This provides the so called *reference sample.* For the reference sample specific data is required on for example the amount of women in a zone, in the age class 16-25, with a household size of four people. The *aggregate data* is mostly available for the whole population and are for example the total number of men or women in a zone, without further specification on the categories age class and household size. For this population of agents, the distribution and correlation of the agents attributes are similar to those in the reference sample. The number of agents within each category matches the aggregate data. Most synthesizers work with a spatial hierarchy, where a region contains several zones and form the study area.

2.2 Algorithms for population synthesis

In the previous section, a general explanation on population synthesis has been given. Multiple algorithms can be found in the literature. In Figure 3 a decomposition is shown with respect to how they relate to each other. In this section an explanation is given on the different algorithms, starting at the left side of the figure to the right side of the figure.

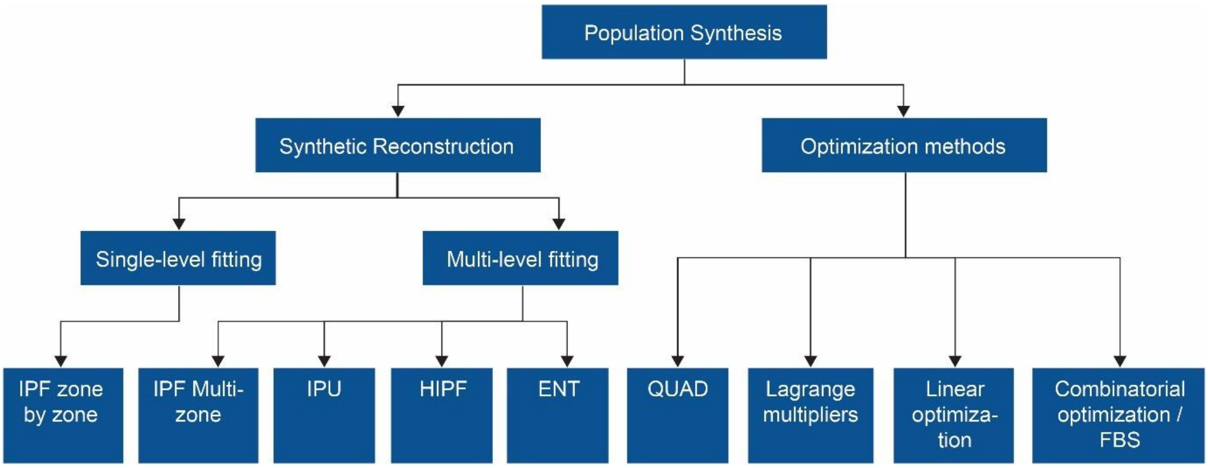


Figure 3 Decomposition population synthesis algorithms

2.2.1 Synthetic Reconstruction (SR)

Synthetic Reconstruction (Beckman, Baggerly, & McKay, 1996) in its origin is best explained as follows: the emphasis is on the creation of synthetic populations, given a known multiway table of conditional probabilities pertaining to population characteristics. Several variations of the Synthetic Reconstruction technique are available which can be both single-level fitting or multi-level fitting. The difference between single- and multilevel fitting can be explained as follows; in single-level fitting the weights obey only household-level constraints. Person-level constraints are considered when selecting households. In multi-level fitting the weights obey constraints at both person and household levels. Household selection is unconstrained (Müller & Axhausen, 2012). According to Müller & Axhausen (2012), the multi-level fitting is performing superior to the single-level fitting. It has a more complicated reweighting algorithm that is fairly new.

Iterative Proportional Fitting (IPF)

The first algorithm that belongs to the SR is the Iterative Proportional Fitting algorithm. First the IPF zone by zone algorithm and second the IPF Multizone algorithm. The IPF zone by zone is applied at zonal basis, while IPF Multizone is applied at all zones at the same time. Zone by zone only takes into account the correlation between segments on a zonal scale while Multizone also takes into account the correlation between segments and the whole population. As long as there are no 0 values in the data the algorithm always converges, which means that is gives the optimal outcome. In IPF zone by zone, the algorithm iteratively scales values in the disaggregate data, one variable at a time, such that the marginals for that variable will match those of the aggregate data. While in this procedure one zone at the time is fitted, in IPF Multizone the algorithm is used twice. First it adds up the marginals for all zones and uses the algorithm for the whole area. Marginals are created at a disaggregate level. Then these marginals divided by the total number of zones are used as reference sample. After this the IPF zone by zone algorithm is applied per zone.

Other SR multi-level fitting methods
The iterative proportional updating algorithm is created by Ye et al. (2009). It uses the following adjustment at person level: Proportionally rescale the weights to the aggregate data of all households that contain at least one person that belongs to the current category. This gives a scaling factor for a household which does not depend on the number of persons that fall into the current category (Müller & Axhausen, 2012). The HIPF algorithm is created by Müller & Axhausen (2011). This algorithm works slightly different than the IPU algorithm. Also the Entropy maximization method suggested by Bar-Gera et al. (2009), is comparable to the two other synthetic reconstruction methods.

2.2.2 Optimization methods

The algorithms that have been discussed up until here, were all synthetic reconstruction methods. All algorithms are iterative processes of rescaling the data towards the known aggregate data and the reference sample. Optimization methods are used to solve a specific optimization problem. The algorithm allows a deviation between the actual aggregate data and the marginals, but also takes minimization of this deviation into account (Brederode & Waanders, 2013). Several optimization methods have been found in literature and will be discussed in this subsection.

Quadratic Optimization (QUAD)
Quadratic optimization method (QUAD), is used in the national transport model of the Netherlands for population synthesis. The QUAD method is used to solve a specific optimization problem. It allows a deviation from the marginals, but also takes minimization of this deviation into account. With two weight factors, one for both the allowance of the deviation from the marginals and the minimization of this deviation, give a weight to which one is more important. When performing the QUAD algorithm over multiple zones a problem occurs, the total population of the synthesized population generally defers from the actual total population of the entire area as in each zone, not an exact match is required with the marginals (Brederode & Waanders, 2013).

Lagrange multipliers
With the Lagrange multipliers method, an optimization problem with constraints can be reformulated to an optimization problem without constraints. When the derivative of this optimization problem is equal to zero, an optimum is found. The advantage of the Lagrange multiplier method is that it will always converge to a global optimum. The main problem is that once the type of objective gets difficult, it might be way more complex to find a solution for this problem. Therefore the applicability of this method is less convenient (Brederode & Waanders, 2013).

Linear Programming
In linear programming, the optimization function and the constraints are reformulated to a linear problem. Since most programs can solve a linear problem way faster than a non-linear problem, this is a very interesting method. Once the optimization problem gets more complicated, it also gets more complicated to reformulate the problem to a linear problem (Brederode & Waanders, 2013).

Combinatorial optimization / Fitness Based Synthesis (CO / FBS)
The combinatorial optimization algorithm or fitness based synthesis constructs a synthetic population by subtracting a subset from a reference sample distribution, which complements

the marginals and the reference sample best. FBS searches for the best reference sample. The method replaces agents in different sets and evaluates if the objective gets smaller. This process keeps on running until the minimum for the objective is found or the number of permutations is reached (Voas & Williamson, 2000).

2.3 Which algorithm for population synthesis

To make a decision on which algorithm to use as a basis in this research, a literature review was conducted. In subsection 2.3.1 criteria are defined to assess the algorithms. In subsection 2.3.2 the algorithms are compared and in subsection 2.3.3 an algorithm is chosen.

2.3.1 Definition of criteria

The different algorithms were assessed based on three criteria. The first criteria is the score in (Brederode & Waanders, 2013) & (Ryan, Maoh, & Kanaroglou, 2009), this is a very important criteria since this reflects the performance of the method in practice. The second criteria is the applicability, the applicability is the most important criteria since the main goal was to find the most applicable algorithm. The third criteria is reaching convergence, convergence must be reached, without convergence there will be no end result.

2.3.2 Algorithms compared

In the literature, two studies have been found that compare algorithms for population synthesis. In this subsection both studies are presented.

Combinatorial optimization vs. Synthetic reconstruction algorithms

The Combinatorial Optimization techniques (Voas & Williamson, 2000) is compared to Synthetic Reconstruction in Ryan, Maoh & Kanaroglou (2009) and Huang & Williamson (2001). In Ryan, Maoh & Kanaroglou (2009), both methods are used to test their ability to re-estimate a small, complete population of firms for Hamilton, Ontario in the year 1990. Since earlier in this literature review, multiple algorithms in the Synthetic Reconstruction group were found, it is valuable to mention that in this case the IPF algorithm is used. The overall conclusion of this study was that the accuracy of the population was highly dependent on the size of the reference sample, but differences per method were found. When using the Combinatorial Optimization method, an increase in the accuracy of the population was found between a reference sample size of 1% and 5%. With a bigger reference sample than 5%, no real difference was found. For the Synthetic Reconstruction method, the largest difference was found between 1% - 2,5% and between 20% - 50%. Between 2,5% - 20% there was hardly no difference in the accuracy of the population. In the end, the research found that neither of the methods was better than the other. Combinatorial Optimization needs a slightly smaller reference sample to obtain the same results, while Synthetic Reconstruction can obtain a better result, but will need a higher reference sample size. Since data collection is costly, the Combinatorial Optimization is recommended by this study.

IPF, QUAD, FBS, Lagrange multipliers and linear programming compared

In 2013, a study is performed in the Netherlands by Breederode & Waanders (2013). Herein several methods are compared based on literature and partly in practice. The algorithms identified are the already mentioned IPF algorithm, the quadratic optimization algorithm (QUAD), Fitness Based Synthesis (FBS), Lagrange multipliers and linear programming. FBS is another name for the already mentioned Combinatorial Optimization method. This study presents and explains these algorithms. They exclude the other algorithms for different reasons. The FBS software did not work. Besides, based on theory there is known that FBS is not efficient and outcomes are hard to interpret. The Lagrange multipliers and the linear programming are in theory very powerful methods, but are also hard to interpret. For IPF there are two versions available. IPF can be applied zone by zone (IPF zone by zone) and on more zones at the same time (IPF Multizone). According to the comparison between the three algorithms, IPF Multizone performs best. The performance of IPF zone by zone is close to the IPF Multizone. The quadratic optimization algorithm performs worst. The total deviation of the population of QUAD compared to the reference sample is more than twice as large as the IPF algorithms.

2.3.3 Score card

The first study by Müller & Axhausen (2010), concluded that in practice the Combinatorial Optimization can give the best population taken into account the costs necessary to collect a reference sample. Though the difference was only small between the two algorithms. In Brederode & Waanders (2013) a study is reported in which methods were compared. In this study the IPF Multizone performs best. To make a choice, a set of criteria was defined. In **Error! Reference source not found.** an overview is given of the criteria and the possible m ethods. A score was given in this table to the different methods from – to ++. According to literature the IPU, HIPF and ENT are almost similar. For this reason they have the same scores in the score card. The same applies for Lagrange and Linear Optimization. As can be seen, the total score for IPF Multizone is the highest. The score in (Brederode & Waanders, 2013) & (Ryan, Maoh, & Kanaroglou, 2009) is a very important criteria, since this reflects the performance of the method in practice. The method that has the highest score on this criteria is the CO/FBS. The main reason that others do not have such a high score is because there is no data found of the methods being used in practice. For this reason they only got an average +/- score. For the population synthesis tool, the applicability of a method is also a very important criteria. Any type of Synthetic Reconstruction method scores high.

Table 1 Score card

Criteria	IPF	IPU/HIPF/ENT	QUAD	Lagrange/LO	CO/FBS
Score in (Brederode & Waanders, 2013) & (Ryan, Maoh, & Kanaroglou, 2009)	+	+/-	-	+/-	++
Applicability	++	++	+	-	-
Reaching convergence	+	+	+/-	++	+/-
Total	+	+	+/-	+/-	+/-

Based on this score card, the IPF Multizone method was selected to use in this research. It scores highest on applicability which represents the practical usability. Besides, there is a positive score on all other criteria.

2.4 Conclusion

In this literature study an overview is given of the main population synthesis algorithms found in literature. The sub question to answer in this chapter was '*Which population synthesis algorithm is most suitable to apply in the population synthesis tool*?'. Criteria were defined to make a choice for an algorithm. Based upon those criteria, IPF Multizone is seen as the most applicable choice and was used in this research.

3. Methodology

Based upon the literature study, a methodology was constructed. This methodology is presented in this chapter. In Figure 4 an overview can be found of the methodology. For every step a more extensive description is given. Every blue block represents a chapter and the arrows show how the steps are related to each other.

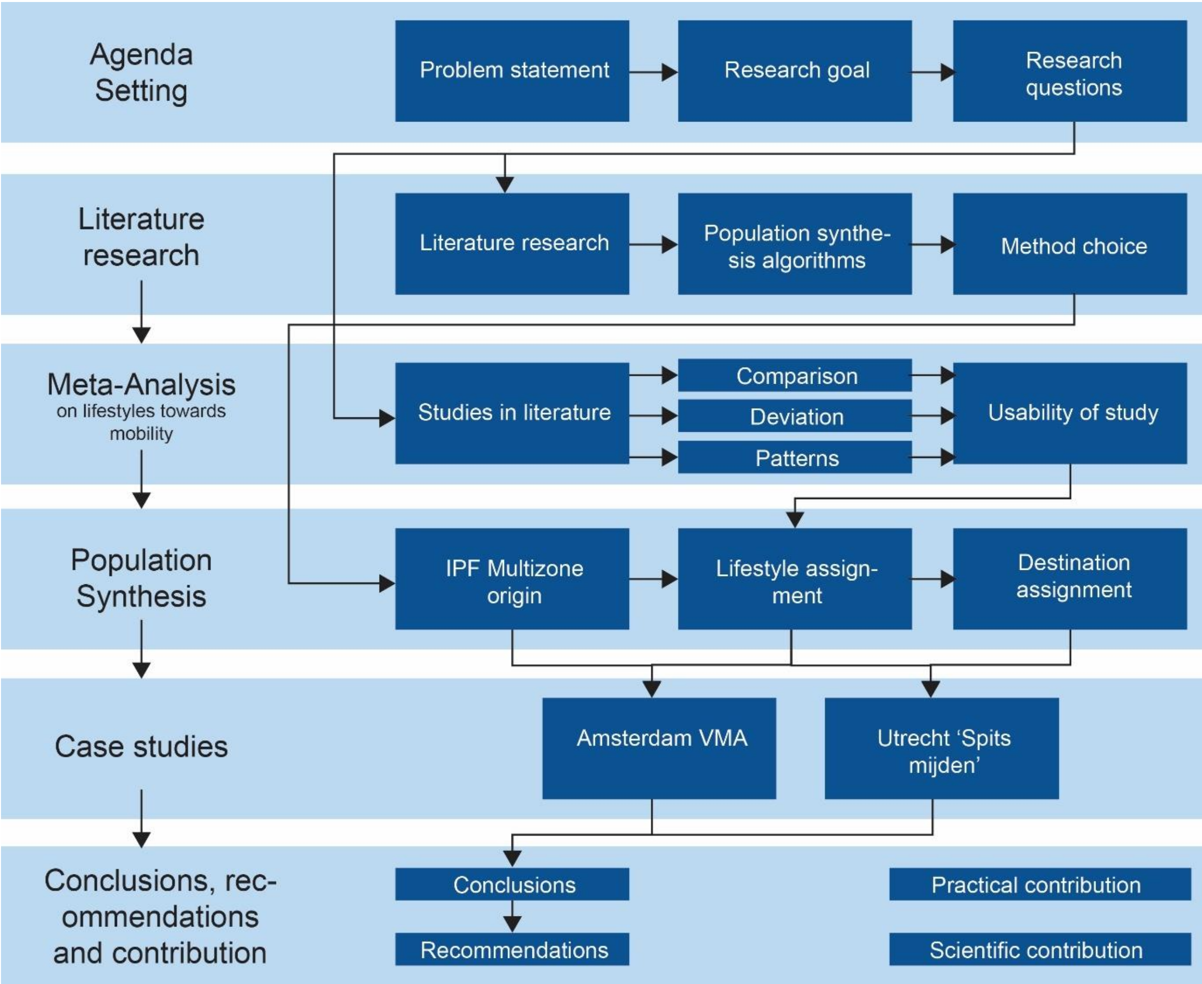


Figure 4 Methodology

3.1 Agenda setting

The first step in this research already has been presented. This was the agenda setting where the problem statement was defined. Based upon the problem statement a research goal was formed and based upon the research goal, the research questions were constructed. The main goal of the research goal and the research question was to obtain a scope for this research, which defined what was studied and what was not in the scope of the research.

3.2 Literature study on population synthesis

The main goal of the literature study was to find additional insight in population synthesis algorithms. With this insights, a method was chosen to use as the basis of the population synthesis tool. To obtain more insights, Google Scholar and Scopus have been consulted. Besides this, a meeting with one of the authors and expert on population synthesis, Luuk Brederode took place. To make a decision for the most applicable algorithm, three criteria were defined. Without any prior knowledge on population synthesis there was not enough information to formulate an approach for the rest of the thesis. For this reason the methodology was defined after the literature study.

3.3 Meta-analysis on lifestyles towards mobility

Data that is commonly used as input for a population synthesizer, in this case IPF Multizone, is data on socio-demographic characteristics. Eventually the output of the population synthesis tool had to be an origin-destination (OD) matrix per lifestyle, based on the distribution over lifestyles per zone, since this is the input required in Fountain. How to get to the distribution of lifestyles per zone was the challenge. A method to do this, is to use the study from Jillian Anable (Anable, 2011). The overall distribution over lifestyles from this study is currently used by Fountain. The study has a dataset with the relation between socio-demographic characteristics and seven lifestyles. Based on this dataset, it is possible to estimate the distribution of lifestyles per zone based on the socio-demographic data that is available on zonal level. The validity of the dataset, provided by a single study, is not guaranteed. Because of the importance of the data, the validity was tested by a meta-analysis. A meta-analysis can be described as analysis that combines the results of multiple scientific studies with the goal to find more specific conclusions. In this research a meta-analysis was conducted in the field of lifestyles towards mobility. The main goal of this meta-analysis was to see if the results in this study are comparable to the same type of studies in the field. Other studies have been found by means of consulting Google Scholar and Scopus. In literature seven studies have been found about lifestyles or attitudes towards mobility. The dataset from Anable (2011) was compared to datasets from other studies.

3.4 Population synthesis

After the literature study, an algorithm for population synthesis was chosen and with the meta-analysis, the method to get from socio-demographic characteristics to lifestyles was chosen. Based on the two methods, a population synthesis tool is been created. The main goal of the population synthesis tool is to replace the population synthesis phase of Fountain and to estimate, based on socio-demographic data, a population for a geographical area for which Fountain is used. In Figure 5 a first proposal for processes in the population synthesis tool can be found.

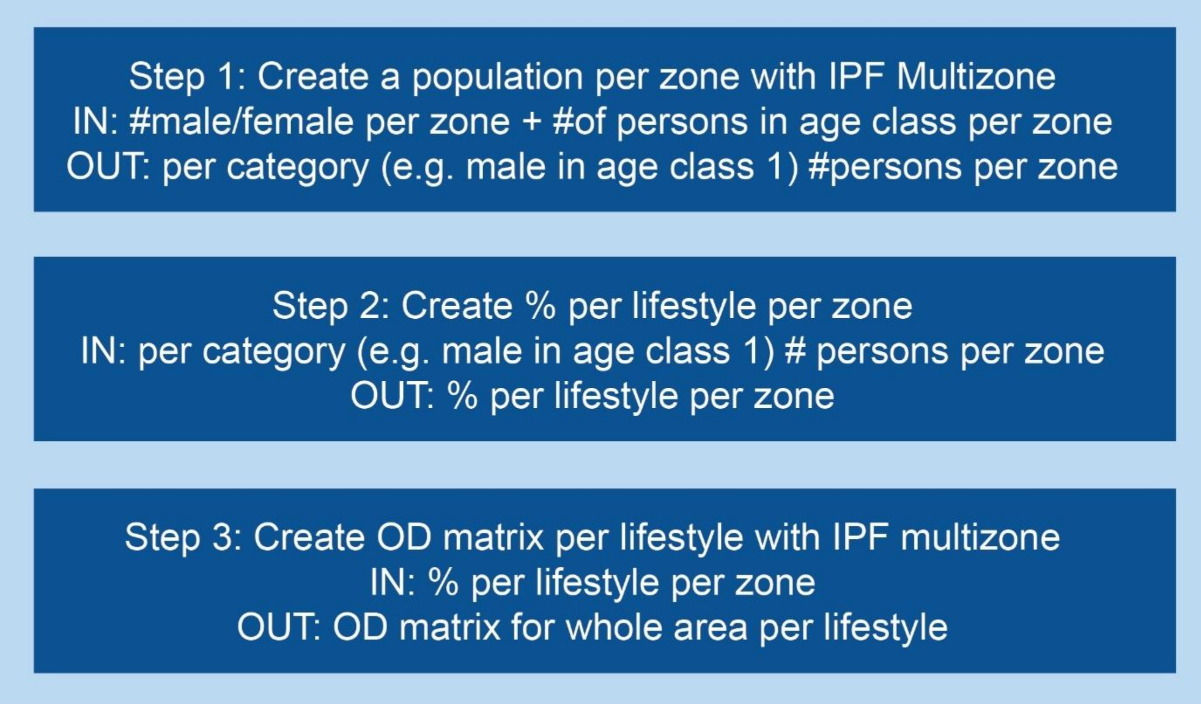


Figure 5 Population synthesis processes

The first step in the process is the creation of a population per zone. This was based on the IPF Multizone algorithm which was the most applicable algorithm according to the literature study. Input in the IPF Multizone method is the total number of males and females per zone and the totals for certain age classes per zone. Output of the IPF Multizone method is, per combination of the above mentioned, the number of persons for every zone. For example, in the category male between 0-15 there could be 4 persons in zone 1. The second step applies the data of the study from Jillian Anable (Anable, 2011) to estimate a distribution over lifestyles per zone, this is the lifestyle assignment. Input required for the lifestyle assignment is the output from the IPF Multizone method plus the data from the study. Output of the lifestyle assignment is the distribution over lifestyles per zone. In the third step, the destination of the agent will be assigned in the destination assignment. Input required for the destination assignment is the output from the lifestyle assignment plus an OD matrix for all travelers that needs to be included in the project. Output of the destination assignment is an OD matrix for the whole area per lifestyle.

3.5 Case studies

Two case studies were chosen. The main goal of both case studies was to measure the performance of the population synthesis tool. In the first case study, the tool is applied for the VMA. VMA is an abbreviation for 'Verkeersmodel Amsterdam', which is the transport model of Amsterdam. The second case study is the project 'Utrecht rush hour avoidance'. The approach for the case studies is described in this section. First the data collection is discussed in subsection 3.5.1 and then the analyses will be described in subsection 3.5.2.

3.5.1 Data collection

An overview of the required data for both case studies can be found in Table 2. The data required are a transport network, the total men and women per zone, the total amount of persons in multiple age classes per zone, a reference sample of the area and an OD matrix.

Table 2 Overview of required data

Data	Location of data Amsterdam VMA	Location of data 'Utrecht rush hour avoidance'
Transport network	In VMA	In VRU (version 3.0)
Total men and women per zone	In VMA	In VRU (version 3.0)
Total person in age classes per zone	In VMA	In VRU (version 3.0)
Reference sample of the area	Not available (but can be created based on data in VMA)	Not available (but can be created based upon VRU)
OD matrix	Not required in this case study	In VRU (version 3.0)

The municipality of Amsterdam allowed the use of the VMA ('Verkeersmodel Amsterdam). In this model all the data that is needed for the IPF Multizone method and the lifestyle assignment is available. The VRU (Traffic model of Utrecht) is used in Fountain before, therefore the data is available. Several versions of the VRU are available, the third version is used in this thesis. The OD matrix that is used, is from the year 2010 (base year) with the modalities Car, PT and bike and the purpose commuter traffic during the morning rush hour.

3.5.2 Analyses to measure performance of the population synthesis tool

In section 3.4 three steps for the population synthesis tools were defined the IPF Multizone method, the lifestyle assignment and the destination assignment. After every step the performance of the tool was measured. In this subsection an explanation is given on how this was done for the two case studies.

Table 3 Analyses in case studies

	Amsterdam VMA	Project 'Utrecht rush hour avoidance'
IPF Multizone	Comparison of actual population with synthetic population by means of measures of fit (TAE/TAD, SAE, MAPE & RMSE).	-
Lifestyle assignment	Analysis on the distribution over lifestyles towards mobility per zone: MAPE.	Analysis on the distribution over lifestyles per zone. MAPE, distribution for specific zones: Uithof (students), Oudwijk, City center and Suburb areas.
Destination assignment (end product of Population synthesis tool	-	Comparison on case 'Utrecht rush hour avoidance' between: the actual data collected in project 'rush hour avoidance', the output of Fountain with the original population synthesis phase and the output of Fountain with the synthetic population.

In Table 3 the different analyses can be found. In the VMA, a dataset of the actual population is available. After the IPF Multizone method, the actual population will be compared to the synthetic population. In literature multiple measures of fit were found which were used to measure the deviation between the two populations. The actual population is only available for Amsterdam. Therefore this analysis cannot be done for Utrecht. The Analysis on the output of the lifestyle assignment was done for both case studies. The MAPE was calculated and specific for Utrecht, multiple zone distributions were compared. The goal of the analyses on the output of the lifestyle assignment was to see if the distribution was different for different type of zones and to see if the distributions were as expected. After the lifestyle assignment a destination is assigned to the population. The output of the destination assignment will be used as input in Fountain. The data collected in project 'rush hour avoidance' was compared to the output of Fountain with the original population and the output of Fountain with the synthetic new population. The goal of this analysis is to see if the performance of the new population is better in terms of the project 'rush hour avoidance' compared to the original population created by the population synthesis phase of Fountain.

4. Meta-analysis on lifestyles towards mobility

In the second chapter a population synthesis algorithm was chosen: IPF Multizone. As input for a population synthesizer it is common to use socio-demographic characteristics (Müller & Axhausen, 2010). The output of the population synthesis tool had to be an origin destination (OD) matrix per lifestyle, which could be estimated based on the distribution over lifestyles per zone. It is challenging to get from socio-demographic characteristics (data available) to the distribution over lifestyles per zone (data not available). A method to do this, is to use the study from Jillian Anable. She identified seven lifestyles towards mobility: status seekers, devoted drivers, reluctant pragmatics, practical travelers, active car drivers, car contemplators and car free choosers (Anable, 2011). The study has a dataset with the relation between socio-demographic characteristics and the seven lifestyles. Based on this dataset it is possible to estimate the distribution of lifestyles per zone based on the socio-demographic data that is available on zonal level. Since this is the dataset of only one study, it might be dangerous to use the dataset without any further notice. To justify the usability of the data from this study, a meta-analysis on lifestyles towards mobility was performed. Based on preliminary literature research the following research question was constructed:

Which studies are available in the field of lifestyles towards mobility and are there patterns that justify the use of this data (Anable, 2011)*?*

In section 4.1 an overview is given of the studies that are included in the meta-analysis. In section 4.2 comparisons between studies are given. In section 4.3 patterns in the data are presented and in section 4.4 conclusions will be drawn.

4.1 Overview of studies for the meta-analysis

In the literature, several studies have been found about lifestyles towards mobility. Every study has a different name for "lifestyles", but the definition is the same. In other studies "lifestyles" are called: attitudes towards mobility, travel behavior segments, market segments for mode choice, attitude-based target groups and mobility styles. In Table 4 an overview is given of the studies that are included in the meta-analysis. More studies that mention lifestyles are available in literature but do not specify a classification, for this reason they are not included in this meta-analysis. In the first column of the table the author and the publication year are given, in the second column the number of lifestyles is given, in the third column the lifestyles are given and in the fourth column theories that are used in the studies are given. In following subsections a short summary can be found on every study.

Table 4 Overview of studies in the meta-analysis

Author	# of lifestyles	Lifestyles towards mobility	Theories
Anable, J. 2005	6	Malcontented Motorists, Complacent Car Addicts, Die Hard Drivers, Aspiring Environmentalists, Car-less Crusaders, Reluctant Riders.	Theory of planned behavior with additional factors added. Namely moral norm, environmental attitudes, efficacy, identity, habit.
Beirao, G.	3	Public Transport Users, Car	Interviews with 24

2007		Users and both.	participants.
Karash et al. - 2008	4	The Transit Loyalists, The Environmental Mode Changers, The Happy Drivers, The Angry Negatives.	Theory of planned behavior
Hunecke, M. et al. 2010	5	Public Transport Rejecters, Car Individualists, Weather Resistant Cyclists, Eco-sensitized PT-users, and Self-determined Mobile People.	Theory of planned behavior, segments based on socio-demographic characteristics
Prillwitz, J. 2011	4	Addicted Car Users, Aspiring Green Travelers, Reluctant Public Transport Users and Committed Green Travelers.	
Haustein, S. 2011	4	Captive Car Users, Captive Public Transport Users, Affluent Mobiles and Self-determined Mobiles.	Theory of planned behavior, Theory of cognitive dissonance
Anable, J. 2011	7	Status Seekers, Devoted Drivers, Reluctant Pragmatics, Practical Travelers, Active Car Owners, Car Contemplators and Car Free Choosers.	Theory of planned behavior

4.1.1 Identifying travel behavior segments using attitude theory (Anable, 2005)

In Anable (2005), an expanded version of a psychological theory of attitude-behavior is used, namely the theory of planned behavior (TPB) (Ajzen, 1991). Scores on factor analyzed multi-dimensional attitude statements were used to segment a population of day trip travelers into potential 'mode switchers', by using cluster analysis. In this study six groups were extracted: Malcontented Motorists, Complacent Car Addicts, Die Hard Drivers, Aspiring Environmentalists, Car-less Crusaders and Reluctant Riders. The goal of this study was to identify groups, the groups need to be serviced in different ways to optimize the chance of influencing mode choice behavior. Socio-demographic factors had little impact on the travel profiles of the segments, suggesting that attitudes largely cut across personal characteristics.

4.1.2 Understanding attitudes towards public transport and private car (Beirao & Sarsfield Cabral, 2007)

The second study is a study from Beirao & Sarsfield Cabral (2007). It is a qualitative study of interviews with public transport users and car users. The goal of the study is to obtain a deeper understanding of travelers' attitudes towards transport and to explore perceptions of public transport service quality. It was concluded that the choice of transport is influenced by several factors, among which individual characteristics and lifestyles. Policies which aim to influence car usage should be targeted at the market segments that are most motivated to change and willing to reduce frequency of car use. In this study only three lifestyles have been identified based upon 24 interviews. The three lifestyles that were identified are: Public Transport Users, Car Users and users of both car and public transport.

4.1.3 Market segment for mode choice (Karash, et al., 2008)

In the study from Karash et al. (2008), four segments for modal change have been identified. The segments are The Transit Loyalists, The Environmental Mode Changers, The Happy Drivers and The Angry Negatives. The research aims at defining target groups that are likely to change modal behavior and target groups that are unlikely to change modal behavior. This study uses the theory of planned behavior (TPB, a more extensive explanation on this theory can be found in subsection 4.2.2). A survey was used with three defined phases; first a pre-intervention application of the full TPB was undertaken concerning one's intention to change personal transportation patterns. In the second phase, an intervention was undertaken in which respondents were first exposed to different messages and then to seven separate potential services that might improve the marketability of the alternative transportation mode. The third phase was another application of the TPB, where participants were allowed to shift to another service that might have occurred based upon the messages of the 2nd phase.

4.1.4 Attitude-based target groups to reduce ecological impact of daily mobility behavior (Hunecke, Haustein, Bohler, & Grischkat, 2010)

The study from Hunecke et al. (2010) analyzes the usefulness of an attitude-based target group approach in predicting the ecological impact of mobility behavior. A survey was used with 1991 respondents living in three large German cities. This study is also based on an expanded version of the TPB. Five lifestyles have been identified: Public Transport Rejecters, Car Individualists, Weather Resistant Cyclists, Eco-sensitized PT-users, and Self-determined Mobile People. A comparison is made between the predictive power for a population based upon socio-demographic characteristic and the predictive power for a population based on lifestyles. This is done based on the use of private motorized modes of transportation. The authors concluded that lifestyles give a better prediction of the actual use of transportation.

4.1.5 Mobility styles, attitudes and individual travel behavior (Prillwitz & Barr, 2011)

The study by Prillwitz & Barr (2011) aims to develop the notion of sustainability-related mobility styles as a context for applying targeted social marketing policies to specific population segments. The research is performed in the South West of England and is based on ten focus group discussions and a survey with more than 1500 participants. Two segmentation approaches are used to identify gaps between different domains of individual travel behavior and the varying role of attitudes for travel decisions. Four mobility styles/ lifestyles have been identified, namely: Addicted Car Users, Aspiring Green Travelers, Reluctant Public Transport Users and Committed Green Travelers. This study supports the idea that certain sociodemographic characteristics are connected to certain lifestyles. Although the correlation is not very high. The socio-demographic characteristics that have been stated are age, sex, income and political view.

4.1.6 An attitude-based segmentation approach for a heterogeneous target group (Haustein, 2011)

In Haustein (2011), a study is reported for lifestyles in only one target group. The assumption is that the elderly are a quite heterogeneous target group with an increasing impact on the transport system. Four mobility types of the elderly were identified in this study: Captive Car Users, Captive Public Transport Users, Affluent Mobiles and Self-determined Mobiles. Elderly are defined in this study as people of 60 years and above. A survey was used by means of 1500 standardized telephone interviews. Several conclusions are presented; a very interesting conclusion is that the resulting mobility types show similarities with lifestyles that have been identified in former studies. Especially when comparing the lifestyles with (Hunecke, Haustein, Bohler, & Grischkat, 2010), a related lifestyle can be identified for all four lifestyles.

4.1.7 Segment baseline survey of new employees - Utrecht (Anable, 2011)

In the study from Anable (2011), a survey was used with 320 respondents from the city of Utrecht. The respondents were all working in the city. The main goal of this study was to classify the respondents in different lifestyles. In this study seven lifestyles have been identified: Status Seekers, Devoted Drivers, Reluctant Pragmatics, Practical Travelers, Active Car Owners, Car Contemplators and Car Free Choosers. The study concludes that the main motivations for choosing a certain mode of transport are as follows: for the status seekers, the status/image is important, besides this they value convenience. The devoted drivers are mostly concerned about how convenient their trip is and are clearly not concerned about the environment. The reluctant pragmatics are practical and very time/cost efficient, they have a preference for travelling by car. The practical travelers are practical and make sensible choices. What is meant with sensible is not mentioned in the paper. The active car owners are very positive towards cycling and walking, worried about climate change and care about their image. It might seem strange to call them active car drivers but this name refers to the fact that this group is driving the car but will try to prevent this by cycling and walking which are both stated as active activities. The car contemplators are the smallest group and do care about status and image. They desire to own a car but do not have one. Finally there are the car free choosers: they do not own a car by choice, think that driving is stressful and care about environmental issues and the costs that come along with owning a car.

4.2 Comparisons between studies

In the previous subsection, an overview is given of the different studies in the meta-analysis. In this subsection, a few observations are discussed.

4.2.1 Same authors

Some of the authors are involved in more than one of the studies. Jillian Anable published a study in 2005 where she distinguished six lifestyles. Six years later in 2011, she published a study which contains seven lifestyles. It seems plausible that the latest study might be more accurate, based on the knowledge of the authors in more recent work. The same applies for Sonja Haustein who published together with Hunecke et al. a study in which five lifestyles were distinguished and one year later in 2010, she published an article in which four lifestyles were distinguished. The difference between the two studies by Haustein can be explained by the target group of the research. The second study was focused on elderly only. The first study did not focus on one target group particularly. All the studies distinguish a different amount of lifestyles and have different expressions for their lifestyle groups.

4.2.2 Theories behind the studies

Five of the seven studies are based upon the theory of planned behavior (TPB) and one study uses the theory of cognitive dissonance. In the remaining study, no theory is mentioned. To understand more about these theories, a short explanation is given.

Theory of planned behavior

The theory of planned behavior is developed by Ajzen (1991). This theory states that behavior can be best explained by the intention a person has, to actually perform the behavior. Intention is based upon three factors:

1. *Attitude towards behavior*: In this case it is about the attitude or opinion that someone has towards certain behavior.
2. *Subjective norm*: People want to belong to a group, therefore opinions and behavior of other people is very important and influences the behavior of the individual. If someone wants to adjust his behavior to the group, this will influences the intention of this person.
3. *Perceived behavioral control*: This applies when someone thinks he can actually perform a certain behavior (there are also studies which have proven that this factor influences the behavior directly).

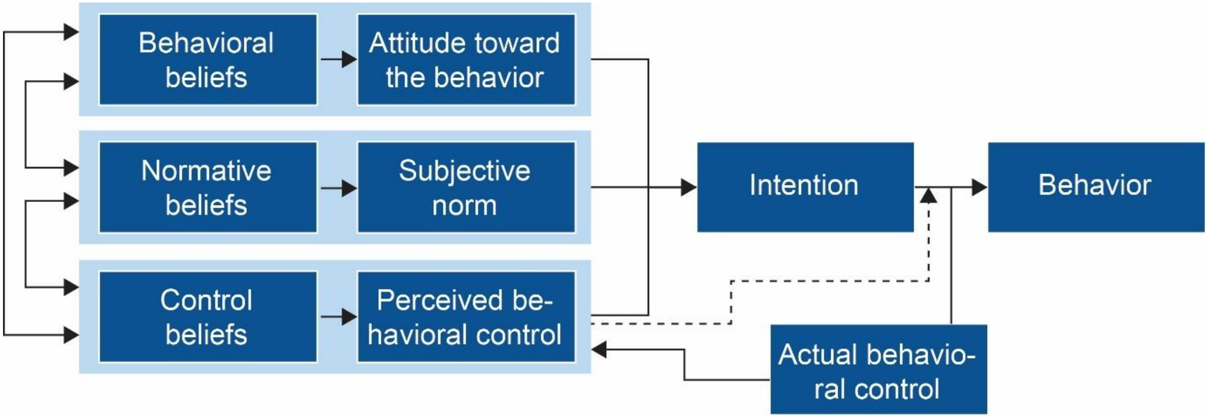


Figure 6 Theory of planned behavior

An overview of the TPB can be found in Figure 6. The theory of planned behavior has a few disadvantages. According to the TPB, the long term benefits or costs are less important (for example the investment costs for a car), than the short term benefits or costs, which might not always be correct (Ajzen, 1991). Besides this, there can be a false consensus effect within a group which implies that the shared attitude of the group can be overestimated. Finally, people often overestimate themselves. Therefore expectations of their behavior do not always meet reality.

Theory of cognitive dissonance

Besides the theory of planned behavior, one of the studies from Haustein is based on the theory of cognitive dissonance. This theory represents the tendency for individuals to seek consistency among their beliefs and opinions and implies that once someone believes his opinion is true, he will do everything to keep it this way. The person will come up with several reasons for believing his opinion is the right one, even if the evidence tells him differently. To change the opinion of the person, something drastic needs to happen (Festinger, 1957).

Both theories give insights in human behavior based on attitudes. The theory of planned behavior is used in most of the studies, the theory of cognitive behavior only in one study. In Fountain the TPB is part of the structure of the model (Weerdt, Hof, & Vonk, 2014).

4.3 Patterns in the data

In the previous section similarities between the different studies are described and several differences have been pointed out. The data of different studies have been compared with each other, to see if they give the same results and the same relations between socio-demographic characteristics and lifestyles. Before the data can be compared, it is important to understand how the data can be compared and if the lifestyles in different studies represent the same characteristics. In subsection 4.3.1 the classification in lifestyles is given. Besides the underlying thoughts of the lifestyle groups, the modal split is analyzed in subsection 4.3.2. Most studies present the percentages belonging to different lifestyles and how they relate to different sociodemographic characteristics of the respondents. This data is analyzed to see if there are any comparisons between the data of different studies in subsection 4.3.3. In subsection 4.3.4 conclusions of the meta-analysis on lifestyles towards mobility are given.

4.3.1 Classification of the lifestyles towards mobility

Since most studies have a different amount of lifestyles, it is important to understand why these classifications of lifestyles were chosen and which characters they represent. For example, the study from Anable (2011) has seven lifestyles of which two of these lifestyles represent a group of people who do not own a car, while for example Karash et al. (2008) only uses one group that does not own a car. Some patterns have been found in these lifestyle characteristics. Four main groups have been identified in such a way that all the lifestyles fit in one group. The four main groups that have been distinguished are:

Segment 1: Full car users. Cars have a certain status in their opinion.

Segment 2: Values its car and has no propensity to like the attributes of a transit-oriented life but if it is more convenient to use another modality this is preferred.

Segment 3: Environmentally friendly people. Care about the environment and see their own behavioral change as worthwhile. Therefore they use other modes more often.

Segment 4: Non-car owners. Do not have a car for different reasons and therefore travel mostly by other modes.

In Table 5, the lifestyles from every study have been classified in one of the main groups.

Table 5 Classification of lifestyles

Study	Segment 1	Segment 2	Segment 3	Segment 4
Anable 2005	Die Hard Drivers Do not see many problems with car use, nor the point of	Malcontented Motorists They need persuasion that reducing their own car use will	Aspiring Environmentalists Pro-environmental behavior is seen as important and worthwhile. The	Car-less Crusaders individuals in this group are slightly more influenced by

	reducing it	make much difference, as they believe other people will not reduce theirs	negative effects of car use enter into the decision making process	personal and social norms
	Complacent Car Addicts Particularly enjoy car travel and believe that all their car use is necessary			Reluctant Riders Do not own a car, Not particularly motivated by environmental issues
Karash et al. 2008		Happy Drivers Values its automobiles and has no propensity to like the attributes of a transit-oriented life	Environmental Mode Changers If certain conditions are improved, they would become transit users	Transit Loyalists Current use and understanding of public transportation services
Hunecke et al. 2010	Car Individualists Quite similar to PT rejecters, extreme high symbolic value to the private car	Self-determined Mobile People Negative on openness to change but high perceived PT control	Eco Sensitized PT-users High ecological norm and positive evaluation of PT	Weather Resistant Cyclists Not highly interested in car use, high bicycle use and weather resistant
	Public Transport Rejecters Lower			

	evaluation of PT, high perceived mobility necessities and low openness to change			
Prillwitz, 2011	Persistent Car Users	Frequent Car Users	Consistent Green Travelers	Constrained Public Transport Users
Anable, 2011	Status Seekers Highest car use and low use of other modes, find other modes difficult. Not motivated be environmental reasons	Reluctant Pragmatics High car use and some use of train. Low intention to reduce car use	Active Car Drivers Strong moral obligation to act on environmental issues, believe they can make a difference	Car Contemplators Very low car ownership, high intention to use walking and above average intention to use PT and cycling
	Devoted Drivers High car use and low use of other modes, choose based on speed, flexibility and cost	Practical Travelers High car use, only PT use to work		Car Free Choosers Very low car ownership, highest intention to walk and cycle and above average intention to use PT

The fact that this classification is possible, shows that the lifestyles of different studies can be compared with each other. For the rest of this chapter, when a group of lifestyles is mentioned, is referred to by segment 1 to segment 4. Now the underlying characteristics of the lifestyles in different studies is found, the modal split will be presented for every study in the following section.

4.3.2 Modal split of different lifestyles towards mobility

In this section the modal split for lifestyles is presented, only for the studies that provided this data. As mentioned, one study does not have the correct data available and one study has a target group of only elderly and will therefore be left out of the analysis. In Figures 8, 9, 10 and 11 the modal split for lifestyles is given. In all the figures the lifestyles in segment 1 are given first on the left side, then segment 2, segment 3 and finally segment 4 on the right side. In (Anable, 2005) a pattern can be found from left to right where the car use starts very high and ends very low on the right side. This same pattern is found in figures 9 till 12.

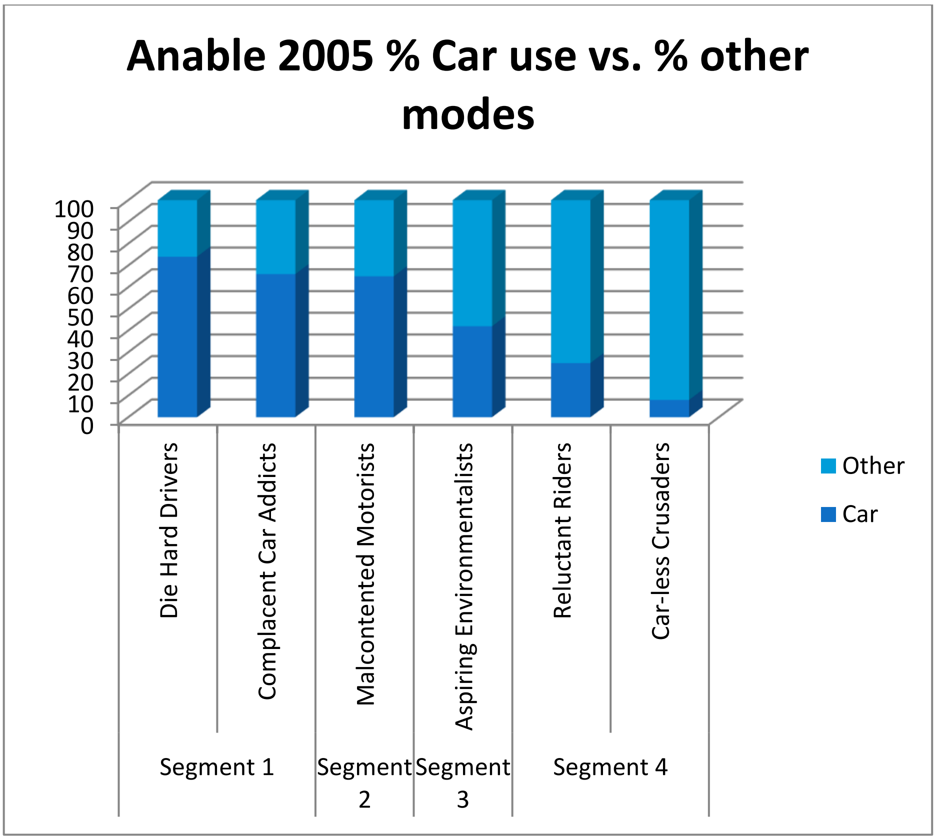


Figure 7 Anable (2005)

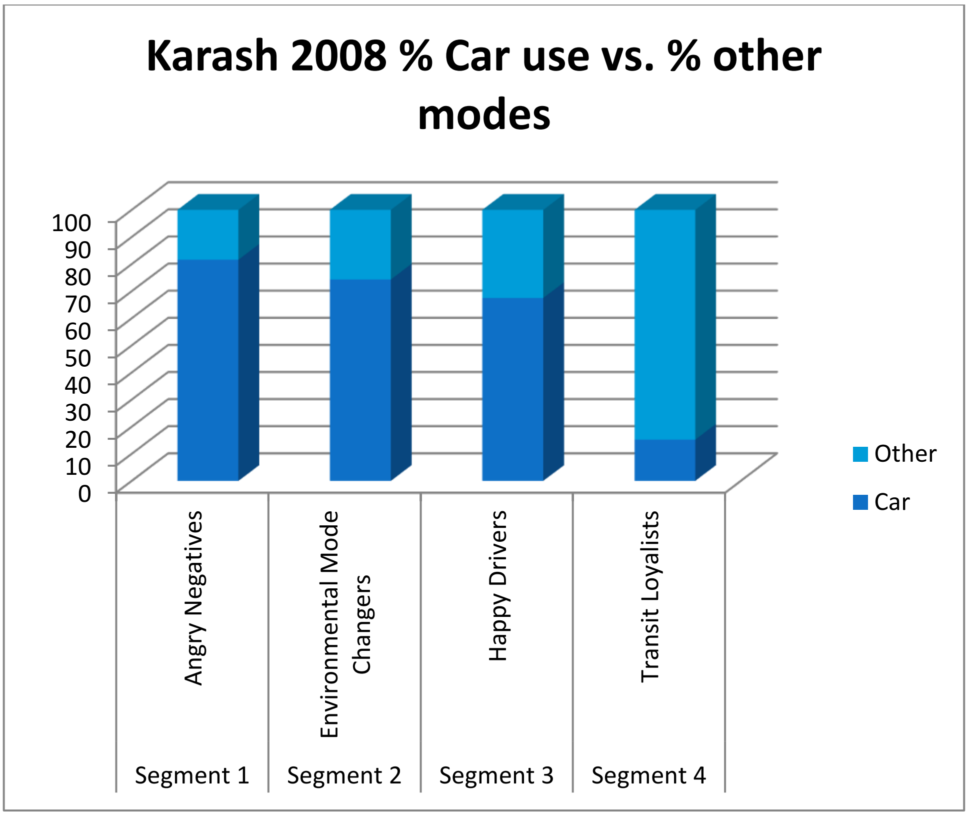


Figure 8 Karash et al. (2008)

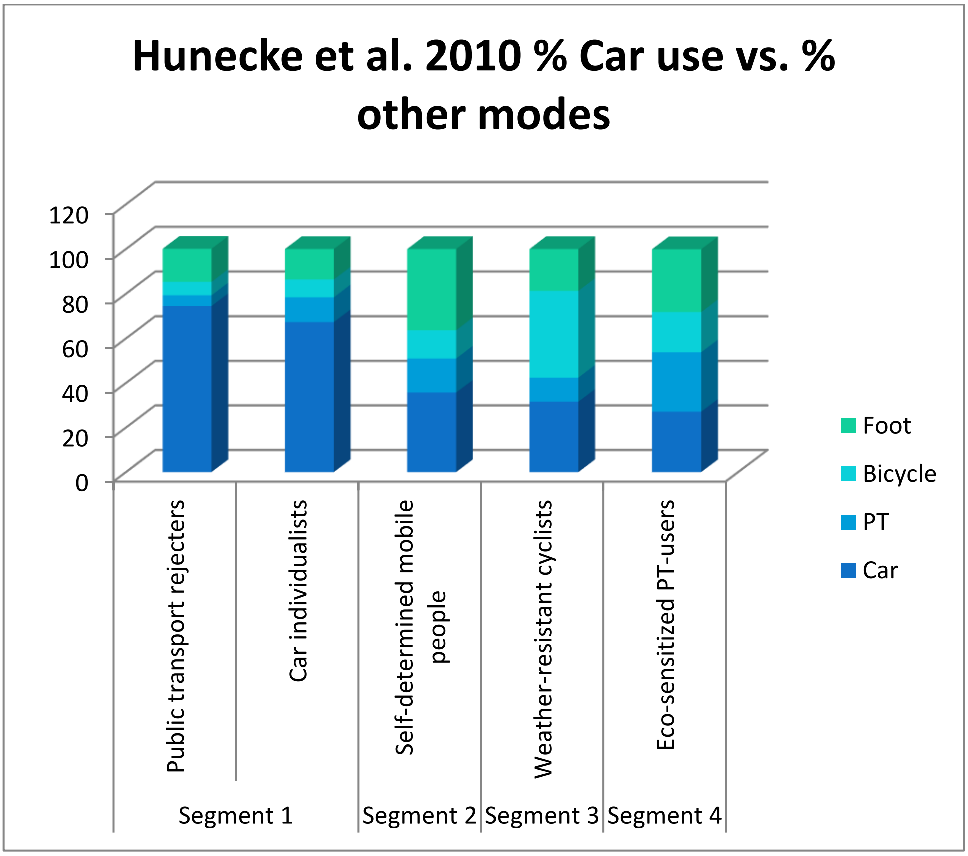


Figure 9 Hunecke et al. (2010)

Other modes are only given in Hunecke et al. (2010), but there does not seem to be any other pattern except for the decreasing line of car users. As expected the weather-resistant cyclists who belong to segment 3 have the largest percentage of the most environmental modes, travelling by foot or bicycle.

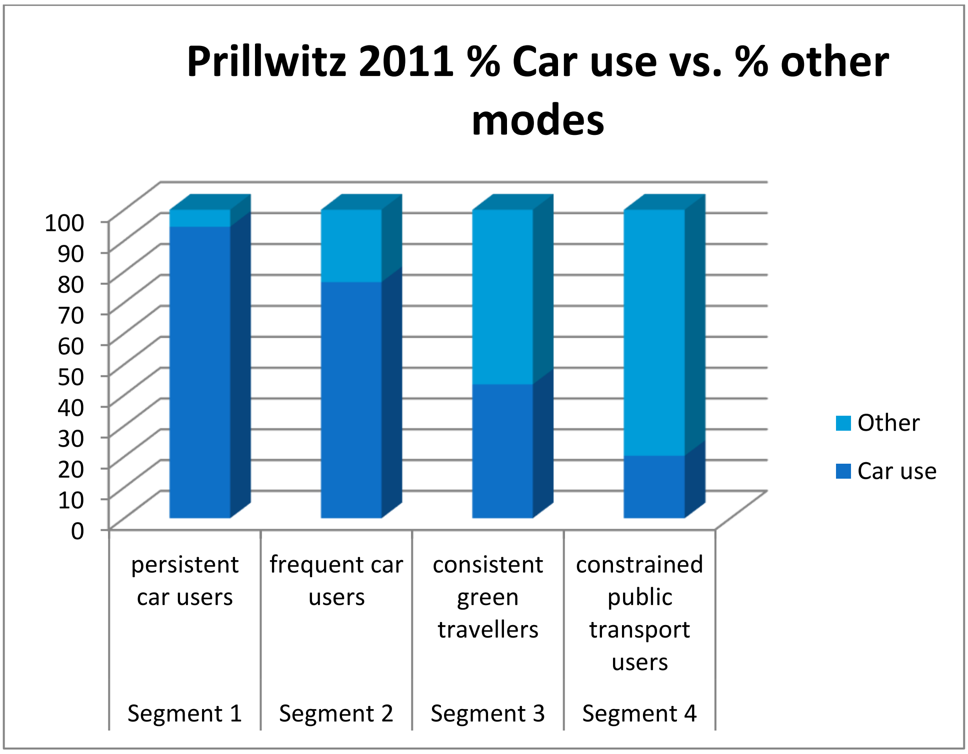


Figure 10 Prillwitz (2011)

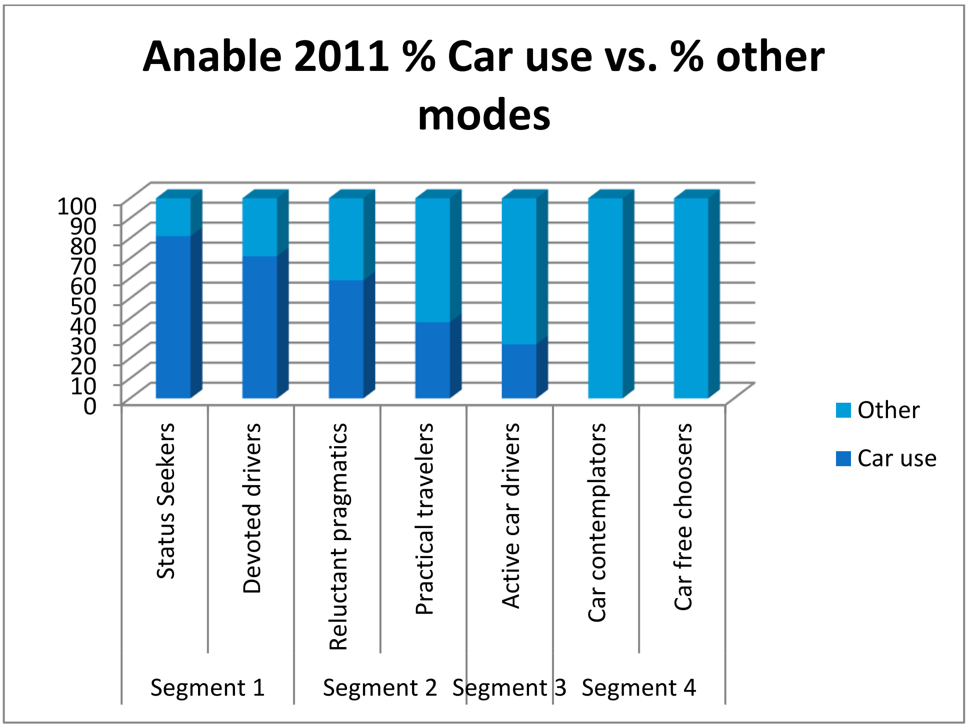


Figure 11 Anable (2011)

The fact that the distribution of the modal split for the different lifestyle groups has the same pattern, shows that different studies classified their lifestyles all according to high car usage to low car usage. Analysis in different socio-demographic characteristics is performed. In the following subsections the relation between socio-demographic characteristics and lifestyles is given.

4.3.3 Distribution of socio-demographic characteristics versus lifestyles towards mobility

In the conclusions of the seven studies in section 4.1, different opinions are given about the relation between socio-demographic characteristics and lifestyles. In Anable (2005) the hypothesis that any changes in attitudes and differences in travel behavior could simply be attributed to personal characteristics was proven wrong. Nevertheless, there were variables that could distinguish the groups, education and age for example. In Karash et al. (2008) socio-demographic characteristics are related to the different lifestyles. For example, the transit loyalists are the youngest group and are disproportionately male. The environmental mode changers are disproportionately female. In Hunecke et al. (2010) almost all (98,3%) individuals could be assigned to one lifestyle segment by means of sociodemographic data obtained. According to Prillwitz & Barr (2011) relations between socio-demographic characteristics and lifestyles are present, especially age, gender, income, household structure and political views can be used. The opinions about the relationship between socio-demographic characteristics and lifestyles are widespread but the most recent studies conclude that relations are present. The goal of this chapter was to see if these relations are the same in different studies. If this is the case, it can justify the relation between socio-demographic characteristics and lifestyles and it can justify using the study from Anable (2011) in population synthesis. This analysis is presented in the following subsection. Not all the studies include data about relations between socio-demographic characteristics and lifestyles. Besides this, one of the study has a target group of only elderly. For this reason only the following studies are used in the following analysis.

Table 6 Studies used for analysis

Author	Name
Anable, J. (2005)	'Complacent Car Addicts' or 'Aspiring Environmentalists'? Identifying travel behavior segments using attitude theory
Karash et al. (2008)	Understanding how individuals make travel and location decisions: implications for public transportation
Hunecke et al. (2010)	Attitude-based target groups to reduce the ecological impact of daily mobility
Prillwitz, J. and Barr, S. (2011)	Moving towards sustainability? Mobility styles, attitudes and individual travel behavior
Anable, J. (2011)	Segment baseline survey of New Employees - Utrecht

In this subsection the following socio-demographic characteristics will be analyzed: age, gender, income, education, household size and political opinion. The characteristics were chosen based on available data in the studies.

Distribution of age versus lifestyles

The first socio-demographic characteristic that was analyzed is age. The data required to analyze this characteristic is the percentage of people in a certain age class with a certain lifestyle. Three of the studies in this meta-analysis have this data available and will be used for comparison. In Table 7 an overview is given of the studies included in this analysis. In appendix B the databases from the used studies can be found.

Table 7 Studies used for the age class analysis

#	Author	Study
1	Karash et al. (2008)	Understanding how individuals make travel and location decisions: implications for public transportation
2	Prillwitz, J. (2011)	Moving towards sustainability? Mobility styles, attitudes and individual travel behavior
3	Anable, J. (2011)	Segment baseline survey of New Employees - Utrecht

The first study from Karash, et al. (2008) uses four lifestyles which include the distribution of age classes for every lifestyle. The four age classes are: under 30 years old, from 30-39 years old, from 40-49 years old and 50-plus years old. An overview can be found in Figure 12. The age-groups with the highest proportion for each lifestyle are distributed as follows: The Angry Negatives in segment 1 have the highest percentage of persons between 30-39 years old. The Happy Drivers in segment 2 have the highest proportion of persons between 40-49 years old. The Environmental Mode changers in segment 3 have the highest proportion of people from the 50-plus class and the Transit Loyalists in segment 4 are mostly persons under 30 years old and especially not people between 40-49 years old. The highest proportions are not significantly different than other proportions.

In the second study from Prillwitz & Barr (2011) four lifestyles are used. Unfortunately there are different age classes. The age classes in this study are 16-19, 20-29, 30-44, 45-59, 60-74 and 75 and older. The proportion for each lifestyle per age class is also given. In Figure 13, a summary can be found. The persistent car users in segment 1 are according this study mostly between 45-59 years old. The frequent car users in segment 2 are also mostly between 45-59 years old but have also many members in the age group of 30-44 and 60-74. The third lifestyle group is the constrained pt users (public transport users). They are mostly between 60-74 years old and also have a high percentage in the group of 75 years and older. The fourth lifestyle group, belonging to segment 4, is the consistent green travelers group. They have the highest proportion 20-29 and the highest proportion 30-44 years old.

In the third study from Anable (2011) another age class classification is used: <24, 25-34, 35-44, 45-55 and >55. In Figure 14, an overview of this analysis can be found. Anable identified seven lifestyles and therefore, more data is available for every segment. In segment 1 contains the status seekers and the devoted drivers. The status seekers have a high proportion of people between 25-34 years old and devoted drivers on people between 35-44 years old. In segment 2 the reluctant pragmatics have the highest proportion on people between 35-44 years old. Segment 2 also includes the practical travelers which have a high score on people from 35 to 44 years old and have many young people up to 24 years old. In segment 3 are the active car owners. This group has a very small amount of young people and a very high proportion of people between 45 and 54 years old. Segment 4 are the car contemplators; this lifestyle has a very high proportion of people up to 24 years old. The car free choosers consist of a disproportional large number of people between 25-34 and 45-55.

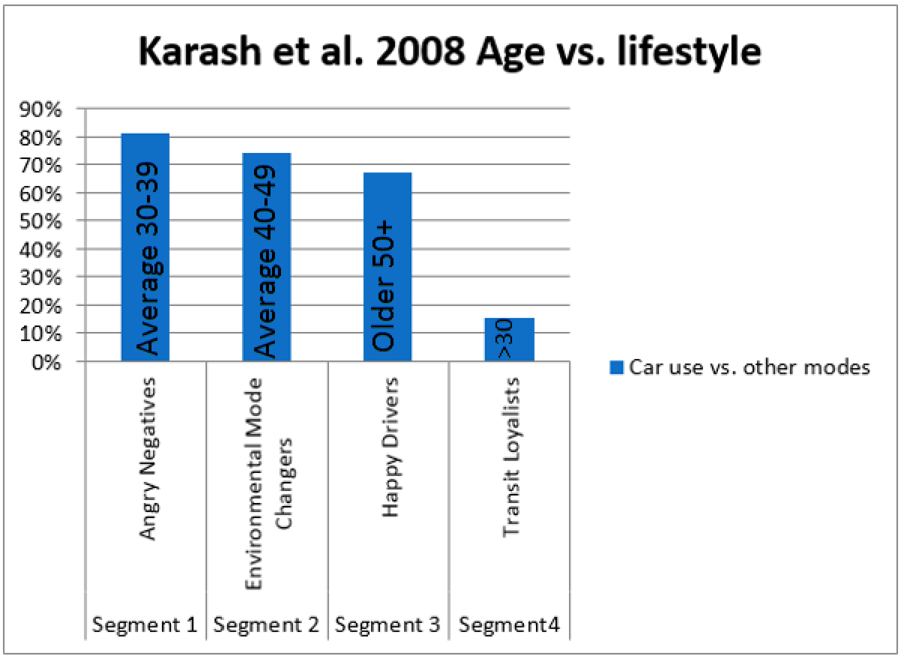


Figure 12 Distribution of lifestyles towards mobility over age classes (Karash et al. 2008)

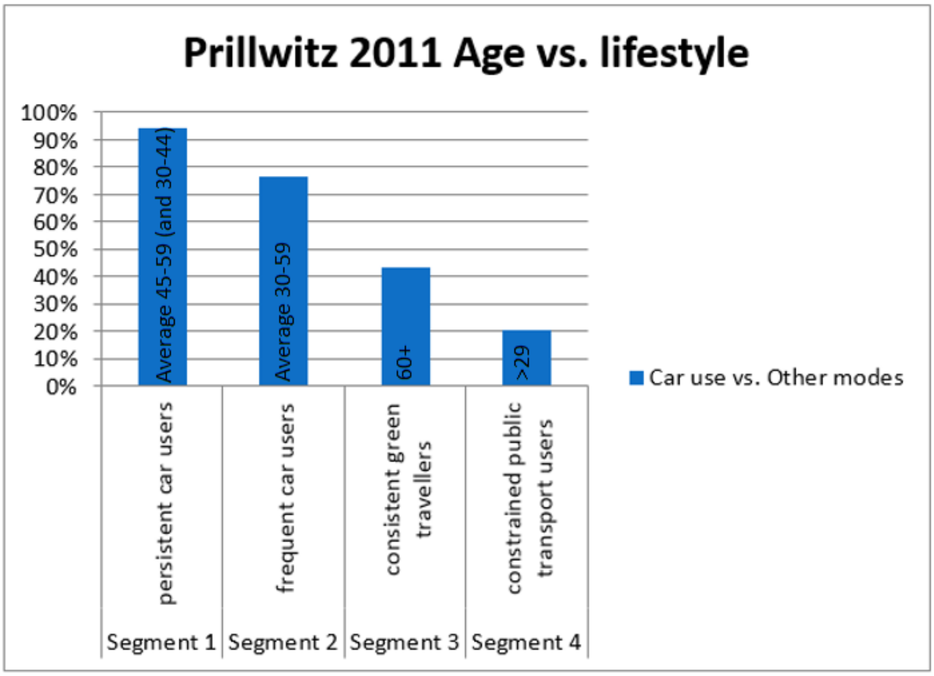


Figure 13 Distribution of lifestyles towards mobility over age classes (Prillwitz, J.; Barr, S. 2011)

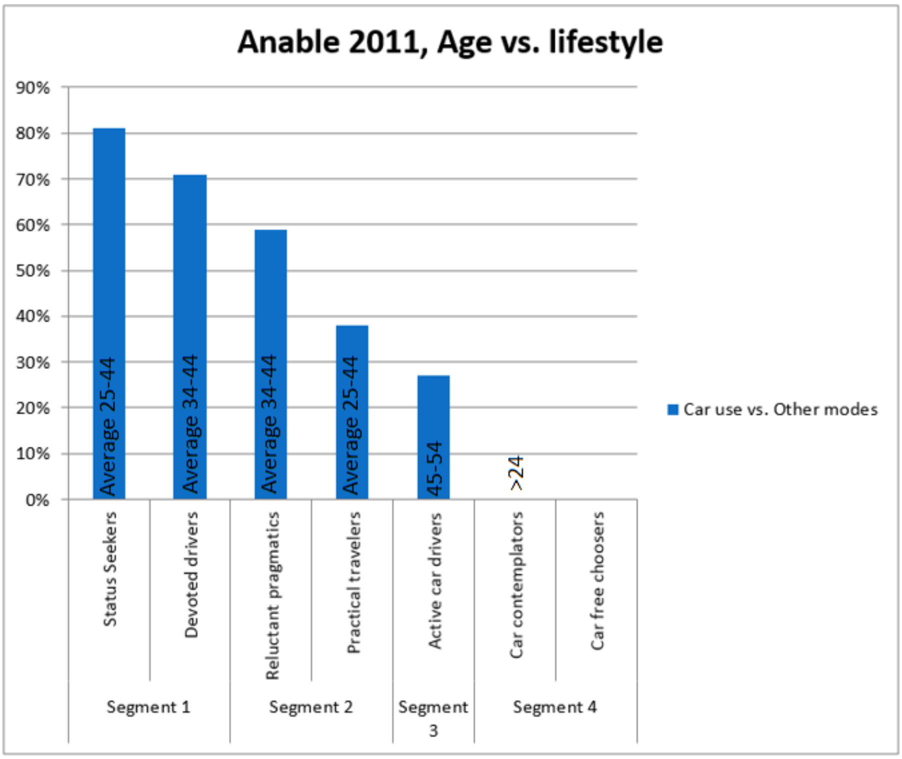


Figure 14 Distribution of lifestyles towards mobility over age classes (Anable, J. 2011)

In every study that was analyzed, the highest proportion per segment was in the same age class. It was hard to compare the studies since the age classes are not the same. However a clear pattern was found:

- The first two segments have the highest proportion of the average age classes, between 30 years old and 50 years old.

- In segment 3 the highest proportion is from the older age classes, 50+ or even 60+.
- In segment 4 the youngest age classes have the highest proportion, >25 or >30.

The studies are not all exactly the same, but this pattern contributes in the idea that a distribution of lifestyles can be obtained by using socio-demographic data.

Distribution of gender versus lifestyles towards mobility

The second socio-demographic characteristic that was analyzed is gender. The data required to analyze this characteristic is the percentage of people with a certain gender and a certain lifestyle. Three of the studies in this meta-analysis provide this data and will be used for comparison. In Table 8 an overview is given of the studies included in this analysis. In appendix B the databases from the used studies can be found.

Table 8 Studies used in the gender analysis

#	Author	Study title
1	Anable, J. (2005)	Complacent Car Addicts of Aspiring Environmentalists? Identifying travel behavior segments using attitude theory
2	Prillwitz, J. (2011)	Moving towards sustainability? Mobility styles, attitudes and individual travel behavior
3	Anable, J. (2011)	Segment baseline survey of New Employees - Utrecht

In the first study by Anable (2005), the first two segments have disproportional more males, while segment 3 and segment 4 have disproportional more females (especially segment 4). In the second study from Prillwitz & Barr (2011), the first segment has more males. The second segment has slightly more females but it must be taken into account that more women (60% women versus 40% men) completed the questionnaire and when correcting for this, there is a larger proportion of men in segment 2. In segment 3 and 4 there are more women, also when taken into account that there are more women who completed the questionnaire. In the third study Anable (2011) segment 1 and 2 have more men and segment 3 and 4 have more women. The differences between men and women in all studies is not substantial but also in this analysis the patterns for the different studies are the same. There is no significant difference, but there can be assumed that the dataset can be used to create a distribution of lifestyles towards mobility based upon the gender data.

Distribution of income versus lifestyles towards mobility

The income class of participants is given in only two studies, in Anable (2005) and Prillwitz & Barr (2011). Both data sets indicate that people in segment 1 and 2 have a higher income, while in segment 3 and 4 there are more people with a low and average income. A contradiction is found where Prillwitz & Barr conclude that segment 3 has many people with a low income (26% under £10k) while Anable concludes that people in segment 3 have a more average income (only 7% under £10k while 56% is between £10k and £40k).

Distribution of education versus lifestyles towards mobility

The second socio-demographic characteristics mentioned in this subsection is education. This characteristic is also only mentioned in two studies. Anable (2005) makes a difference between having a degree and not having a degree. The other study Anable (2011) gives the education by the number of years of education. The first study from Anable shows that in segment 3 there is a disproportional amount of people having a degree, while the rest of the segments have a more average distribution. In the second study from Anable, the car contemplators have a high proportion of people who are still studying. The car contemplators is one of the two lifestyles belonging to segment 4. The active car owners in segment 3 have a very high proportion of people with 20+ years of education, while the status seekers in segment 1 have, compared to the average of the sample, a high proportion of people with 15 years or less education and people with16-19 years of education. Since the classification that is used to show the amount of education, is different between the two studies it is hard to compare. Anable concludes in both studies that people with a higher education have a higher chance at belonging to segment 3 where environmental issues are seen as important. No contradicting evidence is found.

Distribution of household size versus lifestyles towards mobility

The third socio-demographic characteristic mentioned in this subsection is the household size. The household size is not mentioned in any of the studies. However, a distinction is made in several studies between households with children, households without children and the amount of adults. Anable (2005) shows that single adult households represent a high proportion in segment 4. The two lifestyles in this segment are the car-less crusaders and the reluctant riders and have respectively 37% and 42% of the single adult household while the other segment all have about 10%. This also works the other way around; households with a duel income have a low proportion in segment 4 and a high proportion in the other segments. Said differently, less single adult households own a car. Anable (2011) has compared to Anable (2005) a comparable type of dataset; one adult households have in this study a slightly higher proportion of people in segment 4, which is the same as in the first study from Anable.

Distribution of political opinion versus lifestyles towards mobility

The last socio-demographic characteristic that will be discussed in this subsection is the party that would be voted for. Only one study gives this data (Prillwitz & Barr, 2011). The conservatives have a high proportion in segment 1, the liberal democrats have a disproportional amount of people in segment 2, the greens have a disproportional amount of people in segment 3 and in segment 4 the highest proportion votes for the labour party. Since only one study says something about this socio-demographic characteristic this will not be used furthermore.

4.3.3 Conclusions other socio-demographic characteristics versus lifestyles towards mobility

The six socio-demographic characteristics that are discussed in this subsection gave different insides. A pattern that makes a distinction between the segments based on age and gender is found to be the same in all analyzed studies. Four more socio-demographic characteristics were analyzed: income, education, household size and political opinion. There are patterns in the data but there is not enough data available on these characteristics. Therefore this data is not used in the population synthesis tool that was developed. In Table 9 an overview is given of the analyzed socio-demographic characteristics.

Table 9 Overview of socio-demographic characteristics used in the population synthesis tool

Socio-demographic characteristic	Applied in population synthesis tool
Age	Yes
Gender	Yes
Income	No
Education	No
Household size	No
Political opinion	No

4.4 Conclusions

In this chapter a meta-analysis is been conducted. The research question that was answered in this chapter is:

Which studies are available in the field of lifestyles towards mobility and are there patterns that justify the use of this data (Anable, 2011)*?*

In section 4.1 an overview is given of the studies that were taken into account in this meta-analysis. Seven studies were found in literature. In section 4.2 some comparisons between studies have been found. Some of the studies have the same author, which implies that the most recent study gives more and new insights while based on earlier research. The studies are mostly based upon the theory of planned behavior, but also the theory of cognitive dissonance is used in one of the studies. In all studies several lifestyles have been distinguished of which is expected that the underlying characteristics of the studies are comparable. In section 4.3 the underlying characteristics of the lifestyles were compared and result in four segment groups in which all different lifestyles from the studies are classified. It is also expected that there is a certain relation between socio-demographic characteristics and lifestyles. In several studies the relations are given and compared. A pattern that makes a distinction between the segments based on age and gender is found to be the same in all analyzed studies. Four more socio-demographic characteristics were analyzed: income, education, household size and political opinion. There are patterns in the data but there is not enough data available on these characteristics. Therefore this data is not used in the population synthesis tool that was developed. However, the fact that the studies had no real contradictions is promising and therefore gives no reason for not using the data from the original study from Jillian Anable. Therefore the data from Anable (2011) is used as input for the population synthesis tool and is used to create a distribution of lifestyles based on age and gender.

5. Population Synthesis Tool

In chapter two, a literature study towards population synthesis algorithms is described. The conclusion of this chapter was that the algorithm IPF Multizone is the most applicable algorithm based on several criteria. IPF Multizone is therefore used as a basis in the creation of the population synthesis tool. Besides this, in the previous chapter the usability of the study from Anable (2011) is found to be the most usable. For this reason the data from this study is used. The main goal of the population synthesis tool is to estimate a population that can be used in Fountain. The population should give a distribution over lifestyle types (seven lifestyles towards mobility of Anable) per zone. The research question: '*How can a population synthesis tool be constructed that is able to construct the amount of people traveling between zones with a certain lifestyle based on the socio-demographic characteristics age and gender?',* is answered in this chapter. In section 5.1 an overview is given of the three steps in the tool with an explanation. Then, in respectively section 5.2, 5.3 and 5.4 the three steps are explained in more detail. In section 5.5, the impact of the reference sample is addressed and in section 5.6 the main conclusions of this chapter are given.

5.1 Processes in the population synthesis tool

Before designing a population synthesis tool, it is important to have a clear understanding of the desired output of the tool. In chapter 3, a description of this output is given. The tool must be able to estimate a population that specifies the number of people with a certain lifestyle traveling from zone to zone. Since there are seven lifestyles towards mobility, the tool should be able to estimate seven origin destination matrices (OD) in which the amount of people that travel between the zones is given. An example of one of the OD-matrices is the amount of Status seekers travelling from origin zones 1 to 20 to destination zones 1 to 20. To achieve this output, several steps have been undertaken. An overview of these steps is given in Figure 15.

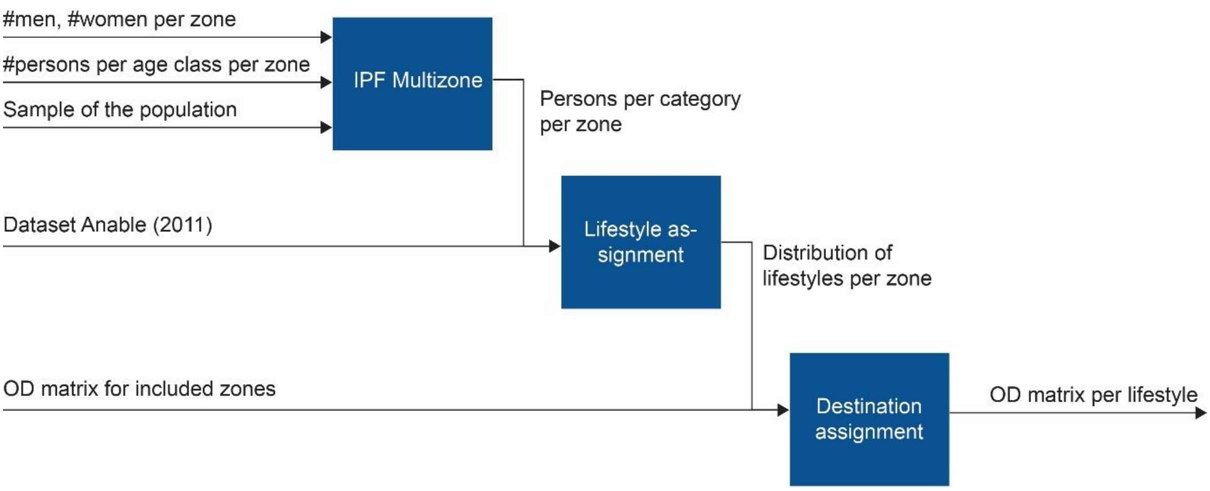


Figure 15 Overview of steps taken to estimate a population synthesis tool

5.1.1 First step in the population synthesis tool "IPF Multizone"

The *first step* was to estimate a population with the correct amount of people and their socio-demographic characteristics age and gender. This was done based on the IPF Multizone algorithm. The required input for this step are the # men, # women per zone, the # people per age class per zone and a reference sample of the population. The output of the first step are the number of persons per category per zone. There are ten categories which will be specified in section 5.2.

5.1.2 Second step in the population synthesis tool "lifestyle assignment"

The *second step* is the lifestyle assignment. Required input for this step are the number of persons per category per zone (this is the output from the first step) and the dataset from Anable (2011). The output of this step is a distribution over lifestyles per zone. With the IPF zone by zone method, the matrices are fitted iteratively to different totals to compute this distribution over lifestyles. For every zone a distribution over seven lifestyles was obtained which can be summed to a total of 1.

5.1.3 Third step in the population synthesis tool " destination assignment"

The third step is the destination assignment. The required input for this step is the output from the second step, the distribution over lifestyles per zone and an OD matrix for the included zones. The output of the destination assignment is an OD matrix per lifestyle. A multiplication of the OD matrix with the distribution over lifestyles will be used to create seven OD-matrices; one for every lifestyle.

5.2 Step 1: IPF Multizone origin per zone

In the previous subsection an overview of the three steps of the tool is presented. In this subsection a more comprehensive description of the first step, IPF Multizone is given. As already mentioned, IPF Multizone uses certain socio-demographic characteristics to obtain the amount of people living per zone in the following ten categories: see Table 10.

Table 10 Categories in IPF Multizone

	Men	Women
Age 0 – 14	Men in age class 0-14	Women in age class 0-14
Age 15 – 24	Men in age class 15-24	Women in age class 15-24
Age 25 – 44	Men in age class 25-44	Women in age class 25-44
Age 45 – 64	Men in age class 45-64	Women in age class 45-64
Age 65+	Men in age class 65+	Women in age class 65+

The data available in the Netherlands is the amount of people living in a zone; the total amount of men living in a zone; the total amount of women living in a zone and the total amount of people in a certain age class living in a zone. The age classes that are available are the classes: 0-14, 15-24, 25-44, 45-64 and 65+. Besides the data about the number of people in a zone, a sample of the population is required that represents the distribution over the categories to perform IPF Multizone. Goal of the method is to obtain a distribution D that specifies the number of people in a zone with characteristics gender and age class. In Figure 16 an overview is given of the first step.

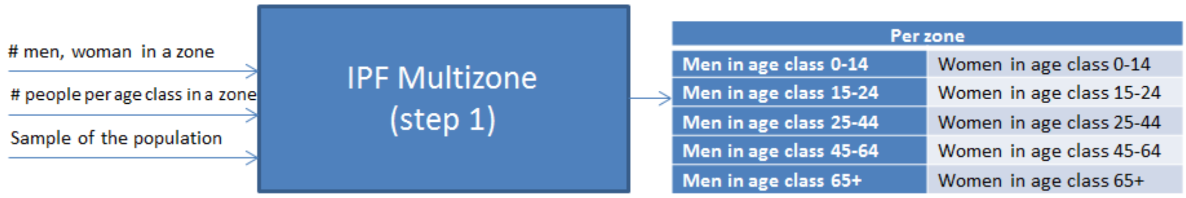


Figure 16 Overview step 1 in population synthesis

below, an extensive explanation of the method is given. In the following subsections the different processes will be elaborated on. Together they form the IPF Multizone method.

5.2.1 Process 1: Calculate totals and averages

The amount of men and women in a zone and the amount of people per age class in a zone needs to be summed per class. This gives a total of the whole area for the amount of men T_{men}, the amount of women T_{women} and the amount of people per age class $T_{ageclass\ k}$. When these totals are calculated, the average amount of people is calculated in a category for all the zones by dividing the totals by the amount of zones T_{zones} in the area. This gives the average amount of men, women and people in age class in a zone $A_{men}, A_{women}, A_{ageclass\ k}$. These numbers are needed in process 4 of this step.

$$T_{men} = \sum_{i=1}^{n} men_{zone\ i} \qquad T_{women} = \sum_{i=1}^{n} women_{zone\ i} \qquad T_{age} = \sum_{i=1}^{n} people\ in\ age\ class\ k_{zone\ i}$$

$$A_{men} = \frac{T_{men}}{T_{zones}} \qquad A_{women} = \frac{T_{women}}{T_{zones}} \qquad A_{age} = \frac{T_{age}}{T_{zones}}$$

5.2.2 Process 2: Use sample reference to find a total distribution over categories

The second process has the goal to find a distribution DW over the categories for the whole area. For this process, the reference sample DW_{ref} is required. An example of a reference sample can be found in Figure 17.

Zone	Reference	man	vrouw	Total
	0-14	25	25	50
	15-24	100	100	200
	25-44	150	150	300
	45-64	100	100	200
	65+	25	25	50
	Total	400	400	

Figure 17 Reference sample

The totals of this reference sample are used to compute a scaling factor S. In the example for men this is done by the following formula:

$$S_{men} = \frac{T_{men}}{DW_{ref}T_{men}}$$

With:

S_{men} =*Scaling factor for men*

T_{men} = *Total amount of men in the area*

$DW_{ref}T_{men}$ = *The total amount of men in the reference sample*

With the same formula a scaling factor for women S_{women} is calculated. This scaling factors are used to compute a first distribution D over categories by using the following formula for every category, in the example this is done for the men in age class 0-14:

$$DW_t = DW_{ref} * S_{men}$$

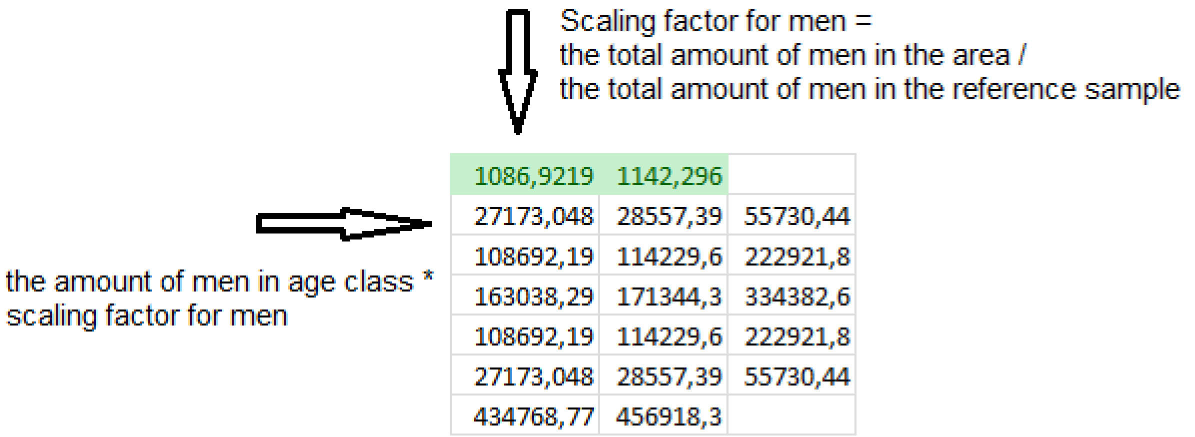
With:

DW_t = *The amount of men in age class 0-14 in the new distribution*

DW_{ref} = *The amount of men in age class 0-14 in the reference sample*

S_{men} = *Scaling factor for men*

The same scaling factor is used for every age class. This computation is also done for women in the exact same way. The visualization of this in excel can be found in Figure 18. The two green cells are the scaling factor for men (left green cell) and women (right green cell). The other cells in the rows are the amount of people according to the first distribution DW_t over age classes. The totals can be found in the lowest and most right cells.



1086,9219	1142,296	
27173,048	28557,39	55730,44
108692,19	114229,6	222921,8
163038,29	171344,3	334382,6
108692,19	114229,6	222921,8
27173,048	28557,39	55730,44
434768,77	456918,3	

Figure 18 Visualization of scaling factor towards men and women

After computing this first distribution DW_t, the totals are correct compared to the actual total amount of men and women in the target area. However, the amount of people per age class is not correct. By scaling towards the totals of gender, the totals for the age classes deviate. For this reason another scaling factor $S_{ageclass}$ is computed by the following formula:

$$S_{ageclass\ k} = \frac{T_{ageclass\ k}}{DW_{ageclass\ k}}$$

With:

$S_{ageclass\ k}$ *= Scaling factor for age class k*

$T_{ageclass\ k}$ *= The total amount of people in age class k for the whole area*

$DW_{ageclass\ k}$ *= The total amount of people in the whole area in age class k according to the last created distribution*.

For every age class a scaling factor S is computed. The visualization of this can be found in Figure 19.

the total amount of people in an age class for the whole area /
the total amount of people in the area according to the last created distribution

Scaling factor	Men	Women	Total
2,891985904	78584,07	82587,57	161171,6
0,518897639	56400,12	59273,45	115673,6
0,759316911	123797,7	130104,7	253902,4
1,193761625	129752,6	136362,9	266115,4
2,344325509	63702,47	66947,82	130650,3
Total	452237	475276,4	

Figure 19 Visualization of scaling factor towards totals of age classes

The scaling factor $S_{ageclass\ k}$ is used in the following formula to obtain a new distribution.

$$DW_{t+1} = S_{ageclass\ k} * DW_t$$

With:

DW_{t+1} *= New distribution for the whole area*

$S_{ageclass\ k}$ *= Scaling factor for age class k*

D_t *= The last created distribution for the whole area*

A new distribution over categories is obtained which matches the exact totals per age class for the whole area, but now the total amount of man and women for the whole area are not correct anymore. The two processes with the scaling factors are called one iteration. The same calculations are used iteratively until the scaling converges close to 1. Once the scaling factor reaches a number close to 1, all the totals are correct and a distribution over the categories for the whole area are found. Five iterations are sufficient to implement until the scaling factor converges close to 1.

5.2.3 Process 3: Create a reference sample per zone

In the previous process, a distribution is computed for the whole area, the area is divided in zones. This distribution for the whole area will be used to create a reference sample per zone. This is done by the following formula:

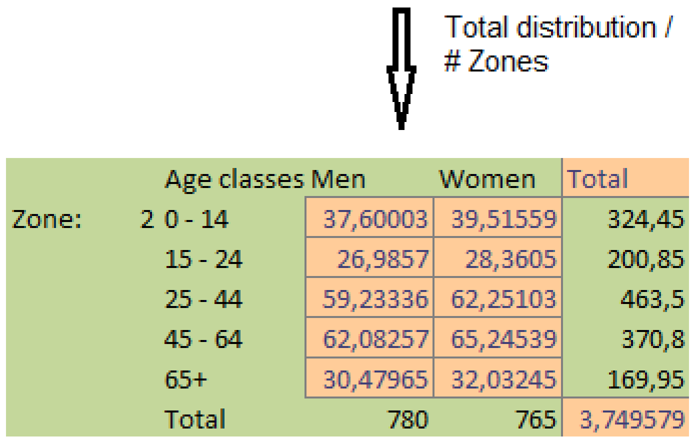
$$DZ_{ref} = \frac{DW_t}{T_{zones}}$$

With:

DZ_{ref} *= Reference sample per zone*

DW_t *= Distribution for the whole area*

T_{zones} *= Total number of zones*



		Age classes	Men	Women	Total
Zone:	2	0 - 14	37,60003	39,51559	324,45
		15 - 24	26,9857	28,3605	200,85
		25 - 44	59,23336	62,25103	463,5
		45 - 64	62,08257	65,24539	370,8
		65+	30,47965	32,03245	169,95
		Total	780	765	3,749579

Figure 20 Expected distibition per zone

The total number of zones depend on the geographical area for which the population will be estimated. The previous formula gives a reference sample on the zonal level. In the next process this will be used in the same way as the reference sample in process 2. The visualization in excel can be found in Figure 20.

5.2.4 Process 4: IPF Zone by zone to compute the distribution over categories per zone

In process 3, an average expected distribution is created for all the zones which can be used as a reference sample. In this process, the same iterations that were presented in process 2 are used but in this process per zone. The data on T_{men}, T_{women} and $T_{ageclass\ k}$ on the zonal is available, to obtain the scaling factors, the average amount of men, woman and persons per age class are used which were produced in the first process ($A_{men}, A_{women}, A_{ageclass\ k}$). Five iterations are sufficient to implement until the scaling factor converges close to 1.

Scaling factor	Men	Women	Total
1,01	27,18195	35,56689	62,74884
1,01	63,42455	82,98941	146,414
1,01	129,8693	169,9307	299,8
1,01	42,28303	55,32627	97,60931
1,01	42,28303	55,32627	97,60931
Total	277,8599	363,5727	704,1814

Figure 21 Final distribution per zone after IPF Zone by zone

Once the convergence is reached, a final distribution over the chosen categories per zone DZ is computed. In Figure 21 an example is given of such a distribution $DZ_{zone\ i}$. In the first column the scaling factor can be found, in the second column the number of men in the reference sample per age class can be found (with in the last row the total amount of men in the reference sample). In the third column the number of women in the reference sample per age class can be found (with in the last row the total amount of women in the reference sample). In the fourth column the total amount of people in age class 1 to 5 in the reference sample can be found (with in the last row the total amount of people in the reference sample). This dataset will be used as input for the next step, the lifestyle assignment.

5.3 Step 2: Distribution of lifestyles towards mobility per zone

In the previous section, the IPF Multizone method is described. This step was based on the Iterative Proportional Fitting method. In this section, the lifestyle assignment is described in full detail. The purpose of the lifestyle assignment is to estimate a distribution over lifestyles towards mobility per zone. The lifestyles are based on Anable (2011) which was originally used in Fountain. Seven lifestyles towards mobility are distinguished: status seekers, devoted drivers, reluctant pragmatics, practical travelers, active car drivers, car contemplators and car free choosers. In chapter four, a meta-analysis is reported towards several studies about lifestyles. Based on the meta-analysis there are strong indications that the dataset from Anable (2011) is reliable, therefore it is used in the population synthesis tool. Up until now, the number of people living in a zone with the socio-demographic characteristics age and gender is available (DZ). Based on this dataset and the data from Anable (2011), the distribution of lifestyles per zone will be computed. In Figure 22 an overview is given of the lifestyle assignment which will be elaborated on in the following subsections.

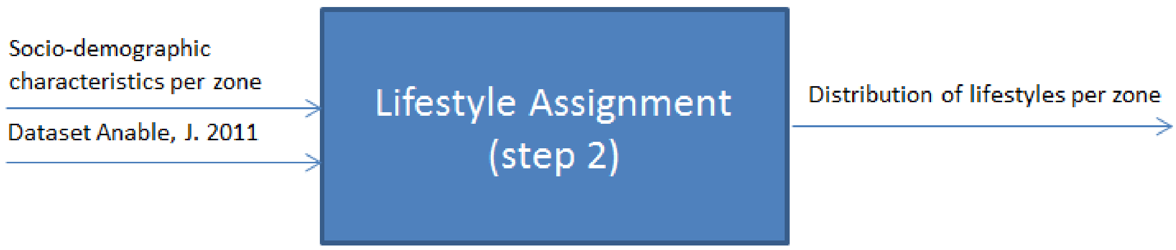


Figure 22 Overview lifestyle assignment in population synthesis

5.3.1 Process 1: Recalculate age classes

The age classes in the database do not have the same age classes as the study from Anable (2011). For this reason some recalculation on the database is required. The available age classes are 0-14, 15-24, 25-44, 45-64 and 65+. These classes need to be converted into the classes from Anable (2011): <24, 25-34, 35-44, 45-54 and >55. In Table 11 an overview is given of the formulas that are used.

Table 11 Recalculation to correct age classes

From age classes	To age class	Formula
A = 0-14	<24	= A + B
B = 15-24	25-34	= 0,5 * C
C = 25-44	35-44	= 0,5 * C
D = 45-64	45-54	= 0,5 * D
E = 65+	>55	= (0,5 * D) + E

In the first column the original age classes are given a letter from A to E. In the second column the age classes that are required are given. In the third column a formula is formulated using the letters given in the first column[1].

5.3.2 Process 2: Segment factors based on dataset Anable, J. 2011

In the previous process the dataset is recomputed based on age classes used by Anable (2011). In this process the dataset of Anable (2011) is used as input.

Segment assignment		Status Seekers	Devoted Drivers	Reluctant	Practical travellers	Active car owners	Car contemplators	Car free choosers
		0,066	0,159	0,244	0,163	0,297	0,025	0,044
	men	0,67	0,51	0,72	0,65	0,58	0,63	0,29
	women	0,33	0,49	0,28	0,35	0,42	0,38	0,71
	0-24	0,03	0,17	0,19	0,27	0,05	0,22	0,08
	25-34	0,65	0,52	0,41	0,45	0,7	0,57	0,64
	35-44	0,12	0,17	0,22	0,18	0,11	0,16	0,112
	45-54	0,12	0,09	0,09	0,1	0,14	0,05	0,12
	55+	0,09	0,04	0,09	0,005	0,005	0,005	0,04
	⇨	0,0013266	0,0137853	0,0333792	0,0286065	0,008613	0,003465	0,0010208
=lifestyle % * men * 0-24		0,028743	0,0421668	0,0720288	0,0476775	0,120582	0,0089775	0,0081664
		0,0053064	0,0137853	0,0386496	0,019071	0,0189486	0,00252	0,00142912
		0,0053064	0,0072981	0,0158112	0,010595	0,0241164	0,0007875	0,0015312
		0,0039798	0,0032436	0,0158112	0,00052975	0,0008613	0,00007875	0,0005104
		0,0006534	0,0132447	0,0129808	0,0154035	0,006237	0,00209	0,0024992
		0,014157	0,0405132	0,0280112	0,0256725	0,087318	0,005415	0,0199936
=lifestyle % * women * 55+		0,0026136	0,0132447	0,0150304	0,010269	0,0137214	0,00152	0,00349888
		0,0026136	0,0070119	0,0061488	0,005705	0,0174636	0,000475	0,0037488
	⇨	0,0019602	0,0031164	0,0061488	0,00028525	0,0006237	0,0000475	0,0012496

Figure 23 Calculations on data Anable, J. (2011)

In Figure 23 the dataset from Anable (2011) can be found, the first row represents the seven lifestyles, in the second row the proportion of people with a certain lifestyle $P_{lifestyle\ l}$ can be found. The third and fourth row show the men/women distribution $P_{gender\ g\ in\ lifestyle\ l}$. The following five rows show the proportions for age classes $P_{age\ class\ k\ in\ lifestyle\ l}$. This data is

[1] For reflection on this part see page 88.

used to develop lifestyle assignment factors LA_{lkg}. With lifestyle l, age class k and gender g. Every lifestyle is multiplied by every gender and every age class. This gives:

*7 lifestyles * 2 genders * 5 age classes = 70 factors*

The factors can be found in Figure 23 in the rest of the rows. Two examples are given:

$$LA_{1,1,1} = P_{Status\ seekers} * P_{men\ status\ seekers} * P_{0-24,status\ seekers}$$

$$= 0{,}066 * 0{,}67 * 0{,}03 = 0{,}0013266$$

$$LA_{1,5,2} = P_{Status\ seekers} * P_{women\ status\ seekers} * P_{55+,status\ seekers}$$

$$= 0{,}067 * 0{,}33 * 0{,}09 = 0{,}0019602.$$

The total of all 70 factors is 1.

5.3.3 Process 3: Segment assignment

The 70 scaling factors that were produced in the previous process, are used to create the data as shown in Figure 24. The table in this figure is based upon the following formula:

$$LD_{lkg} = LA_{lkg} * DZ_i$$

With:

LD_{lkg} = *The lifestyle distribution with lifestyle l, age class k and gender g*

LA_{lkg} = *The lifestyle assignment factor with lifestyle l, age class k and gender g*

DZ_i = *The distribution for zone i*

This could for example be:

The factor for men between 0-14 in lifestyle status seekers * The number of men in zone 1 in age class 0-14. In Figure 24 the proportion of status seekers between 0-14 years old that are men can be found in the first cell in the second row.

In Figure 24 the proportion of men per age class and per lifestyle can be found. The same table is made for women.

Segment assignment men	status seekers	Devoted drivers	Reluctant Pragmatics	Practical travellers	Active car owners	Car contemplators	Car free choosers
⇨	0,01698048	0,17645184	0,42725376	0,3661632	0,1102464	0,044352	0,01306624
= men in zone 1 between 0-14 * factor men between 0-14 in lifestyle status seeker	0,1034748	0,15180048	0,25930368	0,171639	0,4340952	0,032319	0,02939904
	0,01910304	0,04962708	0,13913856	0,0686556	0,06821496	0,009072	0,005144832
	0,03608352	0,04962708	0,10751616	0,072046	0,16399152	0,005355	0,01041216
	0,05253336	0,04281552	0,20870784	0,0069927	0,01136916	0,0010395	0,00673728
	0,2281752	0,470322	1,14192	0,6854965	0,78791724	0,0921375	0,064759552

Figure 24 Segment assignment men

Because of the small factors, a scaling factor is required to reach the correct totals. To obtain a scaling factor, for example for the number of men in age class 0-14 for zone 'one', this needs to be divided by the total of the first row in Figure 24. This gives the total number of men in age class 0-14 with all lifestyles together. This gives 5 scaling factors for the table for men and 5 scaling factors for the table of women, one for each age class. The numbers in Figure 24 are multiplied by the scaling factors with the result that the total number of men in an age class after lifestyle assignment is equal to the total number of men in an age class before lifestyle assignment.

5.3.4 Process 4: IPF Zone by zone

In the previous process, the dataset from Anable (2011) is used to adjust the distribution over lifestyles towards mobility based on age and gender. Now, the Iterative Proportional Fitting method is used again, to fit the population to the dataset of Anable (2011). In section 5.2, an iteration was done. Towards the total men and women in a zone and towards the total number of people in an age class. Now the same will be done, but then in four directions instead of two. One iteration consist in this situation of scaling towards the ratio men and women, towards the ratio of age classes, towards totals of men and women and towards totals of lifestyles. 24 iterations are performed with, as a result the following table per zone. For all zones, a distribution for the seven lifestyles towards mobility is given. In Table 12 the distribution over lifestyles towards mobility can be found. In the first row the seven lifestyles can be found. In the following rows the number of people per lifestyle can be found for zone 1 to 20.

Table 12 Distribution over lifestyles towards mobility in # of persons

Zone	Stat	Dev	Rel	Prag	Acti	Car Con	Car Free	Total
1	7,822099	10,6497	27,86003	7,474506	12,14662	0,252327	3,011375	69,21666
2	56,95676	260,6725	399,9439	277,9486	442,2104	29,00346	45,12773	1511,863
3	7,822099	10,6497	27,86003	7,474506	12,14662	0,252327	3,011375	69,21666
4	7,822099	10,6497	27,86003	7,474506	12,14662	0,252327	3,011375	69,21666
5	4,876137	7,913557	15,87196	4,131863	8,979998	0,192297	2,754068	44,71988
6	12,75258	22,45111	51,28327	16,06371	21,72063	0,637474	6,192199	131,101
7	4,876137	7,913557	15,87196	4,131863	8,979998	0,192297	2,754068	44,71988
8	4,876137	7,913557	15,87196	4,131863	8,979998	0,192297	2,754068	44,71988
9	5,581193	9,835211	36,6152	14,38898	8,596285	0,56828	1,201187	76,78633
10	3,414123	4,957219	13,75882	3,87276	13,87991	0,321222	0,997003	41,20106
11	4,166768	4,647568	9,318159	1,245106	2,862315	0,048862	2,60754	24,89632
12	59,7395	103,6477	191,2543	50,90032	139,122	2,817357	38,21831	585,6995
13	1,003788	3,556306	8,257551	6,410049	4,258967	0,284047	0,451562	24,22227
14	7,686893	7,256263	26,93342	6,085734	15,08834	0,252546	1,486061	64,78926
15	39,48576	10,03992	34,8594	1,169714	13,98962	0,024383	10,88239	110,4512
16	2,327269	2,111021	8,553663	1,630195	2,349496	0,060223	0,426031	17,4579
17	0,352703	1,872606	3,640194	1,923333	0,332606	0,14211	0,26358	8,52713
18	38,20269	10,21216	56,07649	3,936287	50,44143	0,119154	3,681158	162,6694
19	253,1579	437,7053	963,075	299,4481	551,8136	13,64637	122,1455	2640,992
20	942,5472	993,273	2001,37	232,4454	663,2374	11,47705	581,2211	5425,571

5.3.5 Process 5: Recalculate towards proportions

The required output was the proportional distribution over lifestyles. In the last process of the lifestyle assignment, the dataset is recalculated towards percentages instead of totals per lifestyle. In Table 13 an example is given of what the output looks like. In the first row the seven lifestyles can be found. In the following rows the distribution over lifestyles for zone 1 to 20 can be found.

Table 13 Distribution over lifestyles towards mobility in percentages

	Stat	Dev	Rel	Prag	Acti	Car Con	Car Free
Zone 1	0,05	0,23	0,35	0,18	0,13	0,01	0,05
2	0,15	0,18	0,43	0,06	0,11	0,00	0,07
3	0,05	0,20	0,30	0,21	0,19	0,01	0,05
4	0,12	0,11	0,47	0,11	0,17	0,00	0,02
5	0,25	0,16	0,39	0,02	0,06	0,00	0,13
6	0,20	0,12	0,39	0,05	0,20	0,00	0,05
7	0,17	0,19	0,37	0,05	0,11	0,00	0,10
8	0,11	0,15	0,40	0,11	0,18	0,00	0,04
9	0,04	0,13	0,35	0,26	0,20	0,01	0,01
10	0,07	0,20	0,33	0,15	0,20	0,01	0,05
11	0,08	0,21	0,34	0,12	0,18	0,01	0,06
12	0,25	0,16	0,39	0,02	0,06	0,00	0,13
13	0,08	0,12	0,33	0,09	0,34	0,01	0,02
14	0,09	0,16	0,34	0,08	0,28	0,01	0,04
15	0,12	0,11	0,42	0,09	0,23	0,00	0,02
16	0,12	0,11	0,42	0,09	0,23	0,00	0,02
17	0,13	0,12	0,49	0,09	0,13	0,00	0,02
18	0,13	0,12	0,49	0,09	0,13	0,00	0,02
19	0,04	0,22	0,43	0,23	0,04	0,02	0,03
20	0,08	0,12	0,33	0,09	0,34	0,01	0,02

5.4 Step 3 : Destination assignment

In the previous section the distribution over lifestyles towards mobility per zone is created. A dataset with the number of persons with a certain lifestyle per zone and a dataset with only distributions of lifestyles per zone are created. The third step is the destination assignment. Once the destination is assigned the output of the population synthesis tool is an OD-matrix for every lifestyle. To achieve this goal, two methods can be used. Both methods will be shortly discussed.

The first method is to implement an algorithm that assigns all the persons in a population to a working zone. This can for example be done with a gravity model (Isard, 1956). The other option is to use an existing OD matrix as a basis of the destination assignment. Since there is a traffic model with an OD matrix available for most of the cities in the Netherlands, this method will be used. Because this method already has data on the amount of people living and working in zones, the only dataset that is required is the distribution over lifestyles in percentages.

In Figure 25 an overview is given of the destination assignment, the required input for the destination assignment, is the distribution over lifestyles per zone in percentages. This input is created in the previous section. Other input is the OD matrix. The output of the destination assignment are seven OD matrices, one for every lifestyle. In this section the processes in the destination assignment will be explained comprehensively.

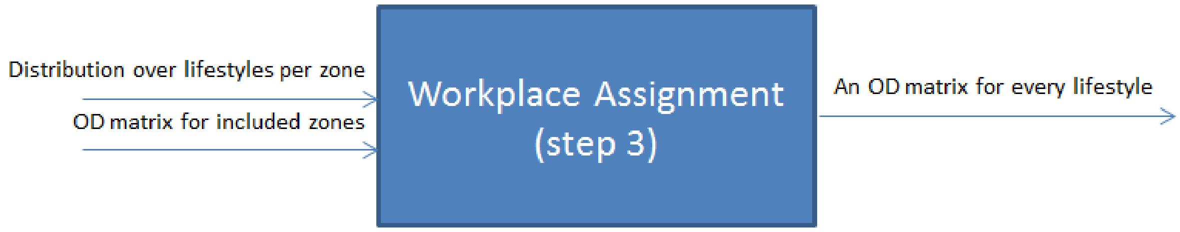


Figure 25 Overview of the destination assignment

5.4.1 Process 1: Multiplication of percentages of lifestyles towards mobility and the OD matrix

The destination assignment has only one process, the OD matrix is multiplied by the percentages of lifestyles. Every zone has a distribution over the lifestyles. In Table 13 the value in the second row from the second column is the proportion of persons living in zone 1 and having the first lifestyle, the status seekers. This proportion is multiplied by the whole first row of the OD matrix. This first row of the OD matrix represents persons living in zone 1 and travelling towards other zones. This is done for all the rows and gives the OD matrix for the status seekers. A visualization of this process can be found in Figure 26. The first matrix is the OD matrix for 10 zones used as input, the second matrix is the created proportion of status seekers in the first 10 zones and the third matrix is the product of both matrices and gives the OD matrix for status seekers in the first 10 zones. For example the value 23,46278 in the third matrix, fourth row and seventh column are the amount of status seekers travelling between origin zone three and destination zone six.

	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	133,701	0	0	0	0
4	0	0	0	0	0	10,143	0	0	0	0
5	0	0	0	0	0	0	0	1,693	0,019	0
6	0	0	96,394	39,579	0	0	0	0	0	0
7	0	0	0	0	0,579	0	0	0	1,971	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0,016	0,001	0	1,925	0	0
10	0	0	0	0	0	0	0	0	0	0

X

Zone	%
1	0,113009
2	0,037673
3	0,113009
4	0,113009
5	0,109037
6	0,097273
7	0,109037
8	0,109037
9	0,072685
10	0,082865

=

	1	2	3	4	5	6	7	8	9	10
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	23,46278	0	0	0	0
4	0	0	0	0	0	1,779964	0	0	0	0
5	0	0	0	0	0	0	0	0,339964	0,003815	0
6	0	0	15,97043	6,557394	0	0	0	0	0	0
7	0	0	0	0	0,116266	0	0	0	0,395788	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0,001791	0,000112	0	0,215505	0	0
10	0	0	0	0	0	0	0	0	0	0

Figure 26 Original OD matrix * Status Seeker proportions = OD matrix Status Seekers

As can be seen in Figure 26 the values in the OD matrix are not all integer values. This might be a problem since agents are integer. To solve this problem the values in the OD matrix are the chance that an agent is created in Fountain. With this method the correct amount of agents is modelled.

5.5 Impact of the reference sample

In the section 5.2, 5.3 and 5.4 the three steps of the population synthesis tool have been described. What might have come to attention, is that the input of every step is based upon the output of the previous step and therefore everything is based upon the reference sample. The reference sample thus has a substantial impact on the output of the tool. In chapter six, section 6.1.3 this is discussed in more detail.

5.6 Conclusions

In this chapter an extensive description of the population synthesis tool is given, three steps have been distinguished. The first step is the creation of a population with IPF Multizone, a population is estimated which presents the amount of people living in a zone and the distribution of those people over gender and age classes. In the second step of the population synthesis tool, the lifestyles towards mobility are assigned to the population based upon the relation with socio-demographic characteristics gender and age. The third step in the population synthesis tool is the workplace assignment. This is done by using an existing OD matrix and multiply the matrix by the distribution of lifestyles that was computed in the lifestyle assignment. The output of the population synthesis tool are seven OD matrices, the matrices together present a detailed population for the whole geographical area that is used with living places, working places (origin and destination) and lifestyle per person. The population can be used as input for Fountain. In the next chapter, two case studies will be conducted were the population synthesis tool will be deployed and tested.

6. Case studies

To measure the performance of the population synthesis tool described in the previous chapter, two case studies are conducted. In each case study, two populations are estimated and the performance is measured. In this chapter both case studies are described. In section 6.1 the case of 'Amsterdam VMA' is described. In section 6.2 the case 'Utrecht rush hour avoidance' is described. In section 6.3 limitations of the population synthesis tool are given and finally in section 6.3 the main conclusions of this chapter are presented. The main goal of this chapter is to answer the following research questions:

'Is the population estimated by the population synthesis tool an improvement compared to the original population in Fountain?'

6.1 Amsterdam VMA

VMA is the traffic model of the city Amsterdam ('Verkeersmodel Amsterdam'). The geographical location that was chosen for this case study is the city of Amsterdam. In subsection 6.1.1 an overview of the data collection is given. In subsection 6.1.2 a case description is given. In 6.1.3 and 6.1.4 the results of several analyses are given. Finally in 6.1.5 the conclusions can be found.

6.1.1 Data collection VMA Amsterdam

Based upon the description of the population synthesis tool in chapter 5, a list can be found in Table 14 with the data required as input for the tool. In the first column the type of data can be found. In the second column the availability of the data is determined and in the third column the step in the population synthesis where the data is required, is given.

Table 14 Data collection VMA

Data	Available	Required for
Transport network	In VMA	Fountain
Total men and women per zone	In VMA	IPF Multizone
Total person in age classes per zone	In VMA	IPF Multizone
Reference sample of the area	Not available (but can be created based on data in VMA)	IPF Multizone
OD matrix	Not required in this case study	Destination assignment

First of all, VMA ('Verkeersmodel Amsterdam') is required including all the data. Permission from the municipality of Amsterdam was necessary. Fortunately this was given and therefore the model can be used in this research.

The VMA has a transport network with a zone classification, the infrastructure and the public transport network. In the database belonging to the VMA, a dataset is available in which for every zone the total amount of men, total amount of women and total amount of persons in the following age classes (0-14, 15-34, 35-64 and 65+) are documented. There is no reference sample, but a database which is available in the VMA of Amsterdam also contains the number of men per age class and the number of women per age class. This is equal to the output of the IPF Multizone method in the population synthesis tool. The totals can be used as a reference sample. Besides the fact that this information can be used to create a reference sample, the dataset can also be compared with the output of the IPF Multizone method from the population synthesis tool. This makes the data and this case very interesting because according to Brederode & Waanders (2013) it is very hard to find a complete database with this type of data.

The analysis that compares the synthetic population from the population synthesis tool after performing IPF Multizone with the actual population is reported in subsection 6.1.3. Finally in the destination assignment, an OD matrix for the area was required. The analysis in this case was only performed on the first two steps of the population synthesis tool, the IPF Multizone and the lifestyle assignment, because there was no data available from a case study in Amsterdam. There is data available in Utrecht from project 'rush hour avoidance'. For this reason a second case study is performed. The OD matrix is for this reason not required in the first case study.

6.1.2 Case description VMA Amsterdam

The model of Amsterdam (VMA) has 1078 zones. The area has 766396 citizens from which 377.520 men and 388.876 women. The dataset of the VMA is categorized in four age classes, 0-14, 15-34, 35-64 and 65+. There are respectively 120.302, 243.433, 317.351 and 85.310 persons in these age classes living in the area. A map of the area can be found in Figure 27.

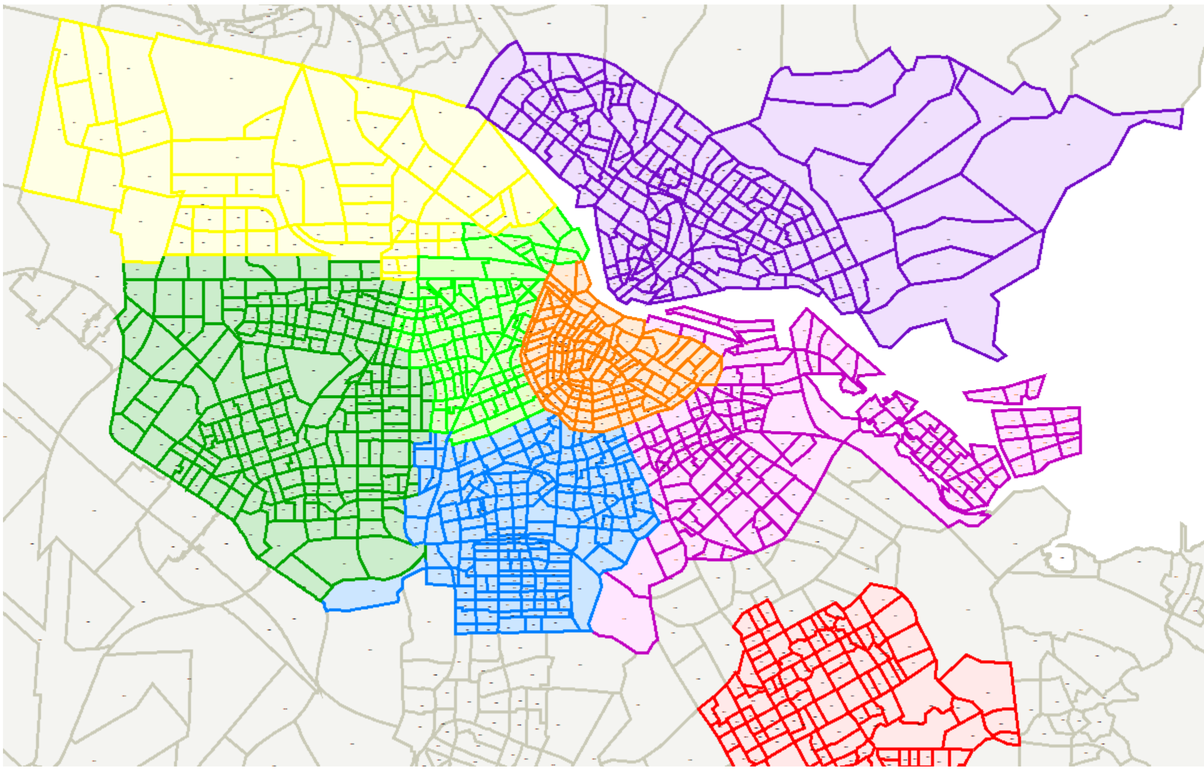


Figure 27 Map of VMA

As can be seen, the area is classified in different colors: every color represents a subarea in Amsterdam. In Table 15 an overview is given with the characteristics of the subareas. For each subarea the number of zones, the area name and the number of people living there are given.

Table 15 Characteristics VMA

Color	Letter	Area name	# Zones	Zone numbers	Number of people
Orange	A	City Center	129	1 - 119, 122 - 131	82448
Lime green	E	Bos & Lommer	104	132 - 205, 212 - 221, 611 - 621, 120, 121, 760, 761, 764, 765, 766, 775, 776	131199
Blue	K	Amsterdam Zuid, De Pijp	182	206 - 211, 222 - 328, 884 - 950, 845, 846	133637
Pink	M	Zeeburg, Diemen	151	329 - 428, 1086 - 1136	116574
Purple	N	Amsterdam Noord	172	429 - 600	86282
Spring green	F	Amsterdam Nieuw-West	205	634, 635, 671 - 675, 677 - 759, 762, 763, 767 - 774, 777 - 844, 847 - 883	134778
Red	T	Amsterdam Zuid-Oost, Gaasperdam, Abcoude	57	601 - 610, 622 - 633, 636 - 670, 676	81478
Yellow	B	Westpoort	135	951 - 1085	407

6.1.3 Analysis: Performance IPF Multizone after step 1 of the population synthesis tool

In order to evaluate the performance of the population synthesis tool, the accuracy of the estimated synthetic population had to be measured. In this subsection the IPF Multizone method in the population synthesis tool is addressed. The deviation between the actual population and the synthetic population in the area of Amsterdam was measured. The actual distribution of persons in Amsterdam over the categories (the categories can be found in Table 16) is available and was used to validate the synthetic population that was estimated with the IPF Multizone method in the population synthesis tool.

Table 16 Categories

Men in age class 0-14	Women in age class 0-14
Men in age class 15-24	Women in age class 15-24
Men in age class 25-44	Women in age class 25-44
Men in age class 45-64	Women in age class 45-64
Men in age class 65+	Women in age class 65+

According to Brederode & Waanders (2013) It is hard to make a statement about the general performance of population synthesis algorithms, since the population can only be tested in a problem where the actual population is known. Since the actual population for this case study is provided by the municipality of Amsterdam, the comparison can be done. The data is coming from the database STIF, which are registrations on addresses in the municipality of Amsterdam. This dataset is aggregated to the VMA zonal level.

The evaluation methods for population synthesis, commonly use a comparison of two populations. An evaluation function or measure of fit can be used. These functions can theoretically be everything and there is no standard function that is considered to be best. In literature, there is a large amount of different measures of fit available. It is common to use more than one measure of fit to test the fit of the populations as they might have different conclusions about how good the fit is, depending on the importance of different properties of the population. Besides this, some of the measures of fit will check the fit at marginal levels while others check it at individual cells.

Generally, the measures of fit can be classified in three categories: general distance statistics, information-based statistics and traditional statistics. The general distance statistics calculates distance between the original population and the synthetic population. The information-based statistics have their origin in the information gain statistics. They reflect how much information could be gained from knowing the actual population. The third category are traditional statistics such as the coefficient of determination, the pearson chi-square and the relative sum of squared Z-scores (Huang & Williamson, 2001), (Fotheringham & Knudsen, Goodness of fit statistics, 1987), (Fotheringham & Knudson, 1986). In Brederode & Waanders (2013), a summary is given of the available methods in literature.

From this list, four general distance methods were chosen and one general statistic method was chosen to apply as a measure of fit in this case study. Another sixth method was used which is an information-based statistics method. With this method the optimal reference sample for this case was calculated based on a minimization function in which the total absolute error or the total absolute distance between two populations is minimized.

Table 17 Measures of fit to measure performance of IPF Multizone in the population synthesis tool

Name	Function	Type
Total absolute error/ total absolute distance (Huang & Williamson, 2001)	$TAE, TAD = \sum_i \lvert y_i - y_i' \rvert$	General distance
Standardized absolute error (Huang & Williamson, 2001)	$SAE = \frac{1}{n} \sum_i \lvert y_i - y_i' \rvert$	General distance
Mean absolute percentage error (Temeurshoev, Yamano, & Webb, 2009)	$MAPE = \frac{1}{n} \sum_i \frac{\lvert y_i - y_i' \rvert}{y_i} * 100$	General distance
Root mean square error (Fotheringham & Knudsen, 1987), (Fotheringham & Knudson, 1986)	$RMSE = \sqrt{\sum_i \frac{(y_i - y_i')^2}{n}}$	General distance
Pearson Chi-square (Fotheringham & Knudson,	$\chi^2 = \sum_i \frac{(y_i - y_i')^2}{y_i'}$	Traditional statistics

1986) (Fotheringham & Knudsen, 1987) and (Huang & Williamson, 2001)		
Minimization of total absolute error	$Optimal\ reference\ sample = MIN \sum_i \lvert y_i - y_i' \rvert$	Information-based statistics

In Table 17 an overview is given of the six methods, y represents the actual population and y' represents the synthetic population. n represents the number of zones in this case. In the first column the name is given, in the second column the function is given and in the third column the type is given.

General distance

The four methods characterized as general distance method have been used on the results of the population synthesis tool and the actual population of Amsterdam. The results can be found in Table 18. In this analysis the yellow area (zone B) that is defined in the case description has a large amount of zones without people living there. For this reason this area was not taken into account in the analysis.

Table 18 Measures of fit VMA

Method	Total	A	E	K	M	N	F	T
TAE, TAD	49595	6789	7913	7816	6134	5886	8609	6448
SAE	57,6	54,8	80,7	55,4	55,8	43,6	54,1	68,6
MAPE	8,38	10,84	7,04	7,77	7,03	9,19	7,90	8,73
RMSE	73,7	71,46	97,88	67,94	72,07	56,87	66,26	89,36

The total absolute error or total absolute distance (TAE/TAD) is the sum of all differences between the actual population and the synthetic population. For every category in every zone the difference between the two values for the number of persons with a certain gender, in a certain age class, is found and the sum of all these values is the total TAE/TAD. Since there are so many zones on such a low scale, the total amount of the TAE/TAD is expected to be high compared to the TAE/TAD of two populations for the same geographical area with less zones. The reason is that with less zones the values are averaged and therefore the total error per zone is lower. This results in a lower total absolute error. The total absolute error of the total area is 49595 persons on a total of 766.396 citizens.

The second method is the *Standardized absolute error* (SAE). This is the TAE/TAD divided by the total number of zones in this area and gives the average deviation per zone. In this measure of fit the same logic in cases of larger zones can be applied. The standardizes absolute error for the total area is 57,6. Area E (Bos & Lommer) has a standardized absolute error of 80,7 which is the highest of them, which means that the performance of the population synthesis tool is the worst on this area compared to the other areas.

The third measure of fit is the *Mean absolute percentage error* (MAPE). This measure expresses accuracy by a percentage. Also in this measure a difference in performance with different zone sizes is expected. The mean absolute percentage error for the total area is

8,38% and is especially high for the city center, area A with a mean absolute percentage error of 10,84%.

The fourth measure of fit is the *Root mean square error* (RMSE). The RMSE is a frequently used measure of differences between population values predicted by a model or an estimator and the values actually observed. It represents the sample standard deviation of the differences between predicted values and observed values. As the measure is scale dependent, it is a good measure of accuracy when comparing forecasting errors of different models for a particular variable but not between variables (Hyndman & Koehler, 2006). For the total area the root mean square error is 73,7. Also in this measure of fit area E has a high root mean square error compared to the other areas; a value of 97,88.

As we are also interested in comparison of the zones using multiple statistics, the values of the different measures of fit are scaled to the interval [0,1]. This is done for each measure of fit, dividing the value by the maximum value of that particular measure of fit. The results can be found in Table 19. Note that not every statistic scales linearly. A lower value means a better fit, but a twice as low value does not always indicate a twice as low fit.

Table 19 Measures of fit scaled to interval [0,1]

Method	Area A	Area E	Area K	Area M	Area N	Area F	Area T
TAE, TAD	0,79	0,92	0,91	0,71	0,68	1	0,75
SAE	0,71	0,68	1	0,69	0,69	0,54	0,67
MAPE	0,77	1	0,65	0,72	0,65	0,85	0,73
RMSE	0,75	0,73	1	0,69	0,74	0,58	0,68

In Figure 28 the measures of fit scaled to interval [0,1] can be found in a graph. The bandwidth with values is small. In a reference case study (Brederode & Waanders, 2013), the values are between 0,2 and 1 while in this case the values are between 0,55 and 1. This shows that the performance for different areas is less diverse.

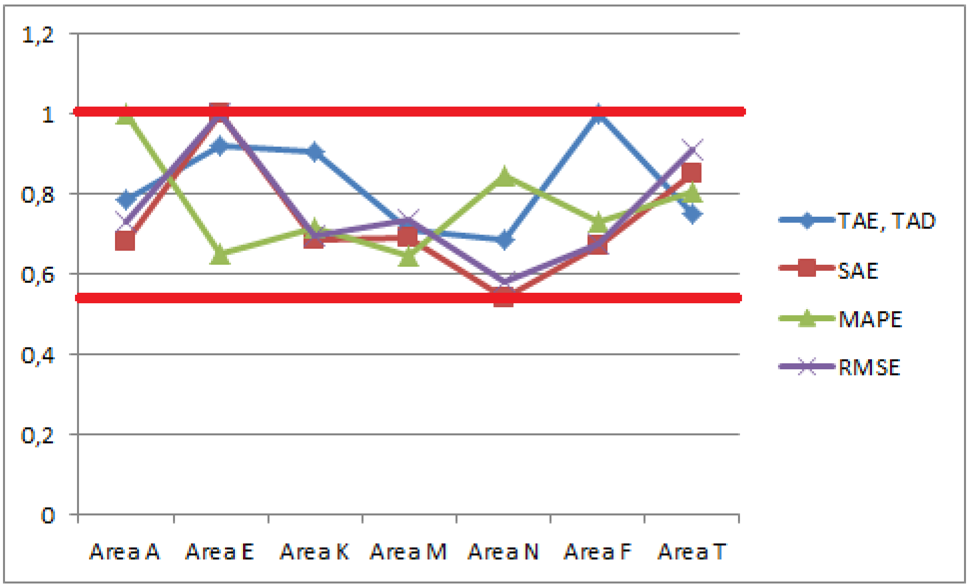


Figure 28 Comparison of performance IPF Multizone between areas

Information-based statistics

The fifth measure of fit is the minimization of the total absolute error and belongs to the category information-based statistics. With this *minimization of the total absolute error* the optimal reference sample can be found. Knowing the actual population, the total absolute error can be minimized by the *excel solver optimization tool*. The optimal reference sample that is found with this optimization gives the best possible result of the IPF algorithm in this case. The minimum total absolute error that was found with the optimization tool is 49595. With a reference sample of the value 1 for all categories, the total absolute error is 139791. The lower the deviation, the better the performance. The effect of different reference samples is high, the difference is about 12% when the error is taken as a proportion of the amount of citizens. Therefore it is important to have an accurate reference sample.

As mentioned, it can be expected that with less zones, the error cancels out because part of the errors are averaged. To test this hypothesis the same analysis has been done per 5 zones. The results can be found in Table 20. As expected, this gives a lower total absolute error/ total persons rate. There can be expected that with even less zones the error cancels out.

Table 20 Results minimization of the total absolute error

Reference sample	Total absolute error	Total absolute error/ Total persons (1 zone)	Total absolute error/ Total persons (5 zones)
Optimized	49595	6,47%	4,96%
Only values 1	139791	18,24%	17,66%

With these results, the performance of the IPF Multizone method in the population synthesis tool can be measured only in respect to each area and how these are related to each other. To see how well the method performs compared to other implementations of IPF Multizone, (the population synthesis algorithm that is used) another case has been found in literature. In Brederode & Waanders (2013) a study is reported towards the Netherlands. The study used several population synthesis algorithms and compared them on a case study with 12 zones, representing the 12 provinces in the Netherlands. This gives 12 zones with a total of 16.165.974 citizens. The total absolute error with the IPF Multizone algorithm was 567758, this gives a 3,51% of deviation rate. Since we just saw that with less zones the rate decreases, the results of this study and our study are hard to compare. Based on the decrease from 6,47% to 4,96% when the number of zones is divided by 5, it is concluded that the IPF Multizone method in the population synthesis performs as expected when compared to the study from Brederode & Waanders (2013).

Traditional statistics

The last method is a traditional statistic and is the *Pearson Chi-square*. The Pearson Chi-square tests a null hypothesis that the frequency distribution of certain events observed in a sample, in this case the actual population, is consistent with a particular theoretical distribution. With this measure of fit, the performance is better when the chi-square value is lower. The Chi-square test can be used to prove that a distribution is not random, therefore this measure of fit can only be used to prove that the outcome of the algorithm is not a random output.

The H_0 hypothesis and the H_1 hypothesis are:

H_0 = There is no significant difference between observed and expected (estimated) population.

H_1 = There is a significant difference between observed and expected (estimated) population.

The chi-square value is calculated by the following formula:

$$\chi^2 = \sum_i \frac{(y_i - y_i')^2}{y_i'}$$

In which:

χ^2 = Chi-square value
y_i = Observed value
y_i' = Expected value

The Chi-square value is 6178,15.

The Chi-square test compares the Chi-square value to the critical Chi-square value. When the Chi-square of this problem is lower than the critical Chi-square value, the chance that the frequency distribution in the actual population is consistent with the synthetic population is over 95% and the null hypothesis will be accepted. The critical Chi-square value depends on two parameters. The first parameter is the significance level which is defined by the 95% certainty and gives an alpha of 0,05. The second parameter is the degrees of freedom, the degrees of freedom are the number of sums in the chi-square formula minus 1. There are 8 variables per zone and 860 zones without zero values (with citizens). Therefore the degrees of freedom are (8*860)-1=6879.

The Chi-square critical value at alpha is 0.05 and 6879 degrees of freedom = 18014.39. The Pearson Chi-square value is 6178.15. The value is lower than the critical value therefore H0 is accepted. There is no significant difference between observed and expected (estimated) population. The outcome of the population synthesis algorithm is not random.

6.1.4 Analysis: Distribution of lifestyles towards mobility

This subsection focusses on analysis in the lifestyle assignment of the population synthesis tool; the lifestyle assignment. The output of the lifestyle assignment is a distribution over lifestyles per zone. On this outcome several analysis were conducted. In the first analysis the distribution of the total population was compared to the original distribution of Anable (2011). It was expected that this is almost the same since the IPF Multizone method scales towards this distribution once in every iteration. In Figure 29 an overview is given of the distributions of lifestyles; they are almost the same. The same analysis is done per color coded area. Most areas still had a similar distribution over lifestyles. There is only a slight difference because there are still many zones per area and therefore the results are averaged.

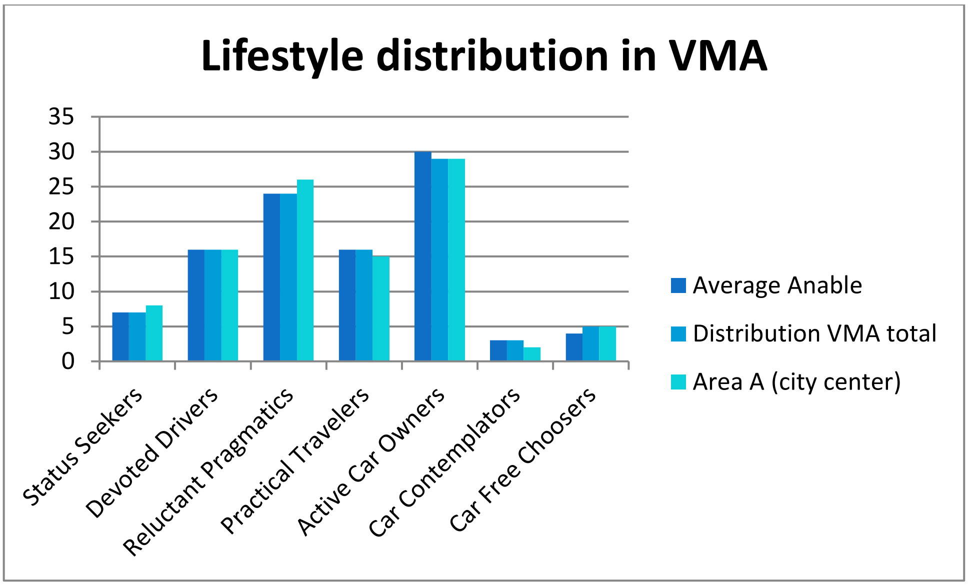


Figure 29 Total distribution of lifestyles towards mobility actual and synthetic

As mentioned, the IPF Multizone assures that the totals of the number of people in a category stay almost the same. Therefore it might be interesting to see what the mean absolute percentage error (MAPE) in this dataset is. While in the performance measurement of the IPF Multizone method, alarger MAPE was a negative output. In the lifestyle assignment it does not say anything about the performance of the tool but, in this case a higher MAPE supports the thought that not all distribution of lifestyles are the same for every zone and that people living in the city center probably have another distribution over lifestyles than for example people living in a suburb. For example a MAPE of 8% says that the average difference between the distribution of lifestyles over zones and the distribution of lifestyles of Anable (2011) is higher. The analysis can be found in Table 21.

Table 21 TAD and MAPE on assignment of lifestyles towards mobility

	Mean absolute percentage error (MAPE)
Total distribution	5,74%
Status Seekers	5,95%
Devoted Drivers	4,38%
Reluctant Pragmatics	7,17%
Practical Travellers	8,50%
Active Car Owners	9,38%
Car Contemplators	2,05%
Car Free Choosers	2,79%

The MAPE shows the averaged differences between the distribution of Anable and the synthetic distribution. When the mean absolute percentage error would be zero, the population would have the same distribution for every zone. This is not the case, the mean absolute percentage errors are higher and range between 2,05% and 9,38%. It can be concluded that the distribution over lifestyles towards mobility is different between zones.

6.1.5 Conclusions

In this subsection the main conclusions on the case Amsterdam VMA will be given. The population synthesis tool is used and the performance of the first two steps of the tool have been measured with several analyses. With the IPF Multizone method a synthetic population is estimated. To measure the performance of this population, six methods were chosen from literature to compare the synthetic population to the actual population of Amsterdam. Based upon this analysis, the algorithm performs as expected (compared to other studies). The six measures of fit are not only used to measure the overall performance of the tool but also to measure performance for different areas with respect to each other. This is done by scaling the values of the measure of fit to the interval [0,1], the bandwidth of all points is between 0,55 – 1. While in the study from Brederode & Waanders (2013), the range was 0,2 – 1. This shows that our algorithm in this case performs less divers compared to the other study. The reason for this could be the fact that the area in our case is smaller and therefore the reference sample is better able to represent all zones. Besides this, the influence of the reference sample is high and could vary till 12% when the error is taken as a proportion of the amount of citizens.

The analysis on the output of the second step of the population synthesis tool, focuses on the lifestyle assignment. The total distribution of Amsterdam VMA as estimated by our tool is almost the same as the distribution of Anable (2011), this was expected since the IPF Multizone method scales towards this distribution once in every iteration. For the city center of Amsterdam a slight difference is found between the distribution of Amsterdam and the distribution of Anable (2011), but this is still small because of the large amount of zones in this area. A mean absolute percentage error was used. When this value is zero, there is no difference between the overall distribution and the distribution of separate zones. The mean absolute percentage error found for every lifestyle in our case study are not zero. They range between 2,05% and 9,38% and therefore it can be concluded that the distribution over lifestyles is different between zones. This supports the thought that the distribution over lifestyles is different in different zones.

6.2 Utrecht rush hour avoidance

In the previous section the case of Amsterdam VMA has been discussed. In this section the second case, ‘Utrecht rush hour avoidance’ is addressed. The geographical location in this case study is the triangle of the cities Utrecht, Hilversum and Amersfoort. The reason for this choice is a case study on the same area that has been conducted before. In our case study, the empirical data collected from the Project ‘rush hour avoidance’ is compared to the output of Fountain in its original format (with the original population). In this case study the population estimated by the population synthesis tool will be used as input for Fountain and the output will be compared to both the empirical output of the project ‘rush hour avoidance’ and the original output of Fountain. In Table 22 the different scenarios can be found.

Table 22 Scenarios in Case study ‘Utrecht rush hour avoidance’

Number	Scenario name	Explanation
1	Project	Collected data from project ‘rush hour avoidance’.
2	Original Fountain population	The population and output of Fountain in its original format.
3	Synthetic population of the tool	The population from the population synthesis tool for Utrecht and the output from Fountain based upon the new population.

In subsection 6.2.1 the data collection for ‘Utrecht rush hour avoidance’ can be found. In subsection 6.2.2 a case description of ‘Utrecht rush hour avoidance’ can be found. After this, the results of several analyses are presented. First in subsection 6.2.3 a comparison is made between the populations of the scenarios ‘Original Fountain population’ and ‘Population of the tool’. In this case study the focus was on the lifestyle assignment and the destination assignment of the population synthesis tool. Since there is no data available of the actual population for this study area, there are no analyses conducted for the IPF Multizone method in the tool. In subsection 6.2.4 analyses are presented on the distribution of lifestyles. Then, in subsection 6.2.5, a comparison is presented between the output of the project, the original output of Fountain and the output of Fountain based on the new population. In subsection 6.2.6 conclusions on this section will be given. How the case is implemented in Fountain can be found in appendix C.

6.2.1 Data collection ‘Utrecht rush hour avoidance’

The checklist that was used in 6.1.1, can be used to obtain an overview of the required data. In Table 23 this overview is given.

Table 23 Data collection ‘Utrecht rush hour avoidance’

Data	Available	Required for
Transport network	In VRU (version 3.0)	Fountain
Total men and women per zone	In VRU (version 3.0)	IPF Multizone
Total person in age classes per zone	In VRU (version 3.0)	IPF Multizone
Reference sample of the area	Not available (but can be created based upon VRU)	IPF Multizone
OD matrix	In VRU (version 3.0)	Destination assignment

To compare the output of scenario ‘Original Fountain population’ with scenario ‘Synthetic population of the tool’, exactly the same zones and areas need to be used. Therefore, the same transport network, with the same zones, infrastructure and public transport networks are required. The original Fountain case was performed based on a cordon from the VRU (version 3.0). VRU is a traffic model from the city of Utrecht. In Figure 30 the map of this area can be found.

Figure 30 VRU (Netherlands)

Utrecht and the surrounding cities are mapped in detail, the zones are very small. When moving further away from Utrecht, less zones are classified per area and areas are larger. The cordon that is mentioned before is only a part of this model. Unfortunately the cordon of Utrecht was not available, therefore a precise copy of the cordon is made with the use of the ‘cordon select’ tool of the program OmniTRANS. This is a multimodal program that is used for modelling interactions between different modalities in urban context (Dat.mobility, 2016). In Figure 31 the map of this cordon can be found with nodes.

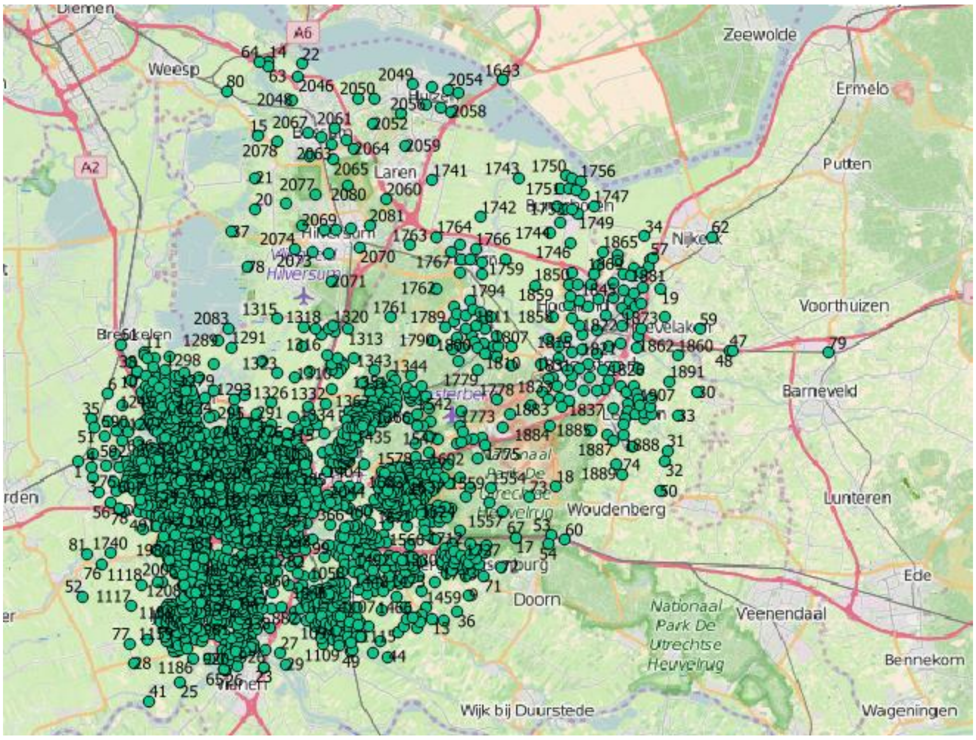


Figure 31 Cutout Utrecht with nodes

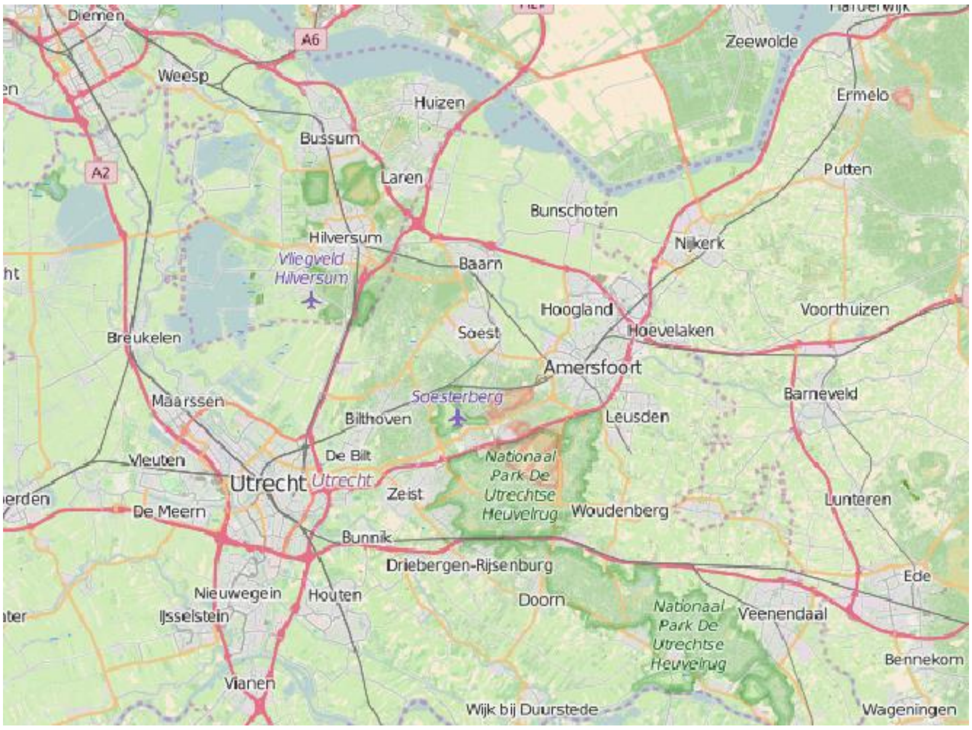


Figure 32 Cutout Utrecht wihout nodes

In Figure 32 the required area for this case study can be found without nodes. The three main cities in this model are Utrecht, Hilversum and Amersfoort. The model in OmniTRANS also contains a database with the required data. In the IPF Multizone method in the population synthesis tool the 'total men and women per zone' and the 'total number of persons in age classes per zones' are needed as input, this data can be found in the database of OmniTRANS. The reference sample is also based on totals created in

OmniTRANS. In the destination assignment, an OD matrix is needed. The OD matrix for the cordon can be found by running a script in OmniTRANS that does the cordon assignment. With this script, an OD matrix of the cordon is created with purpose: commuter traffic, for the modes: car, PT and bike and time: in the morning rush hour. By using this data, the exact same environment as the original run of Fountain is obtained.

6.2.2 Case description 'Utrecht rush hour avoidance'

In this subsection, the case description of 'Utrecht rush hour avoidance' can be found. The cordon of Utrecht has 2092 zones. The area has about 5.8 million citizens from which 630.000 citizens travel between home and work during rush hour. The dataset in the model in OmniTRANS has other age classes then the model of Amsterdam VMA. For this reason some adjustments are made to the population synthesis tool. The age classes in the model of Utrecht are 0 – 14, 15 – 24, 25 – 44, 45 – 64 and 65+. In this subsection, first, the main goal of the project is given, then an explanation of the intervention is given and finally the actual results of the project are given.

The main goal of project 'rush hour avoidance' was to improve the traffic flows on the highways A1, A27 and A28 within the triangle formed by the nodes Rijnsweerd, Eemnes and Hoevelaken. An attempt to achieve this goal is made by realizing behavioral changes. When a participant avoids rush hour, a financial reward was given based on the travelled distance in the participating area. In Table 24 the height of the reward for every travel distance within the participating area is given.

Table 24 Reward per rush hour avoidance

Travel distance within participating area	<5km	5km – 10km	10km – 15km	>15km
Reward per avoidance	1,50 euro	2,00 euro	2,50 euro	3,00 euro

The reward was given to a participant for 15 months, after which 3 months were measured without a financial reward. The project specifically focused on road users on the three highways during rush hour. Therefore several strategies were used to find participants for the project. By the use of the license plates, participants were found. Besides this, a collaboration with several employers was used to incorporate participants. 43.331 invitations were send which resulted in 4.635 participants. Besides this, 1.072 persons signed up on the website without an invitation. During the kick off in October 2011, 4.129 participants met the constraints to receive a reward. In December 2011 this amount grew to 5.445 participants. Once the after measurements started (last three months), there were only 3.121 participants left. In Figure 33 the number of participants over time is given.

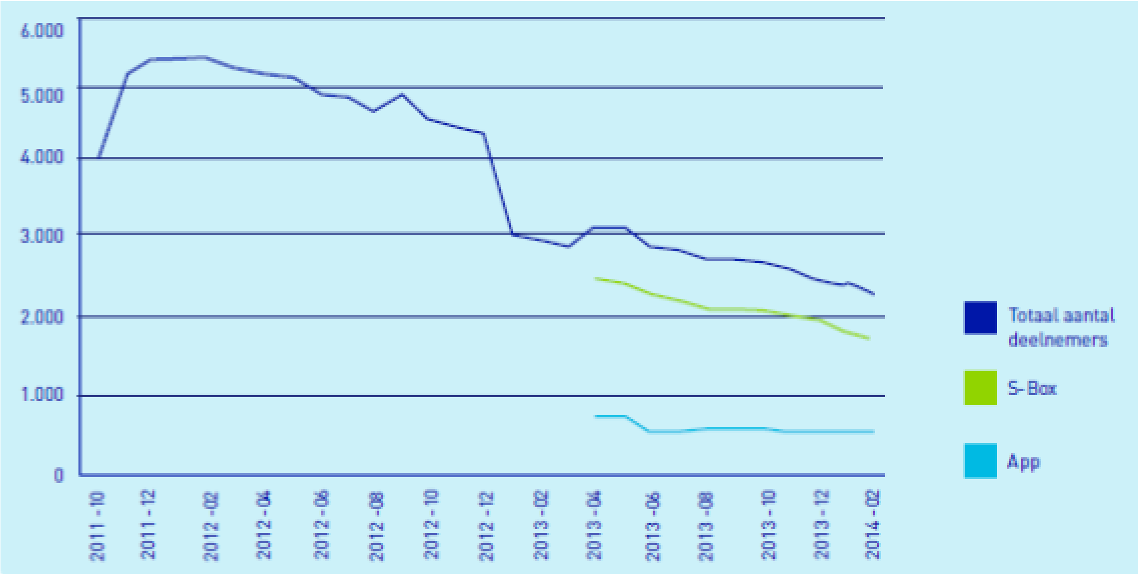


Figure 33 Number of participants over time

The avoidances could be achieved in several ways. The most frequently used alternative was travelling before or after rush hour, other popular alternatives were using public transport, using a bicycle, or work at home. The amount of persons that changed their behavior and avoided rush hour is expressed in number of avoidances over the total number of participants. An average can be calculated over the two phases of the project. During the first phase of 15 months, the average was 46,3% avoidances over total number of participants. During the second phase, three months without reward, this was only 36,6%. This is a decrease of more than 20%. In Figure 34 the visual representation is given.

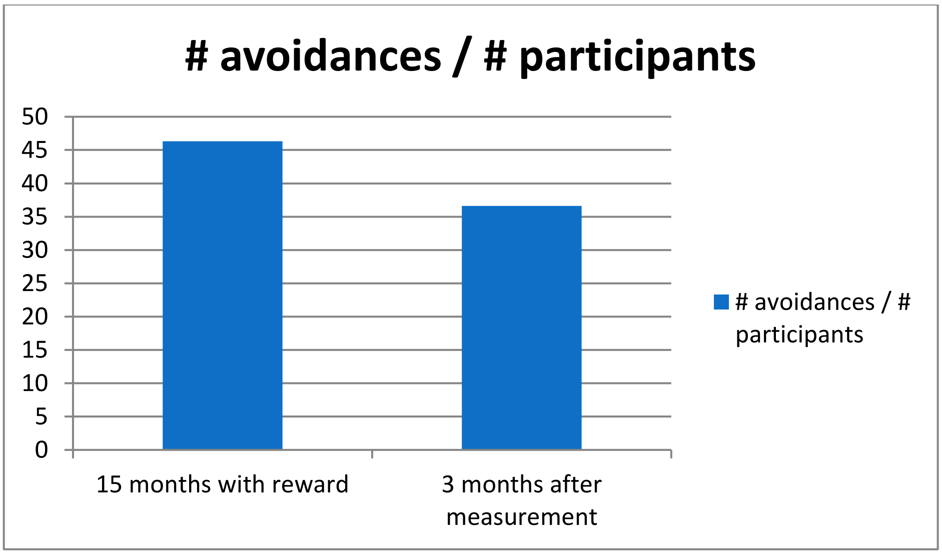


Figure 34 Average project result

From the start of the project until January 2012 the number of avoidances decreased from a 50% rate to a 43% rate. Towards the summer 2012 this percentage increased again towards 50%. After the summer, the percentage decreased again. In the first three months of January 2013 there was no reward. In the last three months the number of avoidances over the number of participants was 36,6%.

6.2.3 Comparisons between the original population in Fountain and the synthetic population of the tool

In this subsection a short overview is given of the population that is estimated by the population synthesis tool, and the population that is estimated by the original algorithm of Fountain. In Figure 35, the number of persons in the different scenarios is given. These are the number of people that travel during rush hour. It can be seen that in the original population of Fountain, the number of agents is only a third of the actual population. *During the research we did not know, but the traffic model in Fountain adds the rest of the missing persons to the model. Therefore this must be taken into account while interpreting the results.* To make a fair comparison, the agents in the population of Original Fountain population are for our case study adjusted from the original study, and scaled towards the actual number of people in the area that travel during rush hour. As a result, the Original Fountain population scenario has the correct amount of agents, representing around 600.000 people. *There must be taken into account during the analysis of the results that there are more persons in the model, which will lead to more traffic jam and therefore probably a larger amount of persons that will avoid the rush hour.*

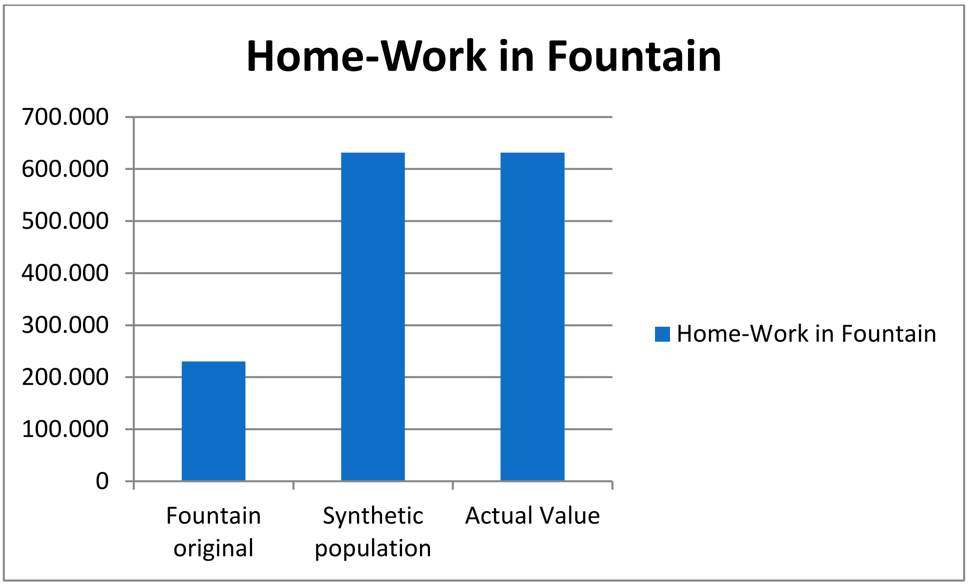


Figure 35 Comparison number of persons in the model

In the original Fountain population, the overall distribution over lifestyles is based upon Anable (2011). Therefore the lifestyle distribution is the same for every zone. In the synthetic population of the tool, estimated by the tool, the distribution is slightly different. Also the distribution on the zonal scale is varied. That is obtained by using the population synthesis tool. The distribution over lifestyles is based on socio-demographic characteristics instead of on an average distribution.

6.2.4 Analysis: Distribution of lifestyles towards mobility

In this subsection the lifestyle assignment in the population synthesis tool is evaluated. The same analyses have been performed for the study Amsterdam VMA. In Figure 36 the overall distribution of lifestyles are presented. The first distribution is the average distribution of Anable (2011), the second distribution is the distribution over lifestyles for the population of

Utrecht (this case study) and the third distribution is the distribution over lifestyles for the population of Amsterdam (previous case study).

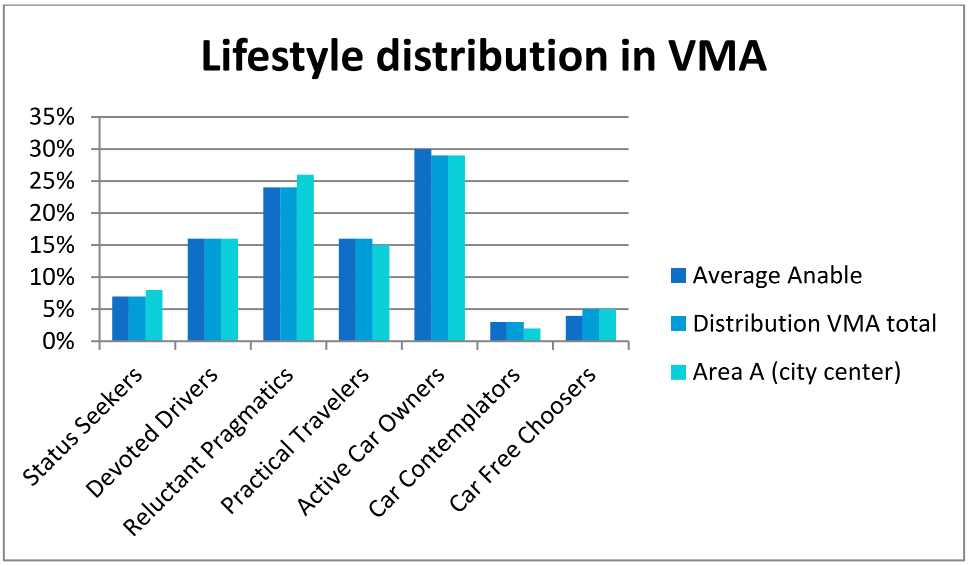


Figure 36 Overall distribution of lifestyles towards mobility

As can be seen, there are some slight differences between the Fountian original distribution and the distribution of the synthetic population of the tool. It can be seen that the distribution of the synthetic population of the tool has more reluctant pragmatics and less active car owners but only with a slight difference. In Table 25 the mean absolute percentage error (MAPE) can be found. While in the analysis of the first step in the Amsterdam VMA case, a larger MAPE was a negative output, in this step a higher MAPE supports the thought that not all distribution of lifestyles are the same for every zone. So, for example the city center has another distribution over lifestyles than for example in a suburb.

Table 25 Mean absolute percentage error

	Mean absolute percentage error (Cordon Utrecht)	Mean absolute percentage error (VMA)
Total distribution	3,5%	5,74%
Status Seekers	5,1%	5,95%
Devoted Drivers	3,2%	4,38%
Reluctant Pragmatics	6,3%	7,17%
Practical Travellers	7,1%	8,50%
Active Car Owners	9,0%	9,38%
Car Contemplators	1,6%	2,05%
Car Free Choosers	2,3%	2,79%

The mean absolute percentage error shows the averaged differences between the distribution of Anable (2011) and the synthetic distribution. If the mean absolute percentage error would be zero, the population would have the same distribution for every zone. This is not the case: the mean absolute percentage errors are higher and range between 1,6% and 9,0%. The distribution over lifestyles is different between zones. In the third column the MAPE for Amsterdam VMA can be found for comparison, the differences for the Utrecht case are slightly lower.

To see how the distribution of lifestyles is on a lower scale, in Figures 37 – 42 the distribution over lifestyles for six areas is presented. In Table 26 an overview is given of the areas that have been analyzed. For the zone numbers of the areas the distribution over lifestyles is compared to the average lifestyle distribution. Every figure will be discussed.

Table 26 Overview areas for distribution analysis

Figure	Area	Zone number
37	City center Utrecht	443,445 & 448
38	Bilt, Bilthoven	
39	Baarn	
40	Soest	
41	Uithof	464
42	Oudwijk	79,80 & 81

For a recap on the seven lifestyles of Anable (2011), The seven lifestyles that were defined by Anable (2011) are presented in subsection 4.1.7.

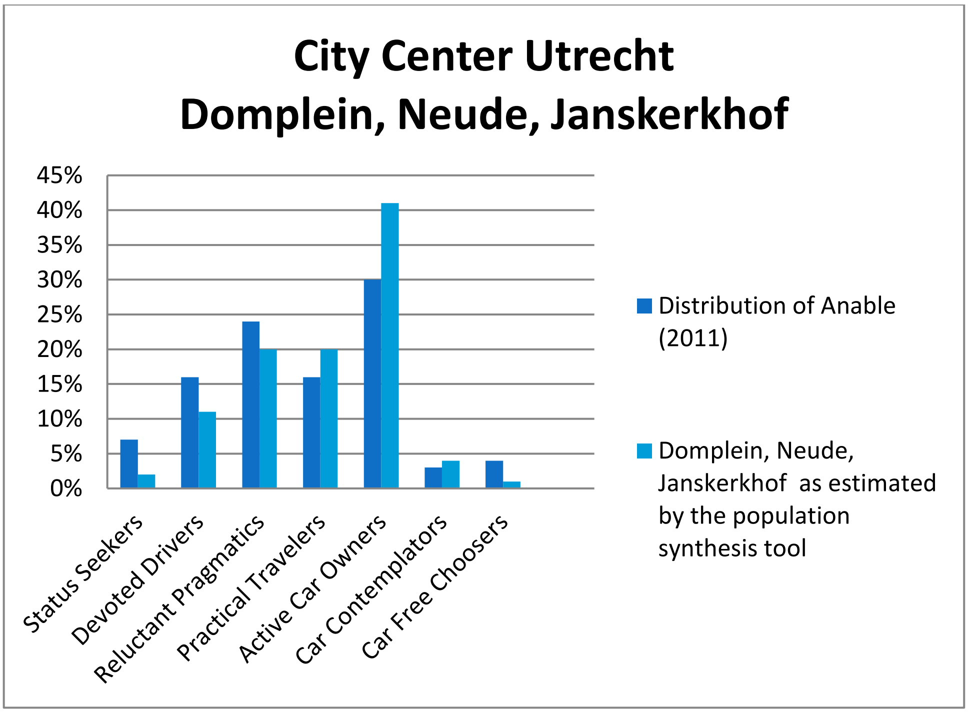


Figure 37 Distribution over lifestyles towards mobility City center

In Figure 37 the distribution over lifestyles for the city center can be found, compared to the average distribution. For the city center, the population synthesis tool results in a population in which the first 3 lifestyles have a lower share, compared to the average from Anable(2011). All three lifestyles have a preference for travelling by car. In the synthetic population of the tool there are more practical travelers who make the most convenient choice without a preference for a specific mode. In the synthetic population of the tool there are also more active car owners who are positive towards walking and cycling. Finally, there are slightly more car contemplators in the synthetic population of the tool, who care about their status and image, desire to have a car but do not have one.

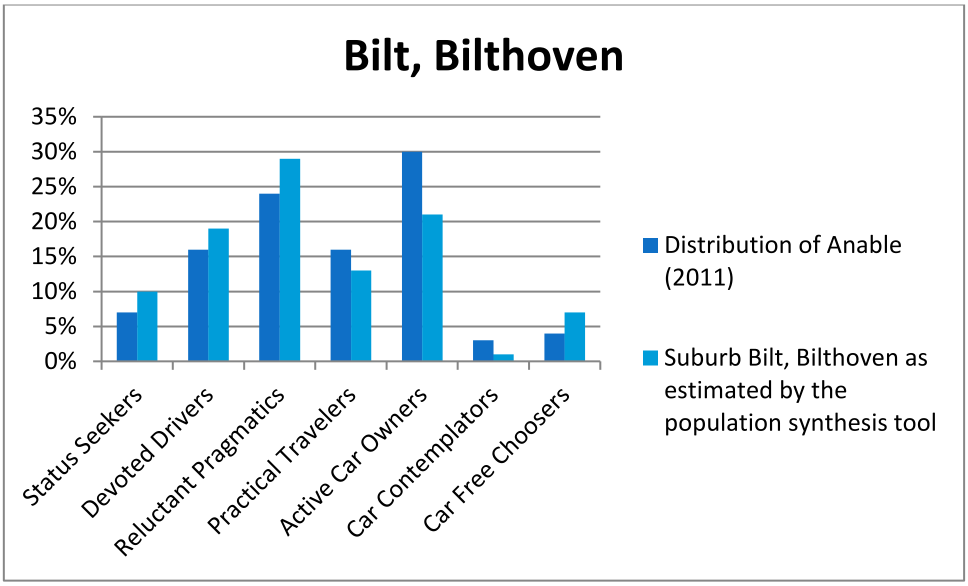


Figure 38 Distribution over lifestyles towards mobility Bilt, Bilthoven

In Figure 39 the distribution over lifestyles for Bilt and Bilthoven can be found. This is a more suburb area compared to the city center of Utrecht. The resulting distribution of the population synthesis tool shows the opposite of the distribution in the city center compared to the average distribution over lifestyles of Anable (2011). There are more status seekers, devoted drivers and reluctant pragmatics in the synthetic population of the tool. The synthetic distribution over lifestyles contains more people who have a preference for travelling by car and less practical travelers, active car owners and car contemplators. There are more car free choosers in the Bilt and Bilthoven in the synthetic population of the tool compared to the average.

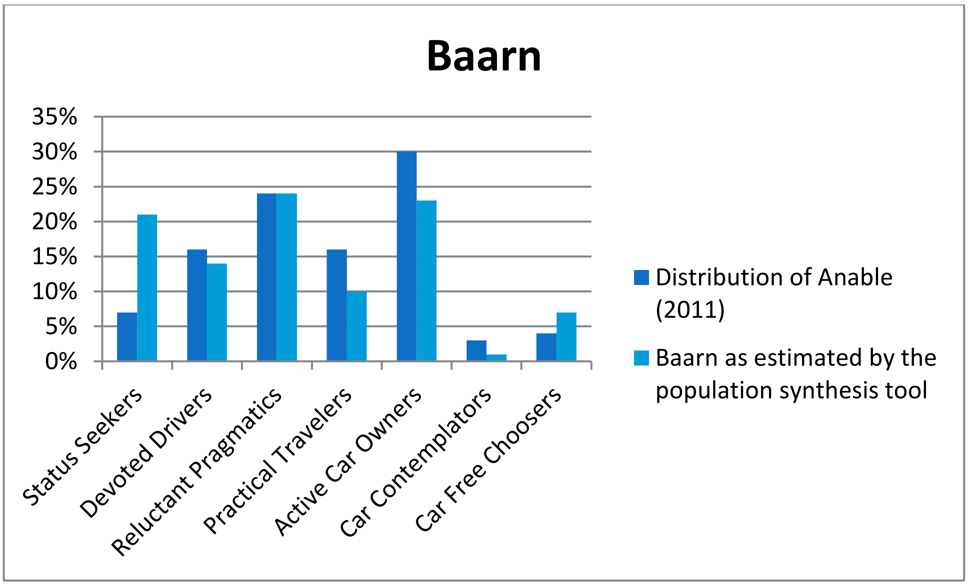


Figure 39 Distribution over lifestyles towards mobility Baarn

In Figure 39 the distribution over lifestyles for Baarn can be found. Baarn is also a more suburb area and has in the synthetic population of the tool a disproportionate amount of status seekers and less practical travelers and active car owners.

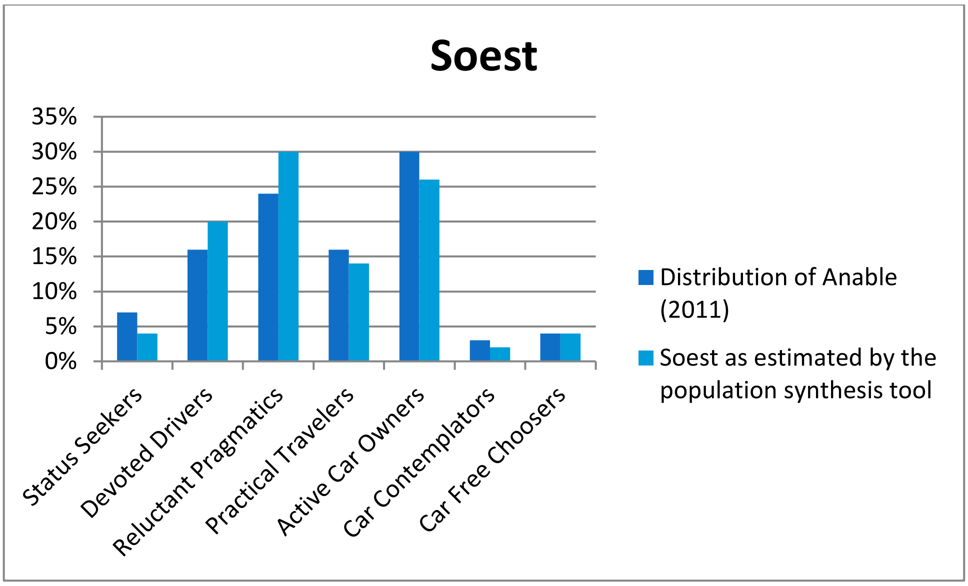


Figure 40 Distribution over lifestyles towards mobility Soest

In Figure 40 the distribution over lifestyles towards mobility for Soest can be found. Soest is also a more suburb area. According to the synthetic population of the tool there are more devoted drivers and reluctant pragmatics and less practical travelers and active car owners.

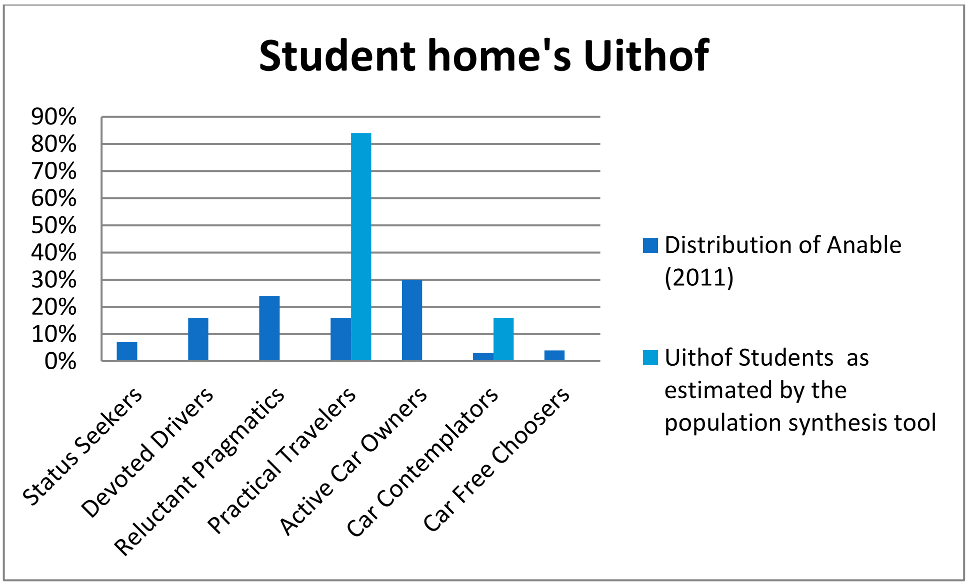


Figure 41 Distribution over lifestyles towards mobility Uithof

In Figure 41 the distribution over lifestyles towards mobility for the Uithof can be found. The only household variant in this zone are student home's. In the synthetic population of the tool, there are many practical travelers, who make the most convenient choice. There is also a higher percentage of car contemplators; they care about status and image. They desire to own a car but do not have one.

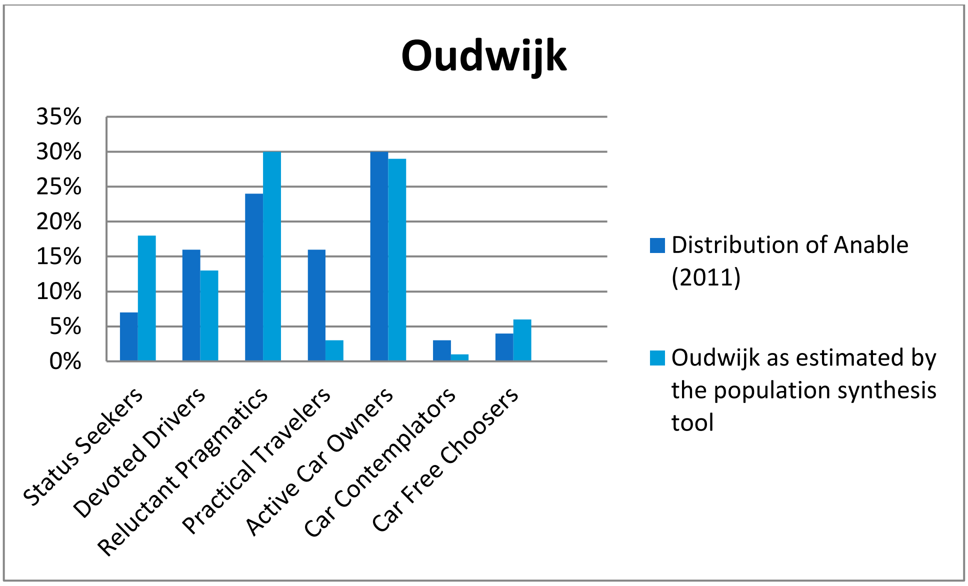


Figure 42 Distribution over lifestyles towards mobility Oudwijk

Finally in Figure 42 the distribution over lifestyles towards mobility for Oudwijk according to the synthetic population of the tool can be found. This is a very expensive neighborhood in Utrecht and is positioned just outside the canal that surrounds the city center. The area has many status seekers (they have a high preference for travelling by car and see their car as a status symbol) and many reluctant pragmatics (time/cost efficient, they have a preference for travelling by car). Less practical travelers and car contemplators.

The original population in Fountain has the same distribution as the average of Anable (2011), therefore no differences between those two populations are found. Based upon the comparison of our knowledge of the different zones and the distribution over lifestyles resulting from the population synthesis tool, we conclude that the synthetic population of the tool makes sense. The new distribution per zone, as a result from the population synthesis tool is closer to the real life expected distribution than the original population in Fountain.

6.2.5 Analysis: Comparison project rush hour avoidance with original population in Fountain and synthetic population of the tool in Fountain

In this subsection the output of the destination assignment in the population synthesis tool is analyzed with the use of project 'rush hour avoidance'. Fountain was set to run the same scenario with the synthetic population of the tool and the original population in Fountain. The results are compared to the results of project 'rush hour avoidance'. An overview of the different scenarios in Fountain can be found in Table 27. A more extensive explanation on the scenarios is given in Appendix C.

Table 27 Overview of scenarios in Fountain 1

Scenario	Population	Persons per agent	Reward
Synthetic 200	Synthetic	200	0,20 euro per km
Synthetic 100	Synthetic	100	0,20 euro per km
Original 200	Original	200	0,20 euro per km

First a short explanation of the differences between the three scenarios will be given. There are two types of populations used as input for Fountain. The synthetic population of the tool is the population created by the population synthesis tool and the original population is the original population used by Fountain. There are also a few parameters that are varied. The persons per agent is the number of persons represented by 1 agent. Ideally, one agent would represent one person. This is unfortunately not possible, since Fountain would need too much memory to run a model with so many agents. To see what the effect is of different settings of the persons per agent variable, there are two synthetic runs with different values for this parameter are included in our experimental design.

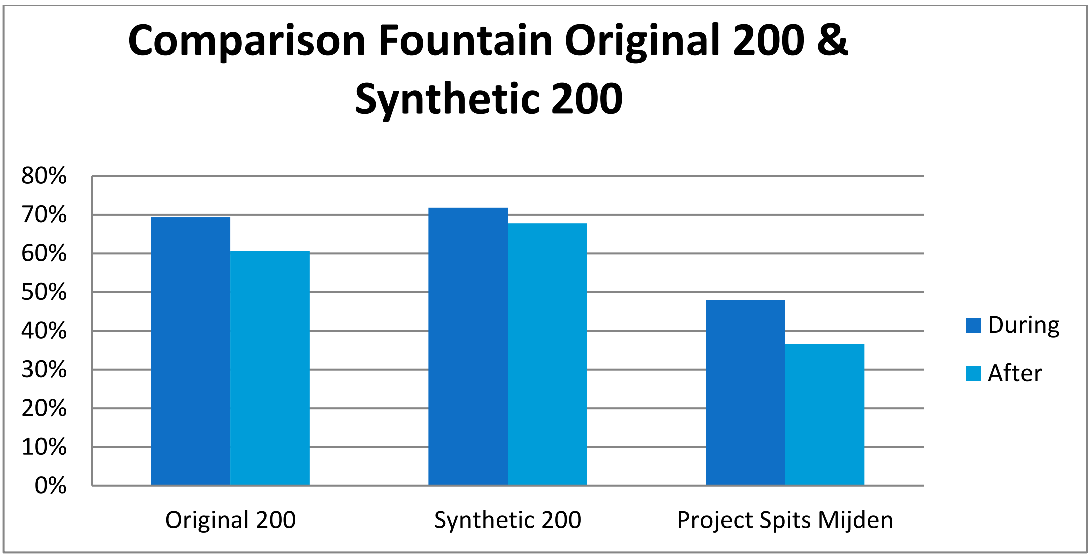


Figure 43 Analysis in Fountain with three scenarios

In Figure 43 the percentage of avoiders over total participants is given on the y-axis per scenario. A percentage of avoiders over total participants is given 'during' the 15 months when the reward was given and 'after' represents the 3 months after measurement without the reward. The original population and the synthetic give almost the same results and overestimate the amount of avoidance compared to the actual results of project 'rush hour avoidance'. This was expected since there are more agents in the model compared to the actual population.

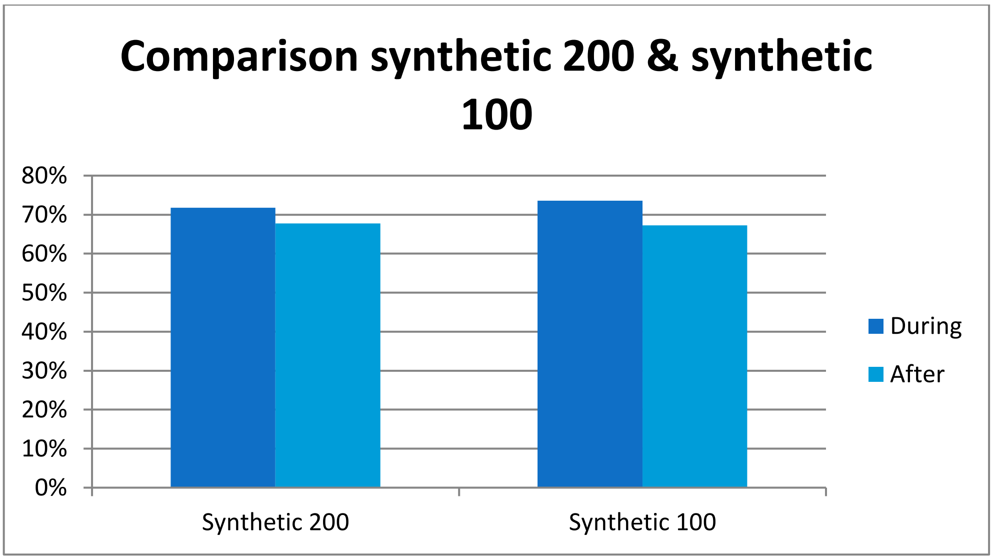


Figure 44 Comparison between synthetic 200 & Synthetic 100

In Figure 44 a comparison is presented between two synthetic populations of the tool; one with 200 persons per agent and on with 100 persons per agent. The amount of avoiders during the reward is higher and the avoiders after the reward is lower in the case of 100 persons per agent. One can conclude that the overestimation is not due to the fact that the population is not a good representation for the actual population in the area but is partly due to the extra amount of agents in the traffic model and partly due to one of the simplifications in Fountain identified earlier in the research, an overview of these simplifications can be found in appendix A. Now the question remains if the results on a lower scale are better due to the new population or that the output of the analyses are random.

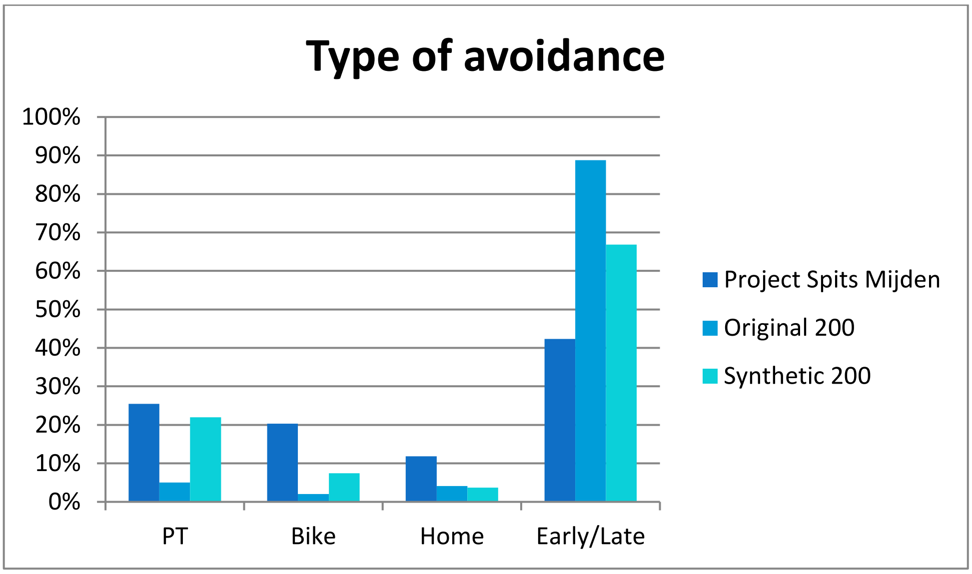


Figure 45 Type of avoidance compared original & synthetic

In Figure 45 an overview is given of the type of avoidance. On the y-axis the percentage per type of avoidance can be found. The synthetic population of the tool performs better compared to the original population in predicting the type of avoidance on the use of public transport (PT), the bike and the travel time. Working at home is the same in the synthetic and original scenario.

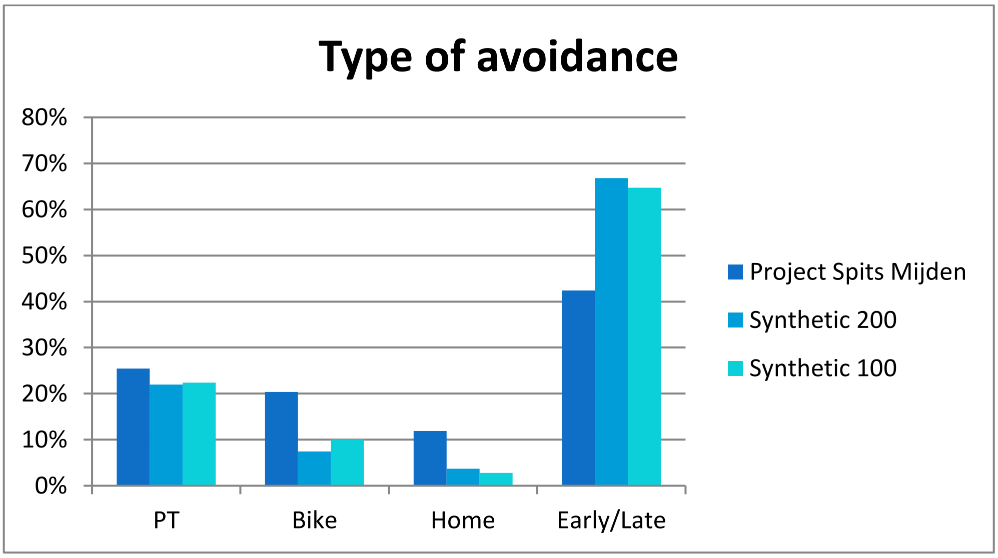


Figure 46 Type of avoidance compared 200 – 100

In Figure 46 the type of avoidance can be found for the two different synthetic scenarios. The type of avoidance is better predicted with a lower number of persons per agent. One could conclude that a lower ratio persons per agent gives a more realistic output.

6.2.6 Analysis: Comparing the output of Fountain for two scenarios, Original Fountain population and Synthetic population of the tool

In the previous subsection several areas have been analyzed on lifestyle distribution. In the following analysis two of those areas are analyzed in more detail. The two areas that have been chosen are the city center of Utrecht, which contains three streets (Domplein, Neude and Janskerkhof) and Oudwijk. In earlier analyses a difference between the original population in Fountain (average distribution) and the new synthetic population of the tool was found. In this analysis a difference in the run results of Fountain is found.

For both areas the modality choice of the overall population and the way of avoidance of the rush hour can be obtained.

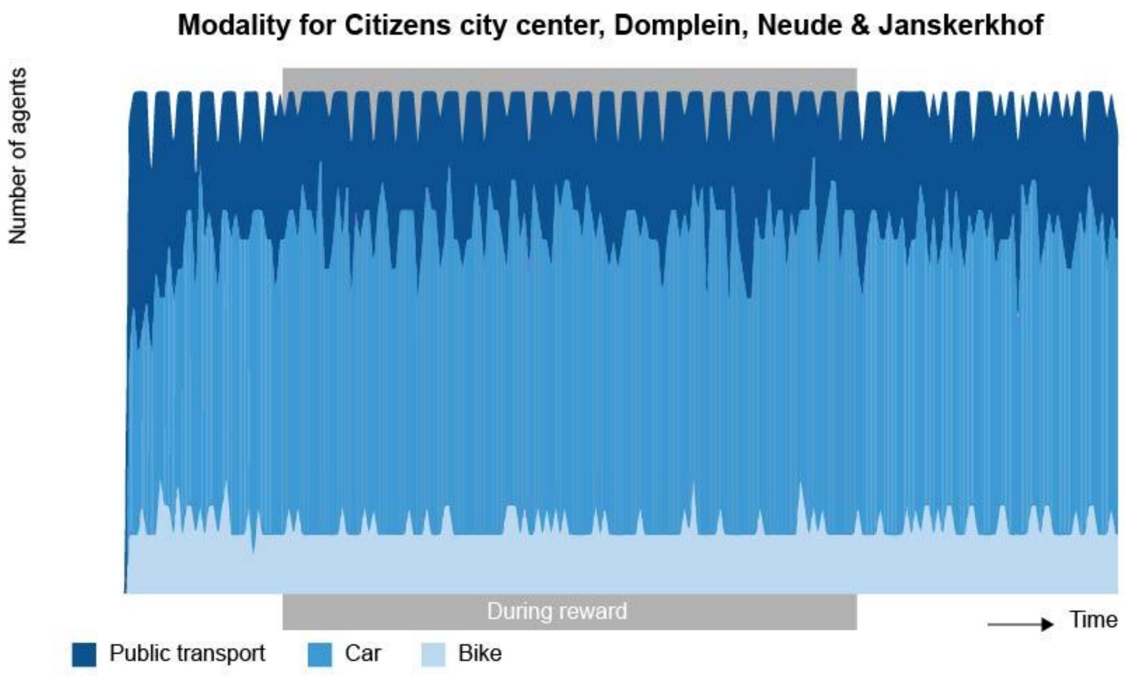


Figure 47 Modality choice of citizens in the city center (synthetic 100)

In Figure 47 the modality choice of all citizens in the city center can be found for the synthetic scenario. In Figure 48 the same graph can be found for the original population. Using the original population, almost all agent travel by car. While using the synthetic population of the tool, a higher percentage of people travel by public transport and by bike. When people in this area avoid the rush hour, they do this by traveling by PT or by bike according to the synthetic population of the tool. In the zones Domplein, Neude & Janskerkhof this meets with the expectations about this area.

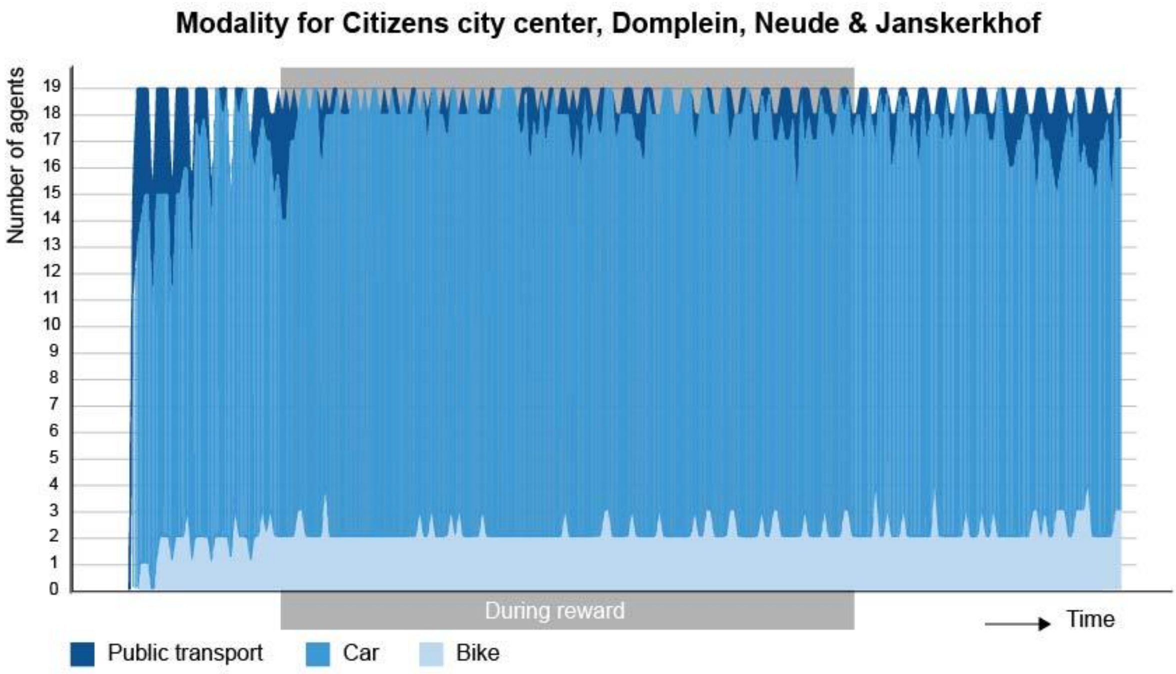


Figure 48 Modality choice for citizens in city center (Original 200)

In Figure 49 an overview is given of the avoiders. In this scenario there are 5 agents participating in the project and two participants that actually avoid the rush hour, the other three agents did had the intention to participate, but did avoid rush hour.

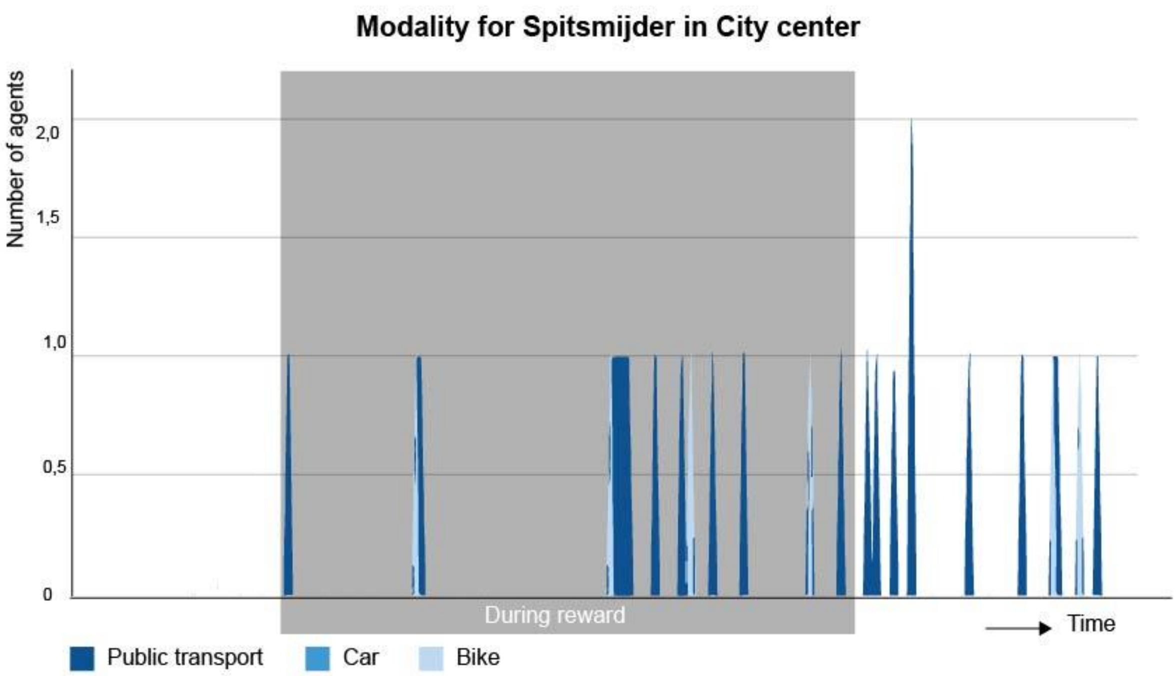


Figure 49 Modality for avoidance of rush hour in city center (synthetic 100)

In Figure 50, with the original population in Fountain, the main modality choice for avoidance is the car. In this scenario there are eight participants and all eight avoid the rush hour.

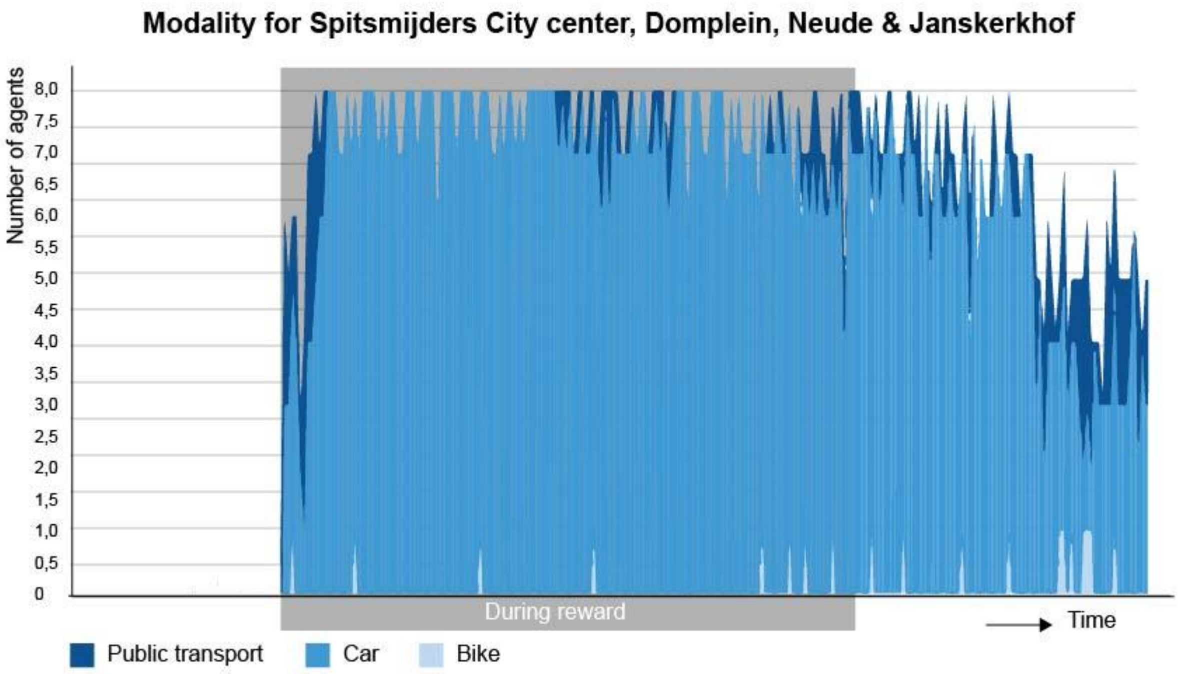


Figure 50 Modality for avoidance of rush hour in city center (original 200)

In the other area that is analyzed, the overall modality choice for citizens in Oudwijk is given in Figure 51. In Figure 52 the modality choice for citizens in Oudwijk can be found for the original scenario. The original population gives an average distribution over modes while in the scenario with the synthetic population of the tool the preference for the car is clearly present.

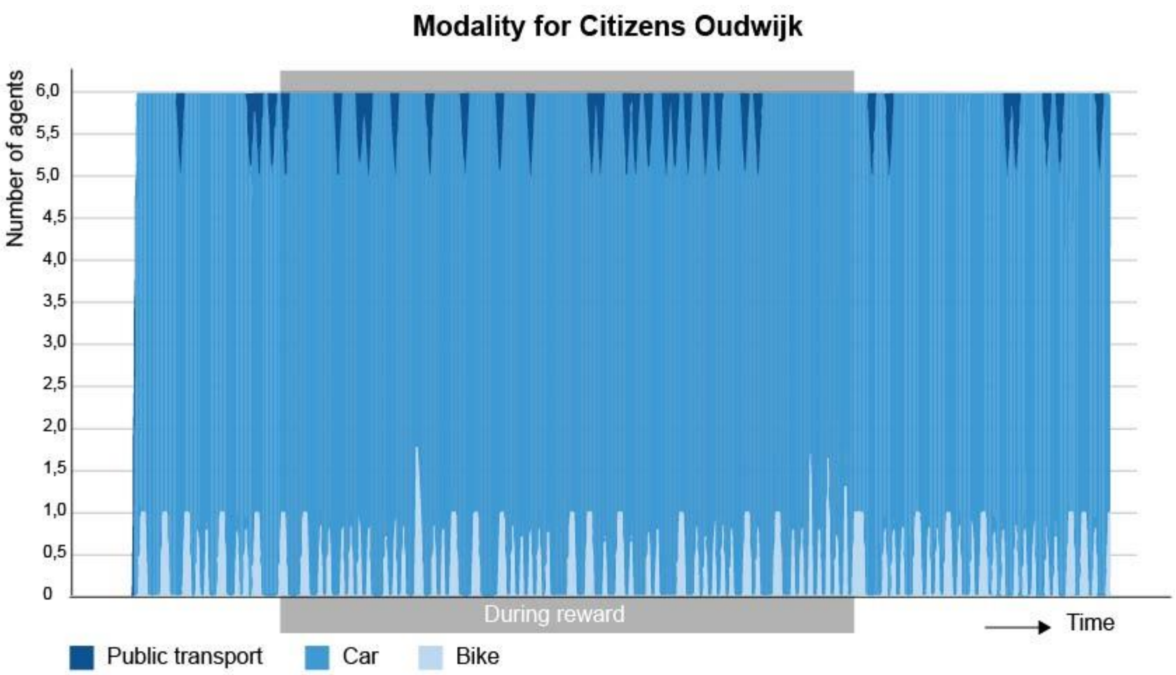


Figure 51 Modality choice for citizens in Oudwijk (synthetic 100)

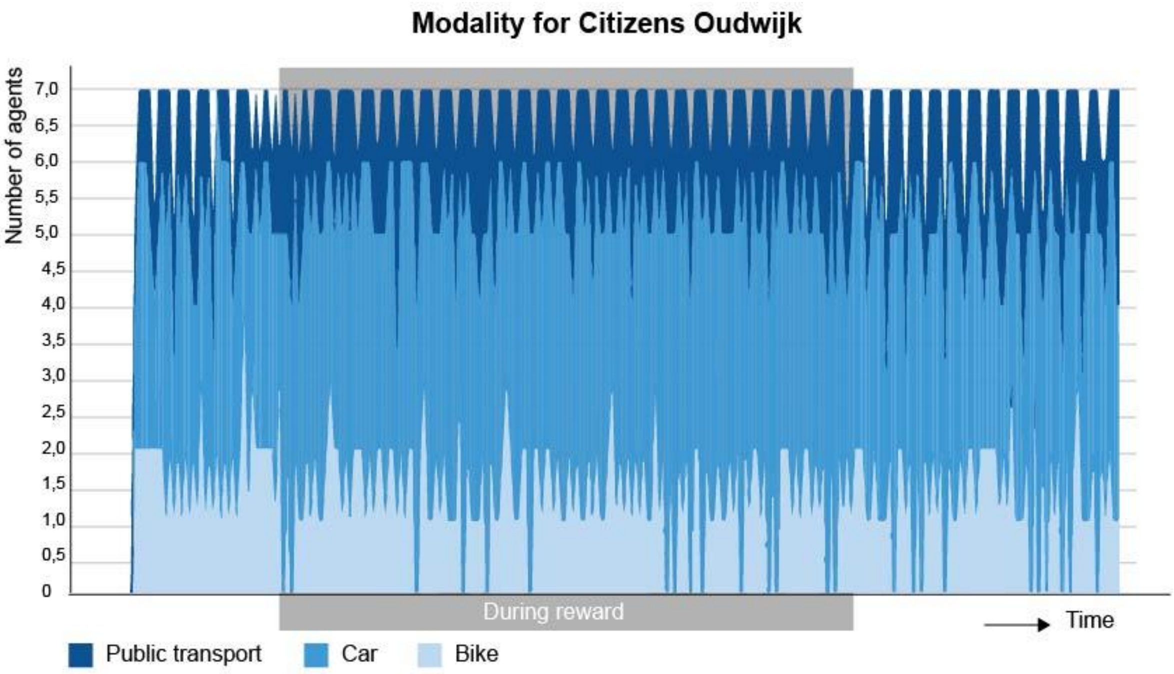


Figure 52 Modality choice for citizens Oudwijk (original 200)

In Oudwijk there are four participants from which no one actually avoids the rush hour in the synthetic scenario, for this reason there is no figure presented (this would be an empty chart). In the original scenario there are two participants that both become avoiders of the rush hour. They travel mostly by car to avoid the rush hour, see Figure 53. The modality choice for avoidance of the rush hour can be found.

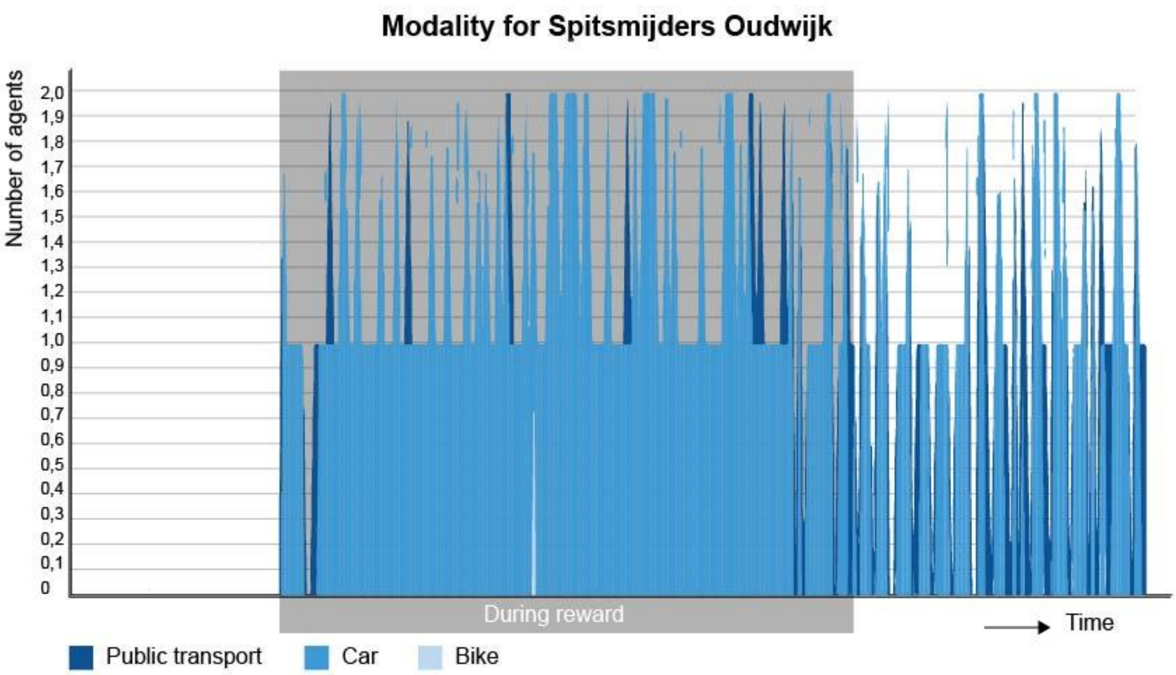


Figure 53 Modality for avoidance of rush hour in Oudwijk (original 200)

In both areas the output of the scenario with the synthetic population of the tool shows a way more realistic pattern based on the comparison of our knowledge of the two zones and the output based on the synthetic population of the tool. Therefore we conclude that the overall overestimation of rush hour avoidance in Fountain is not caused by the simplifications in the original population of Fountain. However, the population synthesis tool performs better on de zonal level in its predictions. We conclude that the usage of the population synthesis tool to estimate a population is better.

6.2.7 Conclusions

In this section the case ‘Utrecht rush hour avoidance’ is presented. Several analyses have been conducted to measure the performance of the population synthesis tool. In the first case of VMA Amsterdam study, the focus of the analysis was on the output of the IPF Multizone method and the output of the lifestyle assignment. In this case study the focus was on the lifestyle assignment and the destination assignment. Since there is no data available of the actual population for this case area, there are no analyses conducted for the first step of the tool. In the analysis on the lifestyle assignment, there was a high MAPE which leads to the conclusion that there are many zones whose distribution deviate compared to the average lifestyle distribution of Anable (2011). The MAPE was in a range between 1,6% and 9,0% which contributes to the thought that not all the zones have the same distribution of lifestyles. A more extensive analysis is performed on the average lifestyle distribution of different areas in the cordon. Based on the synthetic population of the tool, there are explainable differences between the different zones and they are captured with the population synthesis tool.

In the analysis on the output of the third part of the population synthesis tool, the observation was made that when the synthetic population of the tool is used as input in Fountain, it performs the same as the original population in Fountain on the overall amount of avoidance. When looking to the results of different scenarios at a lower scale a difference can be seen between the original population and the synthetic population of the tool. The type of avoidance is predicted more accurate by the synthetic population of the tool. Also the modality choice per zone when using the synthetic population of the tool appear to be more realistic compared to the results of the original population.

6.3 Limitations

Based on the analysis in this chapter, the observation was made that the synthetic population of the tool is more accurate, has a more deviated distribution of lifestyles towards mobility per zone compared to the average and performs better in the case ‘Utrecht rush hour avoidance’. The exception is the prediction of the overall amount of avoidances: this is not better, but equal compared to the original population in Fountain.

Besides these conclusions there are some limitations. In this section the limitations on two different subjects are discussed: first, limitations on the population synthesis tool in subsection 6.3.1, then limitations of the use of Fountain in combination with the population synthesis tool in subsection 6.3.2

6.3.1 Limitations of the population synthesis tool

A first limitation can be found in the IPF Multizone method in the population synthesis tool; where a population is estimated based on age and gender. Both socio-demographic characteristics are used in the lifestyle assignment, to assign lifestyles towards mobility.

Such a population is usable for large scale areas. However, when a smaller area is assessed, more characteristics might be needed. When, for example a city has a green neighborhood like the GWL zone in Amsterdam (Boer, 2016), and the number of persons living in this neighborhood has a substantial amount of influence on the total model, then the distribution of lifestyles based on gender and age will not meet the needs of such a case.

The second limitation are the differences between populations of countries. These differences have been left out of the scope of this research. It must be taken into account that the population tool might need different data on lifestyles towards mobility versus socio-demographic characteristics for different countries. Therefore, the type of country must be taken into account when using the population synthesis tool. When using the tool in another country, the tool can be used, but another dataset than the one from Anable (2011) might be needed.

The third limitation is the output of the destination assignment in the population synthesis tool. The output are seven OD matrices: one for every lifestyle. The disadvantage of this method is that most values in the matrix are not integer. This is now solved by giving a chance at each value that it will be an agent in Fountain. By using this method there is also a chance that some persons disappear from an origin zone in the model while in other origin zones there are too many agents.

The fourth limitation is the environment of the population synthesis tool. The tool is made in Excel which worked for this research but makes it hard to reconstruct the tool to other cases. It is recommended to implement the tool in another software package such as Matlab, Phyton or Ruby. The reconstruction might then be less time consuming.

6.3.2 Limitations of the use of Fountain in combination with the population synthesis tool

For the future use of Fountain also several limitations have been identified. The first limitation is the overestimation of the overall amount of avoidance in the models. We recommend to find out how this can be improved, starting with running the model with the correct amount of persons.

The second limitation that is identified is the scaling in Fountain. The lowest scale that could be used with the capacity of a good computer is now 100 persons per agent. When an agent represents 100 persons, zones with less persons than 100 persons have two options: when by chance there is no agent in this zone, persons are missing in this area, when by chance there is an agent in this zone, the amount of persons travelling from this zones is overestimated, therefore less persons per agent are recommended.

Besides the two limitations we recommend to improve the earlier identified simplifications in the model, see appendix A. The identified points were the parameters in the utility model and the parameters in the rest of Fountain (also contains the change of the default choice that was just mentioned), the threshold, the investment costs in a car and the availability of parking places. Furthermore, more case studies with Fountain are recommended. This will lead to more certainty about the usability of the results after improving the identified points.

6.4 Conclusions

In this chapter multiple analyses have been conducted to measure the performance of the population synthesis tool. This is done in two case studies, ‘Amsterdam VMA’ and ‘Utrecht rush hour avoidance'. The population synthesis tool has three steps; the IPF Multizone method, the lifestyle assignment and the destination assignment. After each step the output is analyzed. In the first case study, Amsterdam VMA, the output of the IPF Multizone method and the lifestyle assignment have been analyzed. The main conclusions in this case study are as follows.

With the IPF Multizone method, a synthetic population based on age and gender is estimated. To measure the performance of this population, six methods have been chosen from literature to compare the synthetic population to the actual population of Amsterdam. Based upon this analysis, the algorithm performs as expected (compared to other studies). The six measures of fit are not only used to measure the overall performance of the tool but also to measure performance for different areas with respect to each other. This is done by scaling the values of the measure of fit to the interval [0,1], the bandwidth of all points is between 0,55 – 1. In other studies the range was 0,2 – 1. The algorithm performs less diverse in our case. The reason for this could be the fact that the area in the VMA case is smaller and therefore the reference sample is able to represent all zones better. Besides this, the influence of the reference sample we found to be high and could vary till 12%. A Pearson Chi-square test was performed and the null hypothesis that there is no significant difference between observed and expected (estimated) population was accepted.

Based on the output of the lifestyle assignment we see that the total distribution of Amsterdam VMA is almost the same as the distribution of Anable (2011). This was expected since the IPF Multizone method scales towards this distribution once in every iteration. For the city center of Amsterdam, a slight difference was found in the distribution, but this difference is relatively small, because of the large amount of zones in this area. A mean absolute percentage error (MAPE) was calculated. When this value is zero, there is no difference between the overall distribution and the distribution of separate zones. The MAPE was found for every lifestyle and ranged between 2,05% and 9,38%. Therefore the conclusion is that the distribution over lifestyles is different between zones.

The second case study is 'Utrecht rush hour avoidance', in this case study the lifestyle assignment and the destination assignment in the population synthesis tool have been evaluated. Since there is no data available of the actual population for this study area, there is no analysis conducted for the output of the IPF Multizone method for this case study. In the analysis on the second part of the population synthesis tool, the MAPE ranged between 1,6% and 9,0%. This contributes to the thought that not all the zones have the same distribution of lifestyles, which seems logical. A more extensive analysis is done on the average lifestyle distribution of different areas in the cordon. The conclusion was that based on this population, there are logical differences between the different zones and they are captured with the population synthesis tool.

In the analysis on the output of the third part of the population synthesis tool, it was observed that when the synthetic population that was estimated by using the population synthesis tool is used as input in Fountain, it performs the same as the original population in Fountain on the overall amount of avoidance. When looking to the results of different scenarios at a lower scale, a difference was observed between the original population in Fountain and the synthetic population of the tool, when comparing the model runs to the actual results of the real project. The type of avoidance is better predicted by a synthetic population of the tool and even slightly more better predicted with a lower persons per agent rate. Also the modality choice per zone made more sense when the synthetic population of the tool was used compared to the results of the original population of Fountain.

Limitations can be found at two different subjects, first limitations of the population synthesis tool and then limitations of Fountain in combination with the population synthesis tool. Based on the results and the limitations, we conclude that the population synthesis tool performs better than the original method that Fountain used to estimate a population. The distribution over lifestyles is deviating from the average distribution that was used and the results from Fountain give a more logical explanation than before.

7. Conclusions & Recommendations

To conclude this thesis, the main conclusions are presented in this chapter. First, the research question and sub research questions are answered in section 7.1. After that recommendations, are given in section 7.2.

7.1 Conclusions

The initial goal of the research was to find a method to assess impacts of new mobility concepts as ‘mobility as a service’ or ‘automated vehicles’. This is done by exploring the usability of Fountain which is a simulation tool that combines agent based modelling, discrete choice modelling and behavioral aspects. In Fountain some simplifications have been identified, the population synthesis was one of those simplifications. The main objective of this research was to improve the population synthesis phase, taking into account lifestyles towards mobility. This resulted in the following main research question:

How can we improve population synthesis, taking into account the lifestyles towards mobility that are used in Fountain as predictor and thereby improve travel demand modelling?

To answer this question, multiple sub research questions have been conducted and answered in this thesis. The sub questions will be answered one by one below.

1. *Which population synthesis algorithm is most suitable to apply as a basis in the population synthesis tool?*

Several algorithms have been found from which the Iterative Proportional Fitting (IPF) and Iterative Proportional Updating (IPU) were the most promising. Criteria were defined to make a weighted choice for an algorithm. The three criteria that were defined were the applicability, the reaching of convergence and the score in the study (Brederode & Waanders, 2013) where different algorithms were implemented and compared. Based upon those criteria, the IPF Multizone is chosen as a basis for the new population synthesis tool for Fountain.

The main goal of the population synthesis tool is to use the available data sources on socio-demographic characteristics to obtain the amount of people traveling between zones with a certain lifestyle. An OD matrix for every lifestyle is required. A method to do this, is to use the study from Anable (2011). This study identified seven lifestyles: status seekers, devoted drivers, reluctant pragmatics, practical travelers, active car drivers, car contemplators and car free choosers (Anable, 2011). The study has a dataset with the relation between socio-demographic characteristics and the seven lifestyles. The validity of the dataset, provided by a single study, is not guaranteed. Because of the importance of the data in the population synthesis tool, the validity is tested by a meta-analysis. This gives the following research question:

2. *Which studies are available in the field of lifestyles towards mobility and are there patterns that justify the use of this data?*

Seven studies have been found and compared. All studies distinguish several lifestyles towards mobility. Comparing the underlying characteristics of the lifestyles resulted in four overall groups in which all different lifestyles from the studies can be classified. Furthermore, a relation between socio-demographic characteristics and lifestyles was found. A pattern in the relationship between lifestyles and age and lifestyles and gender was found the same in all the analyzed studies. Based on this analysis the dataset from (Anable, 2011) was used in the population synthesis tool.

3. *How can a population synthesis tool be constructed that is able to construct the amount of people traveling between zones with a certain lifestyle based on the socio-demographic characteristics age and gender?*

A population synthesis tool is constructed that contains the following three steps:

1. Population synthesis with IPF Multizone
2. Lifestyle assignment (based on the dataset of Anable (2011))
3. Destination assignment

The tool is able to estimate a population when provided with a transport network of the geographical area, data of the socio-demographic characteristics age and gender, a reference sample of the area, the dataset from Anable (2011) and an OD matrix of the area. To measure the performance of the population synthesis tool, multiple analysis have been undertaken. The following sub question has been answered:

4. *Is the population estimated by the population synthesis tool an improvement compared to the original population in Fountain?*

To answer this sub-research question, two case studies have been undertaken. The first case study, ‘Amsterdam VMA’, was chosen based on the available actual data on the population of Amsterdam, this is used to compare the performance of the IPF Multizone algorithm used in the population synthesis tool. The second case study, ‘Utrecht rush hour avoidance’, was chosen based on the earlier implementation in the model Fountain and the available actual data of the project rush hour avoidance. The output of all the three steps of the tool has been analyzed.

<u>Case study one Amsterdam VMA</u>

The first case study is ‘Amsterdam VMA’. In Amsterdam VMA the IPF Multizone method and the lifestyle assignment have been analyzed. The findings from this case study are as follows:

With the IPF Multizone method a basic population with the characteristics age and gender is constructed. To measure the fit of this population, six methods have been chosen from literature to compare the synthetic population to the actual population of Amsterdam. Based upon these analyses, the algorithm performs as expected (compared to other studies). The six measures are also used to see if the tool (as expected) results in different distribution over lifestyles towards mobility for different areas. It is found that the algorithm performs less diverse over different zones compared to other studies, which results in a more robust prediction of the population. The reason for this could be the fact that the area of the case study is smaller than the areas used in the other studies. As a result, the reference sample is a better representation of all zones.

The distribution over lifestyles towards mobility is for Amsterdam VMA almost the same as the distribution of Anable (2011). This was expected, since the IPF Multizone method scales towards this distribution. A mean absolute percentage error was calculated based on the comparison between the distribution over lifestyles towards mobility in the synthetic population and the average distribution of Anable (2011). When this value is zero, there is no difference between the overall distribution of the city and the distribution of the separate zones (such as the city center). The mean absolute percentage error is found for every lifestyle and they range between 2,05% and 9,38%. This indicates that the distribution over lifestyles towards mobility in the synthetic population is, as expected, different between zones.

<u>Case study two Utrecht rush hour avoidance</u>

The project ‘Utrecht rush hour avoidance’ has been used as a second case study. By performing this case study, the lifestyle assignment and the destination assignment of the population synthesis tool have been evaluated. Since there is no data available of the actual population for this study area, no analysis was performed for the IPF Multizone method in the tool. The analysis of the lifestyle assignment resulted in a range of the mean absolute percentage error between 1,6% and 9,0%. Again, this indicates that the distribution over lifestyles towards mobility in the synthetic population is, as expected, different between zones. A more extensive analysis was performed on the average lifestyle distribution of different areas in the study area. Based on the synthetic population, the conclusion was drawn that explainable distributions towards mobility were captured with the population synthesis tool for typical areas.

The analysis of the output of the destination assignment indicated that when the synthetic population of the tool is used as input in Fountain, it performs the same as the original synthetic population in Fountain on the overall amount of avoidance in the case study ‘rush hour avoidance’. When looking into the results of different scenarios at a lower scale, a difference can be seen between the original synthetic population in Fountain and the new synthetic population of the tool. The type of avoidance is better predicted by a synthetic population and even slightly more better predicted with a lower persons per agent rate. The modality choice per zone is better predicted with the synthetic population of the tool when compared to the expectations of the analyzed zones.

Based on the gained experience with the population synthesis tool, several limitations have been identified. The limitations are classified in two groups: first the limitations of the population synthesis tool and second the limitations of the use of Fountain in combination with the population synthesis tool.

The three main limitations of the population synthesis tool are:

- The two socio-demographic characteristics in the tool are age and gender. Not all areas can be caught in statistics, for this reason case studies at a zonal level might not give good results.
- The applicability in other countries is possibly limited, other lifestyles and other statistics about lifestyles are expected.
- The output of the tool are mostly integer values. This results in the chance that agents of persons are left out of the model. In agent-based modelling this remains a problem.

At the use of Fountain in combination with the population synthesis tool scale, three limitations were found:

- In the case study ‘Utrecht rush hour avoidance’, the model overestimates the amount of avoiders. As a result the model can be used to obtain insight in the direction of an intervention, but the exact values given by Fountain cannot be used as reliable information.
- the scaling of agents: 1 agent represents 100 persons. In zones with a smaller amount of persons this might result in a biased view due to the low granularity.
- This thesis has targeted one of the simplifications of Fountain. It should be noted that the other simplifications are still in place when using Fountain with the population synthesis tool. See for more information Appendix A.

Based on the results and the limitations, the conclusion is made that the population synthesis tool performs better compared to the original method that Fountain used to estimate a population. The estimated population by using the population synthesis tool is more accurate, has a more deviated distribution of lifestyles per zone compared to the average and performs better in the case ‘rush hour avoidance’. Population synthesis while taking into account the lifestyles towards mobility is improved. As a final result, a new population synthesis method has been constructed, which can be used as a basis in travel demand modelling.

7.2 Recommendations for further research

Based on the limitations found in chapter 6, several recommendations for further research have been identified. The recommendations are prioritized per type of recommendation. First the recommendations with respect to the population synthesis tool are given. Second the recommendations with respect to the use of Fountain in combination with the population synthesis tool are given. Third recommendations with respect to Fountain as a modelling tool.

With respect to the population synthesis tool:

- The current tool uses two socio-demographic characteristics. It is recommended to insert extra characteristics such as household size of income where possible, especially for microsimulations. More relevant socio-demographic characteristics can be found by further research into the field of lifestyles.
- When using the population synthesis tool in another country than the Netherlands, another dataset than the one from Anable (2011) might be needed. Therefore it is recommended when using the tool in another country, to review studies about lifestyles towards mobility from that country.
- The output of the third step of the population synthesis tool are seven OD matrixes; one for every lifestyle. The disadvantage of this method is that most values in the matrix are not integers, therefore persons in the population might not be represented in a model. This is now solved by giving a chance at each value that it will be an agent in Fountain which has its downsides. A more elegant solution is making sure that the totals stay the same during the process. This solution is recommended to be implemented during future work.
- The tool is made in excel which worked for this research but makes it hard to reconstruct the tool for other cases. It is recommended to implement the tool in another software package such as Matlab, Phyton or Ruby. The reconstruction might be less time consuming.

With respect to the use of Fountain in combination with the population synthesis tool:

- With a decent PC/laptop, the lowest ratio that can currently be used in Fountain with is 100 persons per agent. Reprogram Fountain could result in a lower capacity demand of the PC/laptop and a lower ratio can be used. This will result in a more detailed run of the model and therefore there is expected to gain better results.
- Executing more case studies with a population from the population synthesis tool in Fountain is recommended. This will result in more confidence about the usability of the population synthesis tool in combination with Fountain.

With respect to Fountain as a modelling tool:

- Currently Fountain overestimates the overall amount of avoidance in the case ‘Utrecht rush hour avoidance’. This overestimation can possibly be solved by two changes in the model. In the case study ‘Utrecht rush hour avoidance’, too many persons were modelled. There is recommended to run the model with the correct amount of persons.
- When still an overestimation of the overall amount of avoidance is found, there is recommended to change something in the model structure. More specific; changing the amount of mode choices before changing the default choice. Now the default choice changes after the agents tries a mode for three times.
- Furthermore, several simplifications of Fountain were identified: the parameters in the utility model and the parameters in the rest of Fountain (also contains the change of the default choice that was just mentioned), the threshold, the investment costs in a car and the availability of parking places. The recommendation is to look into these simplifications during future studies to improve Fountain even more.

8. Scientific and practical contribution of this research

In this chapter a discussion is presented on the scientific and practical contribution of this research. In section 8.1 the scientific contribution is given and in section 8.2 the practical contribution is given.

8.1 Scientific contribution of the research

Most travel demand modelling tools use a population based on a population synthesis method, since a full dataset on the actual population is almost never available. In the field of population synthesis, multiple methods of estimating a population can be found in literature. All methods have one thing in common: the input of the methods is a set of socio-demographic characteristics on an aggregated level and the output is a set of socio-demographic characteristics on an disaggregated level. The use of such a population synthesis method results in the use of socio-demographic characteristics as a predictor of travel behavior. According to Hunecke et al. (2011), lifestyles towards mobility have a higher predictive power than socio-demographic characteristics. For this reason a population synthesis method is constructed to obtain the distribution over lifestyles towards mobility instead of generic socio-demographic characteristics.

The contribution of this work is; finding the relation between socio-demographic characteristics and lifestyles towards mobility (based on Anable (2011)) and using these findings to construct a population synthesis tool. Thereby a new population synthesis method is constructed, that is able to obtain a population with the distribution over lifestyles towards mobility per zone. With this new population synthesis method, lifestyles can be used as predictors in travel demand modelling.

The main theoretical contribution of this work is the usability of the population synthesis tool for general travel demand modelling. Lifestyles towards mobility can be used as predictor of travel behavior instead of the socio-demographic characteristics that are currently mostly used.

8.2 Practical contribution of the research

As an exception on most travel demand models, the modelling tool Fountain already used lifestyles towards mobility as a predictor of travel behavior. However, only the lifestyle distribution for Utrecht was available. In order to apply Fountain in another area, a new dataset on lifestyles towards mobility was needed. Furthermore, Fountain was based on the assumption that the lifestyles distribution was the same for all zones in an area.

With respect to Fountain, the constructed population synthesis tool allows for:

1. Broader applicability. A population to be estimated for Fountain for all geographical areas in the Netherlands. The tool can be applied for other countries, but a small review would have to be undertaken.
2. Improved performance. The population synthesis tool estimates the distribution of lifestyles towards mobility at the more detailed level of zonal scales (e.g. neighborhood) instead of area scale (e.g. city).

9. Reflection

In this chapter I will reflect upon my master thesis. For each major research activity reflection is given, along with the choice to graduate at TNO instead of at the university.

9.1 Research proposal

When I started with this master thesis I expected to do something with automated driving, work on a simulation model and report on the results that I had found. After a short while, I noticed that there is no such thing as a simulation tool that gives the impacts of automated driving on a silver plate. I started working with Fountain, a model developed by TNO. In this model I saw a lot of things I could do to improve the model, but if I would do that, the automated car was out of the scope of my thesis.

In this part of the research I learned that I am enthusiastic about many things. When I do not scope what I want to do (in for example this thesis, but also in my work later), I might end up with too many things to do. If I could make my research proposal again, I would choose the same path. Although I would decide to leave the subject of automated driving in an earlier stage, in order to prevent doing a lot of additional work. The whole literature research on automated driving costed me a lot of time.

9.2 Literature study

When I decided to work on population synthesis, it felt like a challenge to start again with a new literature study. Luckily, my supervisor Erik introduced me to Luuk Brederoode from Goudappel Coffeng, who helped me with finding the most important papers on population synthesis. Once I knew which papers to read, I felt more comfortable because of the scope of my work. I learned that it is not only better for me to have a clear scope because of my enthusiasm, but also for my own comfort. Besides learning everything about population synthesis during this literature study, I also learned how to do a proper literature study.

At the end of the literature research, I chose for the IPF Multizone method: this was the most safe choice since it was applicable and the method converges to an optimal outcome. There are some optimization methods that in theory perform better. However I am happy that I chose IPF Multizone as the other methods require advanced mathematics and would require me to invest much time in understanding these methods. I believe that in a thesis like this such a choice is defendable, given the limited amount of time of a thesis.

9.3 Meta-analysis on lifestyles towards mobility

When the next step was another literature review, I felt very comfortable since I knew that I could do this. After the experience of the first two literature review, the third actually went quite fast. I can happily say that I love doing literature studies since my master thesis. Which I absolutely did not expect during the start of my thesis.

I found seven studies in the field of lifestyles towards mobility and unfortunately not all the studies had a database I could use. If I would have had more time I think I could have found more studies or contact the authors to obtain the correct data. In this way not only the socio-demographic characteristics of gender and age could have been used for the population synthesis tool, I might find more patterns in the data and use socio-demographic characteristics as income or household size. I noticed that it worked very well to talk to the authors directly, I would advise other students to be more pro-active in approaching authors in your field of study.

9.4 The population synthesis tool

I already started with the construction of the population synthesis tool during the meta-analysis on lifestyles towards mobility. Once I found the relation between lifestyles and the socio-demographic characteristics age and gender I started with the tool and decided to only use these two characteristics. The main reason was that I did not see how more characteristics could be included when I started with the tool, besides I did not find enough data on the other characteristics. Once the tool was finished I knew how to implement more than two socio-demographic characteristics but it would have taken a lot of time to implement others. And still, it would also have taken more time to find more data on other socio-demographic characteristics. If I could do the thesis again, knowing what I know now, I would spend some extra time on this part and implement extra socio-demographic characteristics. Besides this, there is one part in the population synthesis tool where the age classes are transformed to other age classes by a simple method. If I would do this again I would use the distribution known in the Netherlands for ages, instead of scaling the age classes by ratio.

9.5 Case studies

I started with a case study for Amsterdam. Not because they have a fully detailed database on the population, but because I had just chosen this as the area of study. I had permission from the municipality of Amsterdam to use the model. But, once I knew I wanted to use the case study that was already implemented in Fountain before, I had to work with Utrecht. This was a misfortune, the model was not available anymore and it took a lot of time to obtain the exact same cordon from OmniTRANS. Besides this, the population synthesis tool was also implemented based on the data from Amsterdam. So, it was very time consuming and I had to implement the data from Utrecht in the tool. When looking back, the implementation of the data from Utrecht in the population synthesis tool was actually a way to improve the tool. By doing this, I found the steps in the tool that were not generic. These steps were improved by applying a new input dataset. Besides this, the case study in Amsterdam proved to be very useful too, since there was data available of the actual population. I still think that the reason for this late switch, is because of my late decision to scope out automated driving. During the phase when I was talking to the municipality of Amsterdam about using the VMA model, I still had in mind to work with automated driving. Again my conclusion is that I should have made that decision earlier in the process.

Besides the choice for the study areas, one of my main results were dependent from the model Fountain. In practice this resulted in a lot of trouble, since it was difficult to find a computer that was able to run Fountain with my population. Furthermore, Fountain needed some adjustments to work with the population from the population synthesis tool. Therefore some additional debugging time in Fountain was needed. Although I learned a lot about the model, it was also a bit stressful to work with it. Fortunate Selmar Smit found all the time that was needed to help me. If I could make some recommendations to other students or researchers that want to work with Fountain, I would suggest to use my work as a basis and then try to run the model with less persons. Besides this, the recommendation that is given in chapter 7 about the amount of choices before the default choice changes can be followed.

Third I would suggest to investigate if the other simplifications that were found, need to be improved. Finally some extra case studies should be undertaken.

9.6 Graduation at TNO

Since I know that I do not like working on my own, I decided in an early stage to search for a company where I could work on my master thesis. This pressured me to work every day from 9 until 5 o'clock. This worked out really well. Once I started at TNO I noticed I worked very hard when being at the office and was unoccupied by my thesis during the evenings and the weekend. Besides this, my colleagues helped me with many things during my thesis and I got to know a lot of them very well.

During the internship, I did not only work on my thesis, I organized a day to the Efteling for the department together with Daniëlle, my co-intern. I also found a very nice job at TNO, which brings me to what I have learned from this decision. When I started with my internship at TNO I was working hard during the days and unoccupied during the evenings and weekend. Now I started working at the 1st of October (I thought my thesis would be finished by then), which resulted in a 40 hour working week and working on my thesis during the evenings and the weekends the past six weeks. I learned that the balance between work and private is very important and will keep that in mind when saying yes too soon to anything again. This is actually the same thing as stated in 9.1, I might be too enthusiastic about too many things.

Bibliography

Ajzen, I. (1991). The theory of planned behavior. *Organizational behavior and human decision processes*, 179-211.

Anable, J. (2005). 'Complacent Car Addicts' or 'Aspiring Environmentalists'? Identifying travel behaviour segments using attitude theory. *Elsevier*, 65-78.

Anable, J. (2011). *Segment baseline survey of New Employees - Utrecht.* Aberdeen: University of Aberdeen.

Annema, J., Brink, R. v., & Walta, L. (2012). *Transport technology to reduce transport's negative impacts.* Cheltenham: Edward Elgar Publishing.

Arentze, T., Timmermans, H., & Hofman, F. (2007). Population synthesis for microsimulating travel behavior. *Transportation Research Record*, 85-91.

Armstrong, S. (1985). *Long range forecasting. From crystal ball to computer.* New York: John Wiley & Sons.

Arnaout, G., & Bowling, S. (2011). Towards reducing traffic congestion using cooperative adaptive cruise control on a freeway with a ramp. . *Journal of Industrial Engineering and Management*, 699-717.

Auld, J., Mohammadian, A., & Wies, K. (2008). Population synthesis with control category optimization. *10th International Conference on Application of Advanced Technologies in Transportation.* Athens.

Auld, J., Mohammadian, A., & Wies, K. (2010). An efficient methodology for generating synthetic populations with multiple control levels. *89th Annual meething of the Transportation Research Board.* Washinton D.C.

Bar-Gera, H., Konduri, K., Sana, B., Ye, X., & Pendyala, R. (2009). *Estimating survey weights with multiple constraints usting entropy optimization methods.* Washinton D.C.: 88th Annual meeting of the Transportation Research Board.

Beckman, R., Baggerly, K., & McKay, M. (1996). Creating synthetic baseline populations. *Transportation Research Part A: Policy and Practice*, 415-429.

Beirao, G., & Sarsfield Cabral, J. (2007). Understanding attitudes towards public transport and private car: A qualitative study. *Elsevier*, 478-489.

Boer, J. (2016, 11 5). *10 jaar ecowijk gwl terrein Amsterdam*. Retrieved from GWL: http://www.gwl-terrein.nl/files/artikelen/GWL_Tekst_def_mail.pdf

Bosch, S. (2010). *Transitie-experimenten: Verkennen van maatschappelijke veranderingen richting duurzaamheid.* Rotterdam: PhD Theses, Erasmus University Rotterdam.

Bowman, J. (2009). Population synthesizers. *Traffic Engineering and Control*, 342.

Brederode, L., & Waanders, M. (2013). *Een vergelijking van de synthetische populatieschatting in het LMS/NRM met andere veelgebruikte methoden.* Rotterdam: Colloquium Vervoersplanologisch Speurwerk.

CBS. (2015). *Transport en Mobiliteit.* Den Haag: Centraal Bureau voor de Statistiek.

Chambers, J., Mullick, S., & Smith, D. (1971). How to choose the right forcasting technique. *Harvard business review*.

Corwin, S., Vitale, J., Kelly, E., & Cathles, E. (2015). *The future of mobility: How transportation technology and social trends are creating a new business ecosystem.* Westlake: Deloitte University Press.

CPB. (2015). *Urban Mobility System Upgrade: How shared self-driving cars could change city traffic.* Parijs: International Transport Forum.

Dat.mobility. (2016, sd sd). *OmniTRANS*. Retrieved 9 28, 2016, from Dat.mobility: http://www.dat.nl/nl/producten/omnitrans/

Department of Transport. (2015a). *The pathway to driverless cars.* London.

Fagnant, D., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: opportunities barriers and policy recommendations. *Elsevier*, 167-181.

Fagnant, D., & Kockelman, K. (2015). Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations for Capitalizing on Self-Driven Vehicles. *Transportation Research Part A: Policy and Practive*, 1-20.

Festinger, L. (1957). *A Theory of cognitive dissonance.* Stanford: Stanford University Press.

Forrester, J. (1971). *Counterintuitive behavior of social systems.* Cambridge: Massachussetts Institute of Technology.

Fotheringham, A., & Knudsen, D. (1987). *Goodness of fit statistics.* Norwich: W.H. Hutchins & Sons.

Fotheringham, A., & Knudson, D. (1986). Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modeling. *International Regional Science Review*, 127-147.

Grimm, V. (2005). *Individual-based Modeling and Ecology .* Princeton: Princeton University Press.

Gruel, W., & Stanford, J. (2015). *Assessing the long-term effects of autonomous vehicles: A speculative approach.* Massachusetts: Massachusetts Institute of Technology.

Guo, J., & Bhat, C. (2007). Population synthesis for micorsimulation travel behavior. *Transportation Research Record*, 92-101.

Haustein, S. (2011). Mobility behavior of the elderly: an attitude-based segmentation approach for a heterogeneous target group. *Transportation*, 1079-1103.

Huang, Z., & Williamson, P. (2001). *Comparison of synthetic reconstruction and combinatorial optimisation approaches to the creation of small-area microdata.* Liverpool: Department of Geography, University of Liverpool.

Hunecke, M., Haustein, S., Bohler, S., & Grischkat, S. (2010). *Attitude-Based Target Groups to Reeduce the Ecological Impact of Daily Mobility Behavior.* Lyngby : Technical University of Denmark.

Hwan Kim, K., Hyung Yook, D., Seok Ko, Y., & Han Kim, D. (2015). *An analysis of expected effects of the autonomous vehicles on transport and land use in Korea.* New York: Marron Institute of Urban Management.

Hyndman, R., & Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 679-688.

Isard, W. (1956). *Location and space-economy.*

Isard, W. (1956). *Location and Space-Economy.* New York: The Technology Press of Massachusetts Institute of Technology AND John Wiley & Sons, Inc. .

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

Kahn, H. (1962). *Thinking about the Unthinkable.* New York: Avon Books.

Karash, K., Coogan, M., Adler, T., Cluett, C., Shaheen, S., Aizen, I., & Simon, M. (2008). *Understanding How Individuals Make Travel and Location Decisions: Implications for Public Transportation.* Washington: Transit Cooperative Research Program.

KiM. (2015). *Chauffeur aan het stuur?: Zelfrijdende voertuigen en het verkeer en vervoersysteem van de toekomst.* Den Haag: Ministertie van Infrastructuur en Milieu.

Levinson, M. (2008). *How the Shipping Container Made the World Smaller and the World Economy Bigger.* Princeton: Princeton University Press.

Litman, T. (2014). Autonomous Vehicle Implementation Predictions Implications for Transport Planning. *Transportation Research Board Annual Meeting*, 36-42.

McFadden, D. (1975). *Urban Travel Demand.* Amsterdam, Oxford, New york: North-holland Publishing Company and American Elsevier Publishing Company Inc.

McNally, M. (2000). *The Activity-Based Approach.* Irvine: Department of Civil and Environmental Engineering and Institute of Transportation Studies University of California.

Milakis, D., Arem, B. v., & Wee, B. v. (2015). *Policy and society related implications of automated driving, a review of literature and directions for future research.* Delft: Transport & Planning, TU Delft.

Milakis, D., Snelder, M., Arem, B. v., Wee, B. v., & Correira, G. (2015). *Development and transport implications of automated vehicles in the netherlands: scenarios for 2030 and 2050.* European Journal of Transport and Infrastructure Research.

Müller, K., & Axhausen, K. (2010). *Population synthesis for microsimulation: State of the art.* Zürich: ETH .

Müller, K., & Axhausen, K. (2011). *Hierarchical IPF: Generating a synthetic population for Switzerland.* Washington D.C.: 90th Annual meeting of the Transportation Research Board.

Müller, K., & Axhausen, K. (2012). *Multi-level fitting algorithms for population synthesis.* Zürich: ETH.

NHTSA. (2014). *Human Factors Evaluation of Level 2 And Level 3 Automated Driving Concepts.* Washington: US department of transportation.

Prillwitz, J., & Barr, S. (2011). Moving towards sustainability? Miblity styles, attitudes and individual travel behaviour. *Elsevier*, 1590-1600.

Pritchard, D. (2008). *Synthesizing agents and relationships for land use / transportation modelling.* Toronto: University of Toronto.

Pritchard, D., & Miller, E. (2009). Advances in agent population synthesis and application in an integrated land use and transportation model. *88th Annual Meething of the Transportation Research Board.* Washington D.C.

Puyleart, S. (2016). *Social desirability and mobility impacts of Social desirability and mobility impacts of.* Delft: TU Delft.

Ryan, J., Maoh, H., & Kanaroglou, P. (2009). Population synthesis: Comparing the major techniques using a small, complete population of firms. *Geographical Analysis*, 181-203.

Salvini, P., & Miller, E. (2005). ILUTE: An operational prototype of a comprehensive microsimulation model of urban systems. *Networks and Spatial Economics*, 217-234.

Schultz van Haegen, M. (2014). *Grootschalig testen van zelfrijdende auto's.*

Schultz van Heagen, M. (2015, 1 23). *Besluit van tot wijziging van het Besluit ontheffingverlening exceptionele transporten (ontwikkeling zelfrijdende auto).* Retrieved 5 15, 2016, from Rijksoverheid.

Shladover, S., Dongyan, S., & Lu, X. (2012). Impacts of Cooperative Adaptive Cruise Control on Freeway Traffic Flow. *91st annual meeting of the Transportation Research Board*, 1-17.

Snelder, M., Arem, v. B., Hoogendoorn, R., & Nes, v. R. (2015). *Methodische Verkenning Zelfrijdende Auto's en Bereikbaarheid.* Delft: Civiele techniek en geowetenschappen.

Srinivasan, S., & Ma, L. (2009). Synthetic population generation: A heuristic data-fitting approach and validations. *12th International Conference on Travel Behaviour Research.* Jaipur.

Srinivasan, S., Ma, L., & Yathindra, K. (2008). *Procedure for forecasting household characteristics for input to travel demand models.* Gainesville: Transportation Research Centre, University of Florida.

Suurs, R., Lindt, M. v., Vecht, B. v., Scheepstal, P. v., Clignett, L., & Jonkers, E. (2013). *FOUNTAIN (Framework fOr Upscaling Transitions And INnovations): doelstellingen, achtergronden en ontwikkelperspectief .* Delft: TNO.

Taylor, J. (2011). *Open-Source NeXTA Version 3 User's Guide.* Utah: Free Software Foundation.

Temeurshoev, U., Yamano, N., & Webb, C. (2009). *Projection of Supply and Use tables: methods and their empirical assessment.* Working Paper.

TNO. (sd). *Implementatie Spitsvrij - interventie in Fountain.* Delft: not published.

Transportation Research Board. (2016). A Comprehensive Review of "Mobility as a Service" Systems. *Transportation research board 95th annual meeting* (pp. 16-5401). Washington D.C.: Transportation Research Board.

van Arem, B., van Driel, C., & Visser, R. (2006). The Impact of Cooperative Adaptive Cruise Control on Traffic-Flow Characteristics. *IEEE Transactions on intelligent transportation systems*, 429-436.

Voas, D., & Williamson, P. (2000). Evaluating Goodness-of-fit Measures for Synthetic Microdata. *Geographical and Environmental Modelling*, 177-200.

Wagenaar, W., & Sagaria, S. (1975). Misperception of exponential growth. *Perception & Psychophysics*, 416--422.

Weerdt, v. d., Hof, T., & Vonk, T. (2014). Agent based gedragsmodellering als tool voor de vorming van mobiliteitsbeleid. *Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk.* Eindhoven: TNO.

Ye, X., Konduri, K., Pendyala, R., Sana, B., & Waddell, P. (2009). *A methodology to match distributions for bots household and person attributes in the generation of synthetic populations.* Washington: 88th Annual Meeting of the Transportation Research Board.

Zmud, J., Tooley, M., Baker, T., & Wagner, J. (2015). *Paths of automated and connected vehicle deployment: Strategic roadmap for state and local transportation agencies.* Texas: Texas A&M Transportation Institute.

Appendix A: Fountain

In this appendix a description of the simulation model Fountain is given in more detail.

In this version, however, this part has been removed as it has been considered confidential by TNO.

Appendix B: Data meta-analysis

In this appendix the data used for the meta-analysis can be found. In Table 28 the personal characteristics of each segment in (Anable, 2005) can be found. The study distinguishes six lifestyles towards mobility, the malcontent motorists, the complacent car addicts, the diehard drivers, the aspiring environmentalists, the car-less crusaders and the reluctant riders. The data on socio-demographic characteristics that is available is gender, age, employment, income, education, and the household size. The data is used for the gender analysis, the income, education and household size. For the age analysis the study is not used since there is missing an age group and there is not stated clear how this should be interpret.

Table 28 Data overview Anable, J. 2005

Personal characteristics of each segment

		1. Malcontented Motorists (%)	2. Complacent Car Addicts (%)	3. Die Hard Drivers (%)	4. Aspiring Env'talist (%)	5. Car-less Crusaders (%)	6. Reluctant Riders (%)	Sample average
Gender (females)		55	41	56	50	59	84	50
Age	<34 years	16	17	14	21	8	0	16
	>65 years	17	8	19	12	35	63	17
Employment	FT+PT	64	63	62	70	39	21	62
	Retired	28	23	29	18	50	68	28
Income	<£10k	8	3	6	7	20	47	8
	>£40k	35	40	27	37	24	6	33
Education	None	6	6	9	1	7	32	7
	>Degree	53	48	53	69	37	32	49
With kids still at home		30	31	35	35	4	5	30
Single adult household		9	9	7	15	37	42	12
Duel income household		53	48	58	44	17	11	48

In Table 29 the percentage by age for each segment from (Karash, et al., 2008) can be found. There are four lifestyles distinguished in this study, the transit loyalists, the environmental mode changers, the happy drivers and the angry negatives. This study only had data available on age versus lifestyles. Therefore the data is only used in the age analysis.

Table 29 Data overview Karash et al. 2008

	Percentage by Age, for Each Segment				
	Under 30 years old	30–39 years old	40–49 years old	50-Plus years old	Total
Transit Loyalists	**29.4**	29.4	14.7	26.5	100.0
Environmental Mode Changers	23.3	22.7	23.3	**30.7**	100.0
Happy Drivers	27.3	26.5	**27.3**	18.9	100.0
Angry Negatives	19.2	**32.5**	20.5	27.8	100.0
All Respondents	24.0	27.5	22.4	26.1	100.0

Age-groups with the highest proportion for each market segment are shown in bold.

In Table 30 the socio-demographic and –economic characteristics of daily travel clusters can be found. The study from (Prillwitz & Barr, 2011) distinguishes four lifestyles towards mobility, the persistent car users, the frequent car users, the constrained pt users and the consistent green travellers. The dataset contains data on age, gender, income, the amount of children in a household and political opinion. For all these analysis the data is used except for the household compilation analysis. There is no data on how many adults are in the household and this was the most important variable that is discussed in chapter 4.

Table 30 Data overview Prillwitz, J. 2011

Socio-demographic and -economic characteristics of daily travel clusters.

Variable	Value	Cluster 1 Persistent car users (%)	Cluster 2 Frequent car users (%)	Cluster 3 Constrained pt users (%)	Cluster 4 Consistent green travellers (%)	No. of Answers
Age (in years)	16–19	0.0	0.3	2.9	4.0	1008
	20–29	6.8	12.0	10.9	19.6	
	30–44	21.3	27.7	12.6	41.5	
	45–59	30.8	28.9	18.1	24.1	
	60–74	28.5	26.2	36.1	10.3	
	75 and older	12.7	4.9	19.3	0.4	
Gender	Male	51.1	43.3	34.5	41.0	991
	Female	48.9	56.7	65.5	59.0	
Annual gross income	Below £5000	0.6	3.1	5.4	7.4	746
	£5001–£10,000	8.5	5.4	21.1	6.2	
	£10,000–£12,500	4.9	5.0	14.3	2.8	
	£12,501–£15,000	8.5	8.1	9.5	3.4	
	£15,001–£17,500	9.1	5.8	8.8	6.8	
	£17,501–£20,000	8.5	9.7	6.8	9.7	
	£20,001–£25,000	13.4	9.7	12.2	10.2	
	£25,001–£30,000	7.3	8.9	8.8	13.1	
	£30,001–£35,000	6.7	12.4	8.2	11.9	
	£35,001–£40,000	14.0	9.3	2.0	7.4	
	Over £40,000	18.3	22.8	2.7	21.0	
Children (under 16 years old) in household	0	79.1	62.5	87.5	57.8	1051
	1	9.8	16.4	4.8	20.7	
	2	8.1	16.4	4.4	15.9	
	3 or more	3.0	4.8	3.2	5.6	
Party that would be voted for	Conservatives	51.9	46.2	35.3	29.3	716
	Labour	18.1	14.8	22.4	20.1	
	Liberal Democrats	10.0	18.6	13.5	16.5	
	Greens	2.5	2.1	5.1	10.4	
	Other	4.4	4.2	7.7	4.9	
	Would not vote	13.1	14.0	16.0	18.9	

Assympt. significance for all variables (Pearson Chi-square): $p < 0.001$.

In Table 31 an overview is given of the data in the study (Anable, 2011). The seven lifestyles towards mobility that have been distinguished are status seekers, devoted drivers, reluctant pragmatics, practical travelers, active car drivers, car contemplators and car free choosers. The data that is available in this dataset are on gender, age, education and the household compilation. The dataset is used in all those analysis.

Table 31 Data overview Anable, J. 2011

DEMOGRAPHICS									
GENDER	Male	67%	51%	72%	65%	58%	63%	29%	61%
	Female	33%	49%	28%	35%	42%	38%	71%	39%
AGE	< 24 years	3%	17%	19%	27%	5%	22%	8%	15%
	25-34 years	65%	52%	41%	45%	70%	57%	64%	56%
	35-44 years	12%	17%	22%	18%	11%	16%	12%	16%
	45-54 years	12%	9%	9%	10%	14%	5%	12%	10%
	> 55 years	9%	4%	9%	0%	0%	0%	4%	4%
	Average age (years)	2.76	2.24	2.54	2.88	2.77	1.63	2.50	2.60
EDUCATION	15 years old or less	5%	0%	0%	0%	1%	0%	0%	1%
	16 – 19 years old	24%	14%	13%	17%	8%	13%	21%	13%
	20+ years old	71%	80%	82%	75%	88%	63%	71%	81%
	I am still studying	0%	6%	5%	8%	2%	25%	7%	5%
	Average education	2.67	2.92	2.92	2.90	2.92	3.13	2.86	2.91
NO. OF ADULTS	One adult only	39%	23%	28%	26%	23%	13%	45%	26%
NO. OF CHILDREN	Households with no children	67%	71%	60%	62%	58%	88%	64%	63%
	Average no. children	.62	.47	.73	.69	.75	.13	.43	.65

Appendix C: Implementation Case Utrecht rush hour avoidance in Fountain

In this appendix an extensive explanation of the implementation of the case 'Utrecht rush hour avoidance', in Fountain is given. First a description of the intervention can be found. In the second section the definition of groups in Fountain can be found, in the third section the definition of the intervention can be found and in the fourth section the environment files are described.

Description of the intervention

Fountain has a roadmap that can be followed to implement a case. The roadmap exists of a list with questions to define the case and the intervention. In this section the roadmap will be followed and the questions will be answered for the Case 'Utrecht rush hour avoidance'. The answers to the questions are stated in (TNO, sd).

Questions

1. What do you want to change: Human behaviour or infrastructure? Human
   a. Human behaviour: Who do you want to be influenced?
      All road users that travel frequently on the highways A1, A27 and A28
   b. Infrastructure: Which infrastructure do you want to adjust?
2. What is case plan (describe the intervention more specific)?
   Give a financial reward for avoiding rush hour at the highways A1, A27 and A28. The height of the reward depends on the number of km travelled on the highways
3. How long will the intervention be inserted?
   15 months with reward and 3 months without reward
4. For every choice possibility the effect will be described. The possible choice are the modality choice, the travel time choice and the work location. The effect on availability, preferences and expected utility is evaluated.
   a. Modality
      - Availability of modalities: Equal, there are no more modality choice possible
      - Preferences: Equal, a reward is not immediately of influence on preference of a modality
      - Expected utility: There is a change, travelling by public transport or by bicycle bring a financial reward. For the car this is equal because there is no reward or punishment by using the car
   b. Travel time
      - Availability: Equal, (early, normal and late)
      - Preferences:
        - Early: Agent has a more positive opinion towards leaving early because of the financial reward
        - Normal: Agent has the same opinion because there is no reward and no punishment
        - Late: Agent has a more positive opinion towards leaving late because of the financial reward
      - Expected utility:
        - Early: The expected utility increases because of the financial reward
        - Normal: The expected utility remains the same

- Late: The expected utility increases because of the financial reward

c. Work location
- Availability: Equal, availability stays the same (home, work)
- Preferences: The agent has a more positive opinion towards working at home because of the financial reward
- Expected utility: The expected utility for working at home increases

5. Execute choices: The agents have the possibility to make a considered choice for the modality, the travel time and the work location.
6. Calculate utility: In the next step, Fountain decides the utility. This depends on two values:
   a. Pay Offs: The utility of the modalities, travel times and work location in comfort, time and costs.
   The implementation of Spitsvrij changes something about these base values. The base value for costs decreases for several options. In Table 32 an overview is given.

Table 32 Pay Offs, comfort, time and costs

	Car/ Normal	Car/ Early	Car/ Late	PT	Bicycle	Work/ Work	Work/ Home
Comfort	Equal	Equal	Equal	Equal	Equal	Equal	Equal
Time	Equal	Equal	Equal	Equal	Equal	Equal	Equal
Costs	Equal	Decrease	Decrease	Decrease	Decrease	Equal	Decrease

   b. Personal values: This is not changed in the case 'Utrecht rush hour avoidance'
7. Evaluate utility
8. Update preferences
9. Execute behaviour

By following these steps, the intervention is described good enough to implement it into the model.

Definition of groups

Three groups are defined in the case 'Utrecht rush hour avoidance'. First group is all agents, the second group are the participants and the third group are the agents that actually avoid the rush hour. In Table 33 an overview is given of the three groups and how they are defined.

Table 33 Definition of groups in Fountain

Group	Definition	Remark
All agents	AgentType >= 0	Every agent meets this requirement
Participants	ChoiceTravel time = NORMAL AND Environment.Modality.SelectedRouteDistance > 0 AND Choice.Modality = Car	At time step 30. When an agent meets the requirements, the variable InSpitsvrij becomes 1
All	InSpitsvrij = 1	Invited agents that

Participants		meet requirements
Avoid rush hour	InSpitsvrij = 1 AND Choice.Travel time ≠ NORMAL OR Choice.Modality ≠ Car OR Environment.Modality.SelectedRouteDistance = 0	Actual avoiders of the rush hour
Only avoidance with reward	InSpitsvrij = 1 AND Choice.Travel time ≠ NORMAL at time 156 OR Choice.Modality ≠ Car at time 156 OR Environment.Modality.SelectedRouteDistance = 0 at time 156 AND ChoiceTravel time = NORMAL at time 216 AND Environment.Modality.SelectedRouteDistance > 0 at time 216 AND Choice.Modality = Car at time 216	Avoidance of rush hour when a reward is received, but once the reward ended there was no avoidance behaviour
Avoiders living in zone i	Residential = i AND InSpitsvrij = 1 AND Choice.Travel time ≠ NORMAL OR Choice.Modality ≠ Car OR Environment.Modality.SelectedRouteDistance = 0	Actual avoiders of the rush hour living in zone i
Avoiders living in zone i with lifestyle l	Residential = i AND PersonType = l AND InSpitsvrij = 1 AND Choice.Travel time ≠ NORMAL OR Choice.Modality ≠ Car OR Environment.Modality.SelectedRouteDistance = 0	Actual avoiders of the rush hour living in zone i with lifestyle l

The group all agents are defined by the agent type. Every agent has a value for agent type higher than 0. Therefore all agents belong to this group. The second group are the participants. This are the agents that received a letter and are potential participants. There are three conditions that must all three be achieved by the agent. First condition is that the agent is travelling during the rush hour at the start time. Therefore the condition travel time must be normal. Other options for travel time are early and late. The second condition is the mode choice. The agent must be travelling by car at the start time. The third condition is the selected route distance. In the environment of Fountain, several highways are defined. The agent needs to travel more than 0 km on one of these highways. This actually says that the agent must travel in the project area during the start of the project. When an agent meets all the three conditions, the variable InSpitsvrij will receive a value 1 instead of 0. The third group is the group that actually avoids rush hour. This group is defined as agents that are in the participant group and therefore have a value of 1 for the variable InSpitsvrij. Besides this they have to meet one of the following three conditions. The first condition is that their travel time is not normal, when the travel time is not normal, the travel time is early or late and the agent avoids rush hour in this way. The second possible condition is the modality choice, if this is not the car then the agent avoids the rush hour by using another modality. The third possible condition is that the selected route distance is zero, when this distance is zero, the agent does not use the selected highways and selects another route or works at home. All three conditions are a way of avoiding the rush hour. The last group is the group that did avoid the rush hour during the reward was given, but this not avoid rush hour when the reward ended in the last three months.

Definition of intervention

In the first section of this appendix a description of the intervention is given. In this section the definition of the intervention is given. In Table 34 an overview can be found of the interventions and other events in the model. In the first column the intervention name can be found, in the second column the group is defined which receives the intervention, in the third column the definition of the intervention is given and in the fourth column the time step where the intervention takes place is given. Every time step represents one working day, every week has five time steps.

Table 34 Definition of intervention

Intervention	Group	Definition	Time step
Reward to avoiders	Avoid rush hour	Reward = 0,25 * km on selected highways	OnAt:30:OffAt:156
Is participant	Participants	InSpitsvrij = 1	OnAt:30:OffAt:219
Trigger 1	Participants	Trigger decision mode choice	OnAt:30:OffAt:40
Trigger 2	Participants	Trigger decision travel time	OnAt:30:OffAt:40
Trigger 3	Participants	Trigger decision work location	OnAt:30:OffAt:40

The first intervention is the reward that is given only in the first 15 months, the reward is 0,25 euro per km on the selected highways and is only given to the group that actually avoids rush hour. The second intervention is the status that is given to participants, the variable InSpitsvrij receives value 1 and only participants can receive the value. The other interventions are triggers, triggers are given in the first 10 days of the project. The trigger tells the agents to make a considered choice for respectively the mode choice, the travel time and the work location.

Environment

The environment of the case needs several files of the geographical area, the zone classification, the network and the population. The files that have been adjusted for this case can be found in Table 35.

Table 35 Environment adjustments

File	Content

Input_hb1.csv	OD matrix for Status Seekers
Input_hb2.csv	OD matrix for Devoted Drivers
Input_hb3.csv	OD matrix for Reluctant Pragmatics
Input_hb4.csv	OD matrix for Practical Travellers
Input_hb5.csv	OD matrix for Active Car Owners
Input_hb6.csv	OD matrix for Car Contemplators
Input_hb7.csv	OD matrix for Car Free Choosers
Input_node.csv	Geographical notation of the zones

The seven OD matrices are the output of the population synthesis tool that is created (more information about this in chapter 5). The geographical notation of the zones is created by exporting the cordon in Omnitrans to a shape file. This shape file is used as input for a program called NeXTA, in NeXTA the correct coordination system can be created for Fountain (Taylor, 2011).