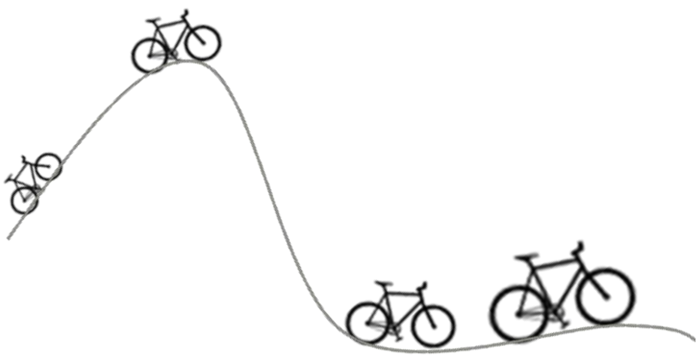
Geographical Information Management and Applications

Route Choice Preference of Cyclists: an agent-based simulation model for the city of Utrecht

Author : Pinar BALCI
Student Number: s6025706 (ITC)
56017889 (UU)
910816 – 032 -100 (WUR)
E-mail : p.balci@students.uu.nl

Master of Science Degree Program of
Geographical Information Management and Applications



Universiteit Utrecht

TU Delft

University of Twente.

Wageningen UR

TNO innovation for life

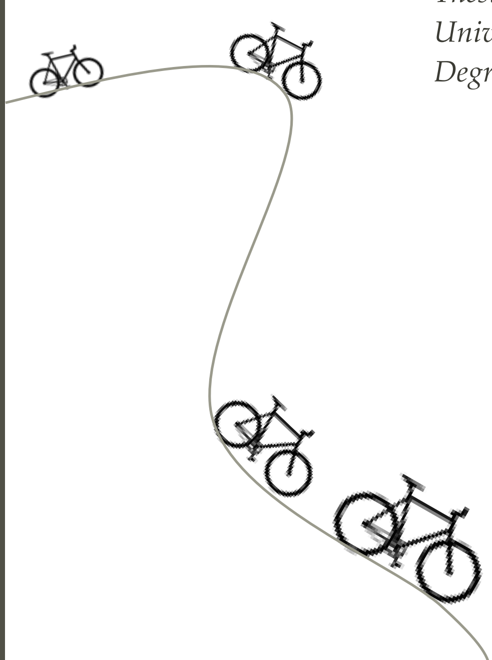
Master's Thesis

Route Choice Preference of Cyclists: an agent-based simulation model for the city of Utrecht

By

Pinar BALCI

Thesis submitted to Utrecht University, Delft University of Technology, Wageningen University and University of Twente in partial fulfillment for the Master of Science Degree Programme in Geographical Information Management & Applications (GIMA).



Thesis Assessment Board

Professor: Prof. Dr. ir. Arnold Bregt
(WUR, Wageningen)

First Supervisor: Dr. ir. Ron van Lammeren
(WUR, Wageningen)

Second Supervisor: ir. Aldo Bergsma
(WUR, Wageningen)

February, 2017

Abstract

This research serves an informative guide and a repeatable approach regarding creating an agent-based modeling of cyclists route-choice preferences. It provides innovative features for the agent-based modeling of cyclists and urban planning domain. The study seeks to determine the underlying reasons for route choices made by cyclists and tries to implement these reasons as the parameters of cyclist agents route choice behavior in the agent-based model, on the given bike road network.

In the first place, it provides a platform where two individual movement frameworks, cycling, and pedestrian movements, are integrated and discussed together within a holistic approach. Due to very limited number of literature and no example models in the field of cyclists route choice preferences, it was relatively a rocky road to execute the research in the field. The combination of cyclist and pedestrian movement fields and elimination of the literature with a set of criteria provided a rich and multi-visional theoretical background in which the main elements of the movement are determined and further studies can benefit from.

Secondly, it presents the analysis of measured GPS data in order to check the applicability of the determined factors from literature to the user group of the study. The case study area was selected as the SSD project area which can be defined as the Central Station area of Utrecht, and the user group was defined as cyclists who daily travels over the Central Station area. The measured GPS data analysis in the field showed that not all factors reflected in previous studies are applicable to the case study. Moreover, the research provides an approach in order to identify more information related to cycling agents and their environment to be modeled by the use of measured GPS data.

The creation, development, calibration and validation of the agent-based model are also elaborated through the report. The high similarity between the revealed intensity pattern of GPS data and the modeled intensity pattern of ABM presented that the parameters used to model the movement of cyclists are the some of the major factors that influence cyclists route choice preference. Also, the travel time distribution of cyclists in the measured GPS data and agents in the model showed quite similarities on the cumulative travel time distribution graphs. These two validation methods clarify that the model represents similar intensity distributions over the bike road network of Utrecht with a real-like cycling movement flow on the given network.

The end-result model of this study may provide limited usage in terms of its applicability to the case study since the creation of such a model requires a longer time than 5-months due to the complexity of cycling behavior and very limited number of studies done in this field. With this

research, the initial foundation model is developed which further developments can be built upon it. For this purpose, the model has been developed by including fundamental elements of cycling movement and having flexibility characteristics in order to be enhanced in future researches. Due to these given reasons, the model has not been applied in the use case of SSD project within this limited time period. However, all elements required are included in the model. It is believed that by the use of listed recommendations given at the end of the report, the application of the model in several use cases should be executed smoothly.

Lastly, it is important to consider the use of such model. Route choice models can be seen as the key elements in the future development of bicycle demand forecasting models. As route choices made by cyclists forms the cycling flow, they shape the intensities and, consequently, demands for future developments on the road network. Understanding the underlying reasons for route choices made and implementing them in a simulation model provides urban planners to observe the possible effects of changes that will be made on the network before putting into practice the urban planning projects in the real world. Hereby, municipalities or responsible stakeholders can develop networks that target the demand and provide tailor made accessibility for cyclists. For that purpose, the model can be used in several urban development scenario analyses. Therefore, it helps to maximize the supply of the limited resources.

Preface & Acknowledgements

This master thesis focuses on the ubiquitous mode of transport in the Netherlands; cycling. The very first day I moved to the Netherlands, I was amazed by the crowd of the cyclists who show a flawless mobility flow in the quite narrow and busy streets of the city of Utrecht. This astonishment and curiosity of mine have been turned into practice in this thesis in order to understand the underlying elements and reasons of this movement.

No later than two years, the same astonishment has been felt again when I have been decided to study this topic and found out how cycling has stayed as such an underexplored topic in the field of individual movements. Admitting the truth, I was initially overwhelmed by the lack of resources in the field of cycling route choice preferences and my lack of experiences in the field of agent-based modeling. Despite that, I felt the responsibility to achieve my goals in order to thoroughly fulfill the opportunity given to me by the organization of TNO.

Now that this thesis has finally reached the completion and lies in front of you, a word of thanks is in place for those who have helped me along this rocky way. First of all, I would like to thank my supervisors Ron van Lammeren and Aldo Bergsma for giving me valuable tips and comments who shared my curiosity about the topic. Your contributions have helped me to upgrade the level of this research.

I would like to specifically mention my technical mentor in TNO, Ernst Jan van Ark. His support, enthusiasm and precious feedbacks have aided me in reaching to this point. Without the discussions made with him, I could not approach to the route choice preference topic through his traffic engineering vision and go through this steep learning curve. His professional manner in data analysis and rich knowledge in traffic engineering forced me to do better works every day. Also, I would like to thank Roel Massink who gave me the opportunity to execute my thesis project in TNO at the first place. By the help of him, I reached the GPS dataset sources, associated stakeholders and the thesis process has been commenced.

Last but not least, I would like to express my gratitude to my fiancee Cagil Mayda who shared the world which was mine for five months in the moments I felt the lowest and be my biggest supporter and helper during this process.

Pinar Balci
February 2017

Table of Contents

Table of Figures

CHAPTER 1
INTRODUCTION

The world is urbanizing at an extremely fast pace which brings along both opportunities and challenges (Zuidgeest, Brussel, & van Maarseveen, 2015). Urban growth has many positive effects in terms of urban economy and social vitality. On the other hand, the daily mobility of these growing populations can be a further burden on the mobility, accessibility, comfort, safety and ultimately the life quality of cities (Municipality of Amsterdam, 2012). Together with the population growth, the number of daily activities occurring in the urban areas is increasing as well as the frequency of the daily mobility. These dynamic and continual changes of travel patterns challenge existing urban forms and physical networks which can occur as a problem in terms of accessibility and safety conditions of urban areas (Bulkeley et. al., 2014).

Mobility provided by the transport systems has a major role in shaping urban forms, the location of social and economic activities and the lifestyle the city offers (Zuidgeest, 1997). Cycling provides fast and inexpensive personal mobility. For many Dutch cities; cycling mobility has been rising rapidly as being the most popular transportation mode (ECECO, 2003). However, cycling traffic in the urban area growing in direct proportion to the population growth. An example of this case can be given from the city of Utrecht, Netherland. Especially the cyclists who travel to and from the city center have been showing a steady increasing (Gemeente Utrecht, 2015). To keep the balance between urban growth and flawless daily mobility of cyclists, and to continue to utilize the benefits of the widespread use of bicycles; a better understanding of cycling' mobility and route choice preference must be gained.

The infrastructural road network of a typical city provides a large number of alternatives to travel between any two locations. ***As most of the time there is more than one route (also can be called as a trajectory) between an origin and a destination location, every trajectory consists of a route, a choice and a route choice decision*** (Usyukov, 2013) (Figure 1).

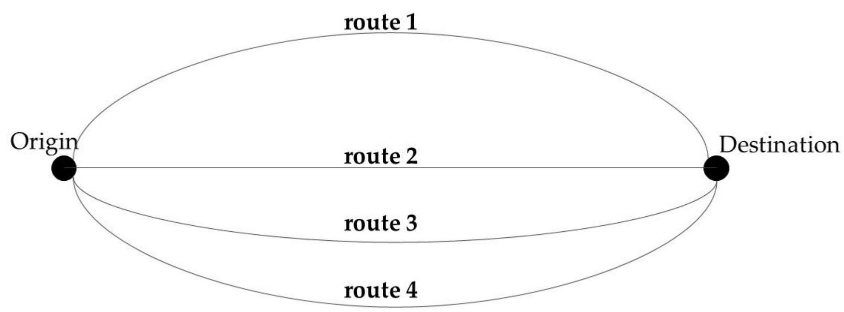


Figure 1 Route choice phenomenon (1)

An understanding of cyclists' route choice may answer questions on how relations between the urban network and urban cycling can be adapted to cater for cycling which leads to more accessible and livable Dutch urban regions (Bulkeley et. al., 2014). Undoubtedly, the result of such a study helps transportation and urban planning domains to invest the limited resources in a most effective manner and to practice urban developments that can provide benefits to the greatest number of cyclists.

There are several ways to analyze and understand individuals activity behavior. Along with the technological developments, automated equipments are available to record the routes taken (**Global Positioning Systems – GPS**) which makes collecting data about cycling activity easier and more accurate comparing to the traditional methods of data collection; i.e. counting data, traditional tracking data (Segadilha & Sanches, 2014a).

Additionally, when GPS used in conjunction with **Geographic Information Systems (GIS)**, it offers the possibility to acquire precise information on the routes taken and activity time of the cyclists. Therefore, it allows researchers to analyze spatial and time dimension of the activity (Broach, Dill, & Gliebe, 2012; Broach, Gliebe, & Dill, 2011; Hood, Sall, & Charlton, 2011; Loidl et al., 2016b; Segadilha & Sanches, 2014a). Several researchers used GPS data show that GPS technology is an effective way of collecting real-time data of cyclists.

While GIS helps to analyze and represent existing time and space pattern of the mobility (Broach et al., 2012, 2011; Hood et al., 2011; Loidl et al., 2016b; Segadilha & Sanches, 2014a), **agent-based modeling (ABM)** allows to simulate continual change and evolution through time and space (A. J. J. Heppenstall & Crooks, 2012). It exhibits an adequate representation of individual's behavior assuming that social and environmental structures strongly influence their travel behavior (Loidl et al., 2016b). Moreover, ***these systems that are disaggregated into individual components can have their own characteristics and rule*** *sets (A. J. J. Heppenstall & Crooks, 2012)* ***or "preferences" as it is called in this research.***

1.1. Background of the Thesis; Key Concepts

- ***Commuter Cyclists: The Dutch Lifestyle***

According to literature, when it comes to analyzing the movement of cyclists; there are mainly two domains of study groups one can focus on: commuter cyclists and leisure cyclists. Commuter cyclists are mostly the local people who cycle through the network on a regular basis to deal with their daily activities such as going to work, coming back to home, shopping for daily needs. This also can be summarized by cycling for utility (for a purpose) (London, 2011). Another group is the leisure cyclists who cycle on a non-regular basis or cycle for pleasure; this group may include tourists and non-local people which may exhibit extraordinary route preferences than daily users of the network.

In this study, the user group is approached by looking at the route choices they made and trying to derive the underlying reasons for these decisions made. As the data used for this study was obtained from the local, daily cyclists; it can be said that the route choices study of this research is focused on commuter cyclists. As the case study is located in the Central Station area of Utrecht (Chapter 3); capturing the revealed route choices of commuter cyclists constitutes a fundamental qualification, since every day over 100.000

cyclists ride to their work, school, public transport, shops or home via the city center between 7 AM and 7 PM (Gemeente Utrecht, 2015).

- ***Route choice preferences***

Each individual perceives the attributes of each route subjectively and each alternative is transformed into an attractiveness scale, which is linked to the traveler's experiences, preferences, or constraints (Usyukov, 2013). "Different individuals may make the same decisions, that is, choose the same route, though on different grounds" (Bovy et al, 1990), and this is what is observed and modeled.

Route choice models take an important place in the transportation domain as they provide these listed features:

- "Predict route choice dependent on the routes' and travelers' characteristics;
- Help in designing and re-designing transportation facilities;
- Assess travelers' reactions to proposed network changes;
- Assess the amount of excess travel, caused directly from route selection criteria" (Usyukov, 2013).

In regards to the first application example, route choice models can be used to predict market segmentation or the share of travelers selecting a particular route depending on road and traveler characteristics.

Other than that, these models help urban planning domain under the topics of supporting the existing traffic demand and on top of that, stimulating the usage of bicycle. Since route choice models have a quantitative feature that contains road and traveler characteristics, they are able to predict which road characteristics are significantly preferred by whom. For example, should the bicycle paths be built along roads or separate than roads? How does this new design effect the flow of this user group of cyclists? ***To sum up, if the cyclists' route choice reasons and behaviors are well known, urban planners can have the opportunity to design cycling facilities in a way that attracts users.***

Route choice models also allow to evaluate traveler attitudes with respect to changes in a transportation system; e.g. which route a traveler will choose if one particular road segment is blocked or congested for a long time? The result of that can be observed through flow distributions (Cambridge Systematics, 1999).

- ***GPS tracking data***

From the ancient days, people have observed several moving entities and investigated their movement behaviors for various purposes (Andrienko et al., 2008). Since the environment where movement takes place has the major influence on the movements, environment and movement need to be considered together within a holistic approach. Moreover, in some cases movements themselves are not the main focus of the study when the main objective is to gain knowledge about the entities that move. Therefore, in the research field of space-time, observing the daily movements of human individuals is the means of investigating activities of different groups of people (Andrienko et al., 2008).

Through the expansion of GPS, new data collection methods to obtain cycling route information have emerged (Jestico, Nelson, & Winters, 2016). While traditional methods of data collection are expensive, time-consuming and lack of spatial and temporal details of movement activities; technological developments based on GPS-equipped devices have started to offer new data collection methods (Andrienko et al., 2008; Jestico et al., 2016). As the availability of GPS devices has grown enormously in the last decade, more people own a navigation system such as TomTom or a mobile phone or any other handheld device with built-in GPS functions (van der Spek, van Schaick, de Bois, & de Haan, 2009). Even though they are mainly used to determine where you are (orientation function) or where to go (navigation function); it is also possible to be used for tracking, such as keeping a track log with the routes traveled. This function makes to collect useful spatial-temporal data to measure activities of people (van der Spek et al., 2009). As the GPS data is easier to obtain comparing to traditional methods, it is as easier to be analyzed by researchers (Andrienko et al., 2008).

As there are many ways to collect GPS tracking data of people; one way is using crowdsourced applications. User-generated data from mobile tracking apps forms a powerful input for GIS to quantify and the spatial and temporal dimensions of the mobility (Jestico et al., 2016).

- ***Agent-based modeling (ABM)***

"In the last two decades, there was a shift from aggregated, regional-scale transport models to simulations on the disaggregated individual activity level" (Loidl et al., 2016a).

Disaggregated models require much more geospatial intelligence than aggregated transport models. Agent-based models aim for an adequate representation of human social behavior (also called as daily patterns) and simulates the actions and/or interactions of autonomous agents by providing a simulation in order to observe the effect of changes made on the system. Hence, the understanding of transport systems and mobility behavior on an individual basis can be significantly enhanced by agent-based models (Loidl et al., 2016).

Agent-based modeling is a computational method that allows researchers to create and experiment a simplified representation of a "target system" with models which are composed of "agents" that interact and influence each other within an "environment" (A. J. Heppenstall, Crooks, See, & Batty, 2013). A model expresses a target system the way in which the developer believes that system operates. ABM allows to simulate continual change and evolution through time and space and offers the disaggregation of systems into individual components which can have their own characteristics and rule sets (A. J. Heppenstall et al., 2013).

Agent-based models fall into the category of individual-based models; which are closely related to Cellular Automata (CA) and Microsimulation (MSM). CA system's space is divided into a grid of regularly spaced cells where each cell has a value of 1 or 0 or between 1 and 0 (Benenson and Torrens 2004). Unlike agents; automata's location does not move or wander in the space and CAs cannot have multiple attributes. For example; it can only be full or empty but it cannot contain building type and date built at the same time. Thus, CA models are mostly preferred to simulate possible land-use changes while ABM is mostly applied to simulate crowd dynamics and traffic flows. As for MSM, the field of modeling the effects of different policy scenarios on individuals is mostly applied. It is also important to note that MSM only able to model one-direction interactions. Therefore, the impact of the policies on the individuals cannot be simulated together with the impact of the individuals on the policies (A. J. Heppenstall et al., 2013; A. J. J. Heppenstall & Crooks, 2012).

Agent-based models can be created through the use of programming languages; Java and C++ are the mostly preferred ones for that aim. However, programming from the ground up requires a substantial time investment on behalf of the researcher to learn how to design and implement a model. Due to time limitations of this thesis study and the fact that not being an experienced programmer, this option is not a very convenient choice. Instead of programming a completely new toolkit, there are several available software for constructing agent-based models which allow user comparatively easier environment to simulate intended scenarios. Most popular ones are NetLogo, GAMA, and AgentSheets. The only major disadvantage of using such a software is that researcher has to learn and understand the language used in the software and research is restricted with the framework supported (A. J. Heppenstall et al., 2013; A. J. J. Heppenstall & Crooks, 2012) (For the details of the chosen software please refer to Section 4.2.)

As the importance of route choice models and usability of GPS data tracks in GIS analysis and practices of agent-based modeling have been mentioned, the route choice model development approach can be elaborated on. Travel behavior is undoubtedly important, and perhaps the most complex emergent, composing to the high complexity of transportation systems (Zheng et al., 2013). Understanding of travel behavior is tied with route choice preferences made. Route choice preferences may change over time because of the interactions between travelers and also the changes in the network topology; e.g. a change in the connectivity of a particular road may cause to not to choose that road to travel next time.

According to Zheng et al., although disaggregated transport models and microscopic traffic simulations have been applied to model route choice of travelers, it is difficult to integrate information-sharing and interaction among them (Zheng et al., 2013). Those behaviors are not applicable to discrete choice models, can be tackled by ABM as it was specifically developed to point this complexity (Zheng et al., 2013). Rather than applying traditional top-down methods that lack of understanding of underlying behavioral elements, a bottom-up manner in behavior-based ABMs are viable to study these complex transportation systems. A bottom-up approach has the flexibility that can be reshaped for several scenarios and the possibility to observe new emergent behaviors of a new environment setup (Zheng et al., 2013).

According to Loidl et al. (2016), further integration of GIS in the domain of transport modeling can significantly contribute to more effective model results. ***Within the thesis, the remark made by Loidl et al. (2016) developed by utilization of GIS in order to explore***

the captured GPS data in terms of modeling route choice preferences in an agent-based model. As Gimblett (2006) and Zheng et al. (2013) indicated, agent-based models linked to GIS have the great competency to research complex systems such as traffic flows (Gimblett, 2006; Zheng et al., 2013)These studies link agents to the real environment and utilize revealed behavior data obtained from the field (i.e. GPS). In conclusion***, the captured data on the actual routes can be interpreted as revealed preferences of people. Moreover, it can be used as main input for the development of an agent-based modeling to mimic individual's activity behavior; especially cycling activity regarding the scope of this research.***

1.2. Background of the Thesis; Practical Relevance

Utrecht is the fourth largest city in the Netherlands which has been growing at a rapid pace. Currently, cycling is the main method of transport in Utrecht with 33% (the average in the Netherlands is 27%) (ECECO, 2003). Therefore, it performs a non-negligible function in the pattern of movement in Utrecht. Cycling provides climate-friendly personal mobility which is also fast and inexpensive comparing to the other modes of transportation. However, cycling traffic is growing in direct proportion to urban population growth. To keep the balance between urban growth and flawless daily mobility of cyclists, a sustainable and comprehensive strategy should be developed with a better understanding of the Utrecht mobility (Sustainable Solutions for Better Environment, 2012).

Smart Sustainable Districts (SSD) is a Climate-KIC Flagship program working for district developments in European cities to enhance less and slower traffic, boost resource efficiency, reduce energy consumption and increase green spaces ("Utrecht Sustainability Institute," n.d.). The project focuses on the needs of inhabitants and the visitors of the city to provide a pleasant place to live, work, study or visit. Utrecht Central Station district is one of the areas chosen for the SSD Project aside from London: Queen Elizabeth Park, Paris: les Docks de Saint-Ouen-sur-Seine and Berlin: Moabit West. Within this project, the Municipality of Utrecht formulated a redevelopment plan together with other stakeholders -Jaarbeurs, USI, Utrecht University, TNO, Deltares, Imperial College (UK), TU Berlin (DE) and TU Munich (DE) - to create a new Centre which has some attractive, alive, safe, green and climate-friendly

Image 1 New Centre of Utrecht (USI, 2015)

characteristics. For this aim, the project has four priorities for the sustainable transformation of Utrecht's new city centre:

1. Hybrid systems for heating and cooling at district level
2. local use of renewable power
3. Green climate robust and attractive
4. **Clean and safe personal mobility**

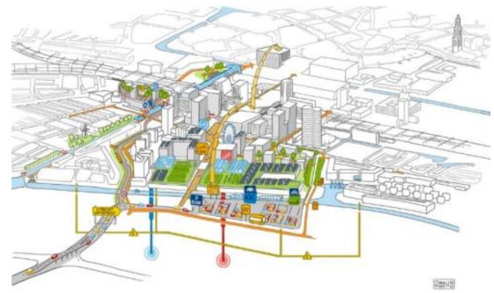


Image 2 New Centre of Utrecht (USI, 2015)

As stated by the fourth priority, clean and safe personal mobility has high importance within the SSD Project and the mode of cycling takes a big attention on itself on this stage due to its high popularity among citizens and visitors of Utrecht. As part of that priority, the bike road network is expected to be developed to supply green, clean and safe mobility which many can benefit from. In order to achieve to provide a sustainable development, it is crucial and inevitable to analyze existing cycling flow on the bike network and understand the cycling movement as a behavior with its characteristics. By this approach, it is possible to detect the demand and maximize the supply with the limited resources. Based on the existing usage of bike network of cyclists; their route choice preference can be simulated by using an agent-based modeling method. Such a model acts as a powerful tool in order to help in designing and re-designing transportation facilities and assess travelers' reactions to proposed network changes within the SSD project area. ***To sum up, if the cyclists' preferences are well known, urban planners can have the opportunity to design cycling facilities in a way that attracts cyclist users. Moreover, it will allow them to develop networks that target the demand and maximize the supply with limited resources.***

This kind of a route choice allows the municipality, urban planners, and traffic engineers to check what affect different use cases have on; e.g. cycling density in certain areas of the city (Chapter 3: Case Study).

1.3. Research Objective & Research Questions

For this research, the following research objective has been formulated:

- **To develop a simulation model of cyclist route choice preference in the city of Utrecht by using an agent-based model.**

To realize this objective, measured GPS routes of cyclists will be analyzed in order to understand their route choice preferences. The captured route choices show the revealed preferences made, which allow us to analyze the underlying reasons for preferences made. These underlying reasons have fundamental importance by means of creating the agent-based model of cyclists route choice preferences.

The research can be formed in a systematical approach that leads to achieve the main objective. For this purpose, research questions must be identified.

- ***RQ1: Which factors from former studies can be implemented in the agent-based model of this research?***

- ***RQ2: What is the added value of measured GPS data in terms of developing the agent-based model?***

- ***RQ3: How to formalize the determined factors in order to implement them in the agent-based model?***

- ***RQ4: How can the simulated behaviors and output trends be evaluated and validated?***

- ***RQ5: To what extent is this research and the model utilizable in urban planning domain?***

1.4. Thesis outline & Reader's guide

The Figure 2 on the below provides the outline of the report. The next chapter will present the theoretical background on cycling and pedestrian movement and modeling. In Chapter 3, the case study will be elaborated on. Chapter 4 will provide the methodology by explaining the steps that will be taken in order to answer research questions and achieve the research objective. Following to the methodology explained, the case study will be implemented by using the measured GPS data captured in the case study city. Chapter 5 will give the results of GPS data analysis, and Chapter 6 will give the results of created input datasets that will be loaded to the model. Throughout the Chapter 7 and Chapter 8, the results of sensitivity analysis, calibration and validation will be presented. Finally, the research will be critically reflected upon by looking back the research questions posed at the very beginning of the research.

Research Question	Method	Chapter	Data
RQ1: Which factors from former studies can be implemented in the agent-based model of this research?	Theoretical Research	2	Literature on cycling and pedestrian movement
RQ2: What is the added value of measured GPS data in terms of developing the agent-based model?	GPS Data Analysis	4.3 & 5 & 6	Measured GPS data of cyclists
RQ3: How to formalize the determined factors in order to implement them in the agent-based model?	Conceptual Framework Creation	4.4	Determined factors from RQ1 and RQ 2.
	Probability Formula Development	4.4.2	
RQ4: How can the simulated behaviors and output trends be evaluated and validated?	Sensitivity Analysis	4.5.1 & 7	Created model
	Calibration	4.5.2 & 8.1	Created model & GPS data measures
	Visual and statistical validation	4.6 & 8.2	Created model & GPS data measures
RQ5: To what extent is this research and the model utilizable in urban planning domain?	Concluding	9 & 10	Results of the research

Figure 2 Flow Chart of the Thesis

CHAPTER 2
THEORETICAL BACKGROUND

The main goal of this chapter is to gain the necessary knowledge in order to answer Research Question 1:

> ***RQ1: Which factors from former studies can be implemented in the agent-based model of this research?***

Despite having an excessive number (34 of them are reviewed) of studies on the **agent-based modeling** (ABM) of pedestrian movement, agent-based modeling of cyclists' movement is a surprisingly underexplored topic. On the other hand, there are an adequate number (54) of studies on the topic of **route choice preferences of cyclists** which enable us, researchers, to conceptualize the determinants of travel behavior of cyclists' in an agent-based modeling structure. However, these route choice studies mostly focused on aggregated level of movement analysis - which gives results on regional or city scale-, rather than disaggregated level of individual movement. Within these two research domains, **the added-value of GPS** data is beyond the doubt as it is elaborated throughout the next sections of this chapter.

This theoretical background allows us to understand the elements that influence cyclists' route choice. Therefore, it provides a foundation ground to creating a conceptual framework of cyclists'. First of all, an introductory section of ABM is provided to grasp a basic knowledge on ABM and agent-based way of thinking. Following to that, reasons to choose ABM as a method is elaborated by relating it to the ultimate goal of the model. After these sections, theoretical background on individuals' route choice preference from the perspective of cyclists and pedestrians are presented. At the end, determination of the main factors that influence route choice preference is made by concluding them from the literature. These defined factors establish a connection between the theoretical background and the conceptual framework of the model to be created.

2.1. Agent-based Modeling

Agent-based modeling is a computational method that allows researchers to create and experiment a simplified representation of a *target system* with models which are composed of "agents" that interact and influence each other within an "environement" (A. J. Heppenstall et al., 2013). A model expresses a target system the way in which the developer believes that system operates and it can take several forms in several domains such as mathematical and statistical modeling. However, an agent-based modeling is the most effective method to model mutual reinforcements and "interactions" between agents (A. J. Heppenstall et al., 2013). Another advantage of agent-based modeling is that it provides a researcher to utilize the model to do *experiments*: an example can be set up several times using various parameters or using the same parameter under different conditions to compare the results. This experimental function of agent-based models is the focus of this thesis since the model to be created within this study must be of use for scenario analyses.

A typical agent-based model consists of two main elements; "agents" that interact with each other within an "environment" they are located (A. J. J. Heppenstall & Crooks, 2012; Macal & North, 2011):

1. Agents: It has attributes and rules. Rules define how and with whom agents interact. Interactions become between other agents and environment.
2. Environment (Spatial organization): Where agents are located, live in and interact with each other (Macal & North, 2011).

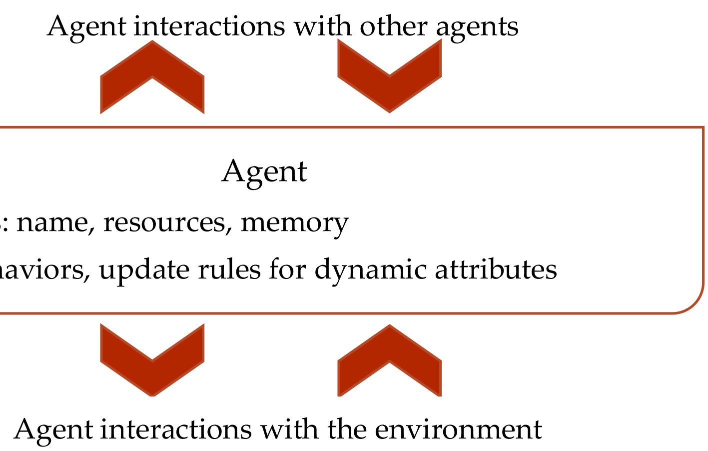


Figure 3 Elements of an ABM

As for activity-based approaches, every individual is a decision maker who confronts a large choice set of several activity patterns in the time-space domain (Zhang & Levinson, 2004). Each combination of activities and their locations, starting points, durations constitute a unique activity pattern. Agents represent people, who have characteristics, goals, and rules of behavior. The environment represents a space in which agents live. Behavioral rules define how agents act in the environment and interact with each other. Vehicles are agents in the simulation, and a static road network is an environment. Vehicles "born" at the start point of the network and "die" at the finish point. "Rules", such as free-flow driving, car-following, and lane-changing, define how a vehicle behaves and interacts with other vehicles and the road network (Zhang & Levinson, 2004).

To apply the agent-based modeling method; the agents involved in the system must be defined with their attributes, characteristics and behavioral rules. Given an initial condition, all agents behave on the basis of these identified elements (Zhang & Levinson, 2004). Before going into the details of theoretical background, the meaning of these required elements should be comprehended in order to filter the literature and have maximized the efficiency to understand given elements within them. On the below, the meaning of these elements from the literature is elaborated.

2.1.1. Agents

For practical modeling purposes, it is assumed that agents have certain attributes. These are:

- Autonomy: An agent has an autonomous nature. Therefore, it can function independently in the environment created and can interact with other agents autonomously.
- Modularity: An agent is modular and identifiable with a set of attributes or characteristics, behaviors capability.
- Sociality: An agent is social as it interacts with other agents.
- Conditionality: An agent has a state that changes over time. These changing states of all agents and the environment compose the state of an agent-based model (A. J. J. Heppenstall & Crooks, 2012; Macal & North, 2006).

Most of the time, agents have additional properties such as explicit *goals* that shape behavior. Also, it may have the ability to *learn and adapt* its behaviors according to its experiences.

2.1.2. Environment

According to Macal & North (2006), spatial relationships to represent agent interactions are:

- Soup: A non-spatial model in which agents have no locational attribute,
- Grid or lattice: Cellular automata represent agent interaction patterns and available local information by a grid or lattice; cells immediately surrounding an agent are its neighborhood. An agent's location is the grid cell index.
- Euclidean space: Agents roam in 2D or 3D spaces. An agent's location is its relative or geospatial coordinates.
- Geographic Information System (GIS): Agents move over and interact with realistic patches of geospatial landscapes. An agents' location is a geographical unit (e.g., zip code) or geospatial coordinates.

- Networks: Networks may be static (links pre-specified) or dynamic (links determined endogenously by relationship creating mechanisms). An agents' location is the relative node location in the network (Macal & North, 2011).

As the content of an agent-based model is grasped its elements; theoretical background can be presented from these required elements' point of view. During the theoretical background, previous studies are discussed in terms of their approaches and used elements that may give some inputs to the agent-based model of this research.

2.2. Choosing Agent-based Modeling as a Method

Before going through the details of literature on the topics of cyclists' route choice preferences to model cyclist disaggregated movement, the reason for choosing agent-based modeling method will be clarified. According to Macal & North (2011), agent-based modeling is especially beneficial when any of these criteria have to be satisfied:

- "When the problem has a natural representation as being comprised of agents
- When there are decisions and behaviors that can be well-defined
- When it is important that agents have behaviors that reflect how individuals actually behave (if known)
- When it is important that agents adapt and change their behaviors
- When it is important that agents learn and engage in dynamic strategic interactions
- When it is important that agents have dynamic relationships with other agents, and agent relationships form, change, and decay
- When it is important to model the processes by which agents form organizations, an adaptation and learning are important at the organization level
- When it is important that agents have a spatial component to their behaviors and interactions
- When the structure of the system does not depend entirely on the past, and new dynamic mechanisms may be invoked or emerge that govern how the system will evolve in the future.
- When arbitrarily large numbers of agents, agent interactions and agent states is important

- When process structural change needs to be an endogenous result of the model, rather than an input to the model" (Macal & North, 2011).

As presented by all of these criteria, agent-based modeling is a powerful method to mimic cycling route choice travel behavior.

Travel behavior is undoubtedly important, and perhaps the most complex emergent, composing to the high complexity of transportation systems (Zheng et al., 2013). Understanding of travel behavior is tied with route choice preference. In general, route choice preference may change over time because of the interactions between travelers and also the changes in the network topology and performance; e.g. due to an incident, a traveler may change their route. Moreover, route choice preference involves experiences, heterogenic traveler groups, and interaction among those.

According to Zheng et al., although disaggregated transport models and microscopic traffic simulations have been applied to model route choice preferences of travelers, it is difficult to integrate information-sharing and interaction among them. Also, the executing changes on the transportation network is more difficult to do by using traditional non-agent models (Zheng et al., 2013). Those behaviors are not applicable to discrete choice models and disaggregated travel demand model, can be tackled by ABM as it was specifically developed to point this complexity (Zheng et al., 2013). Rather than applying traditional top-down methods that lack of understanding of underlying behavioral elements, a bottom-up manner in behavior-based ABMs are viable to study these complex transportation systems. The major reason is that top-down method provides a scenario-specific indicator which requires establishing a new top-down method from scratch when a change in the scenario is made. Unlike top-down methods, a bottom-up approach has the flexibility that can be reshaped for several scenarios and the possibility to observe the new emergent behavior of a new environment setup (Zheng et al., 2013). To conclude, ***"traditional top-down method studies what is the performance of a complex transportation system, whereas the bottom-up ABMS tries to understand why travelers make those decisions and how does the transportation system perform in such a circumstance"*** (Zheng et al., 2013). ***Therefore, it is viable to exhibit how a change made on the network topology can influence the route choice of traveler and to show travelers decision-making under a variety of scenarios by applying ABM.***

In conclusion, the strengths of integrating ABM to study and exhibit travelers' route choice preference, rather than traditional models include these abilities:

- To detect individual's rational and irrational behavior and preferences which are difficult to measure in traditional route choice models.
- To formulate the mechanism of travelers' complicated decision-making process.
- To consider and identify that travelers may have different socioeconomic characteristics, travel habits and thus exhibit heterogeneity.
- To capture the interaction between travelers.
- To consider that travelers may have limited or non-knowledge about the network, traffic, weather and all other factors that can affect the route choice; hence, ABM captures the vagueness of driver behavior in contrast to discrete choice models which assume they are always rational with full access to full information.
- To observe travelers response and emergent behavior to new environment setup.

2.3. Individual Human Movement

Despite having an excessive number of studies on the **agent-based modeling** (ABM) of pedestrian movement, agent-based modeling of cyclists' movement is a surprisingly underexplored topic. Therefore, besides the literature on the agent-based model of cyclists' movement, pedestrian's movement included into this theoretical research as well in order to provide a richer and sufficient vision with a multi-directional approach. Even though the complexity and the methods may differ from one to each other, the information gained from all can be a help to create a successful agent-based model of cyclists' travel behavior.

The pedestrian movement may appear to be more complex, and at the same time quite similar to cyclists' movement. Both transportation modes follow predictable movement patterns and can be identified in a simple framework of elements. As an example, according to Kempers, Borgers & Timmermans (2008) pedestrians walk forward until a choice point is reached, for example, an intersection, and this simple rule obviously applies to cyclists as well.

For both transportation modes, there are three different scales which agents' flow can be observed; micro, meso and macro. These represent respectively; "interactions between vehicles, groups of vehicles sharing common properties and flows of vehicles" (Morvan, 2012). While macro models focus on developing strategies to control flow and prevent congestion in highways, micro and meso models provide the opportunity to simulate networks in inner urban areas (Morvan, 2012). According to Reynolds (1987) level of abstraction of pedestrians modeling can be classified in by this approach:

- "*Macroscopic models* delineate the average or aggregate pedestrian dynamics by densities, flows, and velocities as functions of space and time.
- *Mesoscopic models* are in between the two previously mentioned levels, taking into account the velocity distribution. Mesoscopic models often include individual entities but model interactions between them with common fields.
- *Microscopic models* describe each pedestrian as a unique entity with its own properties."

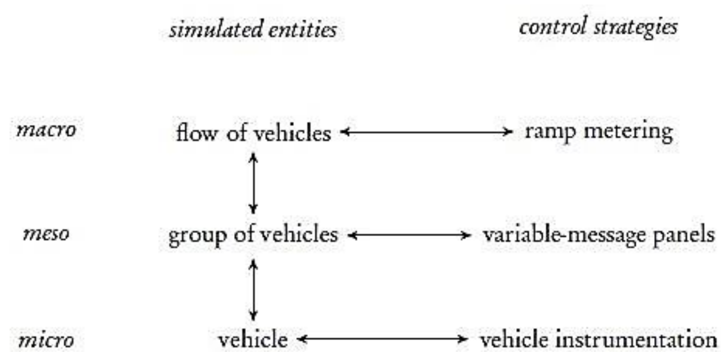


Figure 4 Scales of ABM (Morvan, 2012)

According to these model scales and reflections made before, the most applicable scale for this research is mesoscopic level. Therefore, the models elaborated in here have the same level of scale and spatial relationships.

2.3.1. From the perspective of studies on cycling movement

Papinski and Scott (2011) generated a table with the factors tested in route choice studies (Figure 5). Looking at this, it is clear that the attributes that are mostly studied in the previous studies are travel time and travel distance. According to Hood et al. (2011) and Broach et al. (2011), route choice modeling requires the identification of non-chosen routes. Several methods proposed, but only shortest-path method has been proven in large networks. The comparison between actual routes and shortest routes can reveal the attributes that may affect cyclists' route choice. On the other hand, the study (Papinski & Scott, 2011) showed that individuals have personal preferences and they selected routes that avoided congestion and that are direct.

Study	Explanatory Variables
Selten et al. (2007)	A, B
Peeta and Yu (2006)	A, B, C, D, E, F, G, H, I, J, K, L
Lo et al. (2006)	A, B
Bar-Gera et al. (2006)	G
Li et al. (2005)	A, C, S, W
de Palma and Picard (2005)	A, B
Avineri and Pashker (2005)	A, B
Eby and Molnar (2002)	A, C, F, H, M, U, V
Cascetta et al. (2002)	A, C, M, N
Lam and Small (2001)	A, D, I, O
Chen et al. (2001)	A, B, C, E, K, P, Q, R, S, T, U
Hato et al. (1999)	A, C, E, J, M

A	Travel time	I	Time of day	Q	Volume
B	Travel time reliability	J	Trip purpose	R	Directness
C	Travel distance	K	Route familiarity	S	Number of intersections
D	Toll on route	L	Route complexity	T	Habit
E	Quantitative information	M	Congestion	U	Comfort
F	Qualitative information	N	Level of service	V	En-route stops or delays
G	Number of nodes/ turns	O	Route switching	W	Average speed
H	Weather	P	Safety		

Figure 5 Factors tested in cyclists' route choice studies (Papinski & Scott, 2011)

Segadilha and Sanches (2014) evaluated the factors that might influence cyclists' route choice. The main motivation for this research was to identify the route choice of commuter cyclists to determine what types of policies or infrastructure operations may promote bicycle usage for utilitarian trajectories. The results of the questionnaire revealed that ***cyclists dislike sharing a road with motorized vehicles and this is one of the important factors that influences route choice preferences of cyclists*** (Segadilha & Sanches, 2014b) This reflection made by Segahilda and Sanches (2014b) strengthens the Factor L:Route complexity, P: Safety given by Figure 4. The situation of cyclists' preferring physically separated bike paths or traffic-calmed residential areas is supported by Buehler and Dill (2015). They also stated that if it is necessary to ride together with motorized traffic, lower and slower car speeds are important characteristics to have along the route (Buehler & Dill, 2015).

As mentioned before, GPS tracking facilitates the examination of factors governing route choice decision by capturing actual routes taken precisely in both space and time. Broach, Dill and Gliebe (2012) showed the type of a road is an important determinant of cyclists when they choose their route, by utilizing the collected GPS dataset and the survey data of 164 utilitarian cyclists. The GPS traces were matched with network streets and trajectory-level factors recorded in the surveys were used in a choice set generation algorithm. The results from this two-fold study showed that; in addition to trajectory distance, slope, and traffic volumes, cyclists appear to prefer off-street bike paths and neighborhood bikeways (aka "bicycle boulevards") (Broach et al., 2012).

Law et al. (2004), aimed to reveal the route choice preferences of cyclists by using two different datasets of cyclist movement from two years; 2003 and 2012. Based on the comparison between two different years' route intensities, ***cyclists were observed on the most direct and continuous routes over the routes which have better cycling infrastructure but are less direct*** (Law, Sakr, & Martinez, 2014). ***Directness*** can be identified as deviations from the Straight-line between a start and an end point of a trajectory; the less the deviation is the more the directness of the route (Dill, 2009; Aultman-Hall et al., 1997). As for ***connectivity***, it refers to the density of connections in path or road network and the directness of links. A well-connected road or path network has many short links, numerous intersections, and minimal dead-ends (cul-de-sacs) (Dill, 2009; Aultman-Hall et al., 1997).

These studies show that route choice data can be obtained using stated or revealed preference surveys or through direct observation (e.g., GPS tracking). One of the primary challenges facing route choice research is to empirically determine the factors influencing route choice decisions (Papinski & Scott, 2011). Determining these factors have a significant importance regarding the agent-based model created since it needs identification of the cyclists' characteristics. To apply the agent-based modeling that mimics GPS data of cyclists, the agents (cyclists) and their characteristics need to be defined (Zhang & Levinson, 2004) (More about the construction of agent-based modeling are given in the Chapter 4.) The studies (Papinski & Scott, 2011; (Segadilha & Sanches, 2014; Buehler & Dill, 2015; Broach et al. 2012, Law et al. 2014) provide beneficial inputs as to route choice preferences of cyclists. Nevertheless, all these stay at the aggregated level of movement without deepening into the individual activities. In contrast to a limited number of agent-based modeling studies of cyclists, there is an extensive amount of literature may be found on the agent-based modeling of pedestrians.

As can be seen from these, there are not many studies on the disaggregated agent-based modeling regarding cyclists' movement in particular. Therefore, the framework of the agent-based model of pedestrians is reviewed to provide a richer and sufficient vision with a multi-directional approach. Even though the complexity and the methods may differ from one to each other, the information gained from all can be a help to create a successful model of the movement behavior of cyclists.

2.3.2. From the perspective of studies on pedestrian movement

There are several elements of influence pedestrians on the meso-scale. According to Haklay et al. (2001), pedestrian movement is shaped by two elements: the structure of the road network and the location of attraction points. For the cyclists, the "road network" is one of the elements that has the major effect on the movement, as we saw from the previous literature. However, the approach on the "attraction points" may differ from commuter cyclists to pedestrians. Pedestrians may change their route rapidly depending on the attractors; for example, seeing a new shop opening may cause them to change their route plans rapidly. However, a commuter or a utilitarian cyclist who rides bike purposively does not alter their route plans as rapidly as pedestrians since they follow the traffic route together with the other vehicles.

Another important element that identifies the activity is the destinations. The trajectory can continue until the destination is reached or the person is constrained in time. According to Bierlare, Antonini & Weber (2003), there is not always a certain destination point in people's mind, but more broadly defined activities. However, ***for a commuter cyclist, there is a certain destination point decided in his/ her mind beforehand; such as a train station or a location of work***.

For pedestrians, knowledge of the area defines the route they take. However, it can be expected commuter cyclists who are most likely local people of the city, would have a general idea of the transportation network better than the visitor from the outside of the city. Therefore, the element of knowledge about the the network does not make an obvious heterogeneity between the agents (commuter cyclists) for the model of this research.

For pedestrians, the attractiveness of the streets changes based on several criteria. Having shops at the both sides or a long straight sight or a historical atmosphere is showing a high level of attractiveness to pedestrians which do not have a major effect on cyclists. For cyclists, the green environment has an appealing atmosphere to cycle.

A final element that is of influence on the movement is the walking speed for the pedestrians, which can be translated into cycling speed of cyclists. As ***the speed of the individuals defines the space-time dimension, the speed must be included in the model as well.***

2.4. Determination of the factors from literature

As can be seen from the perspective of studies on cyclists' and pedestrians' movement; there are several aspects that influence pedestrian movement and at the same time cyclists' movement, in real life. From this combined perspective, factors given on the below serve as core notions that can be implemented in the model:

- This thesis focuses on the meso-scale agent-based modeling.
- Cyclists ride forward until a choice point is reached; such as a crossing or an intersection.
- Road network is one of the elements that has the major effect on the movement.
- Road type is another major element. Cyclists prefer to ride along the cycling roads that are physically separated and tend not to choose a road if it is shared with motorized vehicles.
- Connectivity and directness qualities of a route are another major factors that influence route choice decision-making in a positive way.
- A trajectory can continue until the destination is reached or the person is constrained in time.
- For commuter cyclists, there is always a pre-defined destination point to reach.
- The speed must be included in the model as it defines space-time dimension.

These factors will be taken at hand during the GPS analysis, in order to analyse their applicability to the case study and get the results.

CHAPTER 3
CASE STUDY

This chapter provides some insight into the case and presents the relation between the thesis study and the Smart Sustainable Project (SSD) that has been applied in the study area Utrecht, Netherlands.

3.1. Study Area

Utrecht is the fourth largest city in the Netherlands which has been growing at a rapid pace *[1]. Utrecht is the capital of the province of Utrecht. The city is located at the heart of the Netherlands, at the intersection of motorways, railways, and waterways.

Utrecht has a large amount of commuter traffic which causes major traffic jams on the incoming roads during morning times, and the outgoing roads during evening times. Currently, cycling is the main mode of transport in Utrecht with %33 *[2].

---

*[1] According to Statics Netherlands, Utrecht population grew with 1.43 percent in 2015 while the national average is 0.5.

*[2] The average in the Netherlands is %27 in total.

Every day, between 7 a.m. and 7 p.m., over 100,000 cyclists ride to their work, school, university, public transport, shops or home by passing through the city center. This intensive mobility constitutes the 60% of all journeys made through the city center (ECECO, 2003).

The most important Dutch railway lines meet in Utrecht. Therefore, the Central Station of Utrecht is the busiest station in the country (ECECO, 2003). Also, several satellite towns (Maarssen, De Bilt, Zeist, Bunnik, Houten, Nieuwegein, and IJsselstein) situated close to Utrecht makes the road network more crowded.

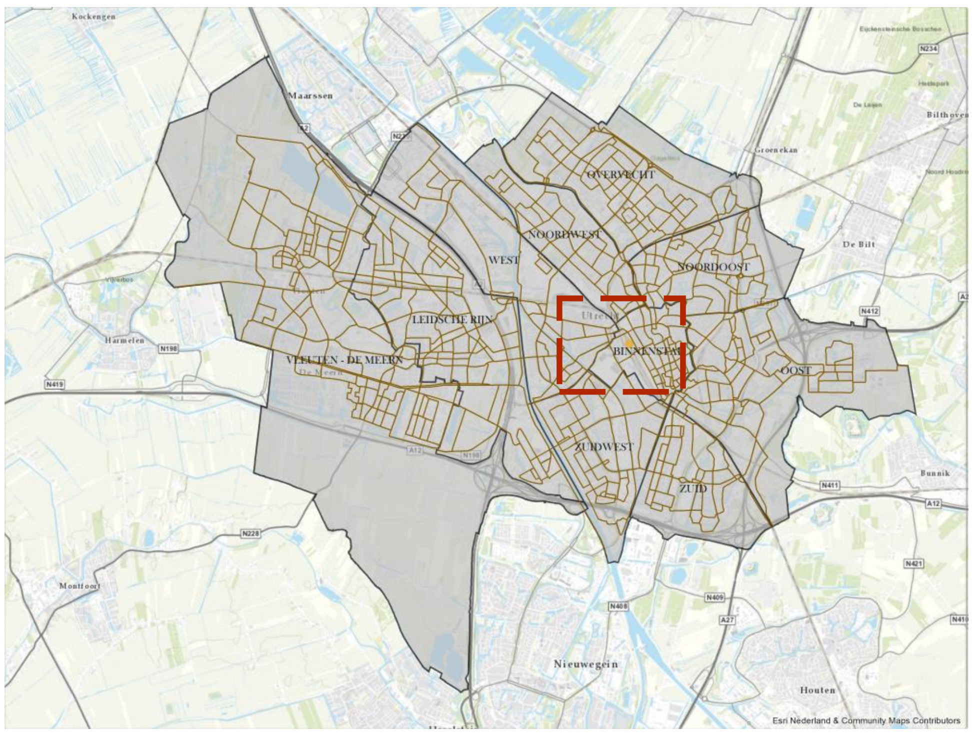


Figure 6 Study area Utrecht and its neighborhoods

The municipality of Utrecht has been making great efforts to make the city a world-class bicycle city as can be seen from the Utrecht Action Plan of 2015 – 2020 (Gemeente Utrecht, 2015).

"To make this happen, the Action Plan has the following aims;

- proving facilities to cyclists,
- making cycling more attractive,
- increasing bicycle use and stimulating the bicycle economy" (Gemeente Utrecht, 2015).

Due to Utrecht's outstanding identity in terms of cycling mobility, its desire to be a bicycle city and the goals aimed by the municipality of Utrecht; the city has chosen as a study area of this thesis study on travel behavior of cyclists. To realize these goals, the Municipality conducts projects and one of them is the SSD project that is elaborated on the below.

3.2. Relation between the research and Smart Sustainable Districts Project

The SSD Project is a Climate-KIC Flagship program working for district developments in European cities to enhance less and slower traffic, boost resource efficiency, reduce energy consumption and increase green spaces ("Utrecht Sustainability Institute," n.d.). The project focuses on provide a pleasant place to live, work, study or visit for inhabitants and the visitors of the city. Utrecht is one of the cities chosen for that project and Central Station district of Utrecht is the project area defined. The area has aimed to developed and transformed to provide a new Centre which has some attractive, alive, safe, green features. One of the priorities this project has which is the one this research focuses on is *"To provide clean and safe personal mobility".*

As a part of that priority, the bike road network is expected to be developed. In order to achieve to provide a sustainable development on the bike road network, it is crucial and inevitable to analyze existing cycling flow and understand the cycling movement as a behavior from the cause and effect relation point of view. Route choices made by cyclists forms the cycling flow, intensities, demands for future developments on the road network. ***If we understand the underlying reasons for route choices made; municipalities or responsible stakeholders can develop the networks that target the demand and provide tailor made accessibility for cyclists and therefore, maximize the supply with the limited resources.***

The aim of this research by means of its relation to the case study is to reveal the main factors that influence route choices of cyclists and to develop an agent-based model that mimics cyclists behavior as close as possible to reality, by using identified factors. The end-result model of this study may provide limited usage in terms of its applicability to the case study since the creation of such a model possibly requires more time than 5-months due to the complexity of cycling behavior and very limited number of studies done in this field. As can be imagined, the initial foundation of the model has a major importance in the development of such complex and advanced model. Therefore, the goal is to provide this "initial foundation model" which further developments can be built upon it. In this manner, the model should have the fundamental elements of cycling movement and has to have flexibility characteristics in order to be enhanced in future researches. Therefore, the end-result of this research will not be applied in the SSD project directly; however, it will provide a re-usable approach to create such a model and the initial foundation to be developed.

Data availability has a major importance in order to realize the research objective and practice the case study. For this research, GPS data which had been measured the GPS traces of cyclists during one week will be made use. More details about the data are given in Chapter 5; Results: GPS Data Analysis.

CHAPTER 4
METHODOLOGICAL FRAMEWORK

This chapter partially tackles with the Research Question 2 and 3:

- ***RQ2: What is the added value of measured GPS data in terms of developing the agent-based model?***
- ***RQ3: How to formalize the determined factors in order to implement them in the agent-based model?***

The chapter illuminates the methodological framework with the steps that will be taken. The aim of these steps and the important dependencies between them such as how one relates to another are also mentioned. The framework provided here serves a repeatable approach that can be applied by other studies for the purpose of developing an agent-based modeling of cycling movement.

4.1. Methodology

The workflow depicted in Figure 7 represents a simplified and summarized visual representation the methodological framework. The numbers adhered to these steps shows their implementation sequence.

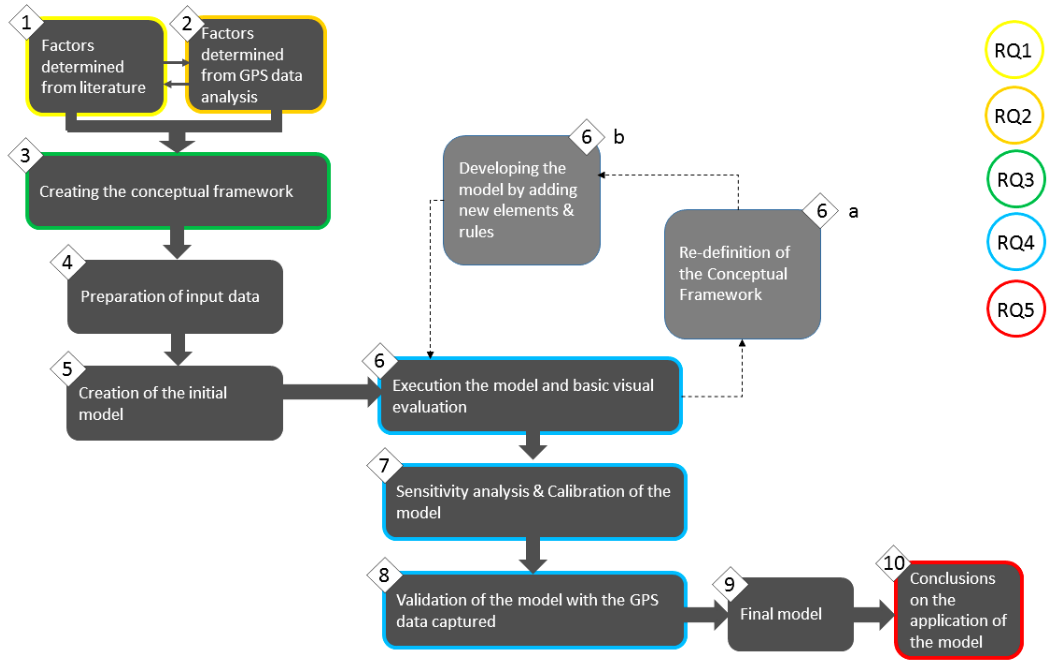


Figure 7 Workflow of the methodology

Firstly, the conceptual framework of the model needs to be defined with the inputs from literature and GPS data analysis. In Chapter 2, some of its elements have already been determined from the literature (1); and they will be checked in GPS data analysis in order to see their applicability to the user group of this study (2). Also, some more elements will be added by the use of the measured GPS data (2). By the use of these two inputs; the conceptual framework of the agent-based model will be created and agent-based model methodology will be started (3). According to the conceptual framework, the input data that will be prepared in order to make the required datasets ready to load into the model (4). Followingly, the model will be created by implementing the conceptual framework in the agent-based modeling toolkit by coding the model and loading the prepared datasets (5). As now the model is created, it can be run to check its performance (6). The model will be run and agents' activity and their movement on the simulation will be visually checked (6). Depending on the occurring errors and the activity of agents, it will be decided if the model needs to be developed more i.e. framework needs to be updated and new element need to be implemented (6.a, 6.b) The outcomes of this iterative process will again be checked and visually validated until it is decided that model can continue to the next step (6). Following to this iterative process, sensitivity analysis and calibration of the parameters will be executed (7). At the final, the model will be validated by the use of empirical data (8). The validated model will be presented as the final model (9) and conclusions on the applicability of the model will be provided at the end of the report (10). As can be seen, this overall methodology fits research questions and their sequence as well (Steps are bordered with the same colors of their corresponding research questions).

4.2. Software

This section will briefly discuss the chosen software that will be used to execute the steps of methodology.

The most important choice in terms of software is for the modeling. Within the several toolkits available for agent-based modeling purposes, NetLogo is chosen for this research. The outstanding features of this toolkit are:

- It is free and open source.
- Includes its own programming language which is simpler to use than Java and C++ (Zheng et al., 2013)
- Provides a comprehensive documentation on the model creations and tutorials, and a large library of models which also includes two examples on traffic simulations
- More websites are available that provides user experiences
- Allows using shapefiles as input data.

Besides the model creation itself, there are more software required to be used. For the data analysis of GPS, ESRI's ArcGIS is decided to be used to do spatial data analysis and geo-referenced visualizations of the results from GPS and the model. Other than ArcFigure, QGIS provides some functions that can be useful i.e. topology checker to fix problems related to the road network. Other than these two, MatLab is mainly used for data handling between several datasets and analyses based on attribute tables. As this software was never used before, it was quite a challenge to learn how to use it in a limited time period. Even though it took time and it required major effort to make; it has been seen that once it is learned, it is an efficient tool to handle data, make the data more concise, make analysis between tables in a short time which may take hours to do in the ArcFigure's attribute table.

4.3. GPS Data Analysis

The main source that will give the main inputs of the agent-based model is the measured GPS data of cyclists. This data shows the revealed preferences of cyclists and as such allows us to analyze the underlying reasons of the preferences made. These underlying reasons or as it is called 'factors' will be included in the model as the parameters of cycling agents and their bike road network environment.

To do this, GPS data needs to be matched with the network roads at first. Following to that, the mobility pattern should be visualized in order to understand the flow of cyclists on the network. By using their individual traces, intensely used roads, in other words intensely preferred roads, will be derived. This intensity map output exhibits the revealed preferences made by cyclists in the real bike road network. The aim of the model is to give such results that shows similar patterns and values to the intensity map of GPS data. Therefore, this GPS data intensity map and its statistical values will be used throughout the GPS analysis and the next steps of the methodology (sensitivity analysis, calibration, and validation) in order to calibrate and validate the model.

As the comparison data is obtained, the next step in the GPS analysis phase is to identify the factors that gives such effects. For that, previously identified elements and factors from the literature will provide a start. Therefore, as a starter, the travels made by individuals need to be obtained as routes between the origin and the destination locations. The time the travel started and the average speed the cyclist has will also provide further information about the characteristics of the travel and the cyclists. Next is to check environmental factors. As determined from literature, there are three main environmental factors: Road type, Connectivity and directness of a road and shortest-path between start and end location of a trajectory. Even though these factors applied by many studies as parameters that influence the cyclists route choice preference, that does not prove that the user group of this study has the same characteristics and influenced by the same factors. Thus, the factors needed to be confirmed by the GPS data. By cross-checking these factors by the use of GPS data, the factors can be adjusted in a way that it fits to the user group of the study.

Moreover, the user group can be divided into sub- user groups based on their similar preferences within the sub-group and differences between the other sub-groups. However, it is important to note that it is not sufficient to only identify the routes that are differently chosen by sub-groups; the underlying reasons in other words factors that affect the cycling group to choose that roads are required. A survey that provides information on the socio-demographical characteristics of the cyclists and/or the objective of the travels made may provide more than enough data on grouping, however, for this research such information was not available.

4.4. Definition of Conceptual Framework

This section partially answers the Research Question 3:

- ***How to formalize the determined factors in order to implement them in the agent-based model?***

The final version of the elements will be determined based on the results of GPS data analysis (Chapter 5).

As the main elements the factors that affect the route choice are derived from literature and measure GPS data itself, a conceptual framework can be created by using them in a holistic approach. Creating a conceptual framework is the starting point of building an agent-based model. These elements need to be defined in order to have a conceptual framework for the model:

- the main objective,
- initial elements,
- outcomes of interest (A. J. Heppenstall et al., 2013). This can be presented schematically as an input/ output procedure.

Given on the first bullet, the objective is the major determinant when defining the conceptual framework of an agent-based model (A. J. Heppenstall et al., 2013) and thereby, it must be defined clearly. For this thesis, the main objective of model is that;

✓ ***To mimic cyclists' route choice preference on Utrecht bike road network***

For that; the main concept is to;

- ✓ ***Simulate cyclist agents' route choice preferences between their trajectories start and end locations on the bike road network environment***

For the second bullet, there are two sets of initial input elements:

- o Structure of the urban space (environmental context)
- o Individual cyclists (agent and its parameters).

The model will take these two main inputs and will generate output or action according to a set of behavioral rules defined. Details on these elements are given on Section 4.4.1: Implementing the factors.

Due to practical use of the model explained in the Chapter 3, the model should enable to apply interventions within the model. Generally, these are the interventions that are made for urban planning and design (van der Spek et al., 2009):

- Adding a road segment to an existing road network
- Removing a road segment from an existing road network
- Changing the road type of a road
- Changing the quality of a road (e.g. making it more cycling friendly).

This states that the model should be able to execute these changes. Fisrt of all, the model should have a network of roads that resembles the actual bike road network in Utrecht in order to execute these interventions. The output of the model will show a number of travels made over each road segments (as it called 'links') of the bike road network and these output should exhibit as close results as possible to the output of GPS data.

4.4.1. Implementing the factors

According to Borgers and Timmersmans (2005), people ride along the road and decide at crossings which road to go along based on the characteristics of a road. This decision-making translated can be into our study as factors and their effect on the probability of a road to be chosen. That means every identified factor has an effect on the probability of a road to be chosen, thereby needs to be engaged with a probability included route choice decision.

4.4.1.1. Environment

Regarding the aimed scope and performance of the model, the bike road network is the main environment of the model. It must be structured as nodes and links to enable agents to make route choices between roads (links) while changing a road. A road between two crossing point will be represented as "a polyline" which is structured from several links. Crossing points will be represented as "nodes" that connects these polylines. In by this approach, links can be removed from the model's network or new links can be added by connecting nodes.

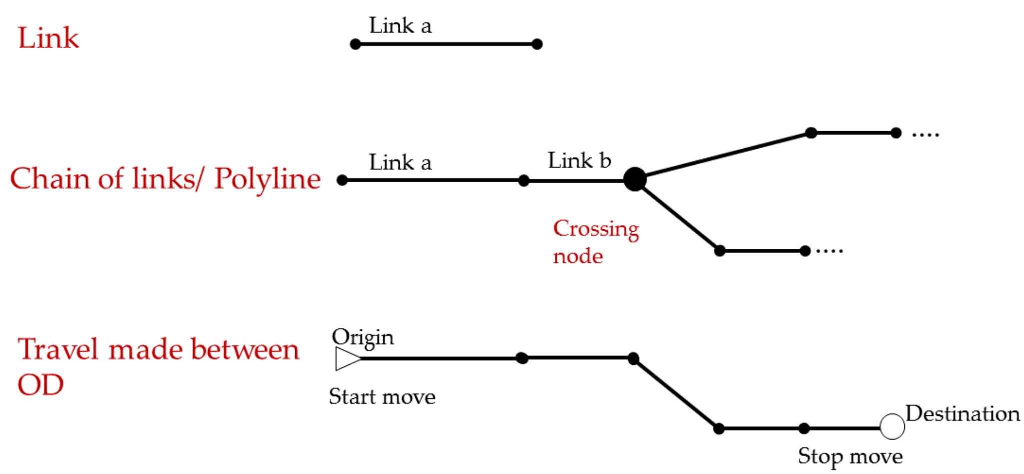


Figure 8 Key terms of route choice modeling

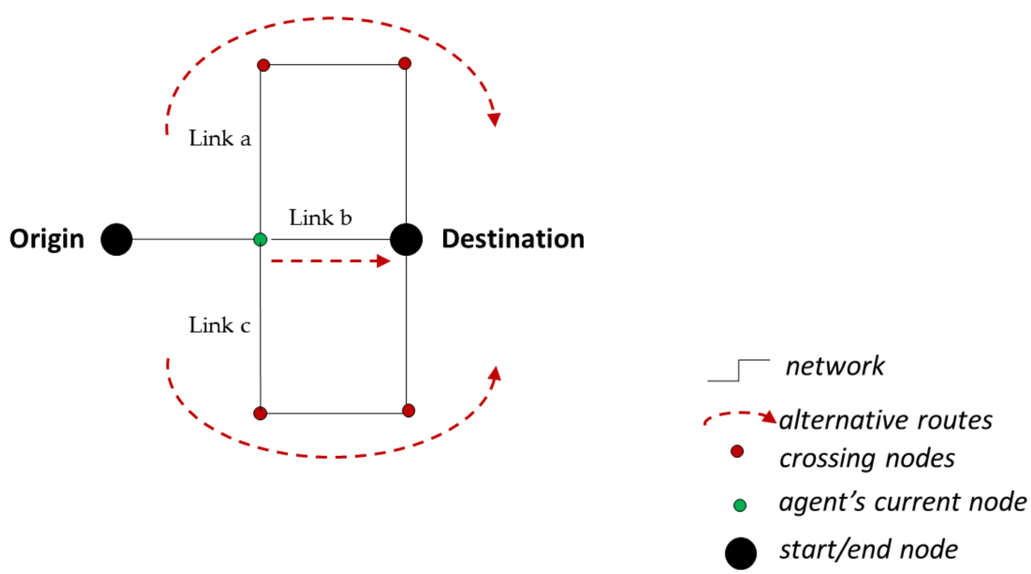
Based the literature review and GPS data analysis, the factors that affect cyclist route choices are defined. Now, the environmental factors will be assigned on each link as probability parameters and each link will get a value based on its characteristics.

Environment: Shortest-path Distance Parameter

First parameter is the shortest-path. The probability of a link to be chosen is affected by the shortest-path distance between one's travel start and end point. The start and end points must be agent dependent.

An important note regarding applying shortest-path parameter in an agent-based modeling is that its binary characteristic; e.g. if the link is 'at the shortest-path'; the risk in this is that it is difficult to determine the sensitivity of the model. Because of its binary characteristic, the effect of the parameter is too high (1 or 0). This was a result obtained from the model when the shortest-path included as a binary element. As an agent-based model is mainly developed for also incorporating user preference in which the decisions are quite subtle, this is not the desired model structure.

To prevent this drawback; the application of shortest-paths is developed. In this method, when the agent stops at a node on the network to choose the next link to ride along; it evaluates the alternative links in relation to each other, if it has e.g. three alternative links (link A, link B, link C) from its standing node. It calculates three shortest-paths while using each of the links as a waypoint, and weights the links based on their shortest-path distances to the destination node. For example, this leads to a result that while using Link a, the distance to the destination node is 700 meter; while using Link b the distance is 500 meter, and finally using Link c is 800 meter. Based on this distances the agent can determine the probability weight for each link for this element by dividing each of the distances by the max or min distance. By using such a dynamic calculation for the parameter, its binary effect is eliminated.



Environment: Straight-line Distance Parameter

The second parameter is about the direction of the roads. Even though such a parameter was not given in the literature, it is obvious that a cyclist needs a sense of direction in order to reach its destination location. It is not realistic to ride through an opposite direction of destination, when the goal is to reach the destination. As the agents do not have a sense of direction operates in their 'mind', such parameter needs to be included since it also affects the probability of a link to be chosen e.g. it is a higher probability to choose a link in the direction of the destination location, than in the opposite direction of the destination location. This direction knowledge can be calculated by measuring the straight-line distance from the end point of the next chain of the links (the polyline) to the destination (Figure X, measurement method b). The reason of not calculating it directly from the next link is that the straight-line distance from the next small links to the destination did not differ much as links are very short and close to each other by means of spatial location. However, the end location of the polyline differs for all and therefore the distance differs more (Figure 9). As a result, the shorter the distance to the destination, the probability weight of next link gets higher in terms of straight-line distance parameter.

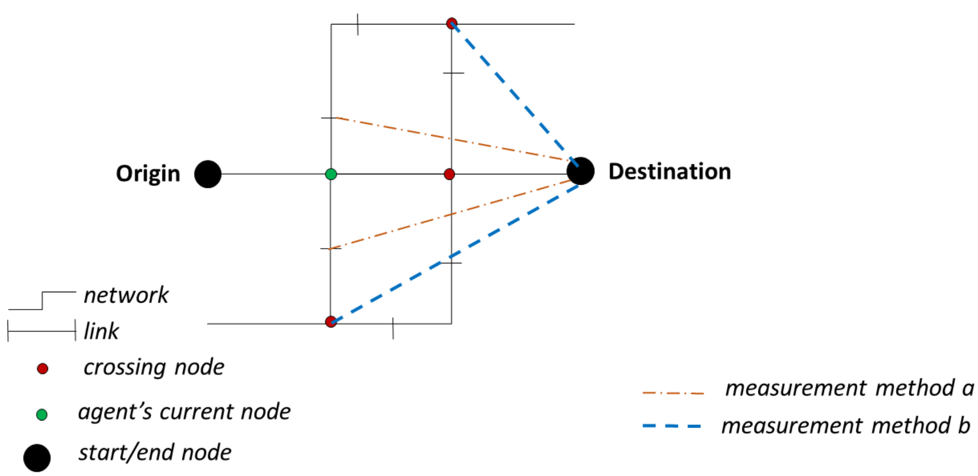


Figure 9 Straight-line distance parameter concept

Environment: Road Type Parameter

The third parameter is related to infrastructural features of the road network which also have an influence on the probability of a link to be chosen; e.g. road type. What is meant by this term is the roads that have shared traffic with motorized vehicles and the physically separated bike paths. Therefore, the road type parameter will be assigned on the environment, so on the links. Another environmental parameter is the directness and connectivity of a road which also again be assigned as on the links. All these parameters need to be analyzed on the GPS data itself, to see their influence on the cyclists' route choices made.

Environment: Boundary

Another thing to decide on the environment is the extent of it. The extent of the environment should be decided by checking the activity area of cyclists from GPS. The network should contain the area that includes all the origin and destination locations and the traveling activities made between them.

4.4.1.2. Agent

The cyclists in the real world will be represented by moving entities as called agents (See 2.1.) The agents of the model will have the same mission; start cycling from the start location until reaching the pre-defined destination location.

Agent: Number and ODs

For cyclists, there is always an end location pre-defined in one' mind (Chapter 2). Therefore, agents should have a start and an end location assigned. Since almost, always there is more than one possible route between the start and end location; every trajectory consists a route, a choice, a route choice decision (Usyukov, 2013). Every route choice contains a decision-making between alternative roads. The agents in the model will ***ride on the bike road network and stop at every crossing node to decide which one of the neighbor links to go (Figure 10)***. The chance of each road to be chosen will be calculated by the use of a probability formula, which calculates the probability of a road (link) to be chosen based on the parameter values assigned to that link. High probability means, there is a high chance of that link to be chosen; while low probability means there is a low chance of that link to be chosen by an agent.

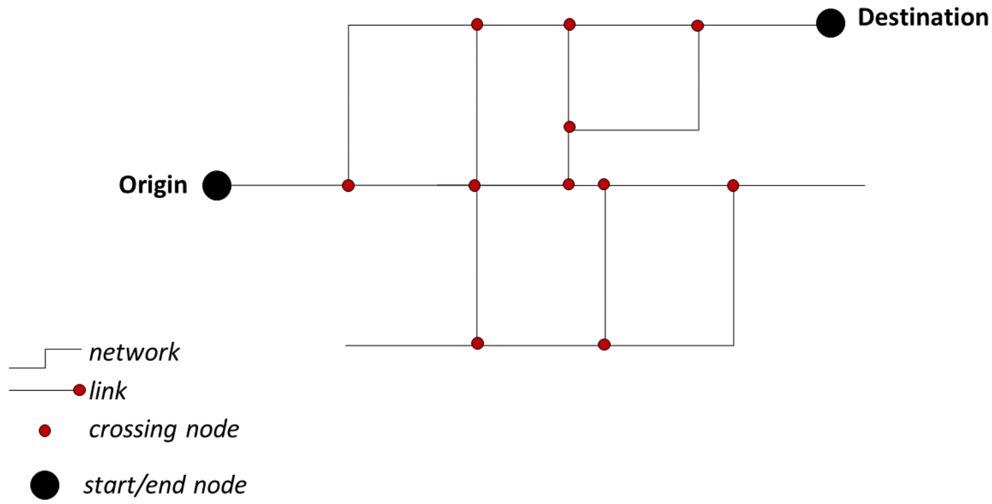


Figure 10 A schematical of a network and its fundamental elements for route choice

Between the start and the end location of a trajectory, the crossing nodes through the network represent "moments" to make a choice on the route. Therefore, crossing nodes should be related to the decision-making process.

Agent: Speed Parameter

As speed defines space-time dimension of a traveler, speed needs to be defined for each of agents as well. As there is no differentiation made between agents, one speed can be defined for all. As it will represent the all, it is better to use an average rather than using very low or very high speeds. This will again be defined via the GPS analysis. This does not have an influence on the probability of a road, but only its traveling space and travel duration.

Agent: Travel starting time parameter

Travel start time of agents also needs to be defined based on the time distribution in the real life. In the model, agents will represent the cyclists flow between 7 am and 7 pm. The release time interval of these agents will be calculated according to the travel time of cyclists in the GPS data.

A travel continues until the end point is reached. The travel and therefore the decision-making between the probability of links will end in the simulation when each agent is reached to the destination location.

These parameters assigned on agents and on the environment will together govern the execution of the movement in the model.

4.4.2. Probability Formula

"When an event occurs in a random way, such as the toss of a coin or the roll of a die, we cannot predict the exact outcome of any single occurrence of the event; but we can often predict the probability that a given outcome will result. Probability is, therefore, the mathematical expression of chance."

(Tuckerman, 2011)

In order to avoid agents to go in tails; probability is included in the model. Every time an agent comes to a crossing node at the network, it has to make a decision to choose one of the next links. This decision-making will be based on the probability formula which calculates the probability of a link to be chosen by the agent. The link which has the highest probability value will most likely be chosen by the agent. On the below, the details of this calculation is given.

As explained, there are three dynamic parameters that affect the route choice decision of a cyclist: direct-angle distance, shortest-path distance and road type which is translated to connectivity and directness of a road. As each of these parameters affect the route choice decision, that means each of them has an effect on the probability of a link to be chosen. In other words, each parameter determines an individual probability value for a link. The sum of all three individual probability values composes a total probability of a link (Equation 1). That total probability value is the probability of an agent *a* to choose link *e*. This total probability of a link can also be seen as between 0% and %100 chance.

Equation 1 Calculation of total probability (1)

$$\boldsymbol{P_{a_e} = P_{Da_e} + P_{Sp_e} + P_{Rt_e}}$$

P_{A_e} = total probability of agent *a* to choose link *e*

P_{Da_e} = probability of link *e* based on *straight-line distance* parameter

P_{Sp_e} = probability of link *e* based on *shortest-path distance* parameter

P_{Rt_e} = probability of link *e* based on *road type* parameter

The details regarding this probability equation, an example case and its implementation in the model are given in Section 7.1.

4.3. Evaluation; Sensitivity Analysis and Calibration

4.3.1. Sensitivity Analysis

The weights given to the parameters affect the route choice decision, and thereby, the outcome of the model. Therefore, it is important to know the impact of changes in the model structure and parameters on the results to be calculated; and this process can be identified as "Sensitivity Analysis".

The goal of the sensitivity analysis is threefold:

- To gain insight about how patterns are generated in the model,
- To examine the robustness of this pattern and emergent properties. (Broeke, van Voorn, & Ligtenberg, 2016)

There are several methods to apply sensitivity analysis. The performance of available methodologies should be compared before deciding which one to apply in this research. According to Broeke et al. (2016), there are three well-known methodologies that can be considered to apply sensitivity analysis for agent-based models. These are; OFAT (one factor at a time), regression-based sensitivity analysis and Sobol' method. These have proven to be popular in terms of sensitivity analysis methods since they can be applied with limited statistical and computational efforts straightforwardly (Broeke et al., 2016). Broeke et al. (2016) applied these three methodologies in a case study and presented an evaluation on the methods. In light of these results, OFAT method is chosen to apply for the model of this research. According to Broeke et al. (2016), OFAT is recommended as being an effective method to gain insight into the mechanism of the model and the patterns that the model produces. By changing one parameter at one time, the effect of individual parameters can be observed. It also shows the robustness of the patterns to changes made. However, this method is not applicable to see interaction of parameters. This drawback of the method can be eliminated through calibration implementation. Therefore, the OFAT method will be applied during the sensitivity analysis. When one parameter's value is changed, others will be kept at 0,5 in order to not to affect the results in a positive or negative way.

Another point to consider when to apply sensitivity analysis is the characteristic/type of the model. As this agent-based model is probabilistic, not deterministic, the results can differ every time the model runs. Therefore, making a conclusion based on results gets complicated. To solve this issue and get a clear understanding on the effects of change in values, the simulation will be run multiple times (100 times). The results (number of travels made over links) will be obtained as shapefiles. These shapefiles will be merged and the average values are calculated by dividing the sum of a number of travels by using a created model in ArcFigure model builder (Appendix 1). In by this approach, the model can be interpreted based on more steady values and the disadvantages of having results from a high probabilistic model can be prevented.

Because of the high probabilistic nature of the model, small changes do not give visible results such as 0.01 steps. Such small changes are difficult to interpret also because it can be a result of the probabilistic nature of the model as well. Therefore, during this analysis steps of 0.1 will be used to change weights of the parameters between 0 and 1.

As the method is defined, which input values will be included in this analysis must be decided. Parameters which do not have any effect on the model do not need to be included in this analysis and will not be calibrated since any change made in them will not change the outcome. For this model, these parameters do not change the outcome of the model:

- Release time of agents:

  The agents represent the cyclists flow between 7 am and 7 pm. It provides a real-like traffic flow and density of the cyclist traffic on the network according to time intervals. This time parameter does not affect the outcome of the model (total number of travels made over a link) as the model does not end until all agents reach their destinations.

- Speed of agents:

  Similar to the release time of agents, the speed also provides a real-like visualisation of cyclists' flow and does not affect the outcome of the model. The speed can be of influence of the model when a time limit is defined for the model run, for example 1 hour. This is discussed in Chapter 10.

- Number of agents:

The number of agents does influence the number travels made in total and over links. Increasing the number of agents means adding more OD pairs into the model, that means more traveling agents on the network. This changes the total number of travels made and at the same time, number of travels made over links. Therefore, the ratio between them stays the same. More about this topic will be discussed in Chapter 6.

The parameters which affect the outcome of the model are the ones which will be included in the total probability formula (Equation 1):

- Shortest-path distance parameter
- Straight-line distance parameter
- Road type parameter

The sensitivity analysis of these parameters will be explained in Chapter 7.

4.3.2. Calibration

With the knowledge gained from the sensitivity analysis, the parameters of the model will be fit to the accepted measurements; and this process can be identified as calibration. The main goal of the calibration is to make the results of this agent-based model most similar to the GPS data, consequently the real life. Again, to minimize the randomness in the results; 100 runs will be collected for every calibration model and the average of the values in that batch will be calculated. The average value will be presented as the results of that corresponding calibration. The results will be calibrated by comparing two aspects of the model results with the empirical data results:

1) Intensity pattern on the bike road network: The total travels made by 350 agents will be logged as a number of travels made per link and the results will be depicted on the Figures. From green to red color scala will represent the intensity classes from lowest to highest. These figures will be compared with the empirical data intensity map (Section 5.2.; Figure 11: Intensity map of GPS) through checking the visual intensity pattern and the intensity classes. The similarity of the model result to the GPS data shows the parameters' are adjusted in a way that the simulation mimics the sample group cyclists' route choice preference correctly -as similar as possible-.

2) Travel time distribution: During the runs of the every calibration, travel time of each agent will be logged which represent how long it took for each agent to travel between its origin to destination location in minutes. The same information will be obtained from GPS data to make the comparison. To have the travel time distribution of cyclists from the GPS data; distance traveled by each cyclist will be measured by summing the length of the links they traveled and this distance will be divided with the speed they had. These graphs are produced by using MatLab as it was quite time-consuming work to do it in ArcFigure or Excel. Even though the author did not have any experiences in the software, the results and the knowledge gained worth the effort being made.

4.4. Validation

The final step of the agent-based model is to visually and statistically validate the model. For that, the measured GPS data will be made use. As the aim of the validation is to show that the developed model could predict comparable real world situations, using the same data measures for calibration and validation is not a reliable method. Therefore, the measured GPS dataset will be split into two halves and one half will be made use for validation purposes. By using this dataset visual and statistical validation will be executed, by using three different techniques: Intensity pattern distribution on the network; Quantification of the difference between two intensity outputs and Time travel distributions of agents comparing to real cyclists.

Firstly, the modeled intensity distribution on the network will be visually validated by the revealed intensity distribution of empirical data. To emphasize on the differences and similarities between two outputs in terms of intensity patterns, the modeled intensities will be subtracted from empirical intensities and, therefore, the validation will be quantified on the network.

Addition to validations made from the aspect of the intensity pattern modeled, the travel time distributions of the model will be validated as well. By comparing travel time distributions of cyclists trajectories measured in the GPS data and cyclist agents in the model; the traveling duration and travel time distribution during a day in the model will be validated. The high similarity between two outputs will prove that the model performs highly similar to real cycling flow on the network by means of daily movement of cyclists.

CHAPTER 5
RESULTS:
GPS DATA ANALYSIS

The main goal of this chapter is to answer Research Question.

- ***RQ2: What is the added value of measured GPS data in terms of developing the agent-based model?***

Further integration of GIS in the domain of transport modeling can significantly contribute to more effective model results.

Loidl et al. (2016)

Agent-based models linked to GIS have the great competency to research complex systems such as traffic flows.

Gimblett, (2006)

The studies (Loidl et al., 2016b; Zheng et al., 2013) state the importance of utilizing the measured mobility data obtained from the field for the aim of developing agent-based models which are linked to the real world factually. Measured GPS data will present the revealed preferences of cyclists and as such will be used as the main input for the development of the agent-based model to mimic cyclists route choice preferences.

5.1. Data Acquisition

In this research, the open dataset of the Dutch event Fietstelweek is used. The concept of yearly event is to collect GPS traces of cyclists at national scale. The first time this national event took place was in 2015, between the dates of 14th and 20th September. To be a participant, the mobile app of the event was required to be installed and must be enabled while riding a bicycle. By this approach, the trajectories made by participants were logged in the app as GPS traces, during one week.

Image 3 Fietstelweek app

As the captured GPS data was made public by the data provider (Keypoint), it was relatively easy to retrieve it from their website. The downloaded data was not raw data; it was pre-processed. To prevent to reveal data that may exhibit some personal info, privacy measures were applied, e.g. first 500 m and last 500 m of trajectories were cut out from the data. This was not stated in the description of the dataset; it was found out during the data analysis.

The preprocessed data was divided into two files. One of them was in the table format providing all the trips made by all cyclists with the trajectories they traveled from their origin to destination. The table does not include any spatial reference of GPS traces but only the link numbers (unique numbers given to the road segments) constitute those trajectories. Each user was saved with a unique number (route ObjectID) and for each user more than one link was logged which shows the links (road segments) that the user was traveled through. The day of a week and the time that the trajectories made are given per every link, separately (For the details; Appendix 2: Routes.csv).

The second part was saved as a ESRI's shapefile. That shows a network consists of the links traveled. Within the attribute table, the unique numbers of these links was again listed. As it can be imagined, a link was traveled by many cyclists, and a cyclist was traveled through many links to reach his/her destination. (For technical details; Appendix 2: Links.shp). For a better understanding on the given data structure, simplified representation of these two files are schematically given on the below.

Routes.csv file (1)

RouteID	LinkNumber	Day	Hour	Speed
1	10	Monday	8	16
1	11	Monday	8	16
2	11	Tuesday	9	12
3	12	Wednesday	8	20

Link.shp file (2)

Objectid	Shapefile	LinkNumber
1	Polyline	10
2	Polyline	11
3	Polyline	12

Other than these, no other information was collected during the event such as demographical or socio-economic characteristics of cyclists. This aspect of the data is discussed in Chapter 10 Discussions & Recommendations.

5.2. Analysis 1: Intensity map of GPS

5.2.1. Obtaining the data for the case city and the focus area

As the event of Fietstelweek was including whole Netherlands, the data was containing the traces made in the whole country. As the case study area of this study is Utrecht, the municipal area of the city is selected and exported as separate data that only contains the travels made in the city of Utrecht. The same was made for the table data as well; by relating the table to shapefile and then selecting the Utrecht area; the trajectories that are only located in this city is seen. These rows are exported as Utrecht routes table and made ready to be used throughout the further GPS data analysis. Unfortunately, it was not possible to join these two files because of the cardinality issues. Such function would make the analysis more powerful in terms of individuals' route choice exploration.

As the case area comprises of SSD project area, the further data analysis is better to focus on the movement flow that occurs around that area. The Central Station (CS) of Utrecht is the main emitter and the attractor of the commuter cyclists' daily mobility in the city and it is located in the SSD project area as being a very important element of the project. By focusing on the cycling movement that occurs through the CS, we can capture the movement around the case study area. This is achieved by firstly relating the routes table to the network that is formed by GPS links and then, selecting the Central Station's polygonal receiving area (500-meters radius circle) on the GPS polyline network and exporting the routes that were used one or more of the selected links within their trajectories. In by this approach, not only the movements that were happened at CS area but also the ones that were traveled around/ though CS are captured (This gave 6936 trajectories made through CS area). By this approach, the analysis can be focused on one specific group of cyclists; the ones who make trajectories through Central Station area and the results of the analysis can provide more useful inputs for the SSD project. These CS routes are again related to links shapefile, and same steps are implemented to see the number of travels made over links but this time not all the trajectories used; only the focus user group's which are the ones traveled through the CS.

5.2.2. Frequency Calculation and Creation of the Final Network

The next step was to understand the spatial pattern of cyclists' movement. For that aim, the number of travels made over road segments needed to be obtained. Such data provides the ratio of a road segment being preferred by this cyclist group. Henceforward, the output of this data will be called as 'Intensity map of GPS'.

To obtain Intensity map of GPS, the road segments cyclists were traveled and how many times they were traveled should be known in order to Figure the intensities occurred. To derive these from the collected GPS data these steps are followed:

First a frequency calculation is made in the Utrecht routes table (5.1. Data acquisition) to count how many times each road segment is traveled Following to that, the output table which has the counts of links used in the all trajectories are visualized by using this count field, on the network that is created by the road segments traveled in GPS. This allowed to having an insight on the intensity pattern, however, not well enough due to the low quality of the network structure. As the GPS dataset already processed by the data holders (Keypoint), it was matched with the Utrecht road network. All roads of Utrecht can be seen on the network, however, not completely but partially because of the very high amount of unconnected links. It was displaying a very low network quality. The road segments are not connected to each other on many spots and many other topological elements are detected (Appendix 3). Also, the size of the data was unnecessarily big because of the way the network constructed. The links between nodes were divided into too many segments which were making the data big and complex. Moreover, no road type classification was assigned on those. Due to the low quality of the network, it was time-consuming to match or intersect it with another network in order to transfer some required information. Using such network in the data analysis and in an agent-based model would not be practical and can cause many errors.

Given the explained reasons, instead of using that GPS' network; another network was decided to be used. As the spatial extent of this study is the bike road network of the city of Utrecht (Section 3.2) the bike road network of Utrecht is obtained from the TNO Database which was created by OSM and used by Utrecht Municipality. As it only shows the bike road network, unnecessary roads (the ones which do not allow bikes) are excluded as it was aimed. The network and topological structure are checked in order to prevent having dead-ends and unconnected links on the network which would cause several data issues in the further analysis and in the model creation. Unconnected road segments and dead-ends are eliminated by utilizing topology and network tools in QGIS and fixing those spots by hand. After eliminating topological errors, some data clearance is made on the network layer to delete unnecessary fields from the network table such as municipality name, road code to make the data compact and clear as much as possible. After fixing several errors, the network is created in a smooth data shape.

Finally, the calculation of how many times a road segment is traveled is inserted to the same road segments of this network. By this approach, the same road segments had the same intensity class.

The output figure; intensity map of GPS, provides the link intensities which shows the revealed preferences of the user group on the bike road network. This figure will be used several times during the data analysis and in the next steps of the methodology, as it shows the revealed preferences of cyclists and the desired output of the model. The data provides which links are preferred at most on the bike road network of Utrecht by the cyclists who traveled through CS area, during one week (Figure 11).

For the validation purposes, same methodology is followed for the half of the GPS dataset, and the output data and the figure will be utilized to validate the model (Figure 36).

5.2.2.1. Visualization of the Intensity

The result is visualized using natural breaks classification in five classes. Classification matters since the grouping of the data into classes is one of the fundamental aspects of mapping. By utilizing most accurate classification method, the data can be made easier to be understood than raw data. The aim of using natural breaks during sensitivity analysis, calibration and also validation phase is to clearly visualize the differences between classes and between result maps. In this classification, class breaks are identified based on similar value groups and maximum differences between these groups. In other words, the method minimizes within-class variance and maximize between-group variance. Where there is a relatively big difference in the data values, a boundary is set. As checked from the histogram of the values; the statistical distribution of the values has a variance and sometimes even gaps. Therefore, this method gives an accurate representation with a choropleth Figures of trends in data (J, Goodchild, & Longley, 2015).

Figure 11 Intensity map of GPS data: The figure shows how many times a road segment was used by the cyclists who traveled through CS area. Therefore, the figure exhibits the revealed prefences of the user group of this study.

By looking at this Figure 11, the activity area of this cyclist group's movement can be depicted. As can be seen, the activity of these cyclists have not distributed over whole bike road network. The user group was traveled on the 10 km radius area while the whole network provides on the 14 km radius area. Based on this, it has seen that it is not necessary to model whole network since some parts of the network were not used by these cyclists. In by this approach, it can be prevented to model a wider area which means losing the focus on the study area. As this model aims to focus on the movement around CS, it would not be practical. To sum up, the extent of the environment should be decided by checking the activity area of cyclists from GPS. The network should contain the origin and destination locations and the trajectories between them.

Other than that, the Figure 11 clearly shows which routes are prefered at most. The links around the Central Station has the highest intensity. This can be a result of the physical structure of the network; railways and Central Station divide the network into two parts as east and west side and, therefore, it limits the access and densifies the mobility on the bridges that connect these two sides. Other than this railway area, the main roads can be defined as well by looking at this figure. This pattern is the ultimate result of the route choice preferences cyclists made, which are based on several factors. As now the visual and statistical comparison data is obtained; now the underlying factors can be analyzed.

5.3. Analysis 2: Time Distribution of the travels

Other than the activity area and the route preferences made by cyclists, it is important to know the activity time of cyclists; time distribution of travels made in time intervals. For that, the number of travels made throughout a day is calculated in percentages by calculating the average of the measured travels made over a week.

Table 1 Number of travels made throughout the day, in percentages (0- 24 hours)

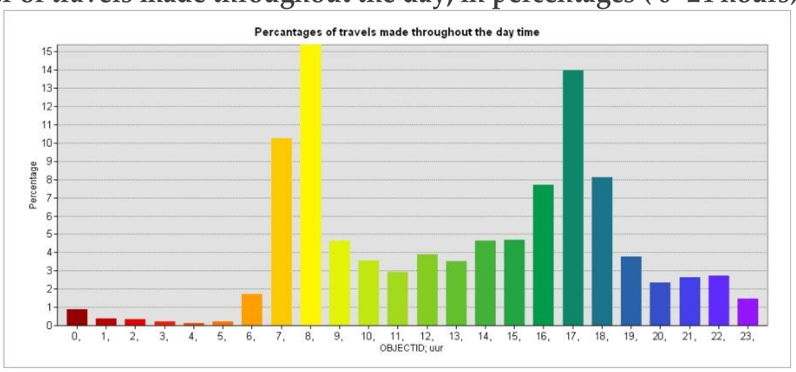
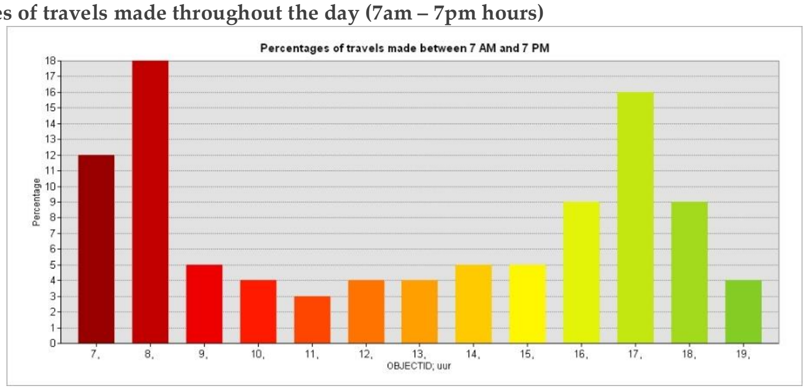


Table 1 shows that cyclists that go over CS area were mostly active between 7 am and 7 pm. This pattern shows that most of them leave their 'origin point' between 7 am – 8 am probably to go to work or to school and come back between 16 pm and 18 pm.

Table 2 Percentages of travels made throughout the day (7am – 7pm hours)



On the above, the percentages of travels made between 7 am and 7 pm throughout a day is represented by the Table 2.

5.4. Analysis 3: Speed Distribution of the travels

Speed is another element that should be analyzed regarding traveling behavior, as the speed of the individuals defines the space-time dimension. According to the data of the cyclists travel through CS, the mean speed is 18 km/hour. As there is no differentiation will be made within the agents, they all can be adjusted to one speed which is the average (18 km/hour).

5.5. Crosschecking the factors from literature with measured data

From literature, several factors are determined that can influence cyclists' route choice preference. By using the measured data, these factors can be checked in order to see how the user group of this study was affected by these factors.

5.5.1. Analysis 5: Road type and Connectivity & Directness Qualities

Road type was one of them that is related to the network. Another factor related to the network was the connectivity and directness qualities of roads. As connectivity refers to the density of connections in path or road network, it can be concluded that the connectivity of a road gets higher in direct proportion to the level of the road; e.g. major roads are more connected than bike paths. At this point, one factor states that major roads are effect cyclists route choice preferences in a positive way, while the other one reflects cyclists do not prefer motorized vehicle shared roads. As these two characteristics (road type and connectivity) are conflicting with each other in this case, it should be known which one affect this study's cyclists group more in order to obtain the underlying factors correctly.

Figure 12 Level of Service for Connectivity Factor analysis

Figure 13 Road Type Factor analysis

Given by the Figure 12, and 13; the intensities showed higher classes at the greater level of service roads, which provides higher connectivity. By other words, user group was preferred to ride along the roads which are shared with motorized vehicles over the only bike paths.

As for another quality of a road network that influences route choice preferences of cyclists; directness can be identified as the deviation of a road between the start and end point of a travel; which means cyclists prefer to ride along the direct routes. The less the deviation is the more the directness of the route. To check the directness quality with the intensities revealed; a subset of origin and destination pairs are obtained and straight-lines between each OD pair is created (Figure 14). The similarity between the Straight-line pattern and intensity pattern showed that this factor indeed influence the user group in a positive way.

Figure 14 Directiveness Factor Analysis

5.5.2. Analysis 6: Shortest-path between Origin and Destinations

According to literature, cyclists tend to choose the shortest-path between their origin and destination point. For that, a network analysis is executed by using a sub-set of the origin and destination pairs on the network and shortest paths are created between these ODs and a number of paths created on the links are compared with the intensity map of GPS. The similarity between these numbers showed that the shortest path parameter was indeed an influencing factor for the user group of cyclists.

5.6. Determination of the factors from GPS data

In conclusion, the group of cyclists (6936) are chosen based on their trajectories' spatial relation to the CS area which is the focus area of this study. Following to that, their revealed route choice preferences on the Utrecht road network are obtained. This obtained data will be made use during the further steps of the methodology as it provides visual and statistical information of the route choice preferences made by the user group.

This data also shows the activity area of the user group, which can be utilized to identify the model environment in further steps. Moreover, time distribution of the travels made and the speed variation of the cyclists are acquired which will be useful inputs to identify cyclist agents in the model.

Lastly, the measured GPS data is made use in order to check the applicability of the factors determined by the literature. It has been seen that negative affect of motorized vehicles was not directly applicable to the user group of this study, as it was reflected in the literature under the category of road types (motorized vehicle shared road type, and only bike road type). The analyses results showed that, the user group was influenced by the connectivity and directness qualities of roads, rather than the road types. Since the motorized vehicle shared roads provide higher connectivity and directness; cyclists preferred to ridde along these type of roads to the contrary of the literature. Other than that, the shortest-path distance factor was indeed one of the factors that influenced the route choice preferences of cyclists in our case.

CHAPTER 6
RESULTS: ABM PREPROCESSING & IMPLEMENTATION

As given by the Section 4.4., agent-based models consist two main inputs;

1. Environment in which represents the urban space of the agents in the modeling structure
2. Agents in which represents the moving entities.

Given by the scope of the model; environment will represent the bike road network of the city of Utrecht while agents will represent the individual cyclists the bike road network.The model aims to mimic cyclists route choice preferences on the given bike road network, by using measured data. This chapter presents the preprocessing steps to prepare the model implementation.

6.1. Input Data Creation: Environment

The network created in the Section 5.2.2. will be used to prepare the environment of the model. To make agents decide at each crossing point; crossing points are created as nodes. Other than that, to provide a smooth movement process and simulation in the model; it is better to have one link between two nodes. As the roads between two crossing points were divided into several segments in the shapefile, the network is edited manually (Figure 15).

Figure 15 Bike road network from Utrecht Municipality

It should be noted that at the later model development process; this node-link creation from the network is added to the model as an extra function of the model. Therefore, it will not be necessary to do this editing manually for the future studies, and they can use different network shapefiles more easily. However, to decrease the data size of the network, the approach mentioned remains as a useful tip that can applied by researchers. After all, the smaller the size of the input data, the faster the model operates. Other than that, there some more aspects that should be taken into account when deciding on the raw network shapefile which are elaborated under the Section 5.2.2. Lastly, to enhance the quality of the network and minimize the possible errors of the model; the operations explained (Section 5.2.2.) can be seen as fundamental steps to apply during the preprocessing phase of agent-based models.

6.1.2. Environment Dependent Parameters

As now the network is structured as nodes and links, the attributes of the environment can be defined in it. The only parameter that is ***directly dependent on environment*** and should be assigned on the network's links is ***the road type parameter*** (Section 5.5.1). The road type categories analyzed in this study were bike paths which are only open bikes traffic and shared roads which have shared traffic with motorized vehicles. This was the approach followed to provide a ground to differentiate road types, however, it can be improved by using these generalized categories. Nevertheless, the road types used in this model provides the fundamental differentiation and flexibility to be detailed, as it was targeted in Section 3.2.

For this parameter, bike paths are given a value of 0,5 while shared bike roads are given 1. The reason of giving such values is related to the results of GPS data analysis. As shared roads were preferred more than bike paths, the value of shared roads must be given higher in order to make those types of roads be chosen by agents more. More about the probability of a link to be chosen and calculations are given under the Section 7.1.

Other than road type parameter, there are two parameters that are ***related to environment: Shortest-path distance and Straight-line distance***. However, these cannot be assigned on the network links with static values as the road type parameter; since these parameters depends on the dynamic calculation of distance measurements. As the distance measurements for these parameters will be calculated through the network; distance calculation is made possible through the network, by identifying measurement method on the network. As these two parameters are dynamic, they are not assigned as values on the links as the road type static parameter.

The last thing to do for the environment is to define the ***boundaries***. The original boundaries were the municipal area of Utrecht. Given by intensity map of GPS (Figure 11), the outer periphery of this area was not traveled through by the user group at all and they were concentrated in a smaller area close to city center (The area size of the whole network was around 80 km^2, and the activities of the cyclists were distributed over around 68 km^2 area). Depending on this information, a polygonal boundary shapefile is created that surrounds this smaller area.

6.2. Input Data Creation: Agents

Agents are the moving entities of the model which represent the individual cyclists in this model. To create agents, these questions should be answered at first:

- How many agents will be created?
- Where will these agents be located at the initialization phase of the model?
- Where will these agent move forward?

All these questions are in fact can be answered by one concept: Origin and Destination pairs. Every traveler in the real world has an origin point where they started their trajectories, and the destination point where they finished their trajectory. Hence, every agent needs an origin destination pair. As the distance between each OD will be traveled by one agent; a number of OD pairs created will be equal to the number of agents needed.

The total number of OD pairs of the travels made through CS was too much to include in the model by means of practicality and model performance and model run durations. Therefore, it needs to be decreased to more feasible number. While decreasing the number of ODs to implement in the model as agents' ODs, there is a crucial aspect to be taken into account; the OD distribution in the measured data must be kept the same in order to implement the intensity distribution over the network similarly and realistically.

To realize such an approach, a method is developed which includes a data processing with a number of steps. The details of the implementation of this method will be shortly presented in here.

To decrease the number of points at the same time keeping the distribution over an area the same, there are two points to think on:

- The distribution of the points is probably not even.
- The decreased number should be related to the original number on factual grounds, to represent the reality.

Taking into account these points, it is decided that a ratio should be chosen to decrease the total number, not a number; e.g. 20% of the points will be chosen. To choose the 20% of the points, the distribution should be divided into equal sized areas (a fishnet grid is used for this research) due to the reason given by the first bullet. Now, all the points (origin and destinations are applied separately) are grouped in equal sized areas. The number of points falls into these grids may not be equal, as the distribution is not even, however, it is not a problem as a ratio will be used calculate the number of points. By dividing the total number of points fall into per square with 20; the decreased number of points are obtained (This number was 350 for this research). To make sure these points are located on the network links, the points are chosen manually. The process is visualized for the origin points by the figures given on the below, from the Figure 16 to Figure 20 (For bigger sizes of the figures, Appendix 4).

Figure 16 Origin points of trajectories made over Central Station (original number)

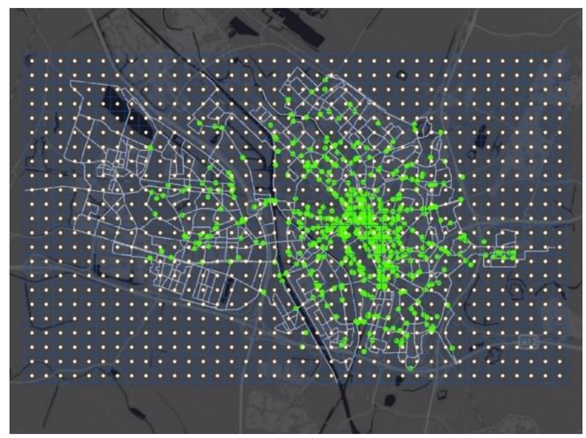


Figure 17 Fishnet grid created to divide points

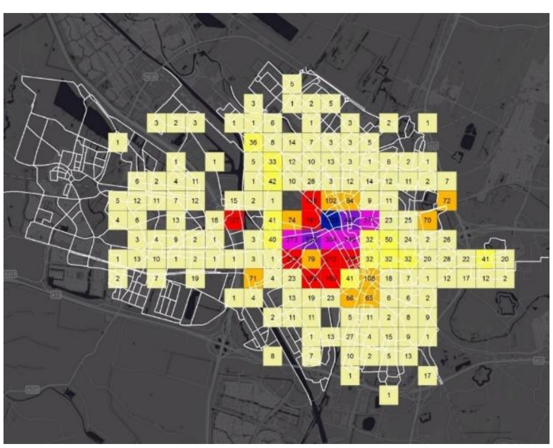


Figure 18 Number of origins fall into squares

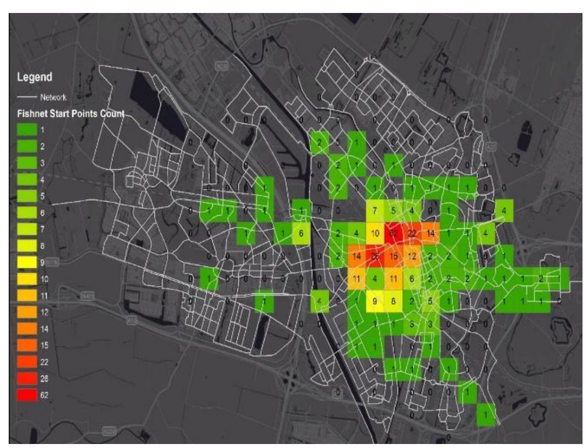


Figure 19 20% of the number of origins

Figure 20 Decreased number of origins (350)

Based on the outcome number of OD pairs, the same number of agents are created in the model; 350 agents to travel between 1 OD pair and OD locations are assigned to corresponding agents. As now the agents are created, next step is to define the agent-based parameters.

6.2.1. Agent Dependent Parameters

One speed value can be assigned to all agents as there is no differentiation made in that senseAs one value will represent the all, it is better to use an average rather than using very low or very high speeds. The mean value of the speeds logged into the measured data within the cyclists traveled through CS is 18km/hour. Such value may be seen as a low speed, however, this can be the result of the dense traffic on the roads that are traveled by this cyclist group. Therefore, the speed of 18km/hour assigned as the speed value of each cyclist.

Travel start time of agents also needs to be defined for each agent. By applying the time distribution of travels made throughout a day into the model, the movement of cyclists can simulate a more realistic traffic flow. Given by the Table 2 (Section 5.3.), the travel start time of the 350 agents are defined accordingly; in which together represents a traffic flow between 7 am and 7 pm. By this approach, the simulation exhibits a denser traffic during peak hours and more calmer traffic during off-peak hours.

Other than these parameters, the agents are programmed to move until they reach their pre-defined destination location, and once they arrived, they 'die' in other words disappear.

With this pre-processing phase, both two elements are created and identified with the parameters in their possession. Herewith, the model is made ready to be executed and adjusted in order to give results that is as similar as possible to the measured data.

CHAPTER 7
RESULTS: PROBABILITY FORMULA & SENSITIVITY ANALYSIS

Chapter 7 and Chapter 8 aim to answer Research Question 4. This chapter will elaborate the probability formula and the sensitivity analysis aspects of the targeted question.

- ***RQ4: How can the simulated behaviors and output trends be evaluated and validated?***

7.1. Probability formula

As given in the Section 4.4.2., the probability formula is generated for the model which calculates the probability of a link to be chosen depending on the parameter weights the link has. As each parameter effects the probability of a link, each parameter determines an individual probability value to be added to the total probability.

The calculation of these individual probabilities is given on the below (Equation 2; Eq. 3; Eq. 4; Eq. 5). By using these equations (Eq. 3; Eq. 4; Eq. 5), it is possible to limit the individual probability by the given weight to its corresponding parameter.

Equation 2 Calculation of total probability (2)

$$\boldsymbol{P_{a_e} = P_{Sl_e} + P_{Sp_e} + P_{Rt_e}}$$
$$\boldsymbol{= \left\{W_{Sl} \times \min\left(Sl_{node_n}\right) * \frac{1}{Sl_e}\right\} + \left\{W_{Sp} \times \min\left(Sp_{node_n}\right) * \frac{1}{Sp_e}\right\} + \{W_{Rt} \times Rt_e\}}$$

Equation 3 Calculation of individual probability for straight-line distance parameters

$$\boldsymbol{P_{Sl_e} = \left\{W_{Sl} \times \min\left(Sl_{node_n}\right) * \frac{1}{Sl_e}\right\}}$$

P_{Sl_e} = probability of link e based on *straight-line distance* parameters
W_{Sl} = Weight given to the *straight-line distance* parameters
$\min\left(Sl_{node_n}\right)$ = Minimum *straight-line distance* within the possible links from node n
$\frac{1}{Sl_e}$ = 1 / *Straight-line distance* of link e

Equation 4 Calculation of individual probability for shortest-path distance parameters

$$\boldsymbol{P_{Sp_e} = \left\{W_{Sp} \times \min\left(Sp_{node_n}\right) * \frac{1}{Sp_e}\right\}}$$

P_{Sp_e} = probability of link e based on *shortest-path distance* parameters
W_{Sp} = Weight given to the *shortest-path distance* parameters
$\min\left(Sp_{node_n}\right)$ = Minimum *shortest-path distance* within the possible links from node n
$\frac{1}{Sp_e}$ = 1 / *Shortest-path distance* of link e

Equation 5 Calculation of individual probability for direct road type parameters

$$\boldsymbol{P}_{\boldsymbol{Rt}_e} = \{\boldsymbol{W}_{\boldsymbol{Rt}} \times \boldsymbol{Rt}_e\}$$

P_{Rt_e} = probability of link *e* based on *road type* parameter
W_{Rt} = Weight given to the *road type* parameter
Rt_e = *Road type* of link *e*

For example; if the weight given to the straight-line distance parameter is 0,5 (W_{Sl}); the link which has the minimum straight-line distance to the destination (($\min(Sl_{node_n})$) will has a probability of 0,5 in terms of this parameter (P_{Sl_e}); and the other alternative links will have P_{Sl_e} value less than 0,5. Same applies for the probability calculation of shortest-path distance parameter as well.

As for the probability of road type parameter(P_{Rt_e}), the calculation will be calculated by multiplying the road type value (Rt_e) with a weight given. Depending on this result of the GPS analysis given by the Section 5.5.1; bike paths are given a value of 0,5 and shared roads are given a value of 1. By this values, the probability of these road type links will be calculated accordingly. For example; if the weight given to the road type parameter is 0,4 (W_{Rt}); P_{Rt_e} of the shared road type links (0,4 * 1 = 0,4 = P_{Rt_e}) will be higher than P_{Rt_e} of the bike path road type links (0,4 * 0,5 = 0,2 = P_{Rt_e}). By this approach, the probability of agents to choose shared roads type links in which provide higher connectivity and directness qualites will be made higher, as it was observed from GPS data.

By implementing this formula, the programmer can have the control on the model by adjusting the weight of every parameter, which means adjusting the effect of every parameter on the route choice decision of agents. These weights will be determined in the calibration phase.

As an example calculation, a sample case is illustrated on the below with three possible links, their values calculated for every parameter tor and three example weights given to the parameters (Table 3).

Table 3 Sample case with three alternative links

	Link a	Link b	Link c
Straight-line Distance *(Sl)*	100	60	80
Shortest-path Distance *(Sp)*	140	120	100
Road Type *(Rt)*	0,5	1	1

Example weights:

W_{Sl} = 0,5

W_{Sp} = 0,3

W_{Rt} = 0,4

Based on the equations given on the given on the above (Eq. 3; Eq. 4; Eq. 5), the probability values of each link for each parameter is calculated (Table 4).

Table 4 Sample case individual probabilities for every link

	Link a	Link b	Link c
$\boldsymbol{P_{Sl_e}}$	0,3	0,5	0,375
$\boldsymbol{P_{Sp_e}}$	0,21	0,25	0,3
$\boldsymbol{P_{Rt_e}}$	0,2	0,4	0,4

As can be seen here (Table 4), the link which has the minimum straight-line distance got the maximum probability value on that parameter, which equals to the determined weight of the parameter; 0,5. Also, the link which has the minimum shortest-path distance got the maximum probability value on shortest-path parameter, which again equals to the determined weight of the parameter; 0,3. And finally, the links which are auto and bike shared (high connectivity and directness) road types got the maximum probability value on road type parameter, which equals to the determined weight of the parameter; 0,4.

As explained by the Equation 1, the total probability of a link to be chosen by an agent will be calculated by summing up all three individual probabilities of a link (Table 5).

Table 5 Total probability for every link

	Link a	Link b	Link c
$\boldsymbol{P}_{\boldsymbol{Sl}_{\boldsymbol{e}}}$	0,3	0,5	0,375
$\boldsymbol{P}_{\boldsymbol{Sp}_{\boldsymbol{e}}}$	0,21	0,25	0,3
$\boldsymbol{P}_{\boldsymbol{Rt}_{\boldsymbol{e}}}$	0,2	0,4	0,4
$\boldsymbol{P}_{\boldsymbol{a}_{\boldsymbol{e}}}$	=0,71	=1,15	=1,075

These probabilities per a link show the probability of an agent to choose one of these links. If needed these probabilities can be translated to percentages by dividing the probability value by the sum of all probability values of alternative links (Table 6).

Table 6 Probability of links translated to percentages

	Link a	Link b	Link c
$\boldsymbol{P}_{\boldsymbol{a}_{\boldsymbol{e}}}$	=0,71	=1,15	=1,075
Percentage	% 24	% 39	% 36

According to Table 6, the agent will choose Link a with a chance of %24; Link b with a chance of %39; and Link c with a chance of %36. The sum of all the percentages will be equal to 100 (%100).

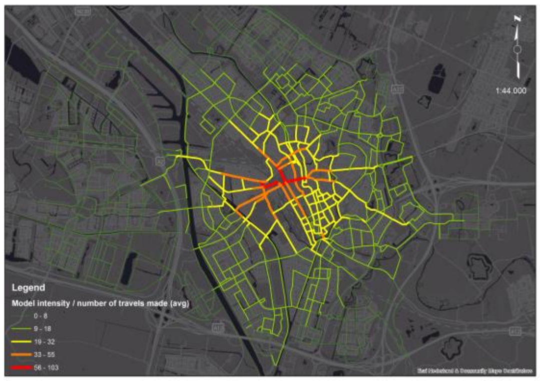
7.2. Sensitivity Analysis

In the Section 4.5.1. the method and the parameters that are going to be used in the sensitivity analysis were explained. For every parameter, a concise information on each parameter and their probable effect to the model are provided. Followingly, results of applied weights are exhibited for each corresponding parameter. For bigger size representation of the results, please refer to Appendix 5.

1. ***Shortest-path distance parameter (*** $\boldsymbol{P}_{\boldsymbol{Sp}_{\boldsymbol{e}}}$ ***)***

Given by the Section 5.5.2., the similarity between the intensity map of GPS data and the shortest-paths between ODs was captured. Therefore, the shortest-path distance is included in the model with having a high importance in order to mimic GPS data. A value of 0,10 makes the importance and effect of this parameter very low. Thereby, the sensitivity analysis is made between 0,25 and 0,90.

Results:



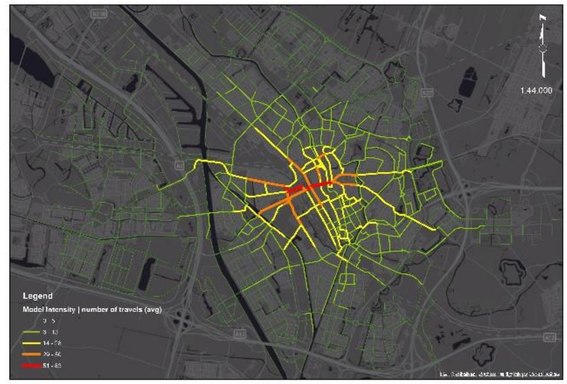
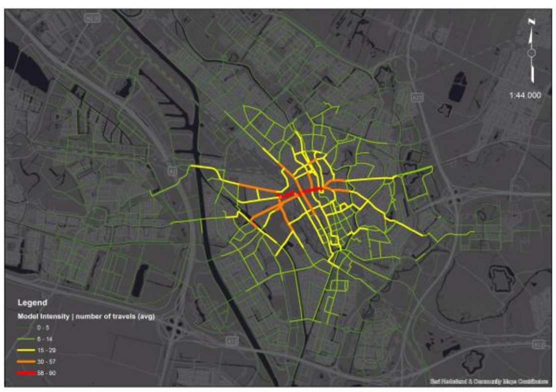


Figure 21 Sensitivity analysis; shortest-path distance with 0.25 **Figure 22 Sensitivity analysis; shortest-path distance with 0.5**



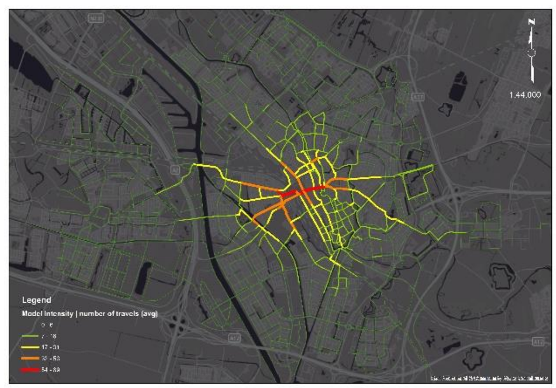


Figure 23 Sensitivity analysis; shortest-path distance with 0.75 **Figure 24 Sensitivity analysis; shortest-path distance with 0.9**

As can be seen from the above figures, a lower value of shortest-path parameter creates a widespread on the network while a high value of 0,9 makes a clustering around the Centrum area. Also, the maximum number of travels made for the value 0,25 is 103 while for the value 0,9 it is 89. That means a low value as 0,25 made agents travel longer times in order to reach destination since a low value makes the effect of shortest-path distance parameter less. A value of 0,90 keeps agents' density on the Centrum area as similar to the GPS data. By looking at these results, it can be said that high values (e.g.; 0,75 and 0,90) provide more realistic results.

The results around the Centrum area are not very different from each other. The links around the Central Station has the highest density. This can be a result of the physical structure of the network; railways and Central Station divide the network into two parts as east and west side. In this case, if an agent has an origin point on the west side and destination point on the east side; it has to travel over those roads that connect to two sides. Therefore, it can be expected that these connective links will give the very similar results in every time as having the highest density. Another conclusion is that; the links that is next to the connective links are also giving very similar results in the model outputs. This shows the agents that traveled over connective links on the railway area are scattering to next links gradually, as in real life. This steady results around the Centrum area proof the robustness of the model in this area, which is a good sign by means of the model performance.

By looking at these results it can be said that a value that is higher than 0,90 can cause to affect the model too much based on shortest-path distances, while a value lower than 0,50 may cause to have too much scattering on the network. Therefore, steps of 0,10 between the values of 0,50 and 0,90 will be used for this parameter during the calibration phase.

2. ***Straight-line distance parameter (P_{Sl_e})***

As the agents do not have a sense of direction, such parameter needed to be inclıded in order to provide agents a direction knowledge. This parameter acts as a direction supporter to prevent agents to go to opposite direction of their destination. By this approach illogical trajectories and non-realistic movements around the links without getting closer to the destination are aimed to be minimized. However, in some cases straight-line to the destination may not be the best direction to go because of the meandering physical structure of the network paths. Hence, the effect and value of this parameter should not be very high such as a value of 0,90. On the other hand, a value of 0,10 makes the direction support very weak. Therefore, sensitivity analysis is made between 0,25 and 0,75.

Results:

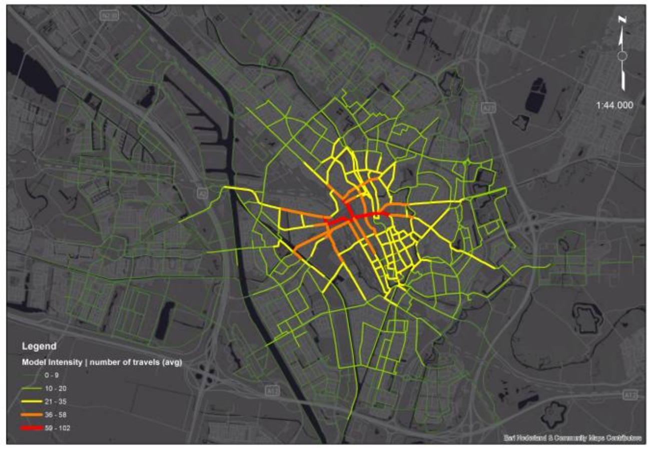


Figure 25 Sensitivity analysis; straight-line distance with 0.25

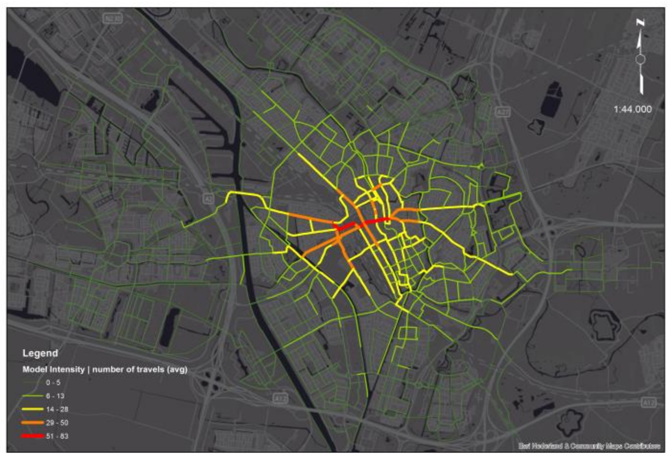


Figure 26 Sensitivity analysis; straight-line distance 0,5

Figure 27 Sensitivity analysis; straight-line distance 0,75

Given by the Figures 25, 26 and 27; a low value of 0,25 created more spread on the network than higher values such as 0,50 and 0,75. This is because agents do not know which direction to go, or they do not have any sense of right/ wrong direction. The only parameter that increases the probability of links that reach to destination points is shortest-path direction parameter. However, as explained before, the model has a probabilistic nature and one parameter is not enough to make agents go towards their destination point. Therefore, a value of 0,25 is not sufficient to get realistic results. The value of 0,50 and 0,75 gave very similar results which represent realistic patterns on the network. Therefore, values between 0,50 and 0,75 will be used for the calibration.

3. *Road type parameter (P_{Rt_e})*

Presented by the Section 5.5.1., user group was preferred to ride along the roads which are shared with motorized vehicles over the only bike paths; this decision was analyzed under the qualities of connectivity and directness. Depending on this result of the GPS analysis, bike paths are given a value of 0,5 and shared roads are given a value of 1. By this values, the probability of these road type links will be calculated accordingly.

As the greater part of the Utrecht bike road network is comprised from motorized vehicle shared bike roads; having a very high value like 0,90 for road type parameter can cause to increase the probability of a great majority of network's links and give a scattered pattern. On the other hand, a low value like 0,10 can cause to neglect a majority of the network's links. Therefore, the sensitivity analysis is made between the values of 0,25 and 0,75. These values are enough to understand the effect of the parameter.
Results:

Figure 28 Sensitivity analysis; road type with 0.25

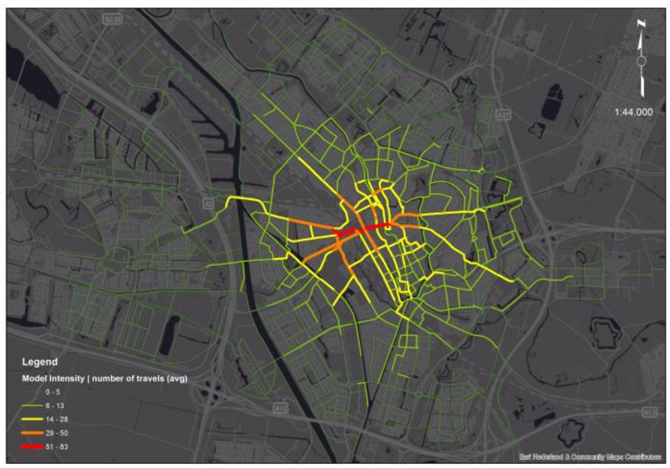


Figure 29 Sensitivity analysis; road type with 0.5

Figure 30 Sensitivity analysis; road type with 0.75

The road type parameter with a value of 0,75 caused a widespread on the network. On the other hand, a low value made cyclists cluster around Centrum area which is a more realistic result. Also, as presented by the Figures 28, 29 and 30; the road type parameter is the least sensitive parameter over all three parameters; the results are not extreme. Therefore, steps of 0,10 would not be effective for calibration. In conclusion, values of 0,20 and 0,40 will be used for this parameter during the calibration.

CHAPTER 8 RESULTS: ABM CALIBRATION & VALIDATION

8.1. Calibration

Given on the below matrix shows which values are checked during the calibration phase (Table 7) (For the method followed Section 4.5.2.). Within the report, only five of the calibration results are given and they are represented in an order from bad to best results. The reason of this is to provide concise yet clear and complete understanding about the calibration process; without confusing the reader.

As explained in Section 4.5.2., the results are obtained in two types:

1) Intensity pattern on the bike road network: These figures are compared with the empirical data intensity map (Figure 11: Intensity map of GPS) through checking the visual intensity pattern and the intensity classes. Similarity of the model result to the GPS data shows the parameters' are adjusted in a way that the simulation mimics the sample group cyclists' route choice preference correctly -as similar as possible-.

2) Travel time distribution: Travel time distribution of agent that shows how long it took for each agent to travel between its origin to destination location is compared with the travel time distribution of cyclists. The result of every calibration is compared with the same information obtained from GPS data. Due to the fact that the total number of cyclists in the GPS data (6936) and the agents in the model (350) are not the same thereby could not be compared individually but in percentages. Thereby, cumulative percentages of travels are calculated for each calibration result; which shows how many of the agents are traveled so far e.g. uptill 10th minutes of the simulation run 60% of the agents are reached their destination; while for GPS data it is 80% of the cyclists. By this approach, the results become easier to read and more explanatory. The graphs which show the travel time distribution according to a fraction of travels are given in Appendix 6.

Table 7 Calibration matrix

Parameter	Calibration values		
Shortest-path distance ($\boldsymbol{P}_{\boldsymbol{Sp}_{\boldsymbol{e}}}$)	0,90	0,80	0,75
Straight-line distance ($\boldsymbol{P}_{\boldsymbol{Sl}_{\boldsymbol{e}}}$)	0,70	0,60	0,50
Road type ($\boldsymbol{P}_{\boldsymbol{Rt}_{\boldsymbol{e}}}$)	0,40	0,20	

Firstly, the shortest-path distance parameter is calibrated by checking the values of 0,90; 0,80 and 0,75. The given Figure 31 and Graph 1 on the below present the results of the Model 1 in which these values are used:

Table 8 Model 1 calibration values

Model 1:	Shortest-path distance ($\boldsymbol{P}_{\boldsymbol{Sp}_{\boldsymbol{e}}}$)	Straight-line distance ($\boldsymbol{P}_{\boldsymbol{Sl}_{\boldsymbol{e}}}$)	Road type ($\boldsymbol{P}_{\boldsymbol{Rt}_{\boldsymbol{e}}}$)
Calibration values	0,90	0,70	0,20

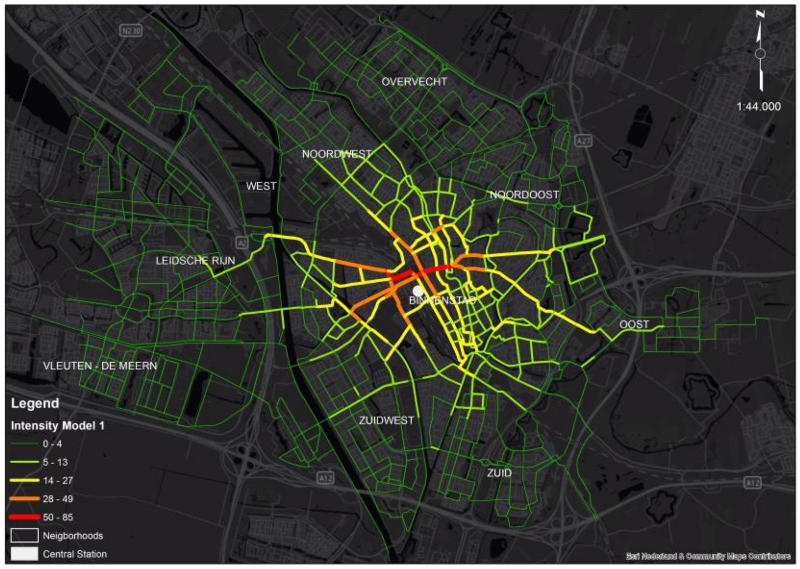
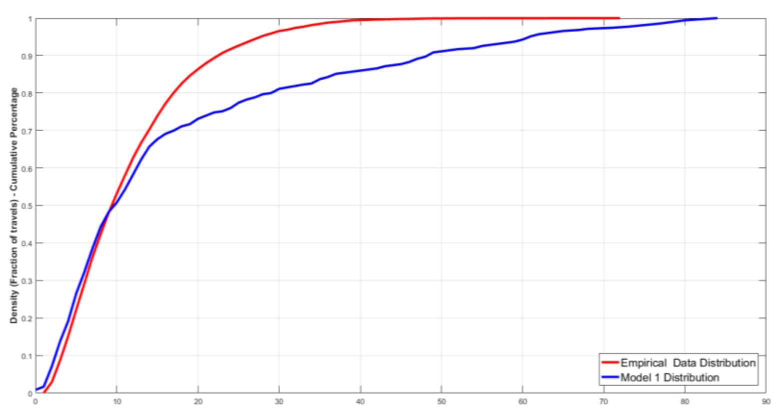


Figure 31 Calibration Model 1

Given by the Figure 31, the first and second highest density classes of agents represent realistic results in the Model 1. However, the spread over the network is too much more than GPS data. The peripheral area of Centum should have a less density than it results in this model's output.



Graph 1 Travel time distribution graph, Model 1

Looking at the travel time distribution Graph 1, it can be seen that the travels in GPS data are completed in 70 minutes while for the model it took around 85 minutes. As the OD locations of the agents are not located as dispersed as the GPS data have in reality (Section

6.2.); the result should have been in the opposite way. That shows the agents spent more time than realistic amounts to find their destination locations, which means the parameters were not effective enough at these weights. As the most of the OD pairs are densely located on 4km radius area, most of the travel should have been completed in around 20 minutes. However, only 70% of the agents reached their destination in 20 minutes with these given weights.

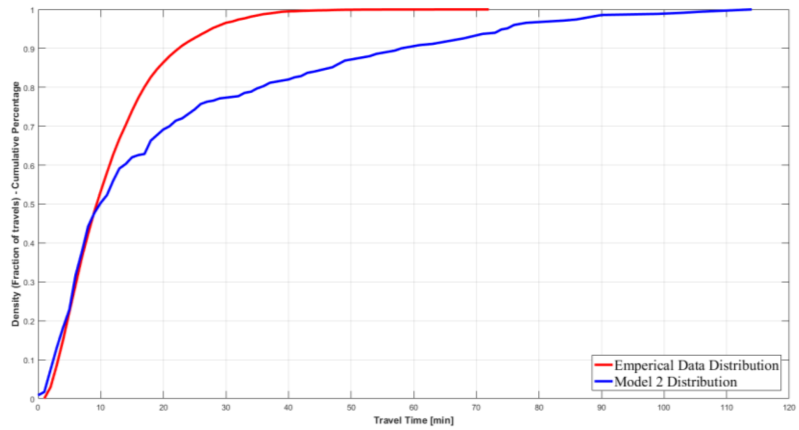
Figure 32 and Graph 2 present results of the Model 2 in which these values are used:

Table 9 Model 2 Calibration values

Model 2:	Shortest-path distance (P_{Sp_e})	Straight-line distance (P_{Sl_e})	Road type (P_{Rt_e})
Calibration values	0,70	0,70	0,40

Figure 32 Calibration Model 2

In Model 2, shortest-path distance parameter is decreased to 0,70 from 0,90. This gave a more spread out distribution on the network which is not the desired result. The distribution should be scattered as it is in Model 2 and it should be concentrated around the Centrum area.



Graph 2 Travel time distribution graph, Model 2

As decreasing the weight of the shortest distance parameter caused more spread over the network, the travel time duration is got longer in direct proportion to that. While the longest time it took between an OD was 85 minutes in the Model 1, it is increased to almost 120 minutes with these given weights.

The third calibration is presented by the Figure 33 and Graph 3. The values used for this model are:

Table 10 Model 4 Calibration values

Model 3:	Shortest-path distance (P_{Sp_e})	Straight-line distance (P_{Sl_e})	Road type (P_{Rt_e})
Calibration values	0,80	0,70	0,40

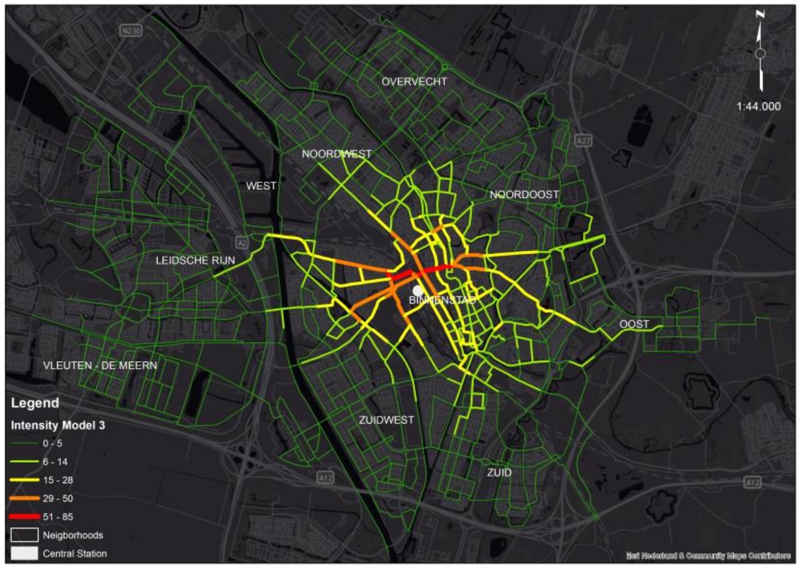
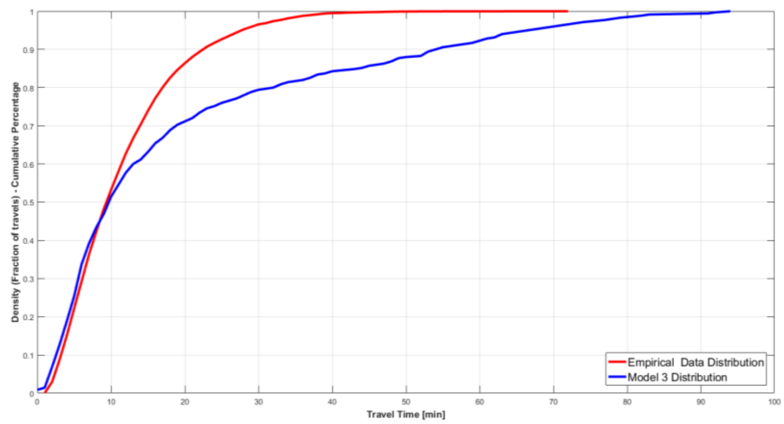


Figure 33 Calibration Model 3

In the third model, the shortest-path distance parameter is increased to 0,80 from 0,70. With this change, the spread-out distribution is decreased which is a better result than first two models. The first and second highest density groups' pattern remain the almost the same. Even though the Central Station area presents quite a similar pattern with GPS data, some differences are detected; the center of the Binnenstad should have a higher density while the road which is parallel and just next to railways should have a lower density. Other than that, the density around southeast of Binnenstad should have a lower number of travel. Overall, the distribution on the southwest, southeast and northeast should be decreased more to have similar results. As the shortest-path distance parameter at 0,80 gave better results than before models; it will be kept at 0,80 from now on.



Graph 3 Travel time distribution graph, Model 3

As the shortest-path distance parameter's weight is increased, the travel time duration is got shorter; decreased from 120 minutes to almost 90 minutes. This is a positive development, however, it needs to be get more shorter than 90 minutes. As it is discussed at the interpretation of the first graph, the longest duration should get even shorter than 70 minutes which is the longest travel duration measured from GPS data.

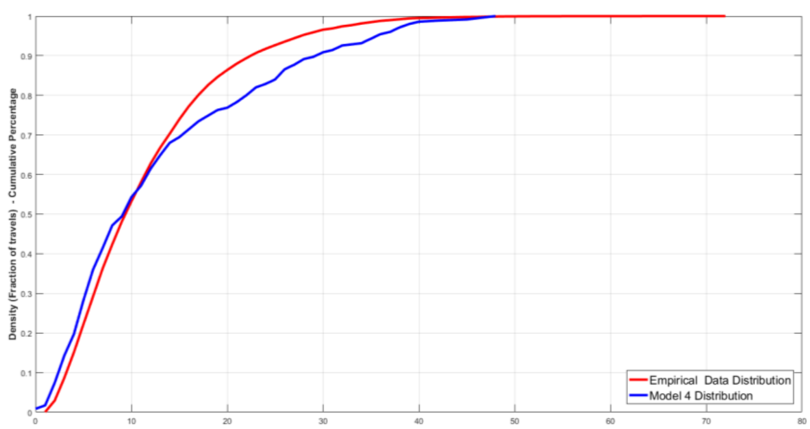
For the fourth calibration which is given by the Figure 34 and Graph 4, the value of shortest-path distance parameter kept at 0,80 and the value of road type parameter is decreased to 0,20:

Table 11 Model 4 Calibration values

Model 4:	Shortest-path distance (P_{Sp_e})	Straight-line distance (P_{Sl_e})	Road type (P_{Rt_e})
Calibration values	0,80	0,70	0,20

Figure 34 Calibration Model 4

As the road type parameter is given a lower value, the probability of a great majority of links are decreased and this caused to have more concentrated distribution around Centrum area. The scattered pattern on southeast and northeast is also decreased. With these results, the road type parameter is decided to be kept at 0,20 as it gave more realistic output.



Graph 4 Travel time distribution graph, Model 4

As the given weights are made the travels more concentrated on the Centrum area as they should be, the travel time duration is got a lot more shorter than Model 3 and GPS data as it was aimed. What is still need to be upgraded is the percentage of agents which were able to reach their destination in around 20 minutes. This ratio still needs to increase from 70% to around %90.

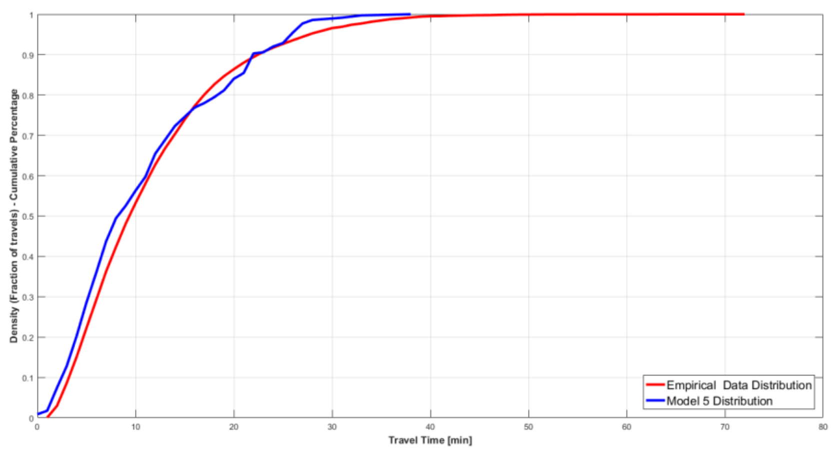
For the fifth calibration, the shortest-path distance and road type parameter are kept at the decided values based on previous calibrations, and the value of straight-line distance is decreased to 0,60. The results are presented by the Figure 35 and Graph 5:

Table 12 Model 5 Calibration values

Model 5:	Shortest-path distance (P_{Sp_e})	Straight-line distance (P_{Sl_e})	Road type (P_{Rt_e})
Calibration values	0,80	0,60	0,20

Figure 35 Calibration Model 5

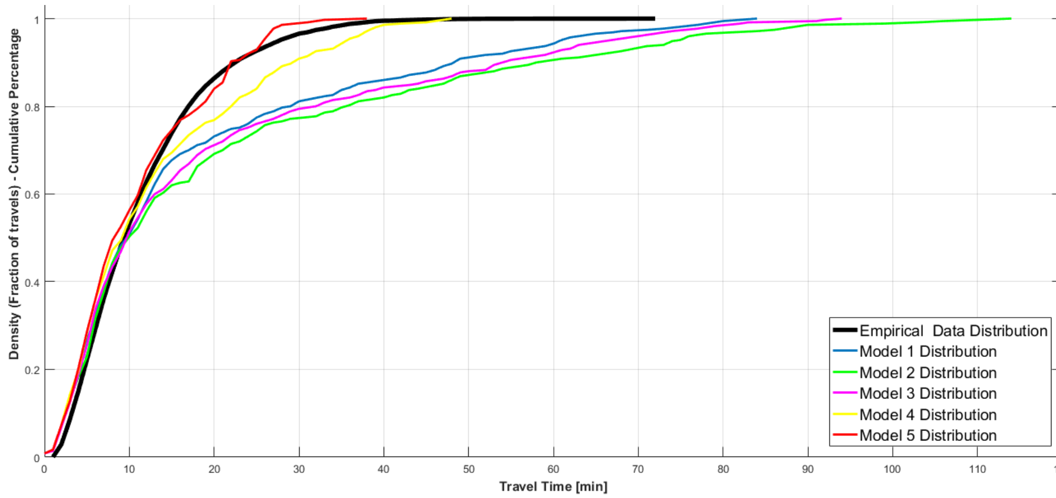
By decreasing the straight-line distance to 0,60 from 0,70 and keeping the other two parameters at before decided values, the most similar output to GPS data is obtained. By decreasing the direction support by the straight-line distance parameter, the agents are showed better results comparing to the ones with 0,70. As explained before, in some cases straight-line to the destination may not be the correct way to go because of the physical structure of the network paths. For example, there not be a link to go on that direction or the road structure is very meandering on that way which extends the distance in meters. Therefore, the impact of this parameter should be adjusted carefully. The value with 0,50 did not provide enough direction support which resulted a scattered distribution. Eventually, a value with 0,70 is checked for this parameter and it gave the most realistic results together with other two parameters.



Graph 5 Travel time distribution graph, Model 5

Given by the Graph 5, with these weights the model gave the most realistic results in the graph as well. The longest travel time duration is decreased to a realistic amount of time regarding the OD distributions. Also, the great majority of the agents have reached their destinations in between 20 to 30 minutes as it should be regarding the OD locations concentration on a relatively small (4 km radius) area.

The Graph 6 on the below shows all the calibration models together in one graph and their similarity to the GPS data. As can be seen, the Model 5 shows the most realistic results within these models and the most similar curve to the GPS data.



Graph 6 Travel time distribution graph, 5 calibration models

The validation of this model will be made in the last phase of the agent-based model methodology; the validation.

8.2. Validation

For the validation of the model, the values are defined based on the results obtained from calibration. Thereby, the values given on the below are used for 500 runs and the average of the results are presented by the Figure 37 Intensity map of the model and the Graph 7 Travel time distribution cumulative graphs.

For the validation purposes, the half of the GPS dataset is used in order to use a different dataset than the one is used for calibration. The intensity values of the measured data is divided in 10, in order to see exact intensity values for the amount of 350 cyclists (Figure 36). The intensity pattern have not changed, only the values did.

Model Validation:	Shortest-path distance (P_{Sp_e})	Straight-line distance (P_{Sl_e})	Road type (P_{Rt_e})
Final values	0,80	0,60	0,20

The Figure 37 shows the number of travels made on the links from the model and represents geographical distribution of intensities on the bike road network of Utrecht.

Figure 36 Intensity map of GPS

Figure 37 Intensity map of model

As can be seen from the Figures of 36 and 37, a great majority of the intensity classes are matching. In terms of the general pattern resulted in the model; the intensity is showing a gradual distribution over the network, without getting scattered on the outer periphery of Centrum area. The hot spots of two maps are the same; Central Station area, connective links (bridges) on the railways. The main roads are clearly represented with the higher and longer intensity patterns; the main roads start from the Binnenstad area and go to East, West and North sides differentiate from other roads and present similar intensities with the empirical data.

The links which surround the railways are also showing high similarities, however, in shorter distances. The intensities on these railway surrounding links resulted in longer distances in the empirical data. The reason for the differences in the distances is the limited number of OD locations. Since the OD locations of the agents have not scattered over all the whole network, the distances agents traveled got shorter. Hence, the intensities occurred in shorter distances. The same distance difference can be seen on the two main routes located between Binnenstad and Oost.

In the measured data, intensities on the grid plan street on Binnenstad show a bit more scattered pattern while in the model results, it is a more dense pattern. This is again related to the cumulation of OD points in the model. In general, the output in this area do present the same high and low intensity roads, but in higher volume.

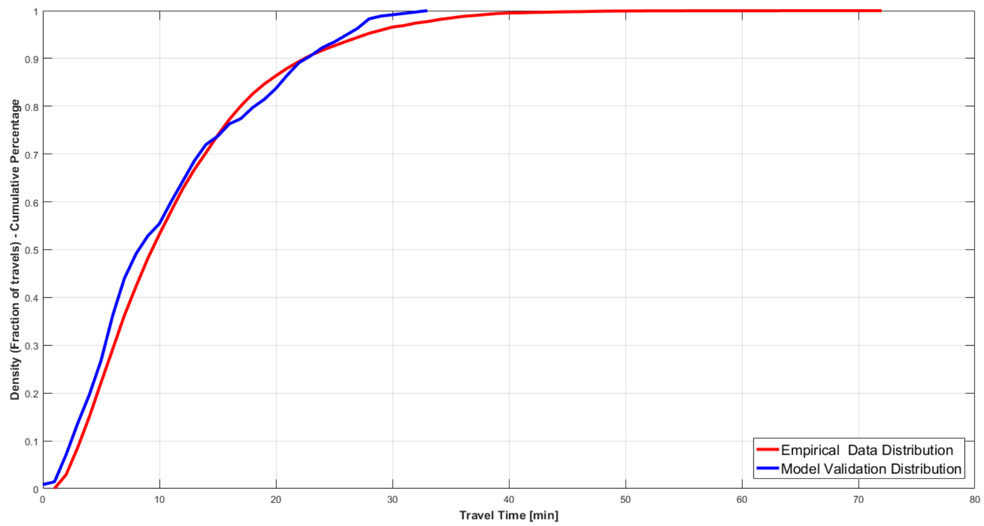
To emphasis on the similarities and the differences between the model and GPS data, model intensity classes are subtracted from the measured data intensity classes. The results are presented in the Figure 38; the smaller the value the less the difference between two results.

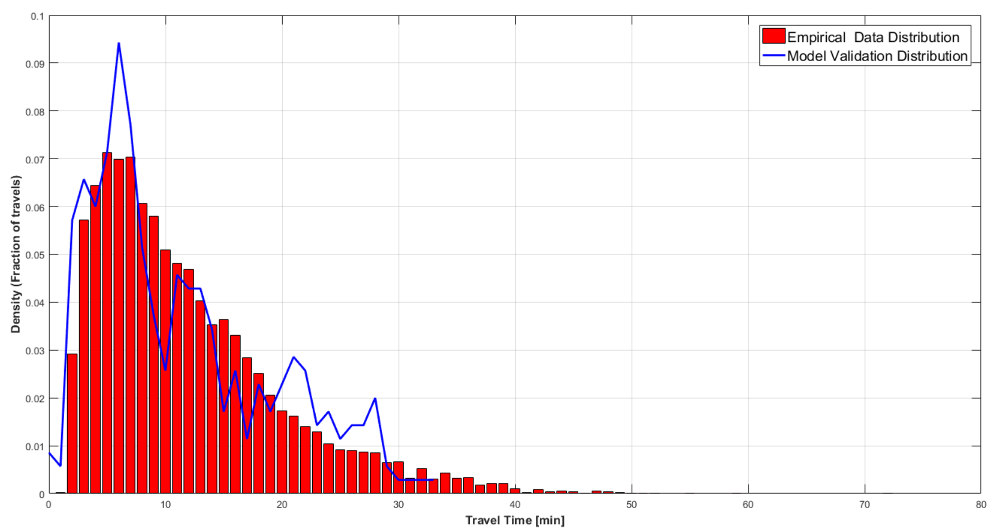
Figure 38 Comparison between GPS and model output

As it can be seen in the above Figure 38, the difference gets higher from light to dark color. As the great majority of the network (84,5% of the total network) has the lightest color, this proofs a majority of the bike road network is mimicked the GPS data intensity level with a zero difference by means of intensity levels.

Following to zero difference, the links which present the intensities with 1 class difference are located mostly around the Centrum area, around the Binnenstad. As mentioned on the interpretation of Figure 37, this can be related to OD locations' cumulation around this area. Increasing agents will allow having more ODs, which means they can be spread over all the network, but still, keep the distribution ratio. This is discussed in Chapter 10: Recommendation.

The 2 and 3 density class differences have a very low proportion in the comparison. The probabilistic nature of the model can be a cause of the differences in low proportions, Lastly, there is no link that has 4 intensity class difference with GPS data.





Graph 7 Travel time distribution graph of the model and the empirical data

As can be seen by the given cumulative percentage of the travel time distribution graphs of the validation model and the empirical data (Graph 7); the model does present empirical data travel time distribution on the simulation. However, the model finishes in a shorter time. There are two reasons for this result; having less number of agents in the model comparing to the number of cyclists in the measured data and the distribution of OD locations being more clustered in the model. These two reasons are directly related to each other, changing one will change the other one in the same direction. This is discussed in Chapter 10.

The output is showing highly realistic results by means of individuals travel time. As the OD locations of the agents are not located as dispersed as the GPS data have in reality (Section 6.2.); the model run time is shorter than empirical data. 90% of the travels are completed in around 20 minutes as the most of the OD pairs are densely located on 4km radius area and the speed was adjusted at 18km/hour for every agent. Looking at the fraction of travels, the individual number of agents travel time also shows high similarities between the individual number of cyclists travel time. That means even though the less number of agents are used in the model, their individual performance on mimicking the cycling movement was realized successfully.

CHAPTER 9
CONCLUSION

The aim of this chapter is to answer the research questions that were posed at the very beginning of the report. So far, each of these research questions is tackled within the corresponding chapters. Herewith, the obtained results and answers will be given concisely.

9.1. Conclusions

The objective of this research was to develop a simulation model of cyclist route choice preference by using an agent-based model and to utilize measured GPS data to realize this goal. In accordance with this objective, the research questions were generated (Chapter 1.2.3.):

RQ1: Which factors from former studies can be implemented in the agent-based model of this research?

The model's conceptual framework has been developed based on the two main sources; previous studies on cyclist route choice preference and analyses results obtained from the measured GPS data. Due to the fact that a limited number of studies are available in the field of cycling movement; studies on pedestrian movement have also been included in the literature research (For more information; Chapter 10.2). From these studies, the ones focusing on the individuals' movement at the mesoscale have been used to study theoretical background.

Several factors determined from the literature are established a foundation ground to identify cyclist route choice preferences. In order to limit the factors to implement, the most important factors have been selected. Later on the methodology process, some of these factors are adjusted according to the user group and some more are added in order to develop the model. It is important to note that, although the elements selected from the literature research have been provided the basic version of the model, the GPS data analysis results are shaped the final versions of the elements (More is given under the RQ2).

The basic model was containing the elements that are determined from the literature: Origin and Destination, Probability concept, Nodes and links relation, Road type, Connectivity and Directness qualities and Shortest-Path. For cyclists, there is always an end location pre-defined in one's mind. Since almost, always there is more than one possible route between the start and end location; every trajectory consists a route, a choice, a route choice decision (Usyukov, 2013) when a cyclist tries to optimize route choice. Between a start and an end location of a trajectory, crossing nodes on the trajectory represent "moments" to make a choice on the route. In other words, nodes to decision-making and links to choose. The infrastructural features of road networks have an influence on the choices made; the shared and dedicated cycling roads is one of them with a major effect. Another feature that influences cyclists is the connectivity and directness of a route. The last one is cyclists' high tendency to choose shortest-paths between their start and end location. To strengthen the elements of the movement; it was noted that a trajectory continues until the end point is reached or the time is constrained. Lastly, speed defines space-time dimension and, therefore, it should be included.

RQ2: What is the added value of measured GPS data in terms of developing the agent-based model?

The added value of measured GPS data can be explained by three aspects:

1. It provides the opportunity to check the factors determined from the literature on the revealed preferences made. By this functionality of the measured data usage, it has been seen that the user group of this study did not negatively affect by the motorized vehicle shared bike roads as the literature was reflecting. Their choices were positively influenced by the connectivity and directness of a route. Consequently, high quality of connectivity and directness features in shared roads made them choose this type of road. Another conclusion made from literature was reflecting that cyclists tend to choose the shortest-paths between their start and destination point. The analysis on this factor showed that this was indeed the case for the user group of this research.

2. It allows further research on the details of the route choice and travels made. As such the origin and destination locations, time and speed details of the travels are obtained by the use of the measured data.

3. It is a powerful source to evaluate and validate the model. This utility was used to validate the model performance in visual and statistical terms. More details on this function of the measured data are given under the Research Question 4.

RQ3: How to formalize the determined factors in order to implement them in the agent-based model?

The elements determined by the use of literature, and developed and adjusted through the GPS data analysis, were formalized under a conceptual framework which divided the factors into small subsets within a holistic approach. By this approach, it was made possible to apply the factors in agent-based modeling software Netlogo. By predicating on the conceptual framework, the formula was developed which presents a mathematical equation of route choice probabilities depending upon three parameters of the model: shortest-path, connectivity and directness qualities of road types and straight-line distance (Chapter 4.4. Conceptual framework and Chapter 4.4.2. & Chapter 7 Probability Formula).

As each of these factors influences the route choice decision, each of them has an effect on the probability of a road to be chosen. Every time a cyclist agent comes to a crossing node at the network, it has to make a route decision within the alternatives. With the probability formula, this decision-making was converted to the mathematical equation that calculates the probability of a link *(a road segment between two crossing nodes)* to be chosen by the cyclist. The link which has the highest probability value has the highest chance to be chosen, but there is also a low chance that lower probability value will be chosen. Developing and using such a formula is a positive side in terms of future developments of the model since it provides a researcher to adjust and control the model. The remaining factors: start and end location, speed and time to start trajectory have been implemented on model's itself as they are not dynamically changing parameters.

RQ4: How can the simulated behaviors and output trends be calibrated and validated?

Following to the creation of the model, next step was to calibrate it in order to make the simulation and its results as close as possible to the reality. To that end, firstly the sensitivity analysis has been executed in order to understand how sensitive the parameters to the changes made and which parameters should be changed in order to obtained realistic results. With the knowledge gained from sensitivity analysis, calibration has been applied with the defined parameters to calibrate. To strengthen the scientific approach, the measured GPS dataset has been split into two subsets; partly to be used for the calibration and partly to be used for the validation. During the calibration, intensity values of the captured trajectories in GPS data (number of travels made over links) and the captured trajectories of agents in the simulation model have been compared. Through the calibration, the final version of the model is shaped up with the values that give the closest exhibition to the GPS results.

At final, the model was run 300 times with the defined values for parameters and the results have been compared with the validation subset of GPS data. To visually validate the model, the intensity pattern on the network is visualized and the similarities and differences between the model and GPS data have been seen clearly. To emphasize the similarities and differences in terms of intensity patterns, the intensities modeled were subtracted from the intensities from the GPS and the results were quantified. Indicated by both visual and mathematical accuracies, the intensity results of the model are largely acceptable by means of its similarity to the reality.

Addition to validations made from the aspect of the intensity pattern modeled, the travel time distributions of the model has been validated too. By comparing travel time distributions of cyclists trajectories measured in the GPS data and cyclist agents in the model; it has been seen that traveling duration and travel time distribution during a day is highly similar to the reality. In conclusion, the model showed a highly acceptable performance with its similarity to real cycling movement on the Utrecht bike road network.

As it can be understood, measured GPS data had a fundamental part of this research in terms of creating, developing and validating the model.

RQ5: To what extent is this research and the model utilizable in urban planning domain?

With respect to the importance of this research question in terms of the model's applicability to the urban planning domain and case study; it is answered in a separate section given on the below (9.1.1. Applicability of the model).

9.1.1. Applicability of the model

This section presents the conclusions made in terms of:

- Applicability of the model in general to support urban planning
- Applicability of the model in SSD project.

In this research, the simulation of cyclist route choice preference on the infrastructural road network is proposed and a modeling approach is developed in which demonstrated by implementing a pilot model for the city of Utrecht, using NetLogo agent-based modeling toolkit. From several aspects, it consists of interesting features for the urban planning domain.

First of all, this research provides a construction of a dynamic model that shows a real-like behavior of the users of a road network of a city. The approach followed in this paper exhibits a guide which is re-usable in future modeling works. Because of the limited number of studies and models produced in this field comparing to the number of pedestrian and vehicle simulations available, it was relatively a rocky road to define an approach and apply it. Even if the end-result of this research is not advanced enough to directly apply the model in decision-makings related to urban planning; it is believed that it displays promising results. With the recommendations given in the next chapter, the model enhancements can be achieved.

The information provided by the model is twofold; firstly, ***it seeks to understand why a cyclist makes those decisions.*** It should be noted that it does not try to mimic the GPS data but the movement and route choices of cyclists. Mimicking the GPS data would be a more easier model to develop as it can be created by adding parameters directly using the revealed intensities from GPS data. However, in that case, the model would only be applicable for that specific area the GPS data captured and for that user group, the GPS traces logged. Also, as the intensities result based on several underlying factors; a model that directly uses intensities cannot use underlying factor at the same time. This causes miscalculations in the model.

Instead of such approach, it was aimed to develop an "initial foundation model" which further developments can be built upon, in compliance with specifications of case studies. Due to the complexity of cycling behavior and very limited number of studies available in the field, it has been difficult to execute a research in the field and identifies the approach to follow in order to create such model. Therefore, the completion of this informative research and creation of the route choice preferences model of cyclists which includes the fundamental elements and parameters of the movement can be regarded as a beneficial addition to this study to literature and urban planning domain.

Currently, the model and the parameters included in the model provide limited understanding and sure can be developed to have a more advanced simulation. For that aim, the model has been created in a way that it is flexible and open to further improvements (10.3 Recommendations). In respect to this, the model provides a limited usage but solid foundation ground for future works.

The second fold of the information provided by the model is that ***it exhibits how does the flow changes in such a circumstance/ scenario.*** Unfortunately, this function of the model has not been tested due to time limitations of the research. However, the elements required to execute the function are included in the model. On the below, the functionality of the model in terms of SSD project bike road developments and example use case is given.

The pilot model itself is applied to the case study city of Utrecht, The Netherlands and utilized the bike road network of Utrecht as the main geographical environment of the model. The model simulates the mobility flow on the bike road network together with its users; individual cyclists. The model allows for flexibility in terms of diversifying or changing the environment in which they operate; in our case bike road network of Utrecht. Therefore, it is a powerful tool in terms of assessing travelers' reactions to the changes made on the network.

Four of the interventions that are generally practiced in urban planning domain are made possible to be executed within the model:

- Adding a road segment to an existing road network
- Removing a road segment from an existing road network
- Changing the road type of a road
- Changing the quality of a road (e.g. making it more cycling friendly by means of directness and connectivity).

These interventions can be applied within the model. Based on the changes made, cyclists flow will be changed on the network. As the route choices of the cyclist agents depends on the features of a road, changing a feature of a road will also affect the mobility flow; such as changing the road type of a specific road segment (links as it called in the model) from shared to dedicated cycling road affects the route choice preference of agents, which affects the mobility flow on the network and the result of the model in terms of intensity of roads. Other than that, creating more directed roads on the network towards the densely traveled destination areas also gives similar results. It is a real-like action-reaction mechanism which focuses on the cycling flow.

By using the model, road development plans can be checked to see probable outcomes of the plan application in the real life. The model is in help for the urban planning department of the Municipality in terms of existing traffic demand and on top of that, stimulating the usage of the bicycle. Since the model has a quantitative feature that contains cyclist and road characteristics, it is able to predict which road and which road characteristic are significantly preferred by whom. For example, it can answer which/ how many of the users prefer to ride along a bike road that will be built in that specific area?

How does the new design effect the flow of these cyclists? In the model, the route choice preference of cyclists changes if any of these parameters effect by the change:

- Road type (Road types are categorized into two types: the bike roads that are shared with motorized vehicles and the ones that are only open to bike traffic. Moreover, these road types are analyzed from the aspect of connectivity and directness qualities. Changing these qualities of the roads should be reflected in the road type categories, which consequently affect the route choice preference of cyclists. For more; Section 10.3 Recommendations).
- Shortest path distance (Changing the physical structure of a road, building more meandering or straightforward routes between the major origin and destination locations, adding or removing a road segment would affect the route choice preferences of cyclists).
- Straight-line distance (Straight-line distance provides agents the sense of direction. The shorter the straight-line distance to the destination means the better direction to follow on the network. Changing the direction of any road affects this parameter and, therefore, the route choice preferences of cyclists).

To understand the use of the model, a use case scenario is given on the below.

Use Case Scenario:

According to the new urban plan of the SSD project area, a road segment is planned to be removed from the existing bike road network. This change can be applied in the model environment to see the effect of this change on the flow of cyclists.

Probable Outcome 1: Flow of cyclists got a denser at some spots on the network which may cause traffic congestions. This proofs removing that road segment will cause negative results on the mobility flow. Decisions and the plan can be revised to prevent these negative results in the real world.

Probable Outcome 2: Flow of cyclists exhibited more balanced distribution on the network. This proofs removing that road segment will provide positive effects on the mobility flow. Decisions and the plan can be applied in real world successfully.

Based on different urban planning scenarios, the network can be edited between the nodes, on the nodes, and on the links. The applied changes exhibit the reactions of cyclists via the changes in the flow; e.g. according to the new urban planning of the SSD area, some particular road segments will be removed. When these required changes are made on the model, the intensity gets higher on particular spots of the network which can possibly cause some negative result in terms of traffic flow. In this case, it can mean that the applied use case may be revised before its application on the real urban area.

By this approach, the model acts as a powerful tool in order to support urban planning and decision-making. To execute the changes mentioned in the previous paragraphs, a basic knowledge about the model creation and the way the model operates in the software together with the loaded input data must be gained. This may cause difficulties in terms of its practicality by urban designers who do not familiar with modeling. Therefore, it is recommended to create a clear GUI interface which automatically executes the required changes. This and more model enhancements are elaborated in the next chapter.

CHAPTER 10
DISCUSSION & LIMITATIONS & RECOMMENDATION

The aim of this chapter is threefold:

- The first is to present this discussion in which looks back on the research and reflects some remarks with respect to internal and external orientation. This section discusses especially in which aspects this research differentiate from comparable studies and what addition it makes to literature in the field of route choice modeling of cyclists (Section 10.1.).
- The second fold is to explain the limitations of this research in order to clarify some aspects of the model and executed research that could be developed more (Section 10.2).

- Lastly, it is to provide recommendations based on given conclusions and discussions. Together with the experiences gained and the results obtained during the research progress, recommendations will be presented in order to lead future researchs (Section 10.3). This is mainly important as this research seeks to supply inputs for possible future works.

10.1. Discussion

The simulation model of route choice preferences of cyclists is created in order to help urban planning domain by presenting an interactive cycling movement flow on the given road network which is shaped by the real users of the real road network. The output is the distribution of trips over the bike road network based on the given a set of origin and destinations. This developed route choice model has been applied to bike road network in Utrecht.

The movement of cyclists on the network is based on choices they made on the routes -route choice preferences-. Therefore, it was crucial to understanding the underlying reasons for route choices made in order to develop a successful model. Since there was very limited number of literature available and no example model developed in the domain of cyclist behavior, it is especially important to reflect on the outcomes of this research in order to fill some knowledge gaps and lead to further researchs in the field.

The results have been showed high similarity to the real movement flow of cyclists. This proofs that the parameters used in the model are indeed some of the major factors that influence cyclists route choice. To this end, it can be said that majority of cyclists choose their routes based on the shortest-paths between the start location of their trip and the destination location in their mind. By knowing the importance of this factor, urban planning of the bike road networks can be a focus on designing shortest paths between the most popular destination and origin locations of cyclists; such as routes between central stations and university campuses or student housing zones. Designing new shortest paths between the major origin-destination zones or developing the existing ones can navigate the great majority of the cycling traffic. Both of this implications can be applied to the model in order to see the outcomes of such operations in the real network. This parameter was mentioned in the most of the literature in the field and showed applicability within this model. From this aspect, it can be said that the theory is applied in practice successfully.

Another important factor that influences route choice preferences of cyclists and, therefore, the urban planning of bike roads is the connectivity and directness qualities of the roads. The meandering routes are unappealing for cyclists to ride along just as the routes that are not well-connected to the destination locations. These qualities should be taken into account in the urban planning projects of road networks. Also, they are important parameters to model the cycling movement in agent-based models. Addition to the studies in this field, it has seen that for some user groups directiveness and connectivity qualities of routes are more important than cycling dedicated routes (only-bike traffic roads).

The last dynamic parameter included, the direction knowledge of cyclists (straight-line distance parameter in the model), should not be ignored while modeling the cycling movement. In the real world, it is obvious that cyclists have a sense of direction and they use the roads in the direction of their destination. However, none of the studies mention such a parameter to be included in the models. Ignoring such parameter results non-realistic movements of individual agents in the simulation; one can move against to the direction of its destination location for hours without getting any closer to it. The results show low similarity to the real world by means of intensity patterns and travel time distributions.

Addition to this unique research and modeling approach applied in order to put into practice these three dynamic parameters; implementation of some of the static parameters showed different methods that make an addition to the literature. The most important one within these was obtaining origin and destination locations by the use of fishnet grid and decreasing the total number of cyclists by keeping the distribution ratio over the network equal with empirical data. It has been seen that the method is quite effective in situations such that the total number of cyclists in empirical data is too much to apply in a model due to time and processor power limitations. On the other hand, the method has critical aspects to be taken into account especially during the interpretation of the model outputs. These are well-explained during the calibration and validation of the model.

The last contribution of the research has a crucial importance by means of the model creation. This research has been tried to understand the underlying reasons for route choices made by cyclists and to apply these identified underlying reasons as parameters of cyclist agents route choice behavior. Through the creation of the model, it has been seen that using revealed intensities as a parameter of the route choices made is not a sound methodology. Because it is the ultimate result, not an underlying reason. Therefore, it should not be used together with the reasons, i.e. shortest-path distance etc. The implication of such revealed result in the model causes doubling the effect of some parameters and causes unsıuccessful simulation models. Although this awareness has a crucial effect on model results, it has not been mentioned in the previous studies before.

10.2. Limitations

The first topic to reflect upon is the absence of the studies done in the field of route choice modeling and agent-based modeling of cyclists. Comparing to a large amount of theoretical and practical studies available on the pedestrian movement behavior; it has been seen that cyclists' movement is a surprisingly underexplored topic. Although pedestrian and cycling activities can be seen as two similar modes of individual mobility, the parameters of the pedestrian route choice preferences do not provide sufficient knowledge for the case of cyclists. The fact that there was no example model that could help the imagination was the first signal of a tough research and methodology process. Nonetheless, the model is successfully developed as it has been presented in these pages. As a result, this thesis research takes an important place as being one of the very few numbers of studies which informs and guides researchers to construct an agent-based model that mimics cyclist route choice preference.

Another point to reflect upon is about the used software for the agent-based model. Because of the restrictions of the NetLogo; some factors could not have been applied in the model; i.e. the decision-tree method, counting intersection locations on the routes and taking into account the traffic light waiting times on the routes. The software allows agents to "sense" the route from standing node to the next node, no longer than that. Not being able to calculate all possible routes between OD pairs caused to cancel implementation of these mentioned factors and caused to spend more time to overcome these issues.

Also, because of the time and computer processor power limits, the number of agents could not have been increased more. Because it was making the model run time duration longer and was requiring more processor power limits. Consequently, this situation was causing crashes during model runs. Since for each sensitivity analysis and calibration, 100 of runs had to be made with the 350 agents on the network that has thousands of links; one run time was taking 4 minutes which takes days to have results of a sensitivity analysis. As time planning of a research is one of the key elements to succeed; it was not feasible to increase it more under the given circumstances.

Another critical reflection is the measured GPS data. As the data was pre-processed, the data was divided into two files which caused several issues during the research. Analyzing and calculating the statistical values of the data was possible such as calculating the number of travels made over a link, or the time the trajectory start, speed it traveled. With some long data analysis, origin and destinations of the trajectories were derived as well. However, it was not possible to visualize individual trajectories made on the network. This data could provide more knowledge on the route choices made which would certainly give important inputs to the model.

Another drawback of the GPS data was about the language used in the documentation and in the data itself (attributes of the shapefile). Since the data is captured in the Netherlands, it is expectable to be documented in the Dutch language as it was. However, providing additional documentation in the English language could provide a higher quality of the “open” data. In the case of this research, this language limitation of the open data caused some problems for the non-Dutch speaking author, such as wrong interpretations of the datasets and spending more time on the data analysis.

10.3. Recommendations

As long as the model shows results that are quite similar to the tracking data, it can be enhanced with the number of recommendations given in this section.

10.3.1. Enhancing existing elements of the model

Here, the existing elements in the model are taken into account to show how they can be enhanced.

- Number of agents:

Increasing the number of agents means increasing the number of travels that will be simulated on the network as new trajectories between new OD pairs. As long as the agents’

OD pairs included in the model show a similar distribution to the empirical data`s OD pairs; the model works sufficiently.

Within this research, the number of agents has been decreased to 110 from the original number of cyclists which was 6936 at the first trial. As a result of this low number, it has been seen that some of the origin and/or destination zones were not included in the model within this limited number and this was affecting the model results negatively by giving dissimilar intensity pattern on the network. This result showed that 110 agents are too less to represent the real pattern. For this reason, the number is increased to 350, and this time almost all major origin and destination zones have been included in the model.

Because of the limitations explained (Section 10.2), this number has not been increased more than 350. However, for future works, it is recommended. Due to a limited number of OD pairs of the model; model's OD pairs are not scattered on the network as mush as GPS data OD pairs. The zones which have a higher number of OD locations in the GPS data have a cumulation of origin or destination points, while some areas have very few number of those. Therefore, the result intensities of the model show more denser patterns in some areas comparing to the GPS data; i.e. Binnenstad (Chapter 8). This issue can be overcome in future research by increasing the number of agents using the same method applied in this research A higher number of agents included in the model would give more similar intensity patterns to the empirical data.

- Road type:

Two types of road types have been assigned on the bike road network: shared roads that are shared with motorized vehicles and dedicated roads that are only open to bikes without any shared traffic with vehicles. As the network data used in the model was only stating auto, bus and bike usage of roads and the literature research was emphasizing the difference between motorized and non-motorized roads on route choice of cyclists; this categorization has been applied in the model. For future studies, it can be enhanced. For the pilot model in Utrecht, the roads can be re-categorized based on three main bike roads types (1.Stroomweg, 2.Gebieds-onstuilingsweg, 3.Erftoegansweg) and the revealed intensity of cyclists from GPS data can be analyzed based on the distribution of these type of roads. The gained knowledge can be applied to the included road type parameter.

- Information related to Nodes:

According to literature, a number of turns, a number of intersections and waiting times at the traffic lights have a negative effect on the route choice of cyclists. These factors could have been assigned as parameters that affect route choice of cyclist agents; however, it could not have due to limitations of the NetLogo software (Explained in 10.2). By assigning waiting times at intersections; and calculating a number of turns and intersection locations on alternative routes; the route decision of agents will be improved.

- Speed of agents:

The speed of all agents is given the same value in the model which is decided based on the mean speed value of the user group obtained from GPS data. For future studies, the speed of agents can be defined differently based on the day and/or hour. Such approach will give more diverse behaviors in the simulation.

10.3.2. Enhancing the model by adding new elements

- User groups:

The current model simulates only one user group; the cyclists that travel over the Utrecht Central Station area between their start and end location. Therefore, there is no differentiation between agents' route choice preference other the parameters that are based on origin and destination locations. The measured GPS data was not including any socio-demographical information about the cyclists. Therefore, grouping the agents according to socio-demographical data was not possible. Another idea was to group cyclists based on the time intervals they started their trajectory and analyze the routes which are more densely used within those time intervals in order to characterize their route choice. However, such a method does not suit with the aim of the research; as the aim is to understand the underlying reasons that affect their route choice preferences, not to use revealed route choice preferences.

To that end, socio-demographical surveys and stated-preference survey data can be very useful to divide the agents into different user groups. By differentiating the agent groups, more realistic and diverse movements will become visible.

- GUI Interface:

For the easy applicability of the model in urban planning use case scenarios; a GUI interface that is easy to comprehend and visually attractive seems essential (Chapter 9.1.1). This interface should include an automated transformation of a geo-referenced road network into the Netlogo toolkit's environment as its own nodes and links. This function has already been included in the model (Section 6.1), however providing it within a GUI interface makes it easier to apply for urban designers who are not familiar with modeling.

Other than that, buttons should enable the user to add and remove a road segment and change the road type of selected road segments. Increasing the number of agents can also be made possible by adding a new origin and destination pairs on the transformed network and assigning a corresponding moving entity to these new pairs.

Lastly, it is important to note that as the topic of route choice preferences of cyclists combines several domains, more can be integrated into sociology, psychology, healthcare, urban health and much more. These might provide different aspects in terms of individual behavior, their decision-making processes, and their differentiation.

Bibliography

- American Institute of Certified Planners. (2009). Technique for bicycle demand estimation.
- Amiri, M. F., Esmaeli, A., & Gholami, M. R. J. V. A. (2007). Application of Intelligent Agent for Developing an Artificial GIS City.
- Andrienko, G., Andrienko, N., Kopanakis, I., Ligtenberg, A., Wrobel, S., Bonchi, F., … Kaya, S. V. (2008). *Mobility, data mining and privacy: Geographic knowledge discovery*. *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*. http://doi.org/10.1007/978-3-540-75177-9
- Aultman-Hall, L. (1996). Commuter Bicycle Route Choice: Analysis of Major Determinants and Safety Implications.
- Aultman-Hall, L., Hall, F. F., Baetz, B. B., Used, D., & Analysis, I. N. (1997). Analysis of Bicycle Commuter Routes Using Geographic Information Systems: Implications for Bicycle Planning. *Transportation Research Record*, *1578*(970168), 102–110. http://doi.org/10.3141/1578-13
- Beale, L., Matthews, H., Picton, P., & Briggs, D. (n.d.). 6TH ERCIM Workshop "User Interfaces for All" Interactive Poster MAGUS: Modelling Access with GIS in Urban Systems: An Application for Wheelchair Users in Northamptonshire.
- Bierlaire, M., Antonini, G., & Weber, M. (2003). Behavioral Dynamics for Pedestrians. *International Conference on Travel Behavior Research*, (August), 10–15.
- Bohte, W., & Maat, K. (2009). Deriving and validating trip purposes and travel modes for multi-day GPS-based travel surveys: A large-scale application in the Netherlands. *Transportation Research Part C: Emerging Technologies*, *17*(3), 285–297. http://doi.org/10.1016/j.trc.2008.11.004

- Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, *46*(10), 1730–1740. http://doi.org/10.1016/j.tra.2012.07.005
- Broach, J., Gliebe, J., & Dill, J. (2011). Bicycle route choice model developed using revealed preference GPS data. *TRB 2011 Annual Meeting*, *5464*. Retrieved from ftp://ftp.hsrc.unc.edu/pub/TRB2011/data/papers/11-3901.pdf
- Buehler, R., & Dill, J. (2015). Bikeway Networks: A Review of Effects on Cycling. *Transport Reviews*, *1647*(May), 1–19. http://doi.org/10.1080/01441647.2015.1069908
- Cambridge Cycling Campaign. (2014). Making Space for Cycling: A guide for new developments and street renewals.
- Cambridge Systematics, I. (1999). Guidebook on Methods to Estimate Non-Motorized Travel: Supporting Documentation. *Technology*, (July), 180. Retrieved from http://www.ibrc.fhwa.dot.gov/tfhrc/safety/pubs/vol2/techrep.htm
- Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, *2*(3), 199–219. http://doi.org/10.1016/S1361-9209(97)00009-6
- Cooper, C. (2015). Modelling pedal cycle usage and flows with spatial network analysis. Retrieved from www.cardiff.ac.uk/sdna
- Deadman, P. J. (1999). Modelling individual behaviour and group performance in an intelligent agent-based simulation of the tragedy of the commons. *Journal of Environmental Management*, *56*(3), 159–172. http://doi.org/10.1006/jema.1999.0272
- Department of Transportation. (2008). Cycle infrastructure design, (October), 92. Retrieved from https://www.gov.uk/government/publications/cycle-infrastructure-design-ltn-208
- Duncan, D. T., Aldstadt, J., Whalen, J., Melly, S. J., & Gortmaker, S. L. (2011a). Validation of Walk Score?? for estimating neighborhood walkability: An analysis of

four US metropolitan areas. *International Journal of Environmental Research and Public Health*, *8*(11), 4160–4179. http://doi.org/10.3390/ijerph8114160

- Duncan, D. T., Aldstadt, J., Whalen, J., Melly, S. J., & Gortmaker, S. L. (2011b). Validation of Walk Score?? for estimating neighborhood walkability: An analysis of four US metropolitan areas. *International Journal of Environmental Research and Public Health*. http://doi.org/10.3390/ijerph8114160
- Dutch Ministry of Transport. (2001). A vision on traffic and transport in 2020. *Water Management*, (October 2000).
- ECECO. (2003). Traffic, transport and the bicycle in Utrecht, The Netherlands, 34.
- FHWA. (1999). Guidebook on Methods to Estimate Non- Motorized Travel: Supporting Documentation.
- Frentzos, E., Pelekis, N., Ntoutsi, I., & Theodoridis, Y. (2008). Trajectory database systems. In *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*. http://doi.org/10.1007/978-3-540-75177-9_7
- Gemeente Utrecht. (n.d.-a). *Action Plan 2015 - 2020*.
- Gemeente Utrecht. (n.d.-b). *Utrecht - We All Cycle - Summary of Action Plan*.
- Gemeente Utrecht. (n.d.-c). *World first for Utrecht: P-route for cyclists*.
- Gemeente Utrecht. (2012). Summary Long-term Bicycle Plan.
- Gemeente Utrecht. (2015). World first for Utrecht : P-route for cyclists.
- Gimblett, R. (n.d.). Where are we and where are we going ?, 5271.
- Goeverden, K. Van, & Godefrooij, T. (2011). *The Dutch Reference Study: Cases of interventions in bicycle infrastructure reviewed in the framework of Bikeability*.
- Griffin, G. (2009). Simple techniques for forecasting bicycle and pedestrian demand. *Practicing Planner*, *7*(3), 15.
- Groeneveld, J. A. (2011). An Agent-based model of bicyclists accessing light-rail stations in Salt Lake City.

- Haklay, M., O'Sullivan, D., Thurstain-Goodwin, M., & Schelhorn, T. (2001). "So go downtown":simulating pedestrian movement in town centres. *Environment and Planning B: Planning and Design*, *28*(3), 343–359. http://doi.org/10.1068/b2758t
- Halldórsdóttir, K., Rieser-Schüssler, N., Axhausen, K. W., Nielsen, O. A., & Prato, C. G. (2014). Efficiency of choice set generation methods for bicycle routes. *1st European Symposium on Quantitative Methods in Transportation Systems*, *4*(14), 332–348.
- Heinen, E., Maat, K., & van Wee, B. (2013). The effect of work-related factors on the bicycle commute mode choice in the Netherlands. *Transportation*, *40*(1), 23–43. http://doi.org/10.1007/s11116-012-9399-4
- Heinen, E., Maat, K., & Van Wee, B. (2011). The role of attitudes toward characteristics of bicycle commuting on the choice to cycle to work over various distances. *Transportation Research Part D: Transport and Environment*, *16*(2), 102–109. http://doi.org/10.1016/j.trd.2010.08.010
- Heppenstall, A. J., Crooks, A. T., See, L. M., & Batty, M. (2013). *Agent Based Models of Geographical Systems*. *Journal of Chemical Information and Modeling* (Vol. 53). http://doi.org/10.1017/CBO9781107415324.004
- Heppenstall, A. J. J., & Crooks, A. T. (2012). Introduction to Agent-based Modelling. In *Agent-Based Models of Geographical Systems* (Vol. 164, pp. 1–759). http://doi.org/10.1007/978-90-481-8927-4
- Hoedl, S., Titze, S., & Oja, P. (2010). The Bikeability and Walkability evaluation table: Reliability and application. *American Journal of Preventive Medicine*. http://doi.org/10.1016/j.amepre.2010.07.005
- Hood, J., Sall, E., & Charlton, B. (n.d.). A GPS-based bicycle route choice model for San Francisco, California.
- Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters-the International Journal of Transportation Research*, *3*(1), 63–75. http://doi.org/10.3328/TL.2011.03.01.63-75

- Hrncir, J., Nemet, M., Hrncir, J., Song, Q., Zilecky, P., Nemet, M., & Jakob, M. (2014). Bicycle Route Planning with Route Choice Preferences, (March 2016). http://doi.org/10.3233/978-1-61499-419-0-1149
- Jestico, B., Nelson, T., & Winters, M. (2016). Mapping ridership using crowdsourced cycling data. *Journal of Transport Geography*. http://doi.org/10.1016/j.jtrangeo.2016.03.006
- Johnston, K. (2013). *Agent analyst: agent-based modeling in ArcGIS*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Agent+Analyst:+Agent-Based+Modeling+in+ArcGIS#2
- Jokar Arsanjani, J., Zipf, A., Mooney, P., & Helbich, M. (2015). OpenStreetMap in GIScience. *OpenStreetMap in GIScience: Experiences, Research, Applications*, (JANUARY), 1–20. http://doi.org/10.1007/978-3-319-14280-7
- Karima, B., & Akdag, H. (2013). Agent-based modeling for traffic simulation, (October). Retrieved from http://dspace.univ-biskra.dz:8080/jspui/handle/123456789/660
- Kemperman, A. D. A. M., Borgers, A. W. J., & Timmermans, H. J. P. (2009). Tourist shopping behavior in a historic downtown area. *Tourism Management*, *30*(2), 208–218. http://doi.org/10.1016/j.tourman.2008.06.002
- Krizek, K. J., Forsyth, A., & Baum, L. (2009). Walking and cycling international literature review: Final report. *Department of Transport Victoria Australia*, 1–104. Retrieved from http://www.designforhealth.net/techassistance/trec.html.%5Cn???,
- Land Transport Safety Authority. (2004). *Cycle network and route planning guide*.
- Langenberg, P. (n.d.). Cycling in Amsterdam Developments and policies.
- Larsen, J. (2010). Beyond the Bike Lane: An Analysis of Cyclists' Travel Behavior in Montreal & A Methodology for Locating New Routes A Supervised Research Project in two parts.

- Law, S., Sakr, F. L., & Martinez, M. (2014). Measuring the Changes in Aggregate Cycling Patterns between 2003 and 2012 from a Space Syntax Perspective. *Behavioral Sciences (Basel, Switzerland)*, *4*(3), 278–300. http://doi.org/10.3390/bs4030278
- Légaré, E., Krizek, K. J., Forsyth, A., Baum, L., & Contacts, K. J. K. (n.d.). Walking and Cycling International Literature Review AUTHORS OF FINAL REPORT Outline of report GLOSSARY 2 SUMMARY OF FINDINGS 4 WALKING AND CYCLING COMMON THEMES 5.
- Link, C., Lopes, J., & Bento, J. (2016). Traffic and Mobility Data Collection for Real-Time Applications.
- Lobben, A., & Bone, C. (2016). Agent-Based Model Simulating Pedestrian Behavioral Response to Environmental Structural Changes. http://doi.org/10.15760/trec.142
- Loidl, M., Wallentin, G., Cyganski, R., Graser, A., Scholz, J., & Haslauer, E. (2016a). Geo-Information GIS and Transport Modeling—Strengthening the Spatial Perspective. *Internatitonal Journal of Geo-Information*. http://doi.org/10.3390/ijgi5060084
- Loidl, M., Wallentin, G., Cyganski, R., Graser, A., Scholz, J., & Haslauer, E. (2016b). GIS and Transport Modeling—Strengthening the Spatial Perspective. *ISPRS International Journal of Geo-Information*, *5*(6), 84. http://doi.org/10.3390/ijgi5060084
- London, T. (2011). Exploring the relationship between leisure and commuter cycling, (October), 1–5.
- Lopes, J., Bento, J., Huang, E., Antoniou, C., & Ben-Akiva, M. (2010). Traffic and mobility data collection for real-time applications. In *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. http://doi.org/10.1109/ITSC.2010.5625282
- Luo, D., & Ma, X. (2016). Modeling of Cyclist Acceleration Behavior Using Naturalistic GPS Data, 14p. Retrieved from https://trid.trb.org/view/1393570

- Macal, C. M., & North, M. J. (2006). Tutorial on agent-based modeling and simulation part 2: How to model with agents. In *Proceedings - Winter Simulation Conference*. http://doi.org/10.1109/WSC.2006.323040
- Macal, C. M., & North, M. J. (2011). Introductory tutorial: Agent-based modeling and simulation. In *Proceedings of the 2011 Winter Simulation Conference (WSC)*. http://doi.org/10.1109/WSC.2011.6147864
- Macbeth, A., & Allen, T. (2007). Cycle Route Network Planning Using GIS.
- Matthews, H., Beale, L., Picton, P., & Briggs, D. (2003). Modelling Access with GIS in Urban Systems (MAGUS): capturing the experiences of wheelchair users. *Area*, *351*, 34–45.
- Ministry of Transport/ Public Works and Water Management. (1999a). The Dutch Bicycle Master Plan. *Water Management*, 131. Retrieved from http://www.fietsberaad.nl/library/repository/bestanden/The Dutch Bicycle Master Plan 1999.pdf
- Ministry of Transport/ Public Works and Water Management. (1999b). The Dutch Bicycle Master Plan.
- Ministry of Transport/ Public Works and Water Management. (2009). Cycling in the Netherlands, 77. Retrieved from http://www.fietsberaad.nl/library/repository/bestanden/CyclingintheNetherlands2009.pdf
- Molina, J. (2014). The Case for Crowdsourcing in Bicycle Planning: An Exploratory Study, (May), 1–135.
- Morvan, G. (2012). Multi-level agent-based modeling - A literature survey, 1–27.
- Municipality of Amsterdam. (2012). Netherlands Bicycle Partnership for Sustainable and Smart Cities.

- National Assembly for Wales. (2013). Active Travel (Wales) Act 2013, (January). Retrieved from http://www.legislation.gov.uk/anaw/2013/7/pdfs/anaw_20130007_en.pdf
- Papinski, D., & Scott, D. M. (2011). A GIS-based toolkit for route choice analysis. *Journal of Transport Geography*, *19*(3), 434–442. http://doi.org/10.1016/j.jtrangeo.2010.09.009
- Pettinga, A., Rouwette, A., Braakman, B., Pardo, C., Kuijper, D., Jong, H. de, … Godefrooij, T. (2009). Cycling-Inclusive Policy Development : A Handbook. *Transport*, (April), 256. Retrieved from www.i-ce.nl
- Pinjari, A. R., & Bhat, C. (2011). Activity-based travel demand analysis. *A Handbook of Transport Economics*, (1), 213–248.
- Preciado, C. R. C. (2012). Spatial Network Analysis for Urban Cycling Networks, (November).
- Pritchard, R. (2015). Overview of methodological approaches Why is route choice important ?, (September).
- Puentedura, R. R., & Ph, D. (1997). A Simple Wolf-Sheep-Grass Simulation Running the Simulation.
- Railsback, S. F., & Grimm, V. (2010). Models, Agent-Based Models, and the Modeling Cycle 1.1. *Agent-Based and Individual-Based Modeling: A Practical Introduction*, 3–13.
- Schneider, R., Patton, R., Toole, J., & Raborn, C. (2005). Pedestrian and Bicycle Data Collection in United States Communities: Quantifying Use, Surveying Users, and Documenting Facility Extent, (January), 199.
- Schuessler, N., & Axhausen, K. W. (2009). Processing Raw Data from Global Positioning Systems Without Additional Information. *Transportation Research Record: Journal of the Transportation Research Board*, *2105*(1), 28–36. http://doi.org/10.3141/2105-04

- Segadilha, A. B. P., & Sanches, S. da P. (2014a). Analysis of Bicycle Commuter Routes Using GPSs and GIS. *Procedia - Social and Behavioral Sciences*, *162*, 198–207. http://doi.org/10.1016/j.sbspro.2014.12.200
- Segadilha, A. B. P., & Sanches, S. da P. (2014b). Identification of Factors that Influence Cyclistś Route Choice. *Procedia - Social and Behavioral Sciences*, *160*(Cit), 372–380. http://doi.org/10.1016/j.sbspro.2014.12.149
- Singh, K., Sjjad, M., & Ahn, C.-W. (2015). Simulating Population Dynamics with an Agent Based and Microsimulation Based Framework. http://doi.org/10.13140/RG.2.1.1809.0964
- Snizek, B. (2016). Mapping cyclists ' experiences and agent-based modelling of their wayfinding behaviour, (October 2015). http://doi.org/10.13140/RG.2.1.1809.0964
- Snizek, B., Sick Nielsen, T. A., & Skov-Petersen, H. (2013). Mapping bicyclists' experiences in Copenhagen. *Journal of Transport Geography*. http://doi.org/10.1016/j.jtrangeo.2013.02.001
- Snizek, B., Skov-Petersen, H., & Sick-Nielsen, T. A. (n.d.-a). *Mappling cyclists experiences and agent-based modeling of their wayfing behvaiour*.
- Snizek, B., Skov-Petersen, H., & Sick-Nielsen, T. A. (n.d.-b). Summary paper: Agent-based model of bicyclists' behaviour, 1–4.
- Snizek, B., Skov-Petersen, H., & Sick-Nielsen, T. A. (2008). Kvintus.org - a choice based agent-based simulation model intergrated with Google Maps. *Reseedings for MMV4 (in …*, 446–450. Retrieved from http://mmv.boku.ac.at/refbase/files/skov-petersen_hans-2008-kvintus.org-a_choi.pdf
- Steenberghen, T., Pourbaix, J., Moulin, A., Bamps, C., & Keijers, S. (2013). Study on harmonised collection of European data and statistics in the field of urban transport and mobility, 149.

- SWOV Institute for Road Safety Research. (2008). SWOV Fact sheet. *October*, *3*(October), 1–6. Retrieved from http://www.swov.nl/rapport/Factsheets/UK/FS_Speed_choice.pdf
- TRB of the National Academies. (2012). Artificial Intelligence Applications to Critical Transportation Issues. *Transportation Research Circular*, *E-C168*(November). Retrieved from http://www.trb.org/main/blurbs/168134.aspx
- Ulmer, J. (2003). Evaluating the Accessibility of Residential Areas for Bicycling and Walking using GIS by Approval Sheet. *Civil Engineering*, (May).
- Ulmer, J., & Hoel, L. A. (2003). Evaluating the Accessibility of Residential Areas for Bicycling and Walking using GIS.
- Urban Movement, & Phil Jones Associates. (2014). International Cycling Infrastructure Best Practice Study: Report for Transport for London, 108. Retrieved from https://www.tfl.gov.uk/cdn/static/cms/documents/international-cycling-infrastructure-best-practice-study.pdf
- Usyukov, V. (2013). Development of a Cyclists' Route-Choice Model : An Ontario Case Study.
- van der Spek, S., van Schaick, J., de Bois, P., & de Haan, R. (2009). Sensing Human Activity: GPS Tracking. *Sensors*, *9*(4), 3033–3055. http://doi.org/10.3390/s90403033
- Van Goeverden, K., & Godefrooij, T. (2011). The Dutch Reference Study Cases of interventions in bicycle infrastructure reviewed in the framework of Bikeability Department Transport & Planning Interface for Cycling Expertise.
- Vine, D., Buys, L., & Aird, R. (n.d.). Experiences of Neighbourhood Walkability Among Older Australians Living in High Density Inner-City Areas. http://doi.org/10.1080/14649357.2012.696675
- Wilensky, U. (2003). The NetLogo 5.3.1 User Manual, (November), 427. http://doi.org/10.1002/ejoc.201200111

- Willis, A., Kukla, R., Hine, J., & Kerridge, J. (2000). Developing The Behavioural Rules For An Agent-Based Model of Pedestrian Movement. *Proceedings of the European Transport Conference*, (September), 69–80. http://doi.org/citeulike-article-id:4074026
- Yeboah, G., Alvanides, S., & Thompson, E. M. (2015). Computational Approaches for Urban Environments. *Geotechnologies and the Environment*, 395. http://doi.org/10.1007/978-3-319-11469-9
- Zhang, L., & Levinson, D. (2004). Agent-Based Approach to Travel Demand Modeling: Exploratory Analysis. *Transportation Research Record*, *1898*(1), 28–36. http://doi.org/10.3141/1898-04
- Zheng, H., Son, Y.-J., Chiu, Hickman, M. (2013). A Primer for Agent-Based Simulation and Modeling in Transportation Applications. *Development*.
- Zuidgeest, M. (1997). *Sustainable Urban Transport Development* (Vol. 4).
- Zuidgeest, M., Brussel, M., & van Maarseveen, M. (2015). GIS for Sustainable Urban Transport. *ISPRS International Journal of Geo-Information*, *4*(4), 2583–2585. http://doi.org/10.3390/ijgi4042583

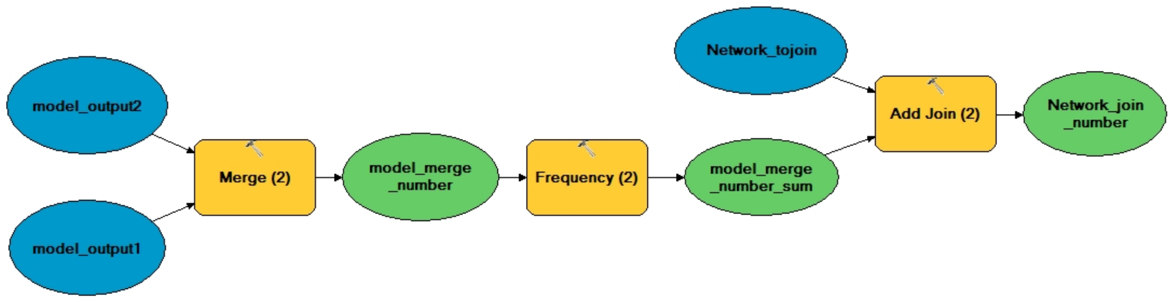
Appendices

Table of Content of the DVD that accompanies the thesis report:

- Report (Word, PDF)
- Midterm Report (PDF) & Midterm Presentation (PPT)
- GIS Datasets used and created
- NetLogo model created together with the input datasets used
- Manual of the model created
- Figures, Maps, Tables given in the next pages of Appendices
- Literature saved in Mendeley (provided as PDFs in the zip file)
- Presentations prepared for the weekly meetings in TNO

Final Presentation (PPT) will be submitted after the Thesis Defence.

Appendix 1: Created model to merge and sum the outputs of ABM



Appendix 2: Fietstelweek data files

1. Routes.csv file

As the data was pre-processed, it was divided into two files. One of them was in the table format which provides all the trips made by all cyclists with the trajectories they used from their origin to destination (Routes.csv). The table does not include any spatial reference of GPS traces but only the link numbers (unique numbers given to the road segments) constitute them. The day of a week and the time that the trips made are given per every link, separately.

Every trip in the file was logged in with a unique Object ID and therefore, the Object ID can be seen as a unique number assigned to a random cyclists' trip. However, the trajectories are not saved in this file with their spatial references, instead they are described with a group of unique number of road segments (as it is called in the file 'link numbers') they rode along during trip, and within this group the link numbers are arranged in a sequential order based on cyclist's trajectory's start link to destination link. Therefore, the file does not include any geographical component related to trajectories but only link numbers constitute them. The day of a week and the time that the trips made are given per every link, separately.

Number of rows in the original routes file: 8903459
Number of rows in the file which contains the trajectories made in Utrecht: 1083900
Number of rows in the file which contains the trajectories made over CS was: 443978

Routes

OBJECTID *	routeid	linknummer	richting	snelheid	uur	weekdag
1	543604536	1550580	t	19,736647	8	4
2	543608349	1346272	t	17,393603	7	4
3	543610424	839906	f	15,083163	8	4
4	543646363	1890915	t	9,141364	8	4
5	543646363	1890914	t	7,481261	8	4
6	543646363	1890913	t	4,63015	8	4
7	543657262	392290	f	14,640018	9	4
8	543657262	391232	t	15,469775	9	4
9	543657262	1692784	t	16,917339	9	4
10	543657262	264202	f	13,199229	9	4
11	543658527	1492563	t	18,123415	11	4
12	543663229	1756802	t	17,638158	8	4
13	543686769	1999149	t	15,740869	19	3
14	543686772	1693365	t	16,045286	9	4
15	543686772	1108051	t	18,700658	9	4
16	543686772	1107834	t	20,517725	9	4
17	543686772	1360016	t	17,504387	9	4
18	543692197	99413	f	17,84928	11	4
19	543699536	1443337	f	12,772295	8	4
20	543703209	106269	t	19,050396	9	4
21	543703209	115811	t	13,446265	9	4
22	543720714	767700	f	18,011382	8	4
23	543721586	1787743	f	22,004285	7	4
24	543721586	103791	f	21,138486	7	4
25	543721586	103792	f	21,510809	7	4
26	543721586	103793	f	21,100371	7	4
27	543721586	103799	f	17,033091	7	4
28	543721586	103795	f	17,039204	7	4
29	543721586	208307	t	19,875243	7	4
30	543736646	79923	t	16,616807	9	4
31	543736646	78726	f	15,823339	9	4
32	543743633	155125	f	18,779914	12	4
33	543743633	155681	f	24,26123	12	4
34	543749012	254079	t	19,215602	14	3
35	543749021	310094	t	14,22355	6	4

1 (0 out of 8903459 Selected)

By grouping the trajectories made according to the route id's and summarizing them; it has been seen that 6936 trips are made over CS.

2. Links.shp

The second part was saved as a ESRI's shapefile. That shows a network consists of the links traveled. Within the attribute table, the unique numbers of these links was again listed. As it can be imagined, a link was traveled by many cyclists, and a cyclist was traveled through many links to reach his/her destination. To decrease the size of the data Utrecht municipal areas is selected within the links that includes the whole country.

Appendix 3: Low quality of GPS network

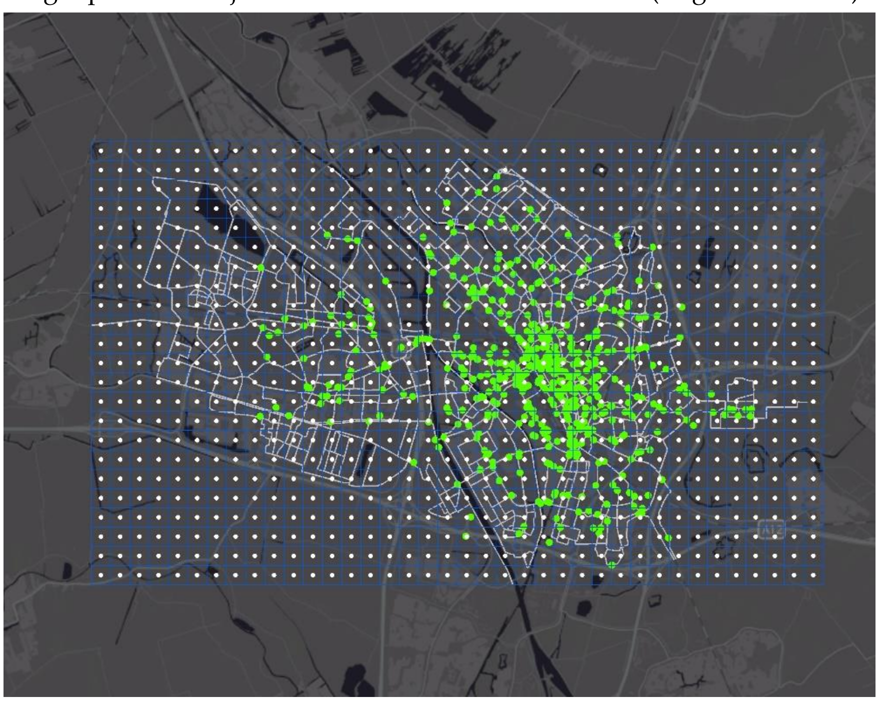
Road network from GPS data

Road network from GPS data; zoomed in

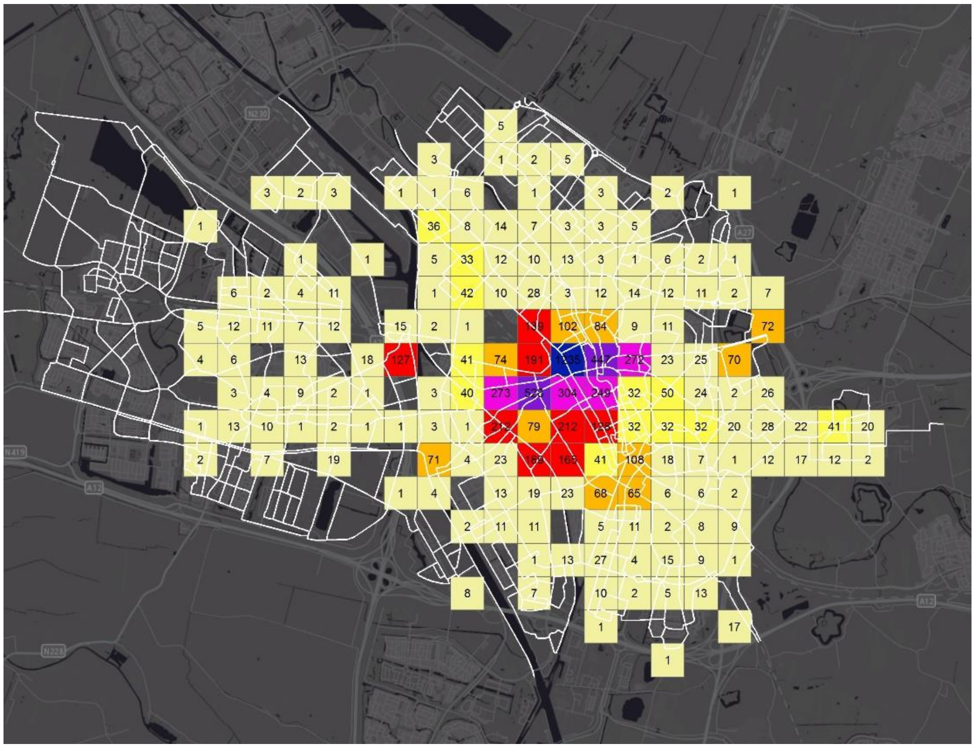
All roads of Utrecht can be seen on the network, however, no road type classification is assigned on those. The other negative side of the GPS network was that it has a very low quality of network. The road segments are not connected to each other on many spots and many other topological elements are detected. Also, the size of the data was unnecessarily big because of the way the network constructed. The links between nodes were divided into too many segments which were making the data big and complicated. Using such network in an agent-based model would not be practical and would cause many errors.

Appendix 4: Decreasing the number of agents

Origin points of trajectories made over Central Station (original number)



Fishnet grid created to divide points



Number of origins fall into squares

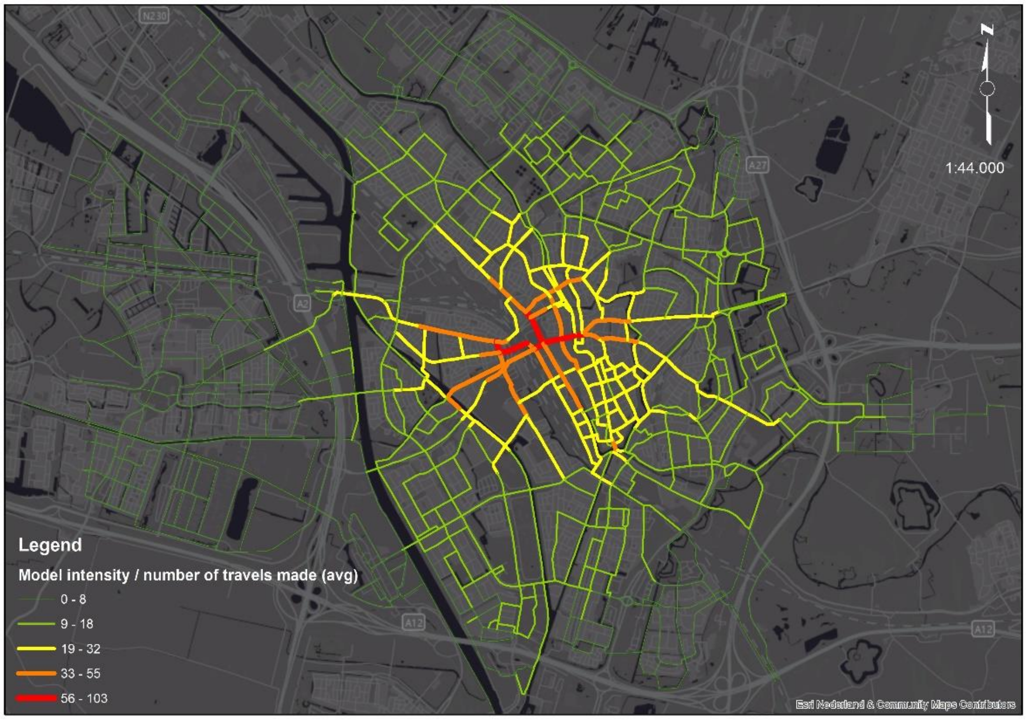
20% of the number of origins

Decreased number of origins (350)

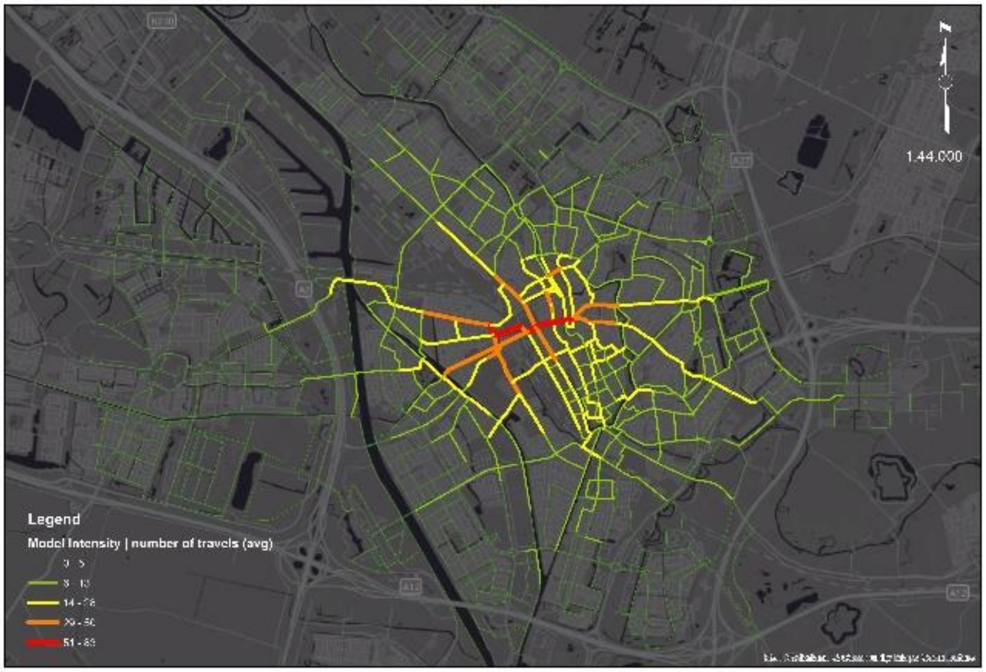
Same methodology that is represented by the previous figures in Appendix 4, has been applied for destination points as well. The result is given by the below figure.

Decreased number of destinations (350)

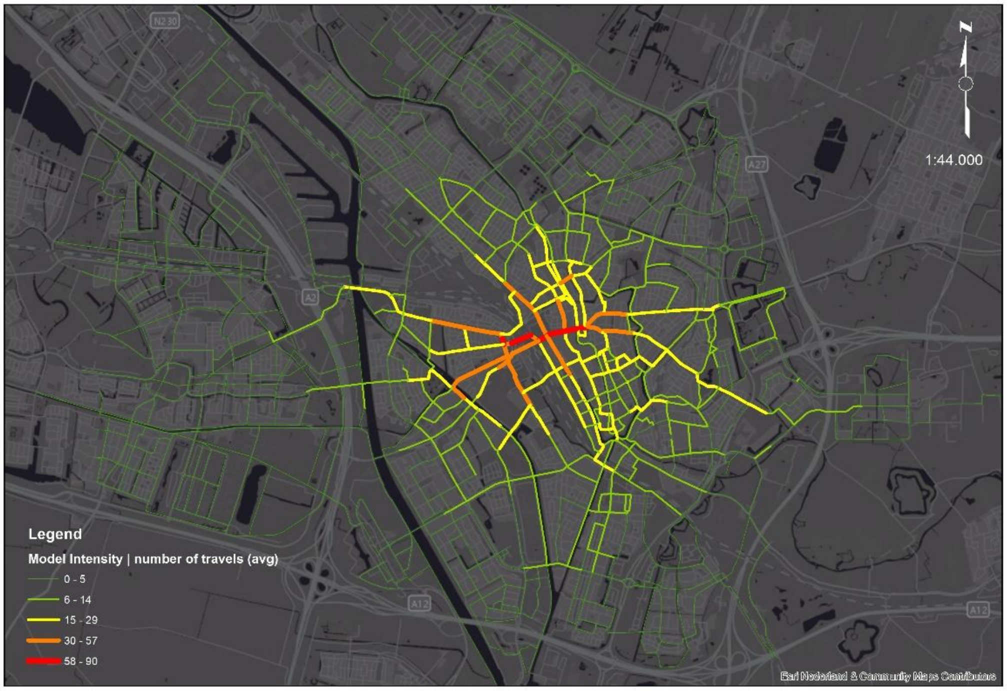
Appendix 5: Sensitivity Analysis maps



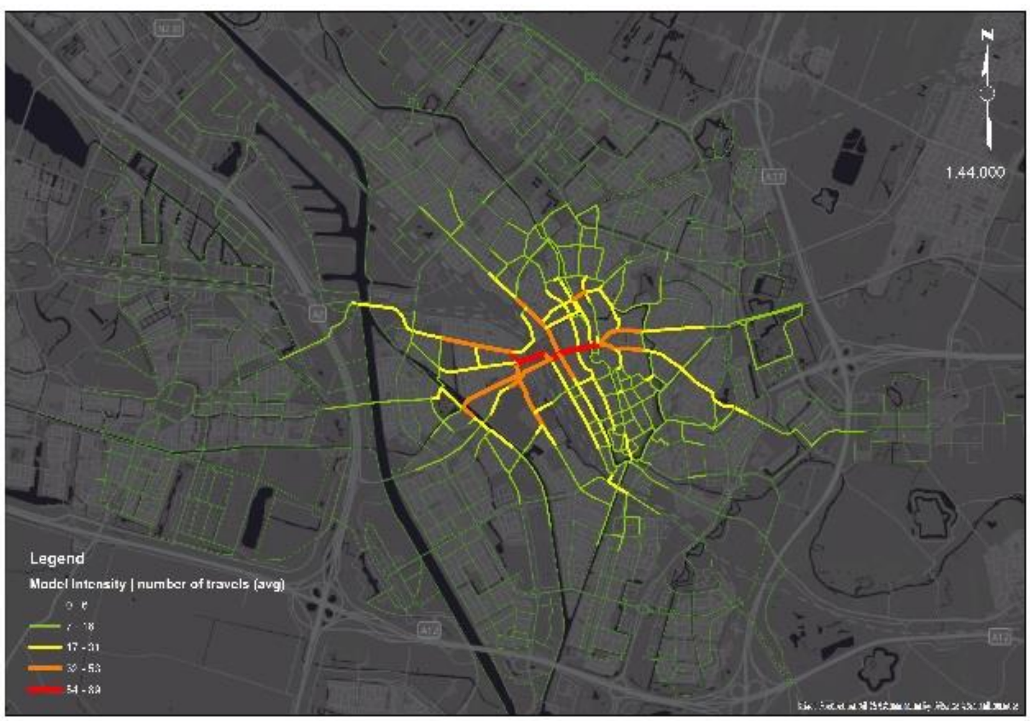
Sensitivity analysis; shortest-path distance with 0.25



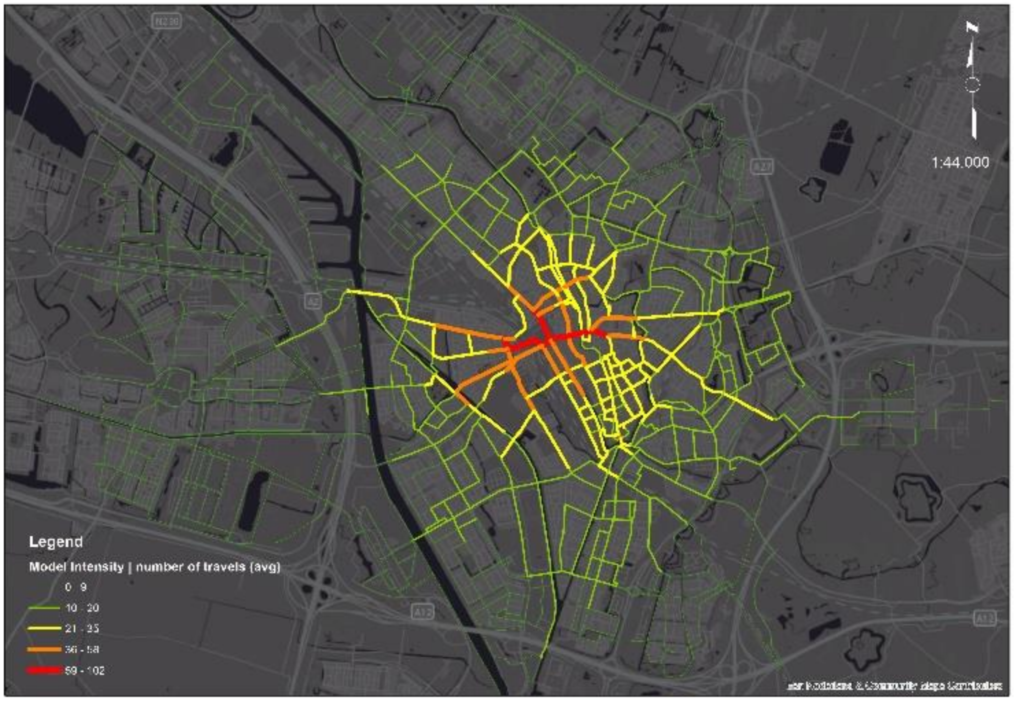
Sensitivity analysis; shortest-path distance with 0.5



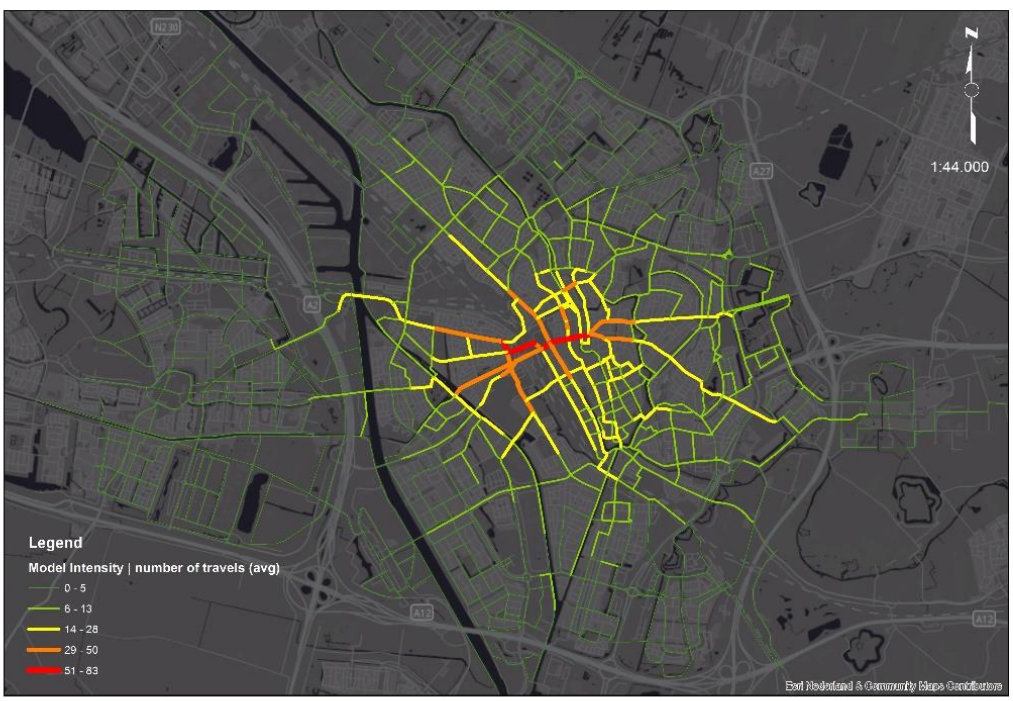
Sensitivity analysis; shortest-path distance with 0.75



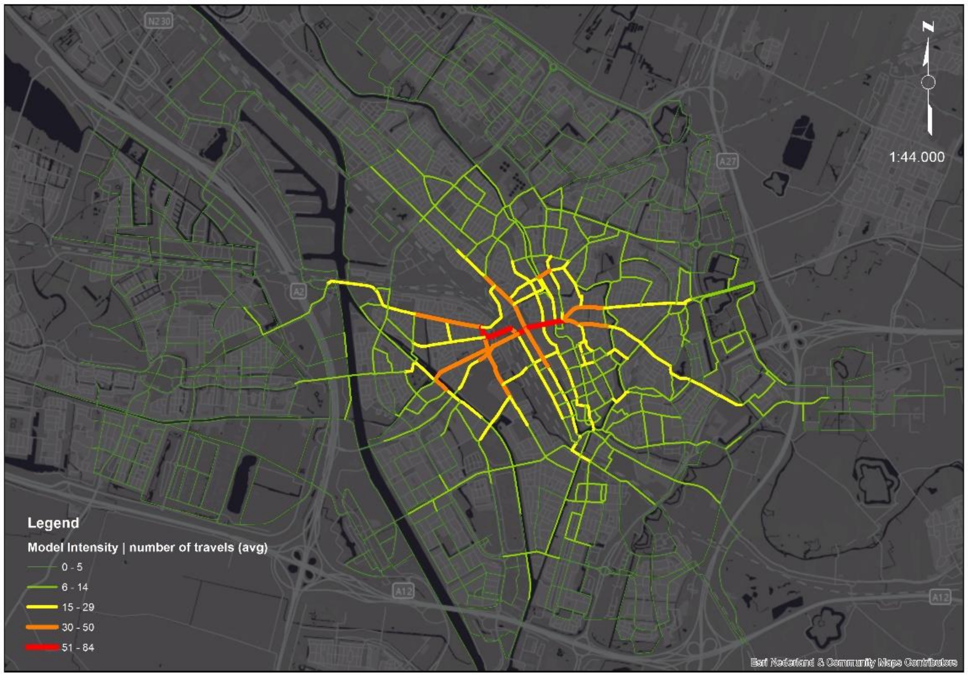
Sensitivity analysis; shortest-path distance with 0.9



Sensitivity analysis; straight-line distance with 0.25

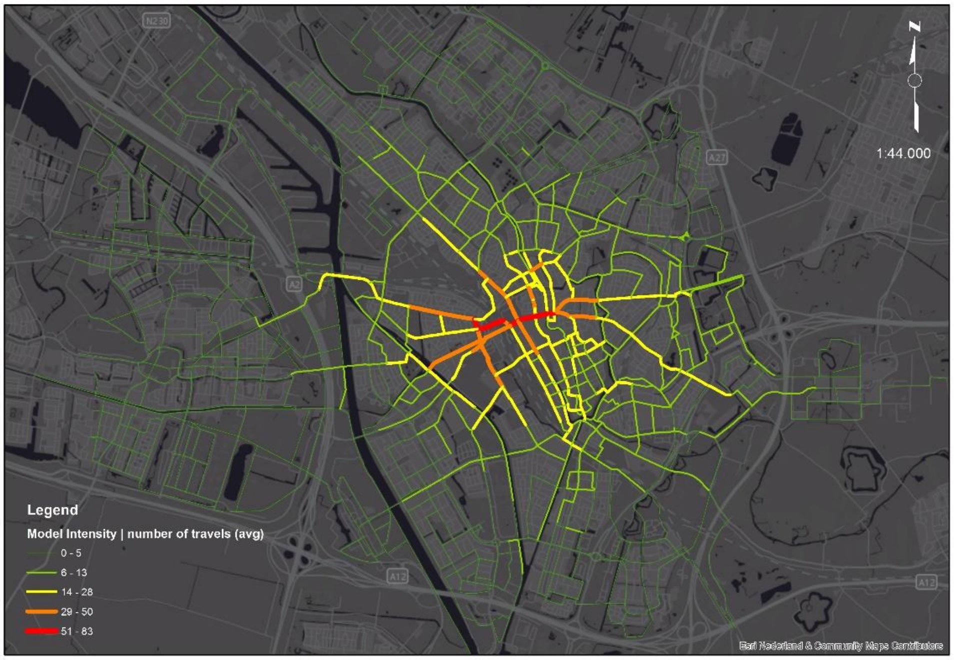


Sensitivity analysis; straight-line distance 0,5

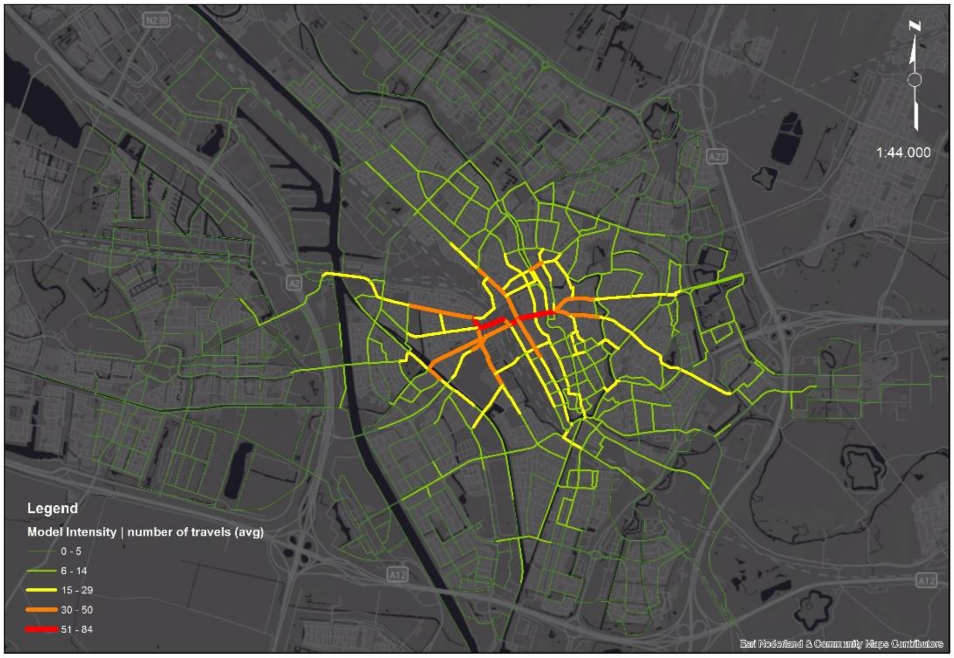


Sensitivity analysis; straight-line distance 0,75

Sensitivity analysis; road type with 0.25



Sensitivity analysis; road type with 0.5



Sensitivity analysis; road type with 0.75

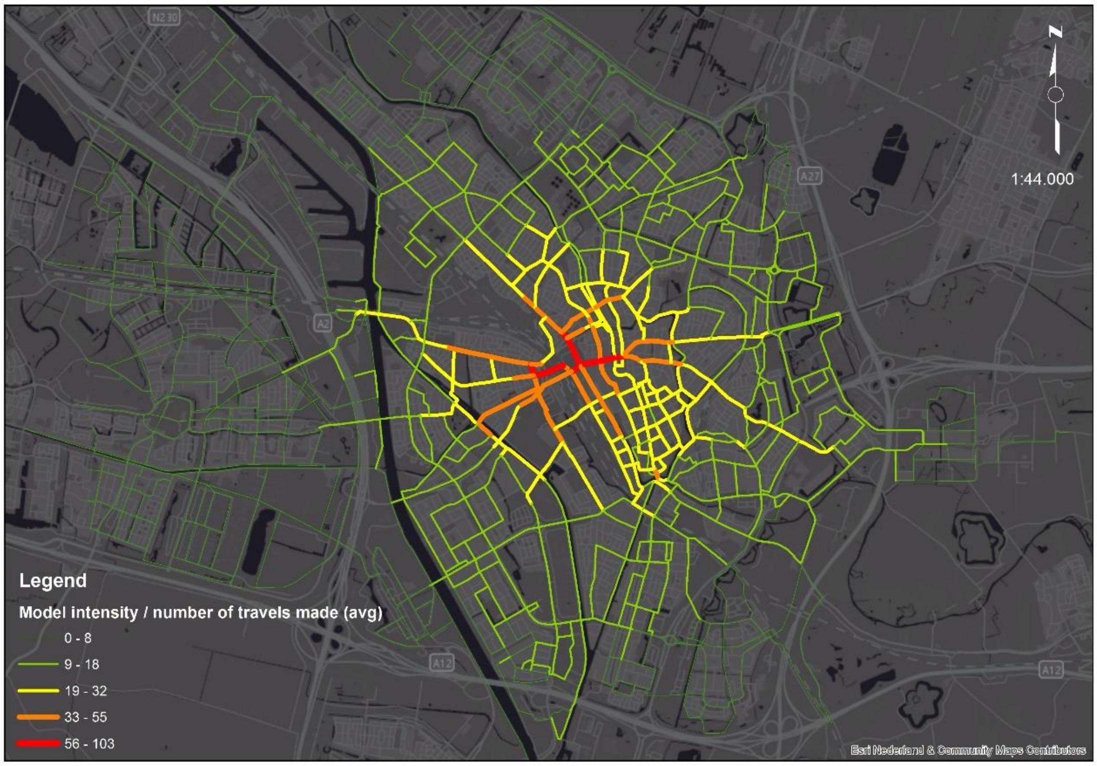


Figure 39 Sensitivity analysis; shortest path distance with 0.25

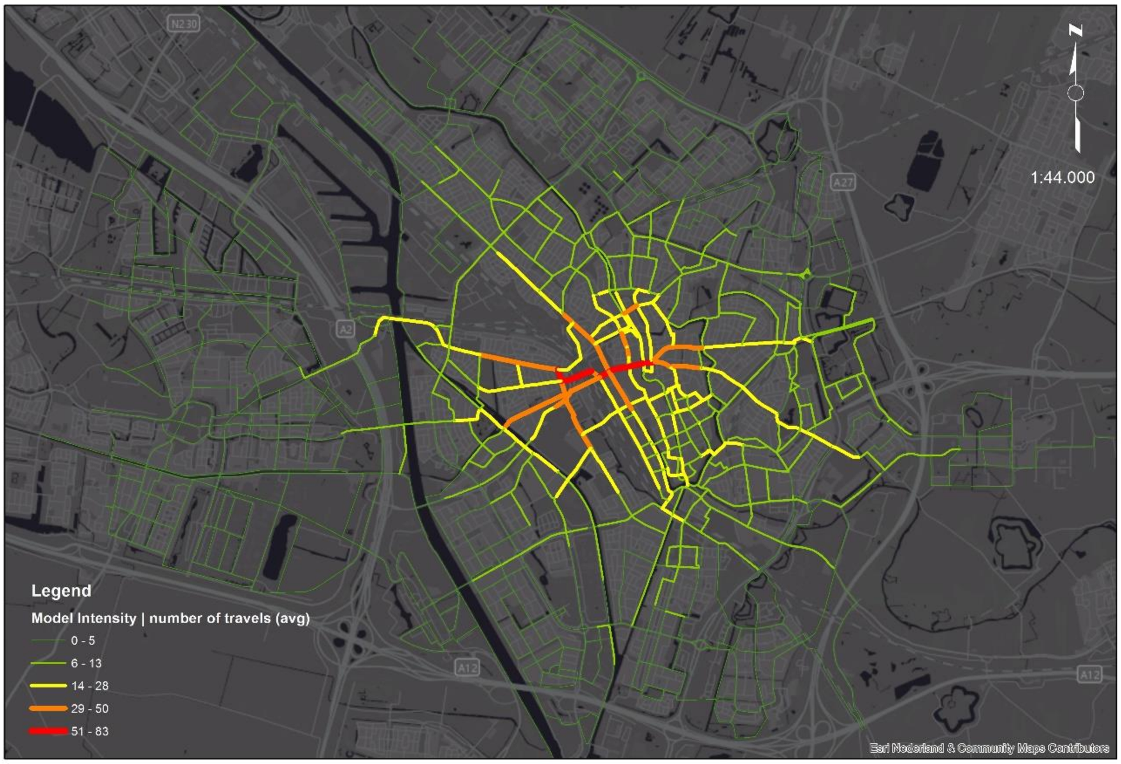


Figure 40 Sensitivity analysis; shortest path distance with 0.5

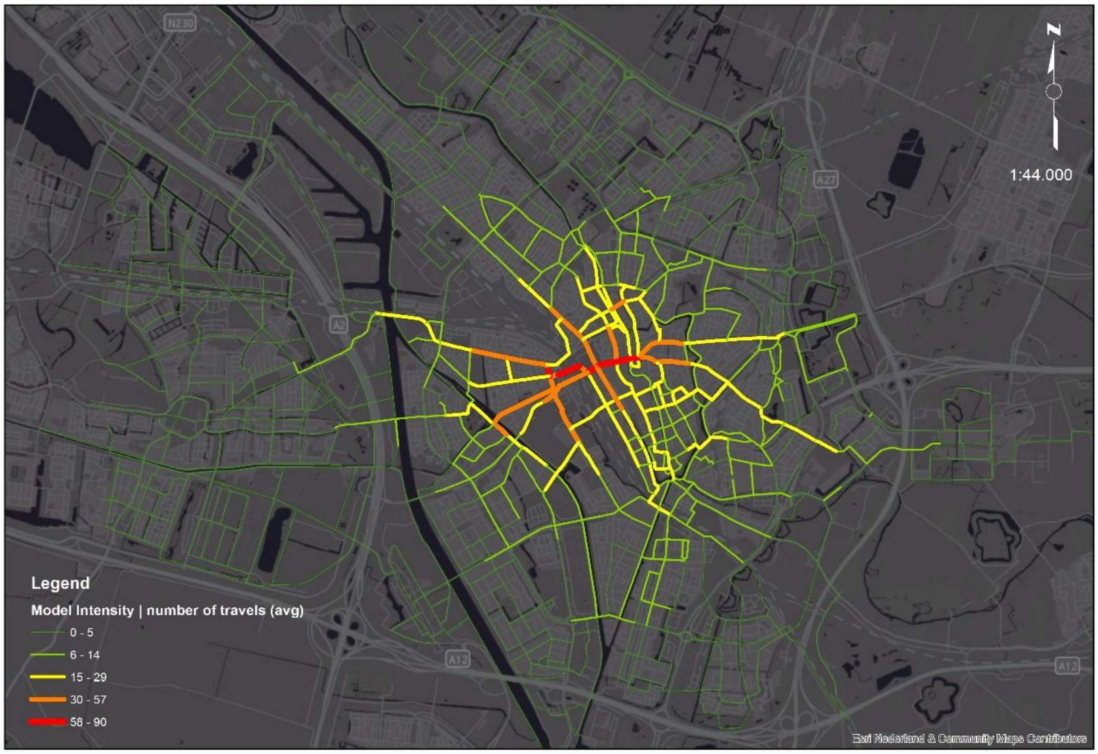


Figure 41 Sensitivity analysis; shortest path distance with 0.75

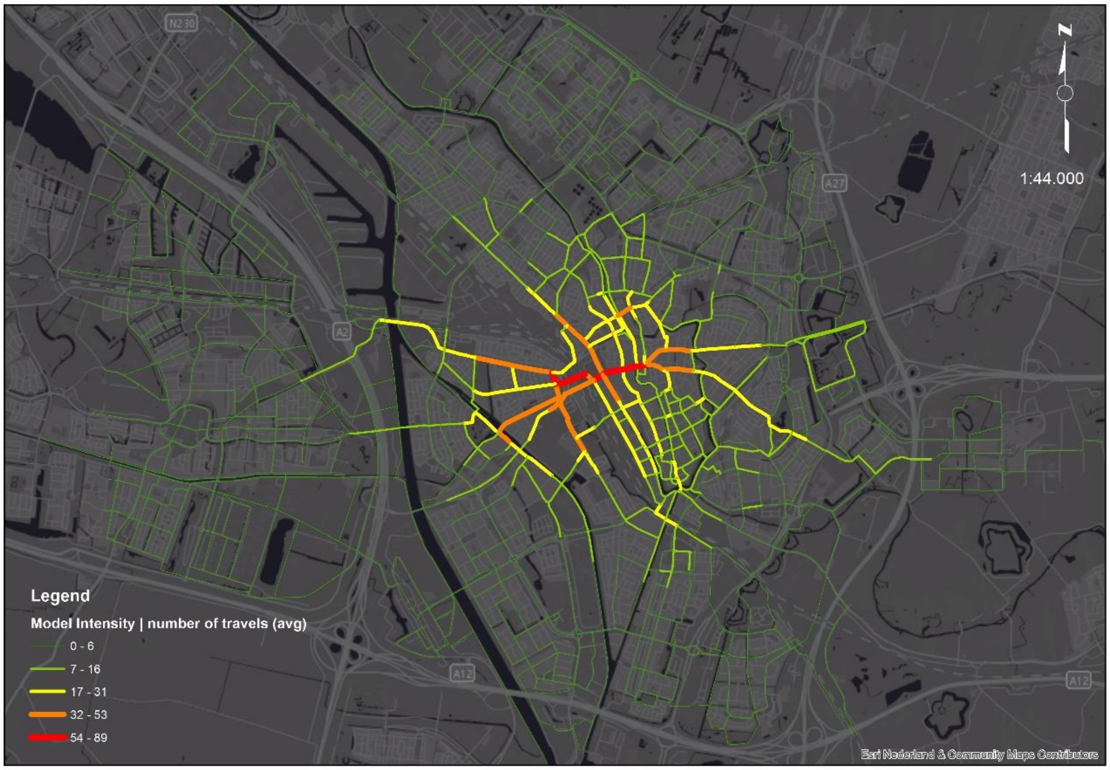


Figure 42 Sensitivity analysis; shortest path distance with 0.9

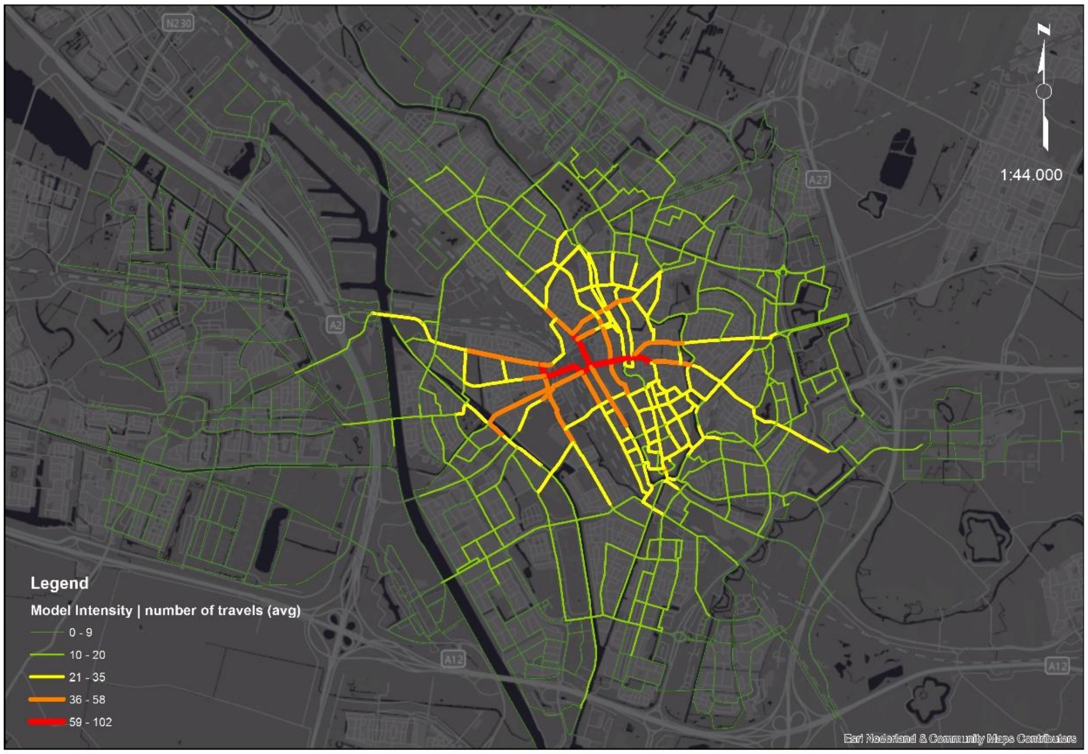


Figure 15 Sensitivity analysis; direct angle distance with 0.25

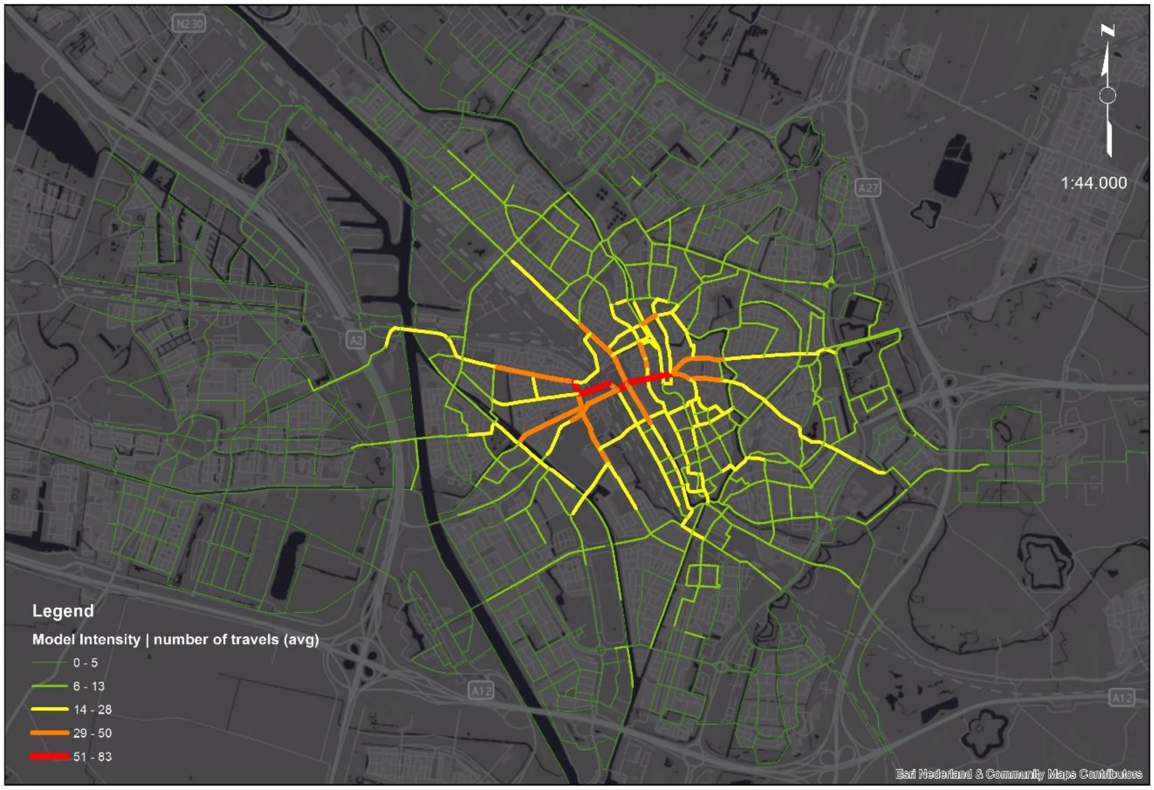


Figure 43 Sensitivity analysis; direct angle distance 0,5

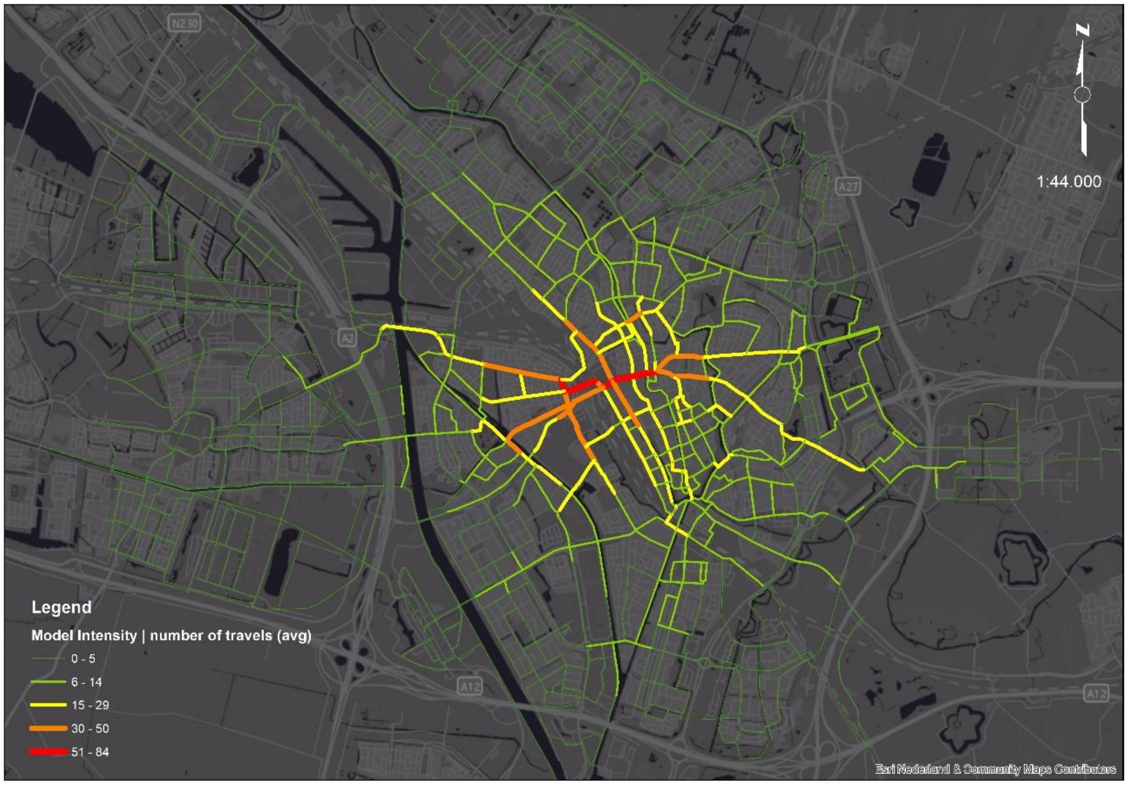


Figure 44 Sensitivity analysis; direct angle distance 0,75

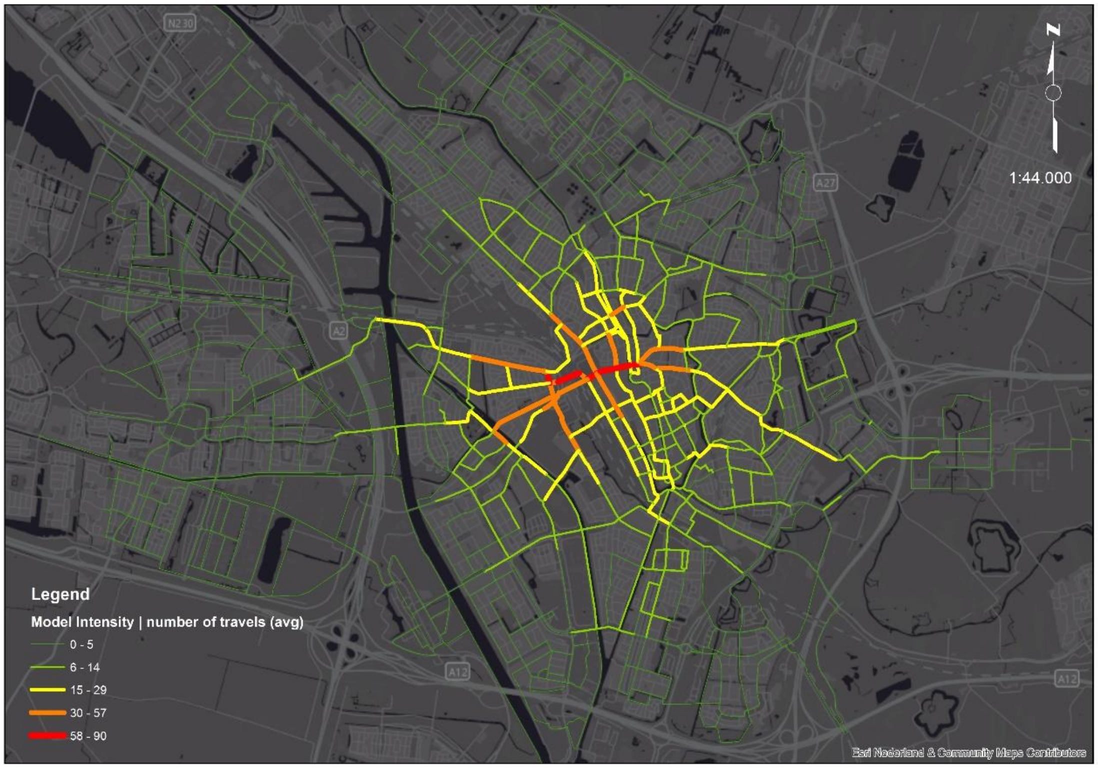


Figure 45 Sensitivity analysis; road type with 0.25

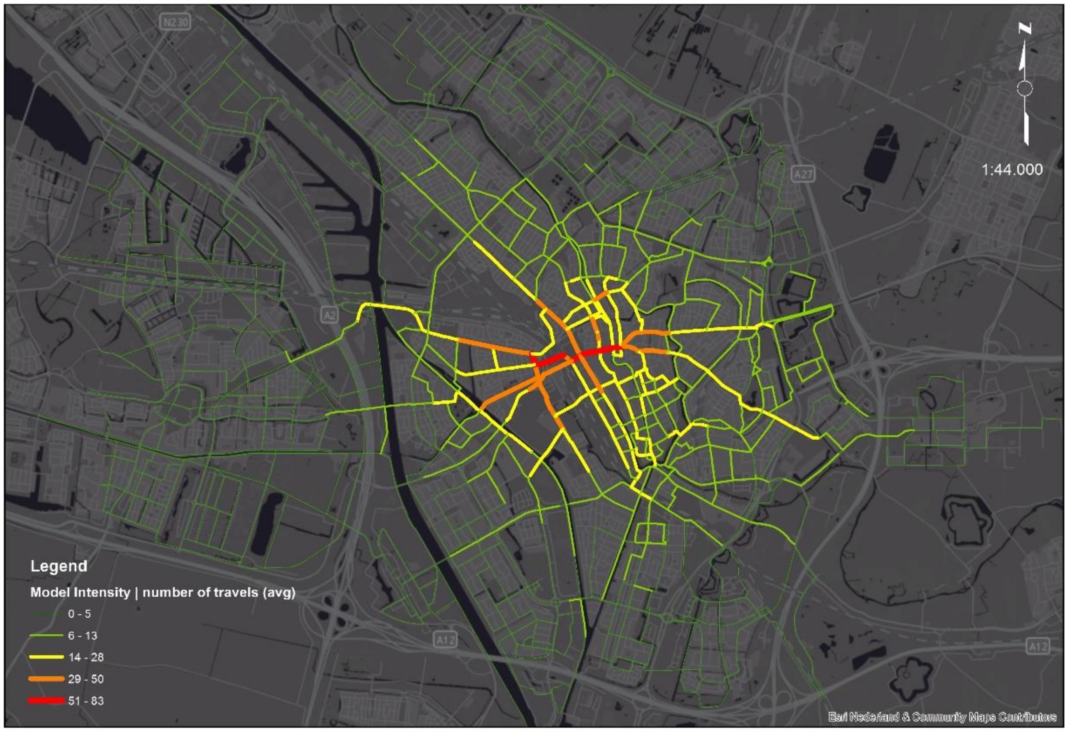


Figure 46 Sensitivity analysis; road type with 0.5

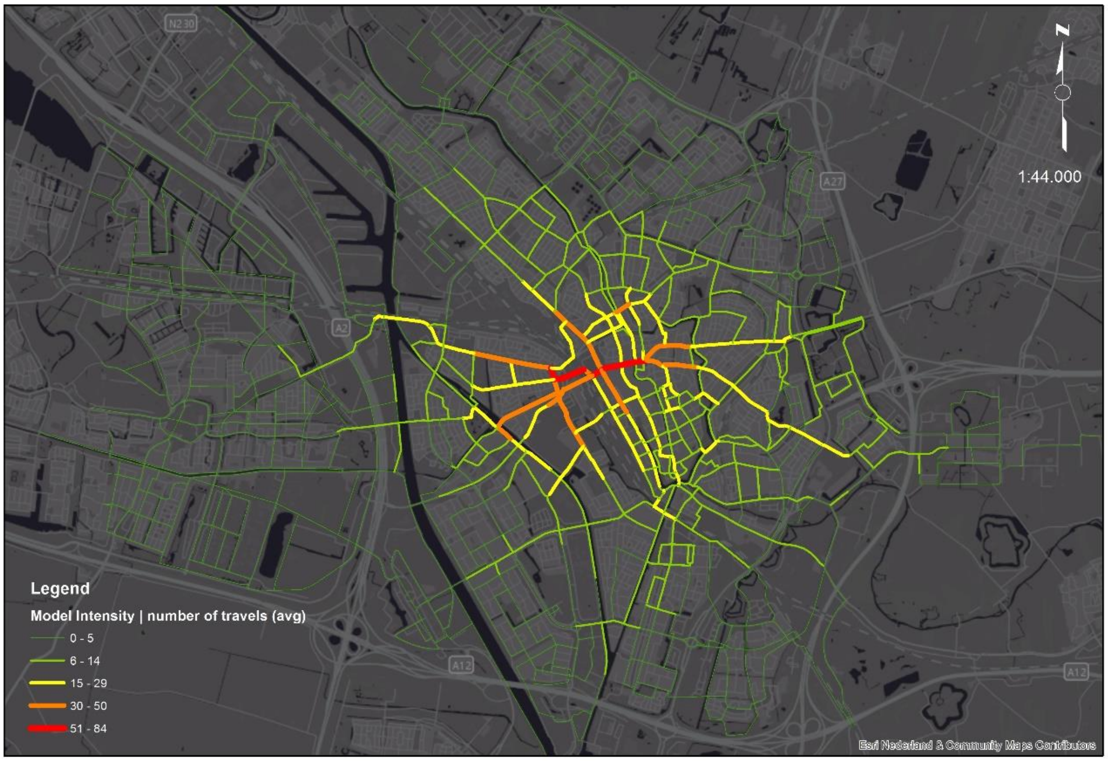
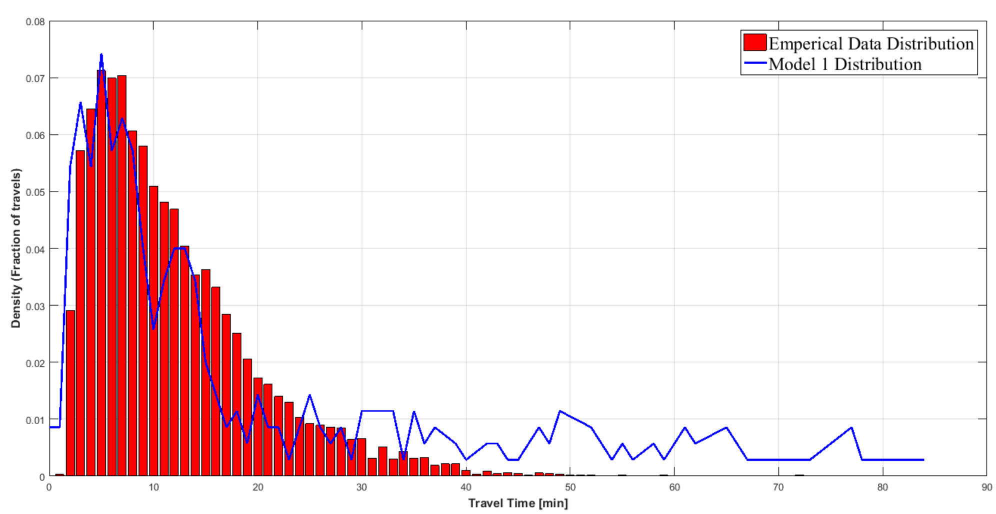
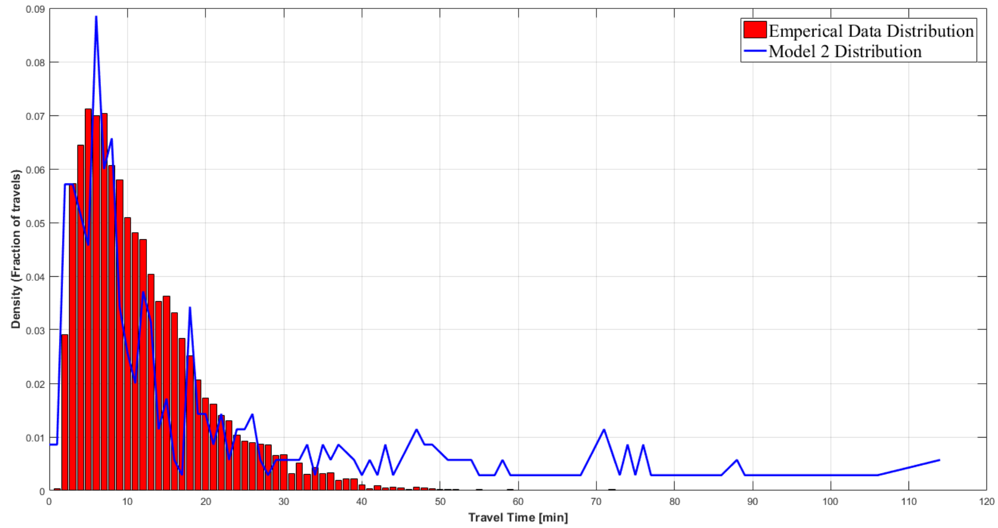
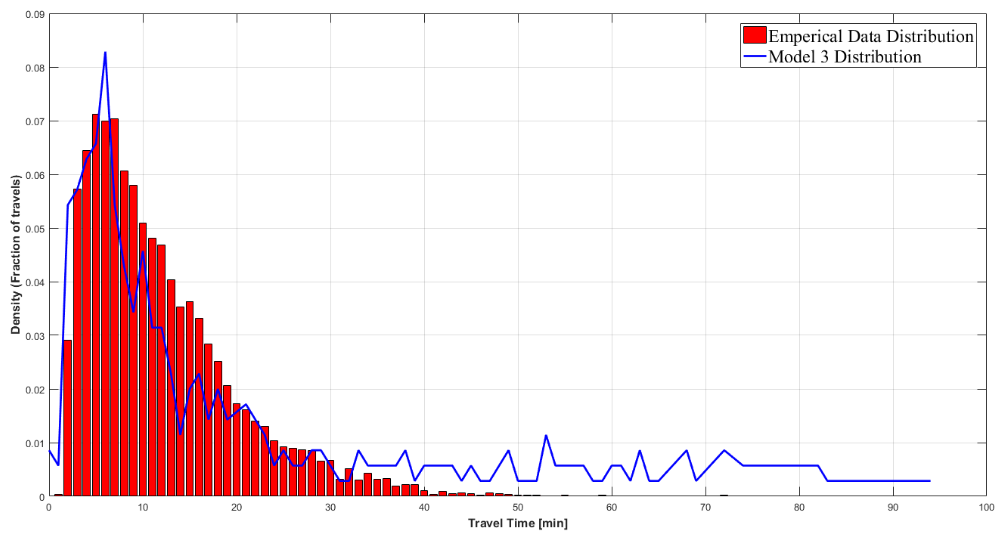


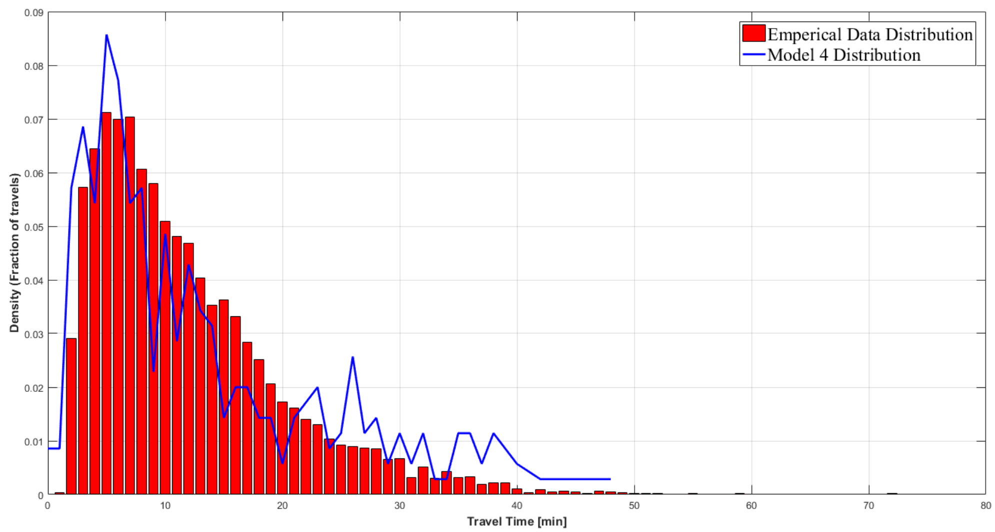
Figure 47 Sensitivity analysis; road type with 0.75

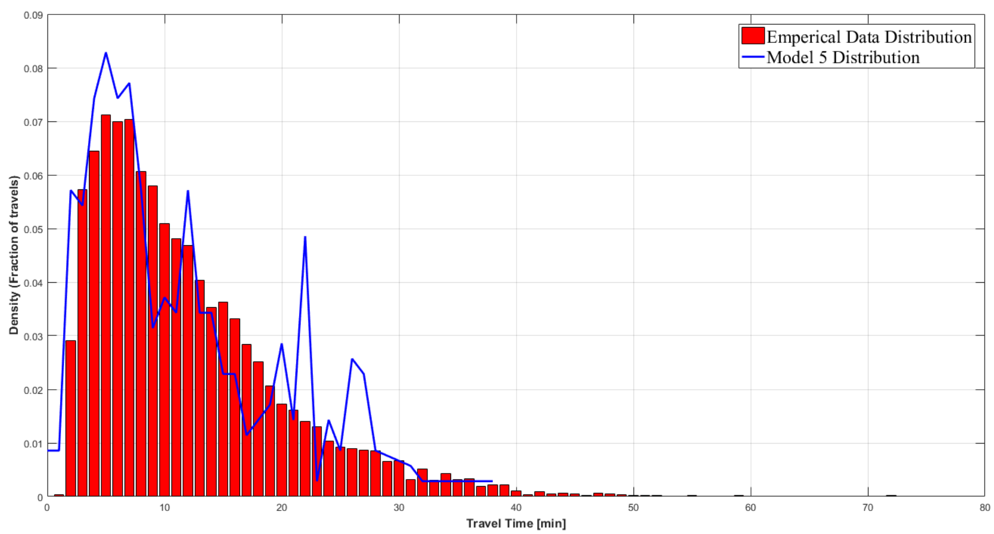
Appendix 6: Travel time distribution graphs of calibration models




Emperical Data Distribution
Model 2 Distribution
Density (Fraction of travels)
Travel Time [min]
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0
10
20
30
40
50
60
70
80
90
100
110
120


Emperical Data Distribution
Model 3 Distribution
Density (Fraction of travels)
Travel Time [min]
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0
10
20
30
40
50
60
70
80
90
100


Emperical Data Distribution
Model 4 Distribution
Density (Fraction of travels)
Travel Time [min]
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0
10
20
30
40
50
60
70
80


Emperical Data Distribution
Model 5 Distribution
Density (Fraction of travels)
Travel Time [min]
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0
10
20
30
40
50
60
70
80