Assignment of cyclists in the Netherlands

Improving the assignment of cyclists in traffic models

Jesse Voorhorst
Goudappel Coffeng

16 April - 6 July 2018
Supervisor University of Twente: O. Eikenbroek
Supervisor Goudappel Coffeng: B. Possel

Preface

In front of you lies my bachelor thesis called *Assignment of cyclists in the Netherlands* which I wrote as part of finishing the civil engineering programme at the University of Twente. At the end of 2017, I approached Goudappel Coffeng about possible assignments and one of them appealed to me: modelling cyclists. Models of other modes are elaborate and well studied, but cyclists are often forgotten. That sounded like a good assignment.

Consequently, I worked on this thesis about assigning cyclists from mid April 2018 to the start of July 2018 at Goudappel Coffeng in Deventer. I enjoyed working at the company and all colleagues were very welcoming. Employees from different teams were always prepared to assist with their knowledge and skills and I would like to thank them for that. In particular, I would like to thank Bastiaan Possel, my supervisor, and Rogier Koopal, who was the acting supervisor while Bastiaan was away. I would not have been able to finish this thesis without their help.

Furthermore, I thank Oskar Eikenbroek, my university supervisor, for his guidance and ideas about the research setup. He was always prepared to give useful feedback and it really helped me in improving the thesis. My peer-review partner, Michiel, helped me a lot with finding alternative ways to phrase explanations. Finally, I want to thank my family for supporting me in the process and my brother for help with creating the front page.

I hope you will enjoy reading this thesis.

Jesse Voorhorst, Zwolle, 5 July 2018

Image front page: Pixabay (2015)

Table of contents

Abstract

In traffic models, cyclists do not always get full attention from the model developers. In most cases an all-or-nothing assignment is used, while previous research showed that cyclists with the same origin and destination take different routes. This research tried to find an alternative assignment algorithm that better describes the routes of cyclists. First of all, a literature review and expert opinions were used to find the most important factors that cyclists take into account in the Netherlands. Based on these results, possible routes were generated using Monte Carlo simulation and Dijkstra's shortest path algorithm. The routes were assigned with a logit model in which the costs depended on characteristics of the routes. The considered characteristics included road type, speed and intersection types. The parameters were calibrated on a detailed network of Tilburg and the surrounding area and validated on a network of The Hague, Delft and the surrounding area. The results showed that this assignment performs a little better on the t-test and GEH-statistic than the all-or-nothing assignments in both networks, but it was able to distribute the cyclists more according to preferences for certain road and intersection types. For example more cyclists were modelled on routes avoiding intersections with traffic signals.

1. Introduction

1.1 Context

The Netherlands are a country famous for its bike use. In 2016, 27% of the trips in the Netherlands were made by bicycle and on average, each inhabitant cycles 2,5 kilometres per day (Centraal Bureau voor de Statistiek, 2017a). The use of cycles has increased since the start of the 21st century. Between 2005 and 2016, bicycle use has increased with up to 3,5% in Dutch cities and bicycles have become the most used transport mode in Amsterdam and Utrecht (Kennisinstituut voor Mobiliteitsbeleid, 2017b). The total distance travelled by bike went up from 14 billion kilometres in 2005 to 15,5 billion kilometres in 2016 (Kennisinstituut voor Mobiliteitsbeleid, 2017c). This increase can be explained by the change of the characteristics of the inhabitants: more students and younger people live in cities like Amsterdam and Nijmegen. E-bikes make it possible to cover larger distances and to stay more mobile when becoming older (Kennisinstituut voor Mobiliteitsbeleid, 2017a).

This increase will probably continue during the following years as the Dutch government is trying to increase the cycle use for short trips to reduce health problems like obesity, to improve the air quality and to reduce congestion on the roads. It tries to achieve this by investing in high-speed cycle paths and creating more parking facilities and cycle paths (Rijksoverheid, 2017). The increased use of the bicycle in the cities can cause problems. For example, on some roads in Nijmegen, traffic jams of cyclists occur ("Nijmegen wil af van fietsfiles," 2016). For these reasons, it becomes more important to have a good method that can predict the number of cyclists on new routes and can help when making policy.

To help in making these policies or design decisions, models can be used. A much-used model is the four-step model which uses four steps to determine the number of vehicles on each road in the network. Advanced methods have been developed for assigning cars and motorised traffic to roads, but cyclists have not been included in them for many years. At this moment, in most cases cyclists are assumed to take the shortest route, but in reality, lots of other factors play a roll as well.

In recent years, multiple studies have been done trying to describe the route choice behaviour of people (see chapter 2) based on surveys or GPS traces, but they have not been translated into an assignment algorithm. Furthermore, most research is focussed on cities abroad, where cycling is not as popular as in the Netherlands. This means there are no cyclist assignments developed in order to describe cyclists' route choice as good as possible. This research tries to find an assignment for cyclists based on the characteristics of how cyclists choose their route.

This research took place at Goudappel Coffeng. Goudappel Coffeng uses models to advise their clients, mostly governments in the Netherlands, about traffic solutions and to predict what effect measures will have. As described, the use of bicycles will increase and they want to improve the handling of cyclists in their transport models to improve their advice. The need for better models is highlighted by the clients who ask specifically for better advise about infrastructure for cyclists. Goudappel Coffeng uses mostly the OmniTRANS software package that they developed themselves. This is based on the standard four-stage model.

Overall, it becomes clear that the routes of cyclists have to be modelled in more detail in order to predict where problems arise and what the usage of new cycle routes will be. This research will therefore focus on better determining the routes of cyclists in order to improve the four-stage model of Goudappel Coffeng for their clients in the Netherlands.

1.2 Research aim and questions

As explained in the previous section, because of an increasing number of cyclists and an increasing interest in cycling from governments, a better assignment of cyclists is needed in four-stage models like OmniTRANS uses. The aim of this research can be summarised as follows:

The aim of this research is to improve the accuracy of the route assignment of cyclists in four-stage models for use in the Netherlands.

From the literature discussed in chapter 2 it becomes apparent that there is no real agreement on what the most important factors are for cyclists when they choose their route. Furthermore, no assignment and little research has been performed about situations in the Netherlands. Based on these results, the following main research question has been formulated:

How can the assignment of cyclists in a four-stage model realistically be modelled for use in the Netherlands?

Realistic can be explained in many different ways. In this case it will be interpreted as closeness to reality. This was measured by comparing the modelled number of cyclists on a certain road with the observed number of cyclists on the same route. In chapters 4 and 5 about calibration, the method is explained in more detail.

This main question was answered by the following subquestions:
1. What are the most relevant factors in the route choice in the Netherlands?
2. What methods do exist for the assignment of cyclists in a four-stage model?
3. What is the most useful method for assignment in the Netherlands?
4. What combination of variables does describe the route choice in the Netherlands best?
5. How well does the all-or-nothing assignment perform in the Netherlands?
6. To which extent is the proposed algorithm an improvement?

1.3 Relevance

This research is relevant because of different reasons. First of all, there is more interest in cycling from governments as a solution to congestion and health problems. In order to make good measures, it can be useful to predict the number of cyclists on a new or improved road. With the current methods, these predictions are not accurate enough as they do not reflect the route choice of cyclists good enough. This research tried to improve the predictions by creating a new algorithm.

Different assignment algorithms have been developed and tried before, but these did not use real data to calibrate it and did not validate at all. Furthermore, multiple studies have been conducted about the route choice behaviour of cyclists, but these were only used for describing the choice of cyclists. These authors did not do an assignment at all. This research will combine these types of studies, which was not done before. The results from the route choice studies were used to determine the most relevant factors that influence the route choice behaviour. These were incorporated in a model and the model was calibrated and validated using real data. The next section describes in more detail the methodology of this study.

1.4 Methodology

In order to find the answer to these questions, a number of steps have been taken. This methodology greatly depended on the available data, so first the data will be explained shortly. For this research, three networks were available of ’s-Hertogenbosch, Tilburg and The Hague/Rotterdam. The Tilburg and The Hague/Rotterdam networks contained information about possible factors that influence the route choice of cyclists. Examples include the road type, surrounding land use and intersection type. Furthermore, these networks included cyclist counts on many roads within the network. In section 2.3 more information can be found about the available data and its quality.

First of all, the factors that influence the route choice of cyclists were determined by reviewing the available literature. This was expanded with interviews with cycling experts working at Goudappel Coffeng. To find which factors are applicable to the Netherlands the factors mentioned in the literature and by the experts were compared. Of each study, the most and least important factors were summarised while taking the quality and relevance in the Netherlands into account. The experts were asked to rank the factors and if possible, to mention some factors that they think are not relevant. Based on all results, the most important factors that describe the cycling route choice in the Netherlands were selected in order to implement them in the algorithm. This selection also took into account about which factors data is available.

To answer the second subquestion about the available methods to assign cyclists, a literature review was used as well. An overview of the different methods and their properties was created of previously used methods in peer-reviewed journals. These were analysed and their stronger and weak points were described.

Following the exploration of different algorithms, a suitable algorithm was selected. The found algorithms were scored on a number of criteria like speed of simulation, speed of implementation, expected accuracy and possibility to incorporate the relevant factors from the first subquestion. Each option received a score on those criteria and the most useful method was selected.

Subsequently, the selected algorithm was implemented in the OmniTRANS software using the Ruby programming language. Two separate algorithms were implemented: one for the generation of routes and one for the assignment itself. The route generation was calibrated using visual comparisons between a map of the area and the generated routes between certain OD-pairs (a combination of an origin and a destination). Additionally, routes between the origins and destinations were planned using the *Fietsersbond* route planner (Fietsersbond, 2018) and compared to the generated routes. The assignment itself was calibrated using the t-test and GEH-statistic to compare modelled cyclist numbers with cyclist counts. These are standard measures within Dutch and international traffic studies, respectively. The t-test and GEH-statistic treat a 10% deviation for example differently if the number of cyclists is 6 or 600, while other measures do not (e.g. MSE or percentage of deviation).

The algorithms were calibrated using the Tilburg network that has great detail in its road network and contains multiple cyclists counts from 2015. The route generation was additionally tested on the network of neighbouring 's-Hertogenbosch, because this network is smaller and was therefore quicker and more useful for testing purposes. The calibration used a OD-matrix (matrix with the number of cyclists between all origin and destination zones) of the morning peak hour.

The performance of the all-or-nothing assignment was tested using the Tilburg and The Hague networks. The closeness to reality was measured with the t-test and GEH-statistic, and with the same counts used for the new proposed algorithm. The results were compared against the standard quality requirements for model performance belonging to the used tests (see section 5.2).

Finally, the parameter values that resulted from the calibration were validated using the The Hague/Rotterdam network. The generated routes were again compared with a map and routes in the route planner. The results of the assignment were compared with the counts using the GEH-statistic and the t-test. The outcomes of the t-test and GEH-statistic were compared with the performance of an all-or-nothing assignment applied to the same network and to performances of other algorithms on other locations as found in literature.

1.5 Reading Guide

This thesis starts with an exploration of the factors that influence the route choice of cyclists in chapter 2. That chapter includes a literature review, opinions of experts and a comparison between all previous research in order to establish the most important factors. The third chapter focusses on the choice of algorithm. Again, literature has been studied, and the different options are scored. Chapter 4 describes the calibration of the first part of the algorithm, the route generation, while chapter 5 describes the calibration of the assignment itself and the performance of the all-or-nothing assignment. Furthermore, a validation has taken place which is described in chapter 6. Finally, chapters 7, 8 and 9 discuss the results, draw a conclusion and give recommendations for further research.

2. Route choice of cyclists

Cycling from home to work, there are in most cases many different options to reach your destination. Some are quick, some are more comfortable, others have a different landscape or are perceived as safer. The choice of the best route in your situation depends on a great number of factors including the goal of your trip and the time. This chapter discusses which factors play a role in choosing a route and tries to establish the most important factors that can be used in the improved algorithm. The first section starts with a description of the most and least important factors according to the literature and assesses the quality of the research and applicability in the Netherlands. This is followed by expert opinion about what factors are most important in the Netherlands in the second section. The third section gives a description of the available data and its quality and limitations and in the final section of this chapter, the most useful factors were selected for implementation in the algorithm, based on the results of the literature review and the available data.

2.1 Comparison of factors in literature

As mentioned in the introduction, there is an increasing interest in the route choice behaviour of cyclists. Two main types of research are used to establish the most important factors: stated-preference (SP) and revealed-preference (RP) studies. The first type uses surveys to ask people what they think is most important when choosing a route, while the second method observes past behaviour. Most research is stated-preference, based on surveys in which people can choose between two or more hypothetical routes with different properties. During last years, more revealed-preference research based on GPS traces is conducted.

Much of this research is conducted abroad and consequently not always applicable to the Dutch situation, because of differences in cycling culture and infrastructure. Furthermore, stated-preference research is less useful for predicting the real choice of people and might only give an indication. This section compares the results of the studies and their quality. In table 2.1 the most and least important factors per study are shown together with some characteristics of that research, like the type of study and location. The least important factors are included as well to indicate which factors should not be implemented in an algorithm, because they do not seem to have any influence.

Some studies in table 2.1 distinguish between different groups of cyclists. In case the most important factor is different between the groups, both most important factors are mentioned. Furthermore, the influence of the factor on the utility of a route is mentioned if data is available. This additional information is shown with notes. In table 2.1 “n-c” means non-commuter and “c” means commuter. “sd” stands for short-distance commuters and “ld” for long distance commuters. The colour of the text indicates the effect on the utility of that factor: factors in red have a negative effect, while the factors written in green increase the utility according to the study. The Dutch assignment includes all types of cyclists, but commuters and persons on their way to school account for more kilometres than recreative cyclists (Centraal Bureau voor de Statistiek, 2017b), so if needed, the focus is on that group.

Table 2.1. Most and least important factors according to different studies. Red indicates a negative influence on the utility of a route, green indicates a positive influence on the utility of the route.

	Broach et al. (2012)	Hood et al. (2011)	Hunt & Abraham (2007)	Sener et al. (2009)
Most important factor	proportion up-slope gradient ≥ 6% (n-c) / proportion of roads with traffic volume ≥ 30000 without bike lanes (c)	proportion wrong-way roads	secure parking at destination or origin	Heavy traffic volume
Second most important factor	proportion of roads with traffic volume ≥ 30000 without bike lanes (n-c) / proportion up-slope steeper ≥ 6% (c)	proportion bike lanes	time in mixed traffic	> 5 stop signs or traffic signals
Third most important factor	proportion up-slope 4-6% (n-c) / proportion of roads with traffic volume 20-30k without bike lane (c)	proportion bike paths	time riding on bike paths shared with pedestrians	High speed limit > 35 mph on road (short-distance) / continuous cycle facilities (long-distance)
Third least important factor	right turn on unsignalised crossing	average up-slope	time on designated lanes*	moderate hills (sd) / moderate parking turnover (ld)
Second least important factor	traffic signals	proportion bike routes		Angle on-street parking (sd) / moderate hills (ld)
Least important factor	stop signals	number of turns per km		no bike lane
Type of study	GPS traces	GPS traces	Stated-preference survey	Stated-preference survey
Location	Portland, Oregon, USA	San Francisco, California, USA	Edmonton, Alberta, Canada	Texas, USA
Types of groups distinguished in research	commuters and non-commuters	none	none	long and short-distance commuters

* : This research took less factors into account than the other studies, therefore no three least relevant factors could be found.

Table 2.1 continued.

	Vedel et al. (2017)	Bernardi et al. (2017)	Menghini et al. (2010)	TfL (2012)
Most important factor	cycling track	purpose leisure	maximum rise	journey time
Second most important factor	distance	beautiful shortest route	bike paths	mandatory cycle lane
Third most important factor	segregated cycle path	little traffic nuisance	average speed	advisory cycle lane
Third least important factor	number of stops	types of link (not significant)	number of traffic lights	bus lane
Second least important factor	large roads			residential street
Least important factor	crowding on road			major road
Type of study	stated-preference survey	GPS traces	GPS traces	stated-preference survey
Location	Copenhagen, Denmark	Netherlands	Zürich, Switzerland	London, UK
Types of groups distinguished in research	none	different purposes of trip	none	none

Table 2.1 continued.

	Claasen & Rienstra (2017)
Most important factor	fastest route
Second most important factor	good surface quality
Third most important factor	safety of cycle route
Third least important factor	no slopes
Second least important factor	low number of turns
Least important factor	no tunnels
Type of study	stated-preference survey without choice model (ranking of factors), non-peer reviewed conference paper

	Claasen & Rienstra (2017)
Location	Utrecht, The Netherlands
Types of groups distinguished in research	none

The only peer-reviewed research that was conducted in the Netherlands is Bernardi et al. (2017), which makes it difficult to draw conclusions about the most important factors in the Netherlands. The areas that were studied in the other researches are probably not very applicable to the Netherlands, because cycle use and the quality of the networks in those cities is much lower. That could implicate that cycling is not as safe as in the Netherlands which might influence the route choice of cyclists. The only city that has similar cycle use and facilities is Copenhagen. The modal share of cyclists is for example comparable to the Dutch situation: 29% in Copenhagen (City of Copenhagen, 2017), while it is 27% for the whole of the Netherlands in 2016 (Centraal Bureau voor de Statistiek, 2017b).

The research seems to show in nearly all cases that people prefer to cycle on a dedicated cycle facility, like a cycle path or lane, but the only revealed-preference study in the Netherlands (Bernardi et al., 2017) did not find a significant influence. Some other studies suggest that traffic intensities influence the route choice, of which Bernardi et al. (2017) is one. However, the stated-preference study in Copenhagen (Vedel, Jacobsen, & Skov-Petersen, 2017) found that people state the opposite. In foreign studies (e.g. Broach, Dill, & Gliebe, 2012; Menghini, Carrasco, Schüssler, & Axhausen, 2010), slopes are important factors that cyclists avoid, but in Dutch studies these are not important. This is probably related to the fact that there are very little hills present in the Netherlands.

Traffic signals and stops give a more mixed view. Some studies found that they are one of the least important factors (e.g. Broach et al., 2012), but others find that the perceived journey time is very important and that cyclists tend to avoid traffic signals (e.g. Sener, Eluru, & Bhat, 2009). The number of stops was not studied in the Dutch context, so it is difficult to predict whether it is relevant for the assignment.

2.2 Expert opinion

The previous section showed that the literature could not provide a clear view on the most important factors in the Netherlands. Therefore, a number of experts working at Goudappel Coffeng were asked to give their view on the most important factors that influence the route choice in the Netherlands. These experts have different backgrounds, but all give advice about cycling infrastructure, policy or behaviour. They were asked to give different factors that they think are important and to rank them from the most important to less important. Their opinions can be summarised as in table 2.2.

Table 2.2. Comparison of most important factors according to the experts.

	Expert 1	Expert 2	Expert 3
Most important factor	time	time, distance and safety (equally important)	time
Second most important factor	distance	comfort, slopes and waiting times (equally important)	number of stops
Third most important factor	environmental factors (eg type of surface)		not relevant: type of road

The first expert mentioned a list including type of road surface, traffic volumes, safety, number of HGVs and the weather as important factors. People change their route for example because it protects them from the worst influences of the weather, like a route with trees that provide shadow. The most important factor according to him still is the travel time. A problem however is that people can experience time differently, which means that a long road that is very quick, can be experienced as longer because of the lack of change of the environment. The second most important factor is the distance and finally, factors about the environment. This includes for example the type of road and road surface. Important notes he made were that cyclists can make a different choice for their return trip and that the purpose of the trip can have an influence on the experience of certain factors.

Another expert made a pyramid graph to illustrate his idea about the importance of factors. The most important factors according to him are time, distance and safety. These are the base factors. Only until there are two or more routes that have the same score on the base factors, people take factors from a higher level into account. The second level exists of comfort, slopes and waiting. If that is equivalent for multiple routes as well, the experience of the route will be decisive. In his opinion, the base factors are good enough in the Netherlands in most places, so the factors on the second level describe the behaviour the most. An important aspect he mentioned as well, was that every cyclist experiences everything differently, so in most cases more than one route between an origin and a destination will be used. How a cyclist experiences a route can even vary per moment of the day (e.g. social safety). He thinks the travel time of cyclists is not depending on the number of other cyclists, but depends most on waiting times at intersections.

The last expert agrees with the other experts that the travel time is the most important for the cyclist. However, she thinks that the number of stops is very important too. She says the stops not just increase the travel time, but cyclists experience even more time. Continuous routes are perceived as shorter routes even when that is not true. Other factors are not important when people are in a hurry, only the time is really important, but when the purpose is recreational, the environmental factors become important. The first two experts thought that the quality of the road is the most important factor after time and distance, but this expert thinks that the type and quality of the paving plays a small role. Only if the road quality is really bad, people will tend to avoid that road if possible. She thinks that the most important difference between findings of foreign authors and the situation in the Netherlands is the preference for cycle paths. Dutch traffic users are more used to cyclists and cyclists are more assertive, so the need to use a cycle lane or path is not that strong as abroad.

All in all, the experts agree on that time is the most important factor, but they all stress on the fact that people perceive time differently, so the factual shortest route is not necessarily the chosen route. Research showed similar results, but time is not included as factor in all studies, because time is sometimes taken as a measure to compare the other factors against.

There is less consensus about other factors. Two experts and nearly all foreign research seem to think that road types are important, while one expert and the revealed-preference study in the Netherlands (Bernardi et al., 2017) do not think it is significant. The number of stops is another example of a discrepancy between the different sources. The experts think it changes the route of people, while Vedel et al. (2017) found it has only a small influence. Congestion or crowding is not important according to the experts and Vedel et al. (2017), but Bernardi et al. (2017) did found a influence.

It is not immediately clear which factors will be describing the route choice behaviour in the Netherlands best. Furthermore, some factors are (nearly) impossible to implement in an algorithm due to lack of information or data. The next sections describe the available data and the choice of factors that are implemented in an algorithm.

2.3 Available data

The choice of factors that is included in the algorithm depends not only on the most important factors in literature, but there should also be data available about those factors. For this research, OmniTRANS networks of ’s-Hertogenbosch, Tilburg and The Hague/Rotterdam were available. These networks include the main car road network supplemented by the cycling network supplied by the Fietsersbond. The detail level in ’s-Hertogenbosch is lower than in the other two networks. It includes only major roads (8500 links in total) and less centroids (148) in a similar area. In the Tilburg and The Hague/Rotterdam networks there is a centroid for each postcode-6 area, while in ’s-Hertogenbosch, there is a centroid for each postcode-4 area. That means that all inhabitants with the same numbers in their postcode share one centroid. In Tilburg and The Hague/Rotterdam, only inhabitants that share completely the same postcode are added together in one zone and centroid. In Tilburg, that results in 1319 centroids and 22800 links. The The Hague/Rotterdam networks consisted of circa 10000 centroids which was reduced to 2800 by only including The Hague and surrounding areas (see section 6.1 for more information about the cutting of the network)

Figure 2.1. 's-Hertogenbosch network.

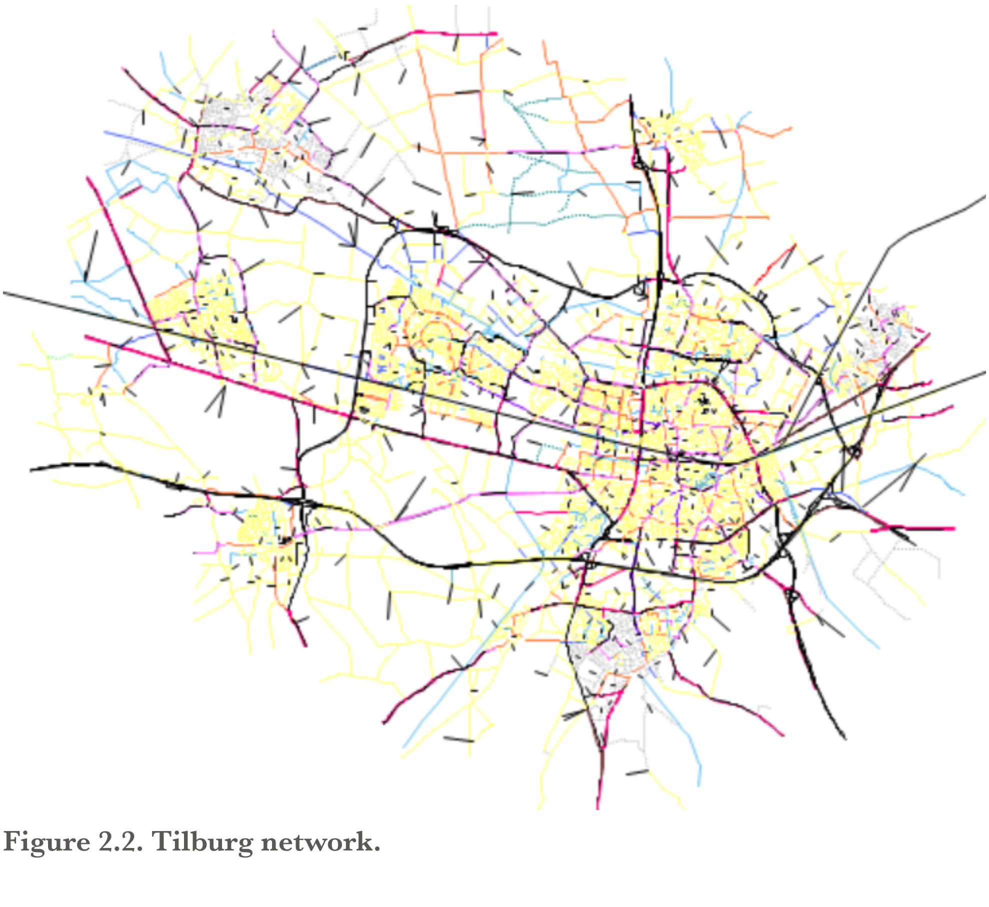


Figure 2.2. Tilburg network.

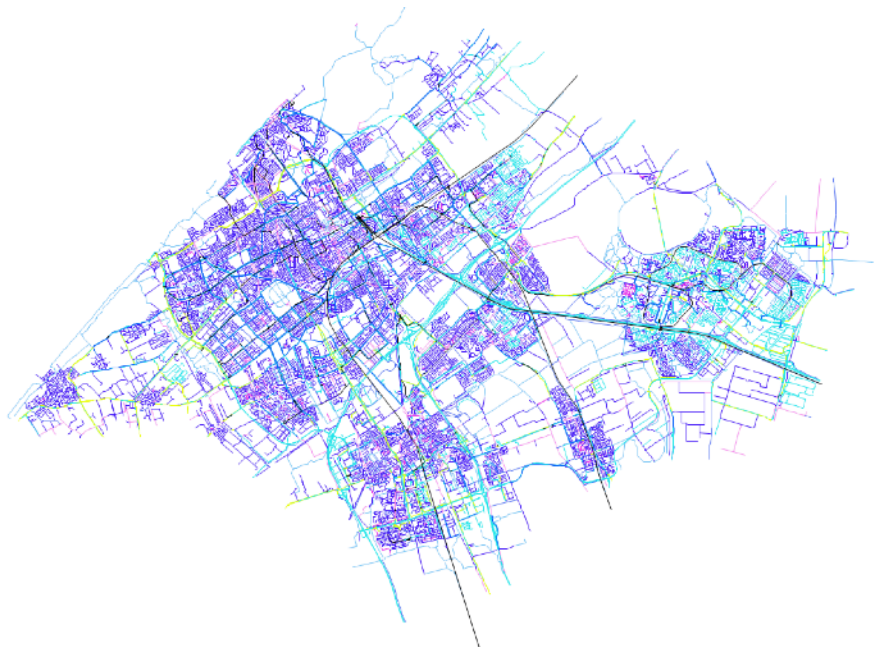


Figure 2.3. Part of the The Hague/Rotterdam network that was used in the validation of the algorithm (see chapter 6).

All networks contain not just the urban area, but also include parts of the countryside around it. The ’s-Hertogenbosch network also includes for example Oss and the area in between. The The Hague/Rotterdam network includes a large area around it. Roughly, it is the area between Leiden and the island of Goeree-Overflakkee and between the North Sea coastline and Gouda. After cutting the network it still included Delft, Zoetermeer and some other towns around The Hague (see figure 2.3). The Tilburg network includes towns in all directions around it.

Every road section on which cyclists are allowed, has additional attributes about characteristics for cyclists. This data is provided by the *Fietsersbond* in the Tilburg and The Hague/Rotterdam networks. These attributes include the type of cycling infrastructure (e.g. cycle path, lane or normal road), type of road surface, type of intersection at the end and start of the road section and whether it is classified as a main cycle route. Furthermore, the Tilburg network has data about the quality of the road surface, lighting, the hinder of cars to cyclists and the type of surrounding land use (e.g. meadows or built environment).

The *Fietsersbond*’s volunteers entered this data in a GIS environment which was coupled to the OmniTRANS network. Information about the type of road and road surface seem to include little flaws and nearly every road in the network has the right attribute. The type of intersection at the end of the road section does not seem always correct and in some cases, many intersections are classified as unknown type. The network of The Hague and Rotterdam seems to have a greater number of unknown types of intersection compared to Tilburg. The land use seems to be correctly inputted in nearly all cases, but the other attributes, like road surface quality are very subjective and the categories might therefore be inconsistently applied. The personal opinion of the volunteer influences the categorisation of those attributes strongly. Another limitation of the lighting, road surface quality and hinder is the fact that this data is only available of existing roads, which makes it impossible to use those for predictions.

Furthermore, the average speed of cyclists is added as an attribute to all roads. The average speed is calculated using a formula earlier found by Hogenkamp (2014), which includes the type of road and the curvature of the road. Also, the length of a road is known from the geometry of the network. The length is an accurate measure, because the geometry is based on maps. The average speed gives only an indication as the formula is a model itself which does not have to be correct in the considered locations.

Other input data for the assignment is the OD-matrix of the area. The OD-matrices were developed by Goudappel Coffeng for the governments in the areas of the network. The OD-matrices were estimated on the basis of the number of inhabitants and the number of jobs in the origin and destination zones with a gravity model. Before applying the gravity model, the resistance to travel for cyclists between areas was estimated based on an all-or-nothing assignment. The number of inhabitants was broken down in different socio-economic and age groups and the jobs in different types. An example of the job types is office jobs to correct for differences between the groups. The modal split and trip length distribution were fitted with data from the *Onderzoek Verplaatsingen in Nederland* (OViN, Research into movements in the Netherlands). This is an annual study by the Dutch statistics office (CBS) on the numbers of trips between areas, time of day and the mode choice. The OD-matrix was not calibrated using traffic counts.

Finally, there are cyclist counts available in Tilburg and The Hague. In Tilburg, these are based on visual counts during one day in September or October 2015. These counts are conducted by a person next to the road noting the number of cyclists passing in a certain direction along that point. The counts were performed on different days in that period with different weather conditions, but all counts were on a Tuesday or Thursday and during one day. There seems to be one minor problem with the data on two parallel bidirectional service roads. All cyclists heading southwards seem to be added together without distinction on which service road they are cycling (see figure 2.4). All other locations seem to be correctly represented in OmniTRANS.

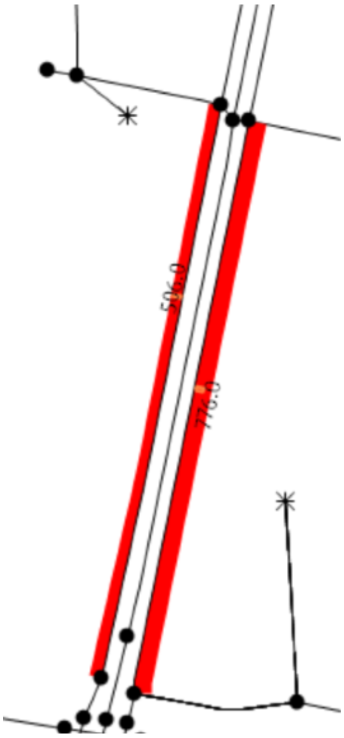


Figure 2.4. Example of a problem in the data around the Goirleseweg in Tilburg. Both service roads only show a number of cyclists in one direction, while in reality they are bi-directional.

The counts in The Hague and surrounding areas were done in 2014, 2015 or 2016. Most counts in The Hague itself are visual counts, while counts in the neighbouring towns and cities are mostly done with pneumatic road tubes. These counts with tubes are mostly from 2016 and around 10% were done in 2015. Tube counts can have a bigger deviation from the real value, because two people cycling next to each other can sometimes be counted as one cyclist. Visual counts have sometimes problems with large groups that make it impossible for the counting person to see them all. Another limitation of the data is the different moments that counts took place, which might implicate that the values of different roads are not comparable. As this network is much larger than the Tilburg network, it includes many more counts, which is an advantage.

2.4 Factors in algorithm

Cyclists in the Netherlands seem to find the travel time the most important aspect of their journey, but the duration of a route can be experienced differently by different cyclists. The stated-preference studies in Utrecht (Claasen & Rienstra, 2017) and London (Transport for London, 2012) confirm this. Other studies focussed more on distance instead of time and distance is in studies in the Netherlands and Denmark one of the most important factors (e.g. Bernardi et al., 2017; Vedel et al., 2017). The distance can be used easily in models as this attribute is always present in networks. The time can in most cases be derived from the distance using an average speed and penalties for intersections. Another option to calculate the time is the use of the results of Hogenkamp (2014), who described the average speed based on characteristics of the roads. Time and distance seem relevant and information is available and these were therefore factors that were included in the assignment algorithm.

The number of stops does not seem to have a great influence in studies with GPS (e.g. Broach et al., 2012), but two of the experts mentioned them as important factors in the Netherlands, because people seem to experience a minute waiting longer than a minute cycling. This factor has some overlap with time, because people do not like to wait, as their travel time increases as a result of that. Therefore, the waiting time at intersections will be used in the algorithm with its own parameters, because the experience of this time is different from the time spent cycling. Data is available about the type of intersections, which makes it possible to include it in the assignment.

Furthermore, in nearly all studies, in cycling and non-cycling countries, the type of road seems to be a very important factor. Cyclists prefer to cycle on separated cycle paths and cycle lanes instead of on a main road, especially in countries without a cycling culture like the Netherlands and Denmark have. Claasen & Rienstra's (2017) survey in Utrecht suggests that cyclists in the Netherlands prefer those routes as well. Two of the experts agreed that people do have a preference for dedicated paths and lanes. However, Bernardi et al. (2017) could not find a significant result and the other expert mentioned that the safety is that high that cyclists do not care on what type of road they cycle. The insignificant results could also be explained by the large number of cycle paths and lanes in the Netherlands which makes that people do not have to take a detour in order to cycle on such a road. Information about the type of cycle facility on a road is available in detail. For example, cycle and combined moped and cycle paths are separated in the data and bicycle boulevards, service roads and lanes are distinguished. Because the type of facility may have an influence and because there is enough data available, this factor was used in the assignment algorithm.

Another factor that has been mentioned in literature is high motorised traffic volumes on a road. This was found to be important in areas without a good bike infrastructure like Portland and Texas, but also in the Netherlands (Bernardi et al., 2017). In Copenhagen however, Vedel et al. (2017) did not find that cyclists see it themselves as important in their choice of route. No expert did mention this factor as particularly important and this factor is not directly available in the traffic model. It would mean that first a car assignment is needed before the bicycle assignment can start. Because of the unclear influence and the fact that the use of the bicycle assignment would become more difficult and time-consuming, this factor will not be included directly. It was however indirectly included in the road type, because normal roads and roads with cycle lanes are distinguished from roads with separate cycle facilities.

Slopes seemed very important in areas with many hills or mountains, like Portland (Broach et al., 2012) and Zürich (Menghini et al., 2010). In the survey in Utrecht it did not seem important and Claasen & Rienstra (2017) explained that with the lack of hills in Utrecht. However, one of the experts (number 2) thought that cyclists in the Netherlands do find them important, for example on bridges and in tunnels. It was not clear whether cyclists in the Netherlands take it into account. Because of that and due to the low number of slopes in the Netherlands, this factor was not taken into account.

The number of curves in a route is important for the average speed as Hogenkamp (2014) found, but no research in the route choice could find that is important. Claasen & Rienstra (2017) found that cyclists do not think it it important in their choice. In San Francisco, Hood et al. (2011) could not find an influence as well in the revealed routes of cyclists. Therefore, the number of curves will probably not be an important factor in the route choice. However, it influences the travel time indirectly, which is an important factor. The speed formula of Hogenkamp (2014) was for that reason used to calculate the travel time.

Finally, the quality of the road surface is important according to some experts, but the last expert did think that most roads are good enough and only a few bad ones will be avoided. The quality has not been mentioned in any literature in any region, so it is unclear whether it has an influence and if so, what influence. Therefore, the quality was not taken into account in spite of the availability of the data.

All in all, time and distance are the most important factors that were used in the algorithm. These were supplemented by the road type and type of intersection, because data about these factors was available and experts and literature thought that these are important. If more data would be available, the number of motorised vehicles would be a good factor too, because multiple researches found that people try to take less busy roads.

3. Algorithm

The factors that have been chosen in the previous chapter needed to be implemented in an algorithm in order to calculate the distribution of cyclists on the roads in an area. Various types of algortihms exist to assign vehicles to a route. Some might be used for cyclists as well, whereas others may be too time consuming or incompatible with cyclists. Therefore this chapter describes the choice for an algorithm.

This chapter starts with a description of the currently most used algorithms to assign cyclists. The second section includes the different possible algorithms to assign cyclists, while the third discusses the requirements and criteria to select an algorithm. Subsequently, the best algorithm was chosen in section 4 and finally, the route generation algorithm was selected in the last section, because in the fourth section the need arises to have such an algorithm.

3.1 Current method

At the moment, most assignments of cyclists are done with an all-or-nothing assignment which is a basic traffic assignment technique. This algorithm assigns all traffic between an origin and a destination to the route with the least costs between that origin and destination. The costs are only based on distance, time or a combination of those. In most cases it is based on distance, which implicates that all cyclists would use the shortest route. In reality, people choose different routes between the same points as is described in sections 2.1 and 2.2. Furthermore, they do not always choose the shortest or fastest route.

As mentioned before, in some assignments the costs are based on time or a combination of time and distance. This only gives a different result from using the distance as cost indicator if the speed is different for each road instead of an average speed applied to all roads. The results of these assignments are not satisfactory as well, because all cyclists are assigned to one route between an origin and a destination, while in reality different cyclists on the same journey can choose two different routes.

Sometimes, a small variation is applied to overcome the problem of just assigning to the route with the lowest costs. In those cases routes are generated and the cyclists are distributed over those routes with a logit model. In the cost function only time is included. This method is an improvement of the basic all-or-nothing assignment, but does not take preferences for routes into account and assumes that cyclists only have a preference for fast routes.

3.2 Route assignment

Besides the basic all-or-nothing assignment that is mostly used at the moment, there exist several other assignment methods. Some of them are specifically made for assignments, others are developed as choice models. Two choice models are widely used: logit and probit models. In both models the costs of the route are calculated, but because the perception of costs varies from person to person and not all elements that influence the costs are known, an error term is added. In logit models, this error term is based on a

Gumbel distribution, while a probit model uses a normal distribution (Han, 2000). The probability that a route is chosen according to the logit model is expressed in formula 3.1.

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \tag{3.1}$$

where:
P_{ni} is the probability that decision maker *n* chooses alternative *i;*
V_{ni} the utility of that alternative according to decision maker *n;* and
j represents the other alternatives.

Probit models are based on a normal distribution, which does not have a closed form. Therefore, Monte Carlo simulation is needed to find the probability that a route will be chosen (Han, 2000).

The utility function in both types of model consists in most cases of multiple factors added up together with scaling factors for each factor. An example is given in formula 3.2. In this example, it is assumed that only the road type and gender determine the utility of an alternative. Because the influence of both factors is different, scaling factors β are used to correct for that.

$$V_{ni} = \beta_{RT} \cdot RT_{ni} + \beta_{gender} \cdot gender_{ni} \tag{3.2}$$

where:
V_{ni} is the utility of alternative *i* to decision maker *n*;
β_{RT} the scaling factor belonging to road type;
RT_{ni} the perceived road type of alternative *i* to decision maker *n*;
β_{gender} the gender scaling factor;
$gender_{ni}$ the gender of decision maker *i* in alternative *n*.

A drawback to logit models is the need for the independence of the alternatives. Routes between two points share in many cases some links, which implies that the alternatives are not independent anymore. Probit models do not have this problem, but they need Monte Carlo simulation and that takes much computational time (Han, 2000).

Most authors prefer to use the logit model (e.g. Vedel et al., 2017), but according to Broach et al. (2012) and Hood et al. (2011), there is a lot of overlap between generated routes in their logit models which implies that the logit model cannot be used as the alternatives are not independent. Because the probit models become too complicated, they opted for a slightly changed logit model: the path-size logit model. To correct for the overlaps, a path-size factor is calculated to distribute the length of those links over the different paths. The path-size factor is added to the utility function just like the functions that were described in the previous section and the utility function is implemented in the Gumbel distribution to obtain the chance that someone chooses that route.

Ryu et al. (2018) used two criteria for selecting the route instead of one cost function. In order to assign routes to vehicles or cyclists, a balance between the two criteria needs to be found. The best scoring routes on both criteria are not chosen, only methods in between. They present five different methods to distribute these routes over the cyclists of which the path-size logit assignment is one. To apply this model, they combine the two criteria, which is not needed for the other methods. Furthermore, they applied their model on Winnipeg (Canada) using all different methods. They found that the logit model worked nearly as good as the less used methods, but they did not compare the results with observations to validate the chosen routes.

Another approach is taken by Christou (2017) for use in Chelmsford, UK. He used a deterministic user equilibrium assignment with the perceived time as costs. The perceived time depends in this assignment on the speed that can be reached on a road section due to congestion and the type of road. The intensities of both car and cycle traffic are included in determining the speed. The free-flow speed is different for each type of road, so separate cycle paths have a higher maximum speed. This model was calibrated on 76% of the counts and validated on the rest of the counts in the same area. This type of assignment has the advantage that no route generation is needed because the costs are attributed to links instead to the routes. However, there are no grounds presented for taking congestion and capacity into account in a cyclists' assignment. No other literature used this method or found that capacity constrains the speed of cyclists.

The logit and probit choice models are used to make a choice between routes, while the user equilibrium and all-or-nothing assignment can assign directly based on costs that are related to links. This means that logit and probit models first need another algorithm to generate possible routes in contrast with the all-or-nothing and user equilibrium assignments. The algorithms needing route generation are therefore called two-stage models, because of the two different steps in assigning the cyclists (Bekhor, Ben-Akiva, & Ramming, 2006; Ryu et al., 2018). A disadvantage of these models is the increased computational time, because more steps are needed.

Assignment of cyclists is fastest with a logit model, but as stated before, this does not take overlapping routes into account. Probit models can be used too, but take much time. A useful compromise is the path-size logit model that does take the overlapping into account. The user equilibrium assignment takes factors only indirectly into account and therefore seems to be less suitable. A downside to this method is the lack of a comparison between the modelled and real situation in the study. It is unknown at this stage whether these models provide realistic results. The next section describes the requirements of the algorithm and the criteria on which the different methods are assessed.

3.3 Requirements and criteria

In the previous chapter the most suitable factors have been determined for application in an algorithm. To find the most useful algorithm, this section discusses the requirements of the algorithm.

As has become clear, in most cases different cyclists take different routes from the same origin to the same destination. This is caused by different perceptions of the travel time and the environment. At the moment, just one route is used for the assignment. In the case that more than one route is used, there is no division based on characteristics of the roads.

The algorithm therefore had to result in multiple routes from an origin to a destination. Furthermore, it was needed to have a better division over the routes with more detail than just for example a 50/50 division. To have a useful algorithm, speed of the assignment algorithm is important. A consequence of that is the need for an effective method without unnecessary calculations. The speed was qualitatively measured by reviewing the literature which used those methods and made comparisons between methods (e.g. Broach et al., 2012).

Furthermore, the algorithm should be able to include the factors that describe the Dutch situation best as were described in the chapter 2. It had to be implemented in OmniTRANS, so the programming language in which it has to be written is Ruby. Finally, a method should be well-substantiated.

These requirements were tested with the following criteria:
- ability to assign cyclists on same O-D pair on different routes
- ability to include the factors from chapter 2
- ability to use the factors to influence the division of cyclists over the different options
- running speed
- ability to implement the algorithm in Ruby and OmniTRANS
- theoretical basis

3.4 Scores of algorithms

The different types of algorithms described in section 3.2 are rated in this section on the criteria of the previous section. Only methods that have been used previously have been compared, because it is not certain whether they work at all. Furthermore, various methods that Ryu et al. (2018) described, were not found in other sources and therefore were not considered as an option.

The requirements were assessed by analysing the literature in section 3.2. Running speed was not measured quantitatively, but it was compared using the comparisons of algorithms in literature (e.g. Broach et al., 2012). The theoretical basis was described by comparing which authors did use it before and in what way they used it. Table 3.1 shows the results of the comparison.

Table 3.1. Scores of different assignment techniques on the criteria set in section 3.2.

Criterion	All-or-nothing assignment	Logit assignment	Probit assignment	User equilibrium assignment
ability to assign cyclists on same O-D pair on different routes	only by applying algorithm more than once	yes	yes	yes

Criterion	All-or-nothing assignment	Logit assignment	Probit assignment	User equilibrium assignment
ability to include the factors from chapter 4	yes, by including them in the cost function	yes, by including them in the cost function	yes, by including them in the cost function	indirectly, Christou (2017) incorporates factors by changing the free-flow speed
ability to use the factors to influence the division of cyclists over the different options	no	yes, by including them in the cost function	yes, by including them in the cost function	partly, speed influences the choice of route
running speed	very quick (no iterations)	quite quick	slow (need for simulation)	quite quick
ability to implement the algorithm in Ruby and OmniTRANS	Standard function in OmniTRANS	Formulae can be implemented in Ruby	Formulae can be implemented in Ruby, more difficult than logit due to need for simulation	Standard function in OmniTRANS
theoretical basis	Good basis, but cyclists do not choose the same routes, so not in the case of cyclists.	Good, e.g. Hood et al. (2011); Broach et al (2012); Vedel et al. (2017); Ryu et al. (2018). Most authors did not use it for assigning however. Standard logit considers alternative routes as independent, path-size logit can correct for overlaps.	Good, most authors did not use it because of the difficulty in applying it, but they agree on the accuracy of the method.	Good for cars, only Christou (2017) applied it to cyclists. No other examples of application for cyclists in literature.

Based on the scores presented in table 3.1, a logit model seemed to be the most useful algorithm to implement. It has an acceptable calculation time, it is able to assign cyclists to more than one route and it has been tested in literature many times before. Furthermore, it is capable of including the different opinions of different people and the characteristics of the alternative routes. A logit model treats the different routes as independent options. That is not the case because it is possible to have overlaps between routes. Therefore, a path-size logit model was applied because that corrects for overlaps.

The probit model is the second best option, but because Monte Carlo simulation is needed, logit is preferred over probit. Simulation takes much more time because of the vast number of iterations and it is not provided as a standard function in OmniTRANS. An all-or-nothing assignment is not useful, because it is not able to create different routes. A variant of this can create multiple best scoring routes, but how many cyclists take which option cannot be established.

Finally, the user equilibrium was considered. This method has only been applied once with cyclists and it is most used in situations with congestion. It assumes that people know what the costs of each route are and take the route with the least costs. As everyone does that, an equilibrium will arise. Christou (2017) used perceived time as cost and the perceived time depended on the speed which on its turn depended on the traffic intensity on the route and the free-flow speed. Normally, congestion is not a problem in assignments of cyclists and no literature describes capacity problems for cyclists. The chosen factors in chapter 2 do not include congestion and capacity. Without including those, this assignment would be equal to the all-or-nothing assignment in which everyone will all take the same route with the least costs. The only way this method would work is modifying it to a great extent. It is not deemed practical because of these reasons and will not be used.

3.5 Generation algorithms

A consequence of the choice of the logit assignment is the need for generating routes, while all-or-nothing and user equilibrium assignments do not need a route set. Therefore, an algorithm needs to be found for this. Multiple algorithms exist to generate routes. The most simple and quick way is to use Dijkstra's (1959) algorithm for the shortest route.

This algorithm generates just one route based on just the criterion of distance. It can be expanded to include a number of shortest routes by randomly changing the lengths of the links in the network (e.g. Yen, 1971), but that is still just based on the criterion of distance. A disadvantage of these algorithms is that a lot of similar routes will be made with for example only a slight variation in the end (Bekhor et al., 2006). Broach et al. (2012) used therefore a method that maximised or minimised multiple variables (see description in section 2.1) within a reasonable detour from the shortest route. With this method the best route based on each variable is found. They calibrated the limit of 'reasonable' with their data of cyclists in Portland.

This a slightly changed version of the labelling technique that Bekhor et al. (2006) describe. Labelling works the same way without the limitation of the reasonable detour. Hood et al. (2011) used a stochastic method based on the utility of routes. In the algorithm, each factor that influences the route choice is calculated using a parameter that is drawn from an interval. Based on the drawn values the shortest route is determined and this process is repeated during a couple of times.

Ehrgott et al. (2012) proposed a method that seeks a balance between two objectives: travel time and suitability. In their method, they assume that cyclists will prefer a route if the travel time is lower, but the level of suitability equal or if the route is more suitable, but the travel time is the same. If better routes can be found, the route is called efficient and they define the set of all efficient routes as the set of routes that cyclists choose from. The travel time is modelled to include delays at traffic signals based on the signal cycle time. The suitability objective is modelled by scoring each road section according to 20 factors. To find all routes, their algorithm starts with finding the route with the shortest travel time followed by all routes that are a bit longer until they found probably most suitable routes. Then the non-efficient routes are taken out and all efficient routes remain. Ehrgott et al. (2012) applied this method on a case in Auckland, New Zealand, but did not compare the results of the model with observations, so it is unknown whether this method works well.

Ryu et al. (2018) propose a similar type of method, but this method is especially made to assign routes to cyclists. They use the distance and the bicycle level of service (BLOS) as the two criteria. The first criterion includes a penalty for turning movements and intersections as well. The second criterion is a score on the perceived safety of the route. It consists out of three measures that have to be calculated themselves too: average bicycle segment score, average bicycle intersection score and the average number of unsignalised conflicts per mile. The first two measures are being calculated on the basis of factors like the vehicle volume and width of lanes. The method first generates routes based on the distance criterion and calculates the BLOS of those routes after that. The routes with a good balance between BLOS and distance are selected as choice set.

Generating routes based on just one objective can be performed not only using Dijkstra's algorithm and Monte Carlo simulation for variation of costs, but also using the slightly different doubly stochastic generation. Instead of only varying the costs of the links using the Monte Carlo simulation, the preference for certain attributes is drawn from another probability distribution as well. After each random draw, Dijkstra's algorithm is applied again to find the route with the least costs (Halldórsdóttir, Rieser-Schüssler, Axhausen, Nielsen, & Prato, 2014).

An alternative for Monte Carlo simulation is the breadth first search on link elimination. This algorithm determines the shortest route and continues with deleting each link in the route once and recalculating the shortest route if that link would not exist. After that, two links at the same time can be deleted and new shortest routes will be found (Halldórsdóttir et al., 2014).

Furthermore, the branch & bound method can be used as an alternative for Dijkstra's algorithm. This method constructs a tree by adding all branches of the route until a certain point. The number of routes can be reduced by adding restrictions, but it still requires much computation time in high-resolution network (Halldórsdóttir et al., 2014).

3.6 Selection generation algorithm

The generation algorithm that was selected had to fulfil a number of requirements. First of all, it had to be possible to implement it in the OmniTRANS software. Furthermore, the computational time should be as short as possible, while retaining enough generated routes. This section describes the reasoning behind the selection of an algorithm out of the options from the previous section.

In OmniTRANS certain standard functions are provided. It has a preprogrammed function to generate the shortest routes using Dijkstra's (1959) algorithm. Furthermore, it can find alternative routes using a Monte Carlo simulation of which the variance can be set. The alternative routes can be filtered to exclude too overlapping routes and too long detours. However, the use of Monte Carlo simulation takes long and it is because of that not the preferred option.

Because the logit model will score each route, it is no problem to evaluate routes that in reality will not be chosen. They will obtain a very low score and no or very little cyclists will be assigned to that route. If a route has not been generated however, the route cannot be assigned at all. Therefore, it is better to generate a few more routes than necessary instead of not generating good options. A disadvantage of generation of more than needed routes is the increasing computational time of the model, especially when using Monte Carlo simulation for generating alternative routes.

According to Halldórsdóttir et al. (2014), the breadth first search on link elimination uses the least computational time compared to alternative methods as doubly stochastic generation method and branch & bound. Furthermore, it provided similar results to the doubly stochastic generation. A downside compared to the Monte Carlo simulation is the that breadth first search on link elimination only determines the route set per OD-pair (Rieser-Schüssler, Balmer, & Axhausen, 2013), while the Monte Carlo simulation in OmniTRANS is capable of determining alternative routes of all pairs from one origin at the same time.

To obtain all possible routes within a reasonable time the shortest route it was decided to generate routes using Dijkstra's algorithm. Other possible routes were generated with a Monte Carlo simulation in which the costs of the links vary. The costs are equal to the length of the link. Unfortunately, there is not another algorithm available that can reduce the computational time. To obtain all relevant routes there can be two main methods distinguished: generating a high number of routes with the Monte Carlo simulation based on distance or applying the shortest route algorithm to different cost functions, for instance based on suitability, complementary to some routes based on distance. Both options have their advantages and disadvantages.

The first option will only generate different routes, but a part of the route set will only be slight variations on other routes. A disadvantage is that time is used to generate and store routes that are not needed, because the route set will be too large for most pairs. However, if this would be reduced, on other pairs, too little routes are generated.

The other option is to change the cost function in order to only include a suitability score. An advantage of this method is the explicit generation of routes that are longer and could be more suitable. A disadvantage is, as some experts pointed out, that the cycle facilities in the Netherlands are good and present on many roads. This implies that many of the generated routes are identical to the generated routes based on distance as costs which means a lot of computational time is taken without a useful result. Furthermore, the influence of the route choice factors will be included twice in the whole process of generating routes and assigning a route, which means that calibration becomes more difficult. When using more Monte Carlo draws based on distance, there is only one factor that should be calibrated in order to generate enough routes: distance.

Considering all options the Dijkstra algorithm was expanded with a Monte Carlo simulation based on varying the costs. The costs are only based on distance, because that is the most important factor in the route choice of people as described in chapter 2 and it would not be efficient to implement other factors that probably have a marginal influence. As the experts mentioned, cyclists will only consider slightly longer routes if they are safer or more suitable. It seemed therefore reasonable to generate more distance-based routes instead of choosing additionally another cost function. The Monte Carlo simulation made sure that more than one route was found.

4. Calibration route generation

In the previous chapter the type of algorithm was selected that has been implemented with the factors described in chapter 2. In order to give a good prediction of the number of cyclists the algorithms need to be calibrated. This chapter discusses the process and results of the calibration of the first part of the algorithm, the route generation. This algorithm needs to generate all routes that cyclists consider so that the assignment algorithm can make the right distribution over the routes. However, it is not possible to generate all routes due to time and storage restrictions. Therefore a balance needs to be found between accuracy and efficiency. The first section of this chapter describes the implementation and the parameters that are involved in the calibration. This is followed by a section about the methodology of this calibration. The route generation algorithm was tested on two networks: ’s-Hertogenbosch and Tilburg. The third section presents the results of the calibration on the former network and the fourth section on the latter.

4.1 Implementation and parameters

In chapter 3 a comparison of the different methods for route generation was done which resulted in the choice for a Dijkstra’s algorithm with Monte Carlo simulation to vary the costs. OmniTRANS possesses a built-in feature to generate routes with these methods that is normally used for dynamic assignments of cars, based on the algorithm of Fiorenzo-Catalano (2007). Parameters of this route generator are the variance, the increment of the variance, maximum variance and the number of iterations. The variance is only increased in case no more new routes are found with the variance at the start and the maximum number of iterations has not been reached yet. The generator uses a normal distribution to draw new costs for each link.

The built-in route generator also has the ability to filter routes that have too much or little overlap with other routes. It was decided to not use this function, because it would mean that an alternative route using a parallel route would be generated, but deleted before it would be added to the route set. This route might be taken, so it is better to let the assignment algorithm judge about whether cyclists take that route.

The objective of the route generation is to provide all routes that people consider in order to make sure that the assignment considers all available routes. Therefore, in an ideal world the best would be to generate a route set consisting out of all possible routes. In reality, it takes too much time to generate all routes and in larger networks, it takes too much computer storage as well. Calibration of the route generation part therefore seeks to find a balance between generating enough routes and reducing the computational time and size of the route set. The parameters that influence these goals are found in the Monte Carlo simulation: the number of iterations and the variance determine how many different routes are generated and therefore how much time and memory is needed.

4.2 Methodology

The routes that are included in a choice set differ from person to person and no data is available about which routes on each OD-pair are considered. To give a good indication of the quality of the route set, different OD-pairs were selected with different characteristics. Some OD-pairs are only on the countryside, some wholly within a city, while others are semi-rural (e.g. rural origin and destination in city). Furthermore, OD-pairs with different distances between origins and destinations were selected, so for example a pair with just one short reasonable route and a pair between two neighbourhoods with many parallel routes.

The quality of the generated route set can only be assessed indirectly. In this calibration the quality of the route set belonging to some selected OD-pairs was compared with a map of the area and it was visually assessed whether all 'reasonable' routes are included. Reasonable means that the most direct route, shortest route and some slightly longer routes are included. The slightly longer route might be more suitable for cyclists which means that cyclists could prefer it over the shorter route. The line between reasonable and not reasonable is very subjective and not measurable. Therefore, the *Fietsersbond routeplanner* (Fietsersbond, 2018) was used to help indicate whether all logical routes were included. This planner gives the option to choose from different types of routes, for example the shortest or the easiest (*makkelijk doorfietsen* in Dutch). It should be noted however that these routes do no have to be realistic routes, because they are model-based, but they seem in general to be more realistic than routes of other planners like Google Maps (Google, 2018). Google Maps tends to plan routes with large detours which does not seem reasonable.

The *Fietsersbond* route planner contains 11 options for planning a route of which 4 are considered (quite) relevant: the shortest route, *makkelijk doorfietsen* (easy cycling), *autoluw* (low numbers of motorised traffic) and *fietsbewust* (routetype with less slopes and preference for traffic lights instead of roundabouts). Routes of type *makkelijk doorfietsen* follow cycle paths along main roads and cycle routes as much as possible and avoid traffic signals if there is a good alternative. These routes are a bit longer than the shortest route, but might be the quickest because minor roads with many curves are avoided. Furthermore, these are easy to remember because they are on main roads. *Autoluw* routes follow dedicated cycle paths as much as possible and in other cases, these routes are on quiet roads. Finally, the *fietsbewust* routes avoid slopes and roundabouts, but prefer traffic signals. These routes avoid small roads as well and take longer straight roads instead (Fietsersbond, 2018). The other types are not considered as these seem too unrealistic for regular cyclists, like the recreational route, race cycling route, nature route and route following *fietsknooppunten* (recreational cycling network junctions).

The route set was considered sufficiently large if the routes that the *Fietsersbond* route planner provides were included for nearly all OD-pairs. Furthermore, detours that seem logical on the map have to be included. However, the *autoluw* routes seem less realistic for regular cyclists and these routes were therefore not required to be in the set, but if they were, the set was judged as better.

4.3 Calibration of ’s-Hertogenbosch network

The ’s-Hertogenbosch network was tested with 5 OD-pairs of which the four selected route types of the *Fietsersbond* route planner (Fietsersbond, 2018) were compared with the generated routes using 4 combinations of parameters. The following OD-pairs were selected:

- Boschveld - city centre of ’s-Hertogenbosch
- De Rompert - Hintham
- City centre of ’s-Hertogenbosch - De Meerendonk
- Rosmalen - city centre of ’s-Hertogenbosch
- Rosmalen - Nuland

The first pair is between a neighbourhood just outside the city centre and the city centre itself and it has been selected because there is only one logical route between these points. The railway station lies in between and forms a barrier which means the tunnel is the only option. The second pair is between two suburban neighbourhoods of ’s-Hertogenbosch and there are many different routes possible between the areas (see figure 4.1).

Figure 4.1. Locations of centroids of the considered OD-pairs (Google, 2018).

The third pair is between the city centre and a neighbourhood just outside it, but this pair has more than one obvious route (see figure 4.1). The fourth and fifth pair start in Rosmalen, a town just outside ’s-Hertogenbosch. The fourth pair results in various routes between a town and a neighbouring city, while the fifth pair focusses on rural routes between a town and a village.

The routes on these pairs were generated using the settings shown in table 4.1. Setting 1 was based on the standard settings that were used for car routes, but the variance and number of iterations were increased a bit because it was expected that they were not enough.

Table 4.1. Settings for route generation

Parameter	Setting 1	Setting 2	Setting 3	Setting 4
minimum number of iterations	10	15	20	20
maximum number of iterations	15	25	30	30
initial variance	0,09	0,15	0,15	0,2
variance grow value	0,05	0,05	0,05	0,05
maximum variance	0,3	0,3	0,4	0,4
threshold for increasing variance	5	5	5	5

The generated routes using these settings are shown in appendix A together with the *Fietsersbond* route planner results of the same OD-pairs. An example of the results is visible in figures 4.2 and 4.3. Figure 4.2 shows the generated routes using setting 1, while figure 4.3 shows the results of setting 2. Compared with figures 4.4 and 4.5, it seems that the easy cycling route was included in the generated routes of both settings, but that the *fietsbewust* route was not included when generating routes using setting 1.

Figure 4.2. Generated number of routes between the city centre and De Meerendonk with setting 1.

Figure 4.3. Generated number of routes between the city centre and De Meerendonk using setting 2.

Figure 4.4. Route (2,0 km) between the city centre and De Meerendonk with setting easy cycling (Fietsersbond, 2018).

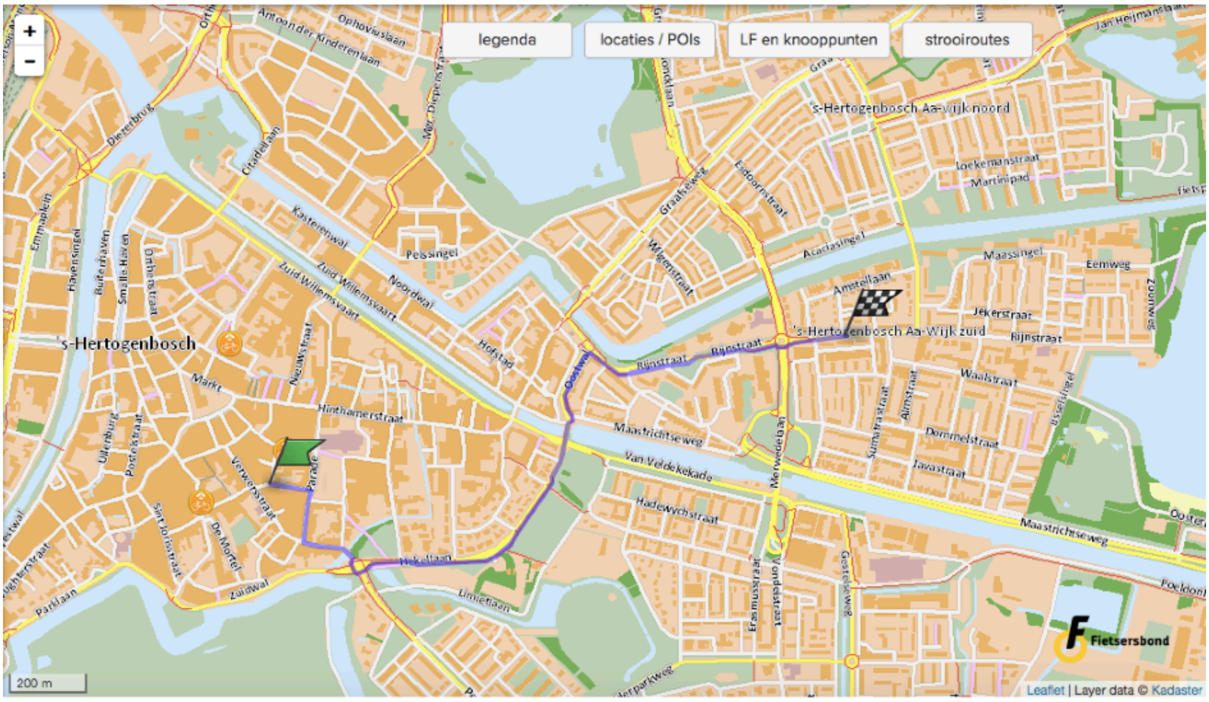


Figure 4.5. Route (2,1 km) between the city centre and De Meerendonk with setting *fietsbewust* (Fietsersbond, 2018).

This example illustrates the use of the route planner to give an indication whether enough routes were available. In this case the *fietsbewust* route seems very reasonable as it is only slightly longer than the other type and could be safer for example. Furthermore, the second setting generates routes along both sides of the Zuid-Willemsvaart canal instead of just one. If one is perceived as being more comfortable, people might choose the other side, so it is necessary to include both sides.

Table 4.2. Number of OD-pairs in which the *Fietsersbond* routes were included using the different settings. Because in some cases a centroid is connected to the road network with more than one connection, multiple routes could be planned and only half of the routes were generated.

Route type in *Fietsersbond* route planner	Setting 1	Setting 2	Setting 3	Setting 4
makkelijk doorfietsen route	3,5	4	4	4
shortest route	5	5	5	5
fietsbewust route	3	4	4	4
autoluw route	0	2,5	2,5	2,5

Table 4.2 presents the number of OD-pairs in which the routes of each types were included using the different settings. Each higher setting included more routes, but also took more computational time. The second setting seemed to have the best proportion between computational time and quality of the route set. Setting 3 does not result in more routes that the *Fietsersbond* route planner provides, while it does take much more time. The extra

iterations result mainly in small variations on the existing routes (see figure 4.6). An example of these variations is a route on a main road that is duplicated with a new route that takes the main road as well, but takes a side street and returns to the same main road using the next side street. Furthermore, the routes generated with setting 3 and 4 did not seem sensible using the map and they added just a few good routes.

Figure 4.6. Generated number of routes between the city centre and De Meerendonk using setting 3.

4.4 Calibration of Tilburg network

The Tilburg network was tested in a similar manner to the 's-Hertogenbosch network, but now not all settings were tried. The first run used setting 2, because it seemed the best result on the 's-Hertogenbosch network. The storage needed for this route set became too large (circa 140 GB). Therefore, setting 1 and a new setting in between were applied to Tilburg as well. This new setting had the following parameters:

- minimum number of iterations: 12;
- maximum number of iterations: 20;
- initial variance: 0,15;
- variance grow value: 0,05;
- maximum variance: 0,05;
- threshold for increasing variance: 5.

Just like the 's-Hertogenbosch network some OD-pairs were selected to give an indication of the quality of the route set. Again, different types of routes were included between towns around Tilburg and the city and between various areas within the city. The following OD-pairs were selected:

- Trouwlaan - city centre;
- Reeshof (Heyhoef area) - Tilburg railway station;
- Berkel-Enschot - Tilburg railway station;
- Goirle - Zorgvlied;
- Korvel - city centre.

The locations of these origins and destinations are visible in figure 4.7.

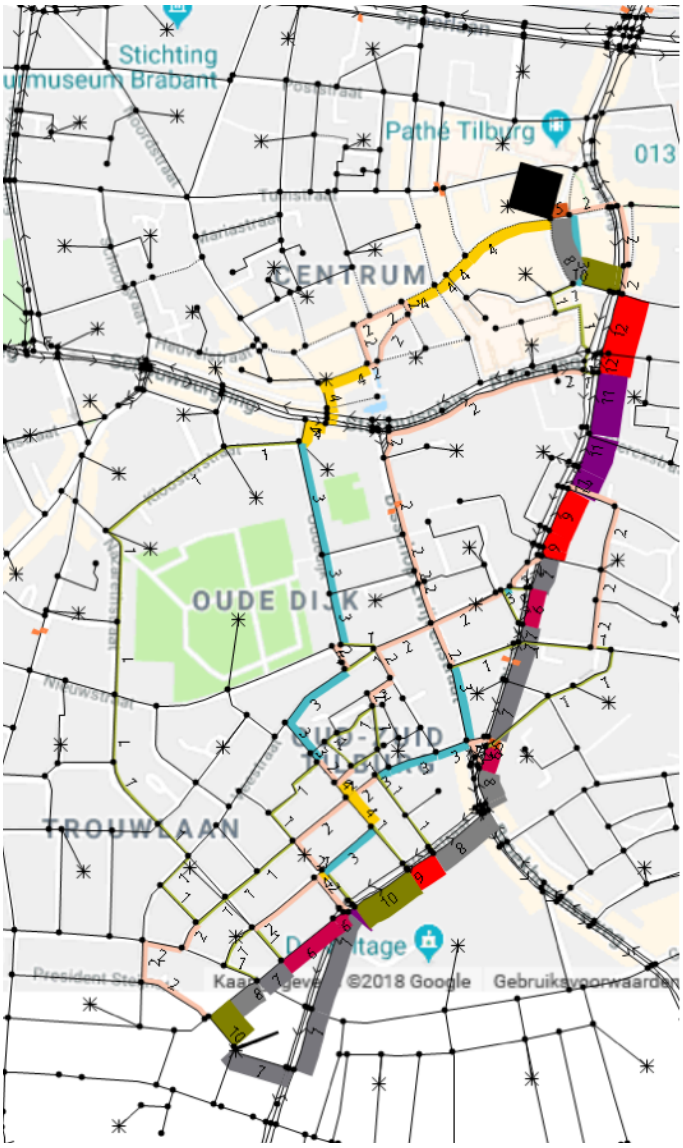
Figure 4.7. Locations of centroids of the considered OD-pairs (Google, 2018).

The experiments revealed that the new setting still provided more than enough routes, but did take less computational time and less storage than setting 2. Setting 1 did not produce enough routes, as parallel roads were sometimes not included, even if the increase in distance was small (a few hundred metres). In all cases, the *makkelijk doorfietsen* route (easy cycling), shortest route and *fietsbewust* routes were included in the route set with this new setting and other routes that seemed reasonable too. Table 4.3 provides an overview of the number of routes that were included using these settings.

Table 4.3. Number of OD-pairs in which the *Fietsersbond* routes were included using the different settings.

Route type in *Fietsersbond* route planner	Setting 1	Intermediate setting	Setting 3
makkelijk doorfietsen route	2	4	5
shortest route	5	5	5
fietsbewust route	4	5	5
autoluw route	2	2	3

The first OD-pair illustrates this well. Figure 4.8 shows the generated routes on this OD-pair with the intermediate setting applied, while figures 4.9 to 4.11 show the different routes that the *Fietsersbond* route planner generates.



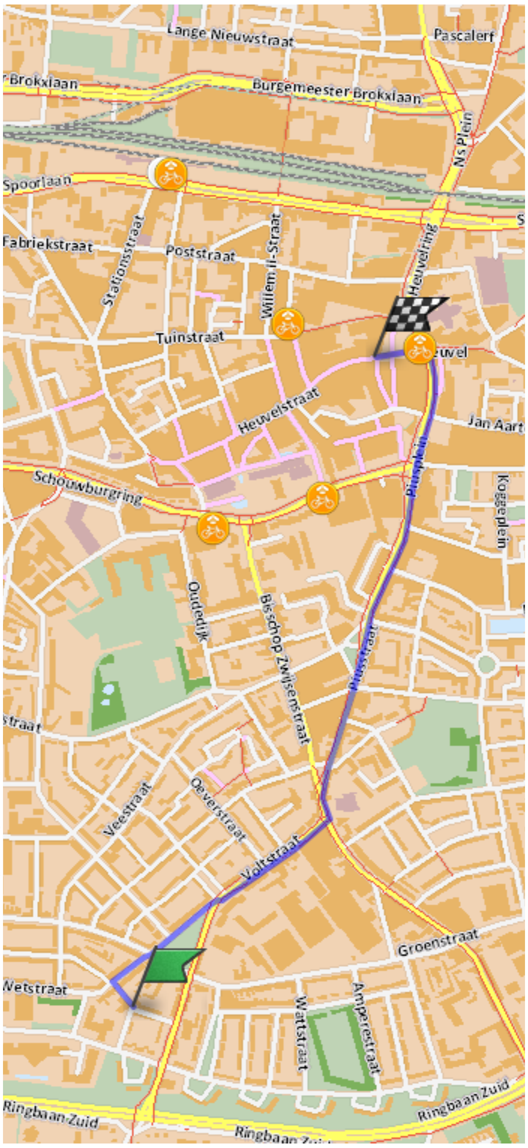
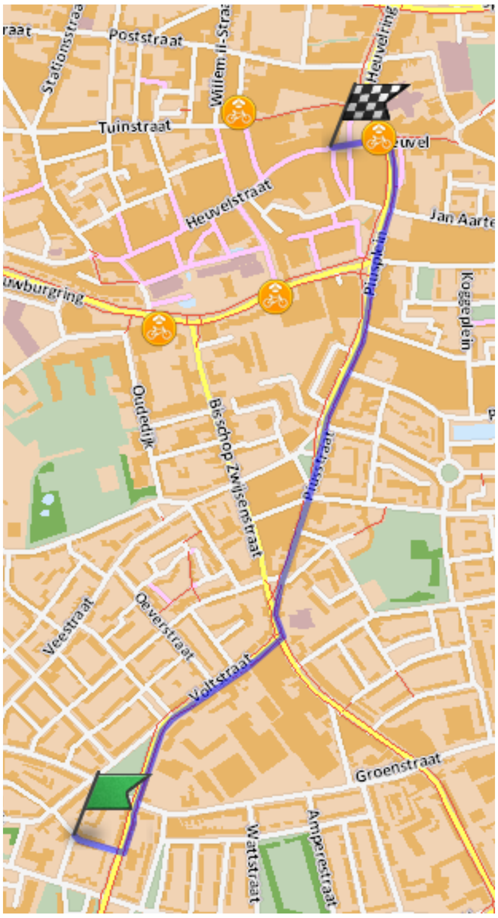


Figure 4.8 (left). Generated routes between the Trouwlaan area of Tilburg and Tilburg city centre with the new, intermediate setting.
Figure 4.9 (right). Shortest route between Trouwlaan and Tilburg city centre (Fietsersbond, 2018).



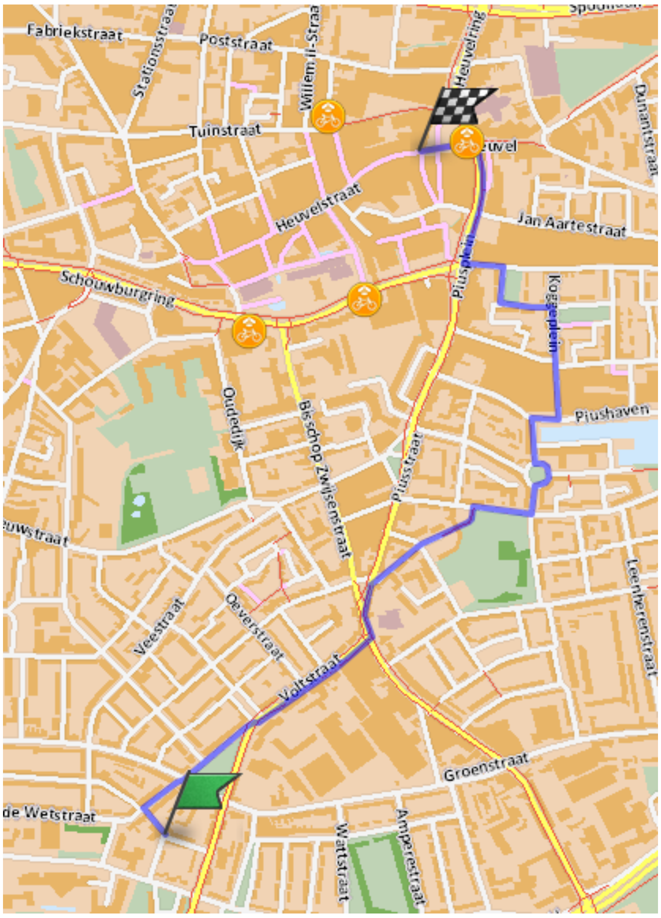


Figure 4.10 (left). ***Makkelijk doorfietsen*** **and** ***fietsbewust*** **route between Trouwlaan and Tilburg city centre (Fietsersbond, 2018).**

Figure 4.11 (right). ***Autoluw*** **route between Trouwlaan and Tilburg city centre (Fietsersbond, 2018).**

The *autoluw* route was not generated, but as mentioned in the previous section about 's-Hertogenbosch, this is not considered a problem, because of the unreasonable detours that are involved and the question whether cyclists do actually use these routes. However, other reasonable routes were included, for example using the Bisschop Zwijsenstraat or taking an alternative route through the city centre. They were just slightly longer than the shortest route (e.g. the route via Bisschop Zwijsenstraat is 300 metres longer) and in case the route is more convenient, people might take that route.

These conclusions could be drawn for each of the OD-pairs. Every time the *makkelijk doorfietsen* and shortest routes were included and the *fietsbewust* routes were included in all but one case where just one short road in the route is excluded from a generated route. Examining the map of the area, most routes that could be an alternative to the shortest route were included, so this setting was used in the following assignment process. The lower needed number of iterations in this network seemed to be caused by the lower number of connections between the centroids and the links. In the 's-Hertogenbosch network, centroids had in most cases two or more connectors, while all centroids in this network only had one. Therefore, it is a route between two points instead of four or more, which reduces the number of routes needed.

5. Calibration of assignment algorithm

With the generated route set and the implemented algorithm, the assignment itself can start. The characteristic of this algorithm is the ability to include other factors than distance and time. The cost of each additional factor is not clear and should be determined with calibration. This chapter first presents the parameters that should be calibrated, followed by the methodology of the calibration and the results. Finally, the results are analysed to see whether the algorithm works as expected.

5.1 Implementation and parameters

The assignment was based on a multinomial logit model which uses the utility of each alternative in order to estimate the number of cyclists that take each alternative. The basic formula is, as explained in section 3.2:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} \tag{5.1}$$

where:
P_{ni} is the probability of user *n* to choose alternative *i*;
V_{ni} the utility of alternative *n* to user *i*; and
j the set of all alternatives.

The utility as it is called in logit models is more accurately a cost function. A longer road increases the value of it and therefore it seems more logical to call it costs. The cost function in the logit model was defined as in formula 5.2:

$$U_k = -\left(\beta_{d,k} d_k + \beta_t t_k + \beta_{is,k}\right) \tag{5.2}$$

where:
U_k is the cost or utility of link *k*;
$\beta_{d,k}$ the distance parameter belonging to the road type of link *k*;
d_k the length of link *k*;
β_t the time parameter;
t_k the travel time on link *k*; and
$\beta_{is,k}$ the intersection type parameter of the intersection type at the end of link *k*.

The total utility of a route was calculated by adding up the utilities of all links that make up the route. Furthermore, the path-size factor that corrects for the overlap between routes was incorporated. So, the utility of an alternative was calculated according to formula 5.3.

$$U_i = \sum_k U_k + \beta_{PS} \cdot PS_i \tag{5.3}$$

where:
U_i is the utility of route (alternative) *i*;
k the set of links in alternative *i*;
β_{PS} the parameter for the path-size factor; and
PS_i the path-size factor belonging to route *i*.

As mentioned before in section 3.2, the path-size factor corrects for the overlapping between routes. It is calculated using formulae 5.4 and 5.5.

$$PS_i = \sum_{a \in \Gamma_i} \frac{l_a}{l_i} \frac{1}{\sum_{q \in C} \delta_a^q} \tag{5.4}$$

$$\delta_a^q = \begin{cases} 1 & \text{if link } a \text{ is part of route } q \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

where:
PS_i is the path-size factor belonging to route *i*;
Γ_i the set of links of which route *i* consists;
l_a the length of link *a*;
l_i the total length of route *i*; and
C the route set;

The time parameter is constant for each link, but the other parameters depend on the type of road or intersection at the end of the link. The following road types are distinguished in the available *Fietsersbond* data with between brackets the Dutch term:

- separated cycle path along road (*fietspad langs weg*)
- separated combined cycle and moped path along road (*bromfietspad langs weg*)
- bicycle boulevard (*fietsstraat*)
- normal road with cyclists on the road (*normale weg met fietsers op de rijbaan*)
- solitary cycle path (*solitair fietspad*)
- solitary combined cycle and moped path (*solitair bromfietspad*)
- ferry (*veerpont*)
- service road (*ventweg*)
- pedestrian short cuts that are sometimes used by cyclists (*doorsteekje*)
- pedestrian area (*voetgangersgebied*)
- cycle lane along road (*fietsstrook*)

The *Fietsersbond* categorises intersections as follows:
- unknown type of intersection
- equivalent intersection (*gelijkwaardige kruising*)
- priority intersection, road without priority (give way sign) (*voorrangskruispunt, voorrang verlenen*)
- priority intersection, road with priority (*voorrangskruispunt, voorrang*)

- intersection with traffic lights (*VRI*)
- roundabout (*rotonde*)
- roundabout with traffic signals (*rotonde met verkeerslichten*)

The distance parameter is different for each type of road, just like the the intersection factor varies for each type of intersection. This implicates that in total 20 parameters can be calibrated. However, the path-size factor parameter was not changed in order to reduce the number of variables and in order to have an accurate correction to overlaps. Other parameters that remained constant are the parameter for ferries and roundabouts with traffic signals as these are not present in the network. Because separated cycle paths and separated combined cycle and moped paths are quite similar, these were allocated the same value in the initial calibration. In no other research, these types are distinguished and it was expected that in general, cyclists do not notice the differences. The same reasoning applied to the solitary cycle path and solitary combined cycle and moped paths.

5.2 Methodology

The remaining parameters were calibrated on the Tilburg network of which the results are discussed in the next section. This calibration started by testing the results of the algorithm with all parameters fixed on 1,0. Subsequently, a set of initial estimated parameters for the different road types and intersection types was created and applied to the network in combination with other parameters fixed on 1,0. An educated guess was used to find these initial parameter values for all road and intersection types. The distance parameters were estimated on the basis of what road types are in general perceived as the most preferable in case two routes having an equal length. The initial parameter values for the different intersection types were determined with the help of an estimated waiting time that cyclists encounter on intersections of these types. In the utility, travel time is taken into account in hours if the time parameter equals 1. Therefore, the waiting times were expressed in hours and rounded to the next round number to become the intersection parameters. All values are shown in table 5.1.

Table 5.1. Estimated parameter values.

Distance parameters	Initial value	Intersection parameters	Initial value
Separated cycle (and moped) path along road	0,2	**Unknown type**	0,01
Bicycle boulevard	0,4	**Equivalent**	0,01
Normal road	0,8	**Give way**	0,02
Solitary cycle (and moped) path	0,1	**Priority**	0
Ferry	1,5	**Traffic lights**	0,1
Service road	0,8	**Roundabout**	0,01
Pedestrian short cut	1,1	**Roundabout with traffic lights**	0,1

Distance parameters	Initial value	Intersection parameters	Initial value
Pedestrian area	1,1		
Cycle lane	0,7		

The time parameter was tested with different values between 0 and 1 and these different values were combined with the distance and intersection parameter combinations in table 5.1. The number of combinations that was tested, was determined by the available time and the running time for one combination. Eventually, 23 combinations were tested (tests 1 to 23).

Furthermore, the initial combinations were changed (tests 24 to 63). Because there are 17 parameters that can be varied, it would result in 10^{17} combinations using 10 values for each parameter. This number is extremely high and 10 values is just an arbitrary number. To reduce the number of experiments needed a number of random draws was used to find different combinations of parameters. The random values were based on an uniform distribution with intervals based on the results of the initial tests.

Finally, some fine-tuning (tests 64 to 97) took place based on the best settings from all previous tests. In these cases all settings were varied individually between the values that were found previously as working well, while all other factors remained constant. This resulted in a sort of sensitivity analysis, which is useful for fine-tuning. Finally, the assignment results were plotted and these were analysed to see whether some parameters needed further fine-tuning.

Performance indicators
The performance was tested with a t-test, which is not the statistical measure with the same name, but a standard measure within Dutch traffic engineering. It uses the following formula:

$$T = ln[\frac{(X_b - X_w)^2}{X_w}] \qquad (5.6)$$

where:
T represents the T-value;
X_w is the observed number of cyclists or vehicles; and
X_b the calculated number of cyclists or vehicles.

The model is normally deemed acceptably realistic if at least 80% of the points on which model and reality are compared, have a value of T lower than 3,5 and 95% a score lower than 4,5. It was probable that this assignment will not reach those values, so this calibration looked for the best scores possible.

In other countries, GEH is a standard method of comparing the model results and the observed number of vehicles. It can be calculated using formula 5.7:

$$GEH = \sqrt{\frac{2(X_b - X_w)^2}{X_b + X_w}} \tag{5.7}$$

where:
GEH represents the GEH statistic;
X_w the observed number of cyclists or vehicles; and
X_b the calculated number of cyclists or vehicles.

In the US and the UK, the standard for a good model is a GEH < 5 for at least 85% of links (Federal Highway Agency, n.d.). GEH is only used as a secondary measure to be able to compare the best scoring parameter combination with the scores of other algorithms. The t-test was used as the main indicator that is calibrated and it was supplemented with some visual inspections of the assignment results. Looking at the results was specifically useful for determining whether the factors are doing their work and which parameters should be changed to obtain better results.

5.3 Calibration results

First of all, the algorithm was tested with all parameters set to 1, which means that all factors have an equal weight. Because the computational time seemed too high to test many settings, the network was reduced in size by omitting some towns and countryside around Tilburg (995 zones instead of 1319 and 11700 links instead of 22800). On this reduced network, the different tests described in the previous section were performed.

Combinations of parameters in table 5.1
The first set of results were based on variation on the time parameter and testing with the initial combinations of distance and intersection parameters in table 5.1. The 20 tested combinations are shown in table 5.2 with their results.

Table 5.2. Number of counts where the assignment has a t-value under 3,5, between 3,5 and 4,5 and more than 4,5 with different parameter combinations.

Test number	Settings distance parameters	Settings time parameter	Settings intersection parameters	T < 3,5	3,5 < T < 4,5	T > 4,5
1	all 1,0	1,0	estimated values table 5.1	30	25	30
2	estimated values table 5.1	1,0	all 1,0	31	22	32
3	all 1,0	0,75	estimated values table 5.1	30	25	30
4	all 1,0	0,5	estimated values table 5.1	30	25	30
5	all 1,0	0,25	estimated values table 5.1	29	26	30

Test number	Settings distance parameters	Settings time parameter	Settings intersection parameters	T < 3,5	3,5 < T < 4,5	T > 4,5
6	all 1,0	0,0	estimated values table 5.1	29	26	30
7	all 1,0	1,0	estimated values table 5.1 * 2	29	24	32
8	all 1,0	0,75	estimated values table 5.1 * 2	29	24	32
9	all 1,0	0,5	estimated values table 5.1 * 2	29	24	32
10	all 1,0	0,25	estimated values table 5.1 * 2	29	24	32
11	all 1,0	0,0	estimated values table 5.1 * 2	29	24	32
12	estimated values table 5.1	0,0	all 0,0	34	23	28
13	estimated values table 5.1	0,0	estimated values table 5.1	39	16	30
14	estimated values table 5.1	0,5	estimated values table 5.1	39	16	30
15	estimated values table 5.1	0,75	estimated values table 5.1	39	16	30
16	estimated values table 5.1	1,0	estimated values table 5.1	39	16	30
17	estimated values table 5.1	0,5	estimated values table 5.1 * 0,5	37	21	27
18	estimated values table 5.1 * 0,5	0,5	estimated values table 5.1	34	20	31
19	estimated values table 5.1	0,5	estimated values table 5.1 * 2	35	20	30
20	estimated values table 5.1 * 2	0,5	estimated values table 5.1	40	18	27
21	estimated values table 5.1 * 2	0,75	estimated values table 5.1	40	18	27
22	estimated values table 5.1 * 2	1,0	estimated values table 5.1	40	18	27

These results seem to suggest that the estimated values work better than only inputting 1, which was expected. Furthermore, it seems that the assignment result is not very sensitive to the time parameter, because while keeping all other parameters constant, a different time parameter results in the same outcomes, independent of the settings of the other parameters. This might be explained by the fact that the speed already partly is incorporated in the distance parameter. Certain road types could have a higher preference because of the ability to cycle faster on those routes. Furthermore, the speed as implemented in this algorithm is just an average and is mainly changed if there are more curves in the route.

Tests using random numbers

The next stage was about calibrating the individual distance and intersection parameters. To obtain variation, random numbers were drawn from a uniform distribution. The range of the distribution was between 0,0 and 1,5 for the distance parameters, between 0,0 and 0,1 for the intersection parameters and between 0,0 and 1,5 for the time parameter. 40 different combinations were created and tested. These ranges were selected based on the first results and the ability to have different ratios between them. The intersection parameter values are smaller, based again on the results of the initial tests.

As expected, very different results emerge. In table 5.3, the best and worst results are shown, but because of the readability, details about the parameter values are only provided in appendix B. Furthermore, appendix B provides an overview of the results of all combinations.

Table 5.3. Number of counts where the assignment has a t-value under 3,5, between 3,5 and 4,5 and more than 4,5 with different random parameter combinations (selection of best and worst results).

Test	T < 3,5	3,5 < T < 4,5	T > 4,5	Test	T < 3,5	3,5 < T < 4,5	T > 4,5
24	38	14	33	49	38	21	26
26	38	18	29	52	30	18	37
32	30	19	36	53	39	17	29
33	36	22	27	55	38	21	26
37	30	25	30	61	40	14	31
46	37	16	32	63	31	20	34
48	30	20	35				

To find the optimal combination, the parameters that belong to these best and worst results were compared. Some parameters did not seem to have a correlation with the performance of the test, while others did seem to have a slight correlation. The distance parameter for bicycle boulevards seemed to have a low parameter in all good results (lower than 1), while all bad results had a high parameter value. In general, the same applied to the solitary cycle and combined cycle-moped path. In good combinations, these were most of the times lower than 1, but some good scoring tests have higher values as well. Normal roads had in general higher parameter values (more than 0,8) in good tests than the bad-scoring results.

These results were expected, because higher values mean a less attractive road and the expectation was that normal roads are less attractive than dedicated facilities. The results of the bicycle boulevard parameter and solitary cycle facilities parameters are conform expectations as well, as these are regarded as safer and more easy to cycle on.

The results of the cycling lane parameter are however not that clear and were unexpected. The good tests had either a very low or very high parameter setting (<0,35 or >1,2), while all worst results are in between. The intersection type parameters do not seem to have a pattern as well. The ‘good’ and ‘bad’ parameter settings are in the same interval and do not seem correlated with the result. This could be because there is no relation between the intersection types, but it could be caused by the change of other parameters as well.

Just like with the initial tests, the time parameter did not seem to have a huge influence. In most cases the value is quite high (>0,8), but one very good result (test 26) had an extremely low value (0,05).

Sensitivity analysis (tests 64 to 97)
To find whether there are better settings than found until now, these results were used to create a new set of parameters. It consisted out of the values of the best results in the previous tests. Because there was no one good value, an average of the best values was taken. These best scoring tests were 49 and 61, because test 49 had the lowest number of counts in the worst category, while 61 had the highest number in the best category (see table 5.4). The next tests used this set of parameters, but in each test, one parameter was changed into a higher or lower value. These higher and lower values were the lowest or highest parameter value that produced a good result in the previous tests. All settings are shown in table 5.4.

Table 5.4. Parameter values derived from previous tests with random settings.

Distance factors	Low value	Base value	High value	Intersection factors	Low value	Base value	High value
Separated cycle (and moped) path along road	0,1	0,8	1,5	Unknown type	0,01	0,02	0,04
Bicycle boulevard	0,3	0,8	1,1	Equivalent	0,01	0,02	0,04
Normal road	0,9	1,3	1,5	Give way	0,0	0,0	0,03
Solitary cycle (and moped) path	0,2	0,2	1,2	Priority	0,02	0,03	0,05
Ferry	n/a	0,9	n/a	Traffic signals	0,03	0,04	0,08
Service road	0,4	0,7	0,8	Roundabout	0,02	0,03	0,06
Pedestrian short cut	0,4	0,9	1,0	Roundabout with traffic signals	n/a	0,01	n/a
Pedestrian area	0,6	0,7	0,7				
Cycle lane	0,4	1,3	1,3				

All tests of this sensitivity analysis were based on the base values in table 5.4. Every parameter, except of the ferry and roundabout with traffic signals, was changed in two tests. In one test it was set to the lowest value and in one to the highest value. The ferry parameter and roundabout with traffic signals parameter were not changed because there are no ferries and intersections of that type present in the network.

All combinations scored equally good or worse than the set with only base values, except for one combination. The set with a high value for the separated cycle (and moped) path along road (test 70) scored better with 40 counts with a T-value under 3,5, 16 between 3,5 and 4,5 and 29 over 4,5%. Appendix B provides an overview of all scores and settings.

Apart from the variations in distance and intersection parameters, different time parameter values were tested while using the base values again for all other parameters. Just like with the initial values, it appeared that the time parameter does not have an influence in itself: all settings (from 0 to 1 with steps of 0,2) resulted in 37 counts with a T-value under 3,5, 18 between 3,5 and 4,5 and 30 counts with a T-value more than 4,5.

Check of OD-matrix

The results until now did not seem very good, because just 44,7% of the counts is within the best category, while 80% is the standard. This is still just a small improvement from the all-or-nothing assignment: 42,4% (36) of the counts have a T-value under 3,5, 21,2% between 3,5 and 4,5 and 36,5% of the counts have a T-value more than 4,5. As visible in figure 5.1, both assignments result in low numbers on two parallel routes crossing the railway. For cyclists in this area, there is no other choice than to cross the railway using one of these roads. Therefore, it seems that the OD-matrix containing the number of cyclists between areas, is not accurate causing too low numbers of cyclists between the north of Tilburg and the south.

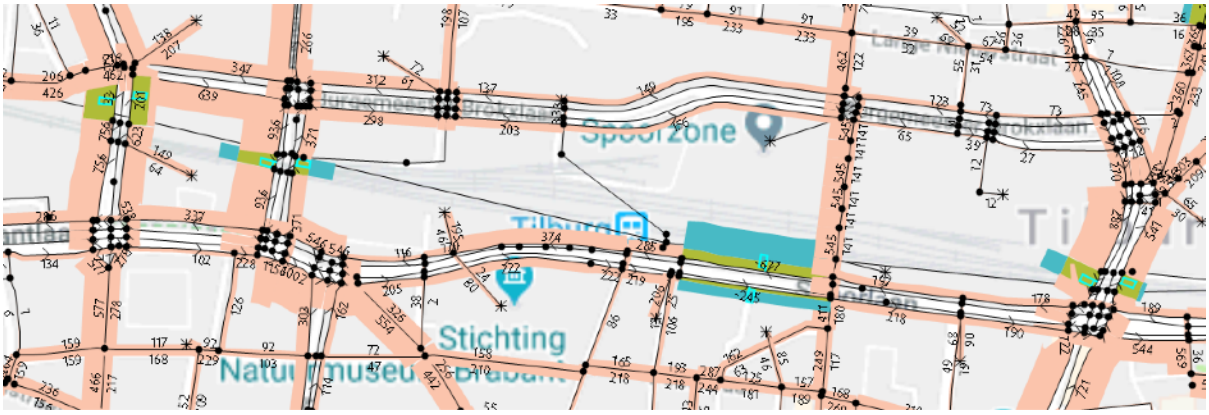


Figure 5.1. Absolute differences between all-or-nothing assignment and counts on the roads around Tilburg railway station. Blue indicates a higher counted number of cyclists than calculated number, green indicates the corresponding part and pink indicates the number of cyclists that is not counted, but is calculated.

To test the hypothesis that the matrix has too low values in it, it was needed to know what the number of cyclists is in the tunnel next to the railway station. In total, there should be an equal number of cyclists going south in every assignment and in the counts. The municipality of Tilburg provided a more recent count in the tunnel (2017) which shows 260 cyclists going south and 113 cyclists going north. In the same year as the rest of the counts, 2015, this tunnel seemed to be not open yet.

In 2015, 3124 cyclists are modelled to cross the railway in southerly direction using one of these roads, while according to the counts, 3539 cyclists passed these roads (excluding the tunnel). That means that there were 405 too few cyclists in the OD-matrix. In the other direction, 374 cyclists were missing in the assignment compared to the counts.

On other locations with a limited number of roads available, similar situations occur. From Tilburg to Goirle, cyclists must cross the A58 motorway, which is only possible on three locations, of which one does not lead to the centre of Tilburg. Figure 5.2 shows the result of the all-or-nothing assignment on these locations.

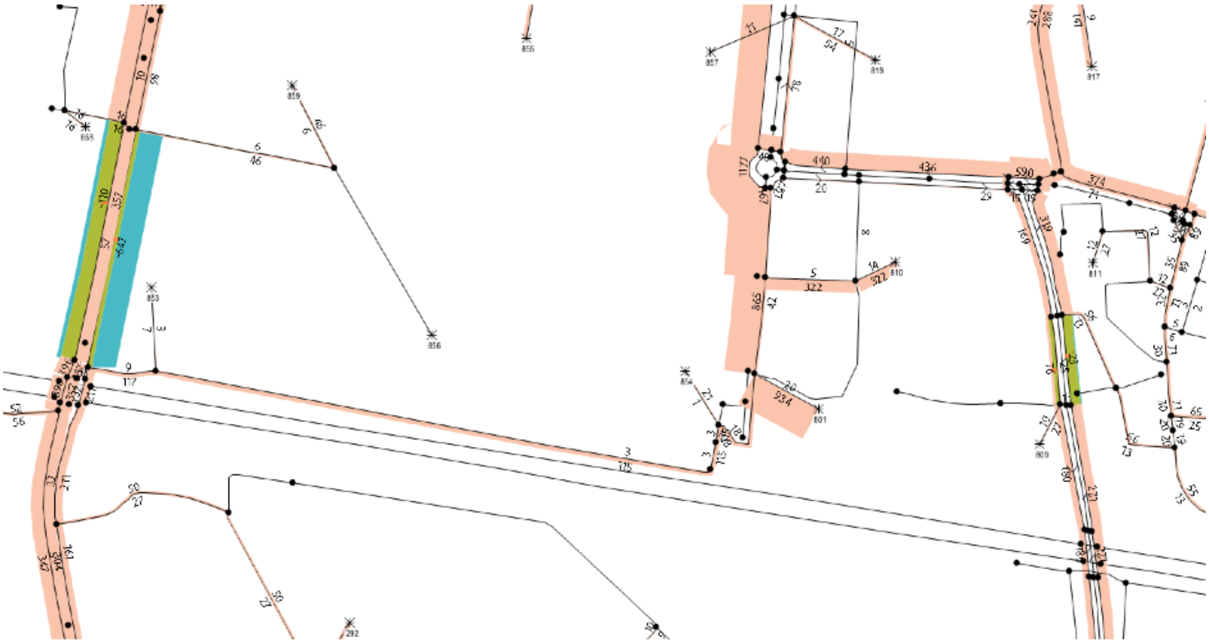


Figure 5.2. Absolute differences between all-or-nothing assignment and counts around the Goirleseweg (left) and Stappegoorweg (right).

In order to compare the results, it should be noted that along the Goirleseweg, there is a service road on both sides of the main road. Cyclists may use both service roads in both directions, but the counts seem to include both roads. The differences are therefore smaller than is shown in figure 5.2. The total difference between assignment and counts is -313 in northerly direction, while there is no significant difference (20) in southern direction.

Because of the too low assigned number of cyclists on multiple roads where there should be more traffic, it seems that the OD-matrix has too few cyclists between certain areas. To test the effect of more cyclists, some tests were repeated with a changed matrix. The matrix was as a whole multiplied with 1,2 in order to add 20% more cyclists on all relations.

Multiplied matrix

After multiplication the total number of assigned cyclists crossing the railway in Tilburg West was more close to the total number of counted cyclists on those roads. For this OD-matrix, tests 26, 33, 49, 55, 61, 70, 71, 90 and 96 were repeated, as these were the best results until now. The scores of these settings are shown in table 5.5 together with the results of a new all-or-nothing assignment. The repeated tests are indicated with * behind the number. It became clear that the multiplication of the matrix does not improve the results, but decreases the scores of each test, especially of the all-or-nothing assignment. The logit assignments did score relatively better than the all-or-nothing assignment, because the latter decreased more.

Table 5.5. Number of counts where the assignment has a T-value under 3,5, between 3,5 and 4,5 and more than 4,5 with different parameter combinations and an multiplied OD-matrix.

Test	T < 3,5	3,5 < T < 4,5	T > 4,5	Test	T < 3,5	3,5 < T < 4,5	T > 4,5
26*	36	14	35	**70***	39	11	35
33*	32	18	34	**71***	35	12	38
49*	38	16	31	**74***	35	13	37
55*	35	12	38	**90***	34	13	38
61*	35	13	37	**96***	35	15	35
AON*	36	18	31				

Because of the new matrix, it is also possible that the earlier found parameter combinations were not optimal at all, because the values were calibrated on a different network. Therefore, 25 extra parameter combinations have been tried (tests 100 to 124) with parameter values in between the values belonging to test 49* and 70*. These were randomly chosen within this range. No combination in these tests could improve the score of test 49* or 70*. The results of these tests can be found in appendix B. Because test 49* had the lowest number of low-scoring counts, this is seen as the final calibration result.

5.4 Analysis of assignment results

The assignments with the settings belonging to test 49* and 70* are still the best-scoring parameter combinations. An overview of the intensities in those assignments is visible in appendix B.

The first impression is that the results are not satisfactory. Only 46% of the counts fall within the best category, while 80% is desired. Possible explanations include that the OD-matrix is not accurate enough, because it was not calibrated using this assignment or using the available counts. Furthermore, it seems possible that more optimal settings exists, but that these were not found. If more time would have been available, these could have been found, but that was not the case.

These assignment results were analysed with the help of the person responsible for the modelling of Tilburg at Goudappel Coffeng. He did not see abnormalities in the results and thought it could be a realistic assignment. However, he thought that the number of cyclists in the cyclists' tunnel near the railway station (Willem II-passage) was very low compared to the number on the alternative roads. He thought that there should be at least an equal number on both routes and probably more in the tunnel dedicated to cyclists. With additional data of the municipality of Tilburg, it became clear that since the opening of the tunnel, this tunnel did attract traffic from the other routes, but the numbers still are much lower (±250 southbound in the tunnel, ±1000 in the existing NS-plein tunnel and ±1400 in the Gasthuisring tunnel). The assignment with the settings of test 70 assigned 520 cyclists to the Willem II-passage, so actually overestimated the number. However, it should be noted that the assignment is based on data from 2015, while this count is from 2017.

In general, the assignments with settings of experiment 49 and 70 had the most deviation from the count values around the railway station. This can be partly caused by the fact that people that use public transport and cycle to or from the railway station are not included in the OD-matrix. In the other counts, no pattern could be found in the types of roads or intersections that had too low or high values.

The use of factors for the type of road and intersection has an influence on the assignment results. In figure 5.3, the relative difference between an assignment with settings of test 49 and the all-or-nothing assignment are shown in the north of Tilburg. More cyclists take the Kraaivenstraat (left) than in the all-or-nothing assignment, while the Midden-Brabantweg loses cyclists. This can be explained by the fact that most of those cyclists want to continue south on the Kraaivenstraat, but they will come cross traffic signals once more than on the other route.

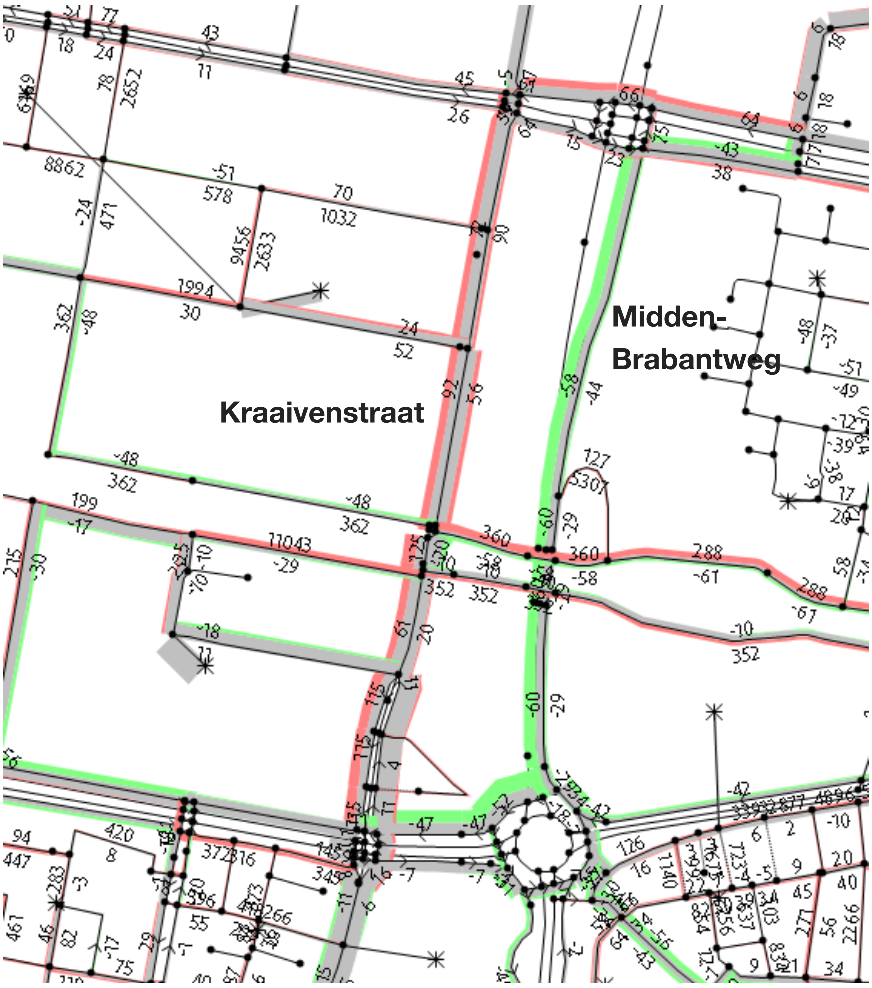


Figure 5.3. Relative difference between assignment results with settings of test 49* and the all-or-nothing test. Green means less cyclists in test 49* compared to the all-or-nothing assignment, red means more.

Results of the settings of test 70* do not show these differences in this area (figure 5.4). This can be explained by the increased parameters of the intersections on the Kraaivenstraat in test 70* compared to test 49* and the slightly lower value for traffic signals. Intersections with priority have a value of 0 in test 49*, while it is 0,03 in test 70*. These are present on multiple locations in the Kraaivenstraat, so change the utility of the routes over it significantly.

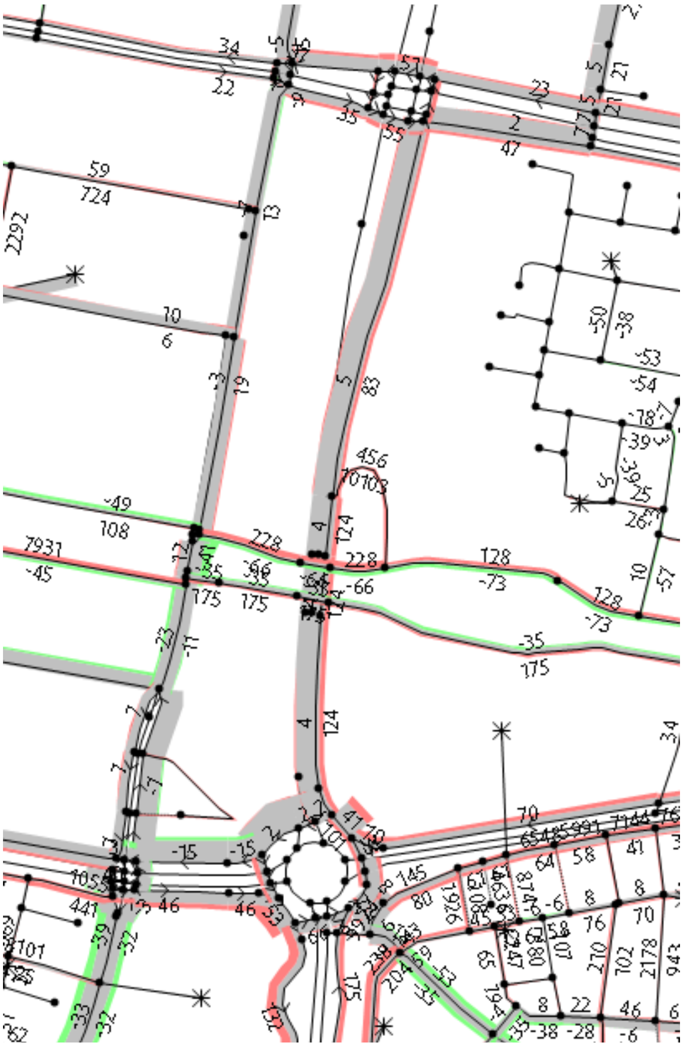


Figure 5.4. Relative difference between assignment results with settings of test 70* and the all-or-nothing test. Green means less cyclists in test 49* compared to the all-or-nothing assignment, red means more.

Figures 5.5 and 5.6 show that both assignments assign more traffic to a slightly longer route to cross the Ringbaan West which is a solitary cycle path instead of the shortest route which is a separated cycle path. Furthermore, the longer route crosses the Ringbaan West with a tunnel (Iepenpad), while the shortest route has traffic lights. This shows that the algorithm exactly performs how it should: it will allow cyclists to take longer routes if they are more liked by cyclists.

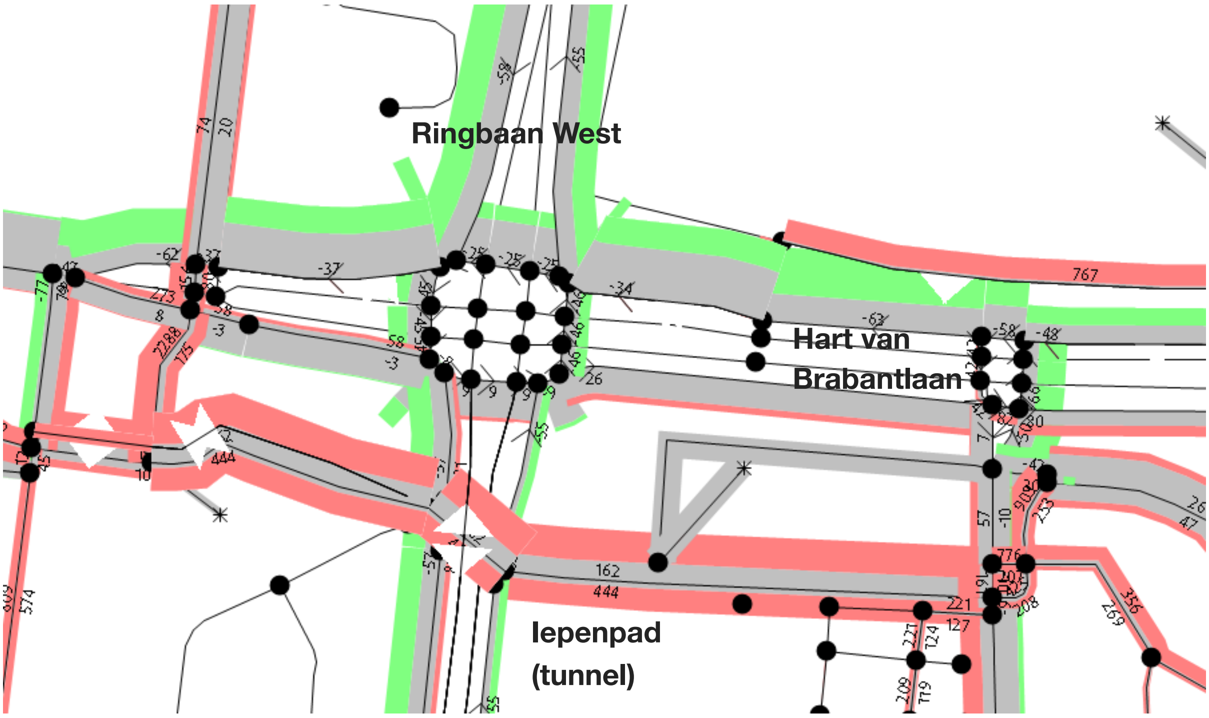


Figure 5.5. Relative difference between assignment results with settings of test 49* and the all-or-nothing test around the intersection between Ringbaan West and Hart van Brabantlaan. Green means less cyclists in test 49* compared to the all-or-nothing assignment, red means more.

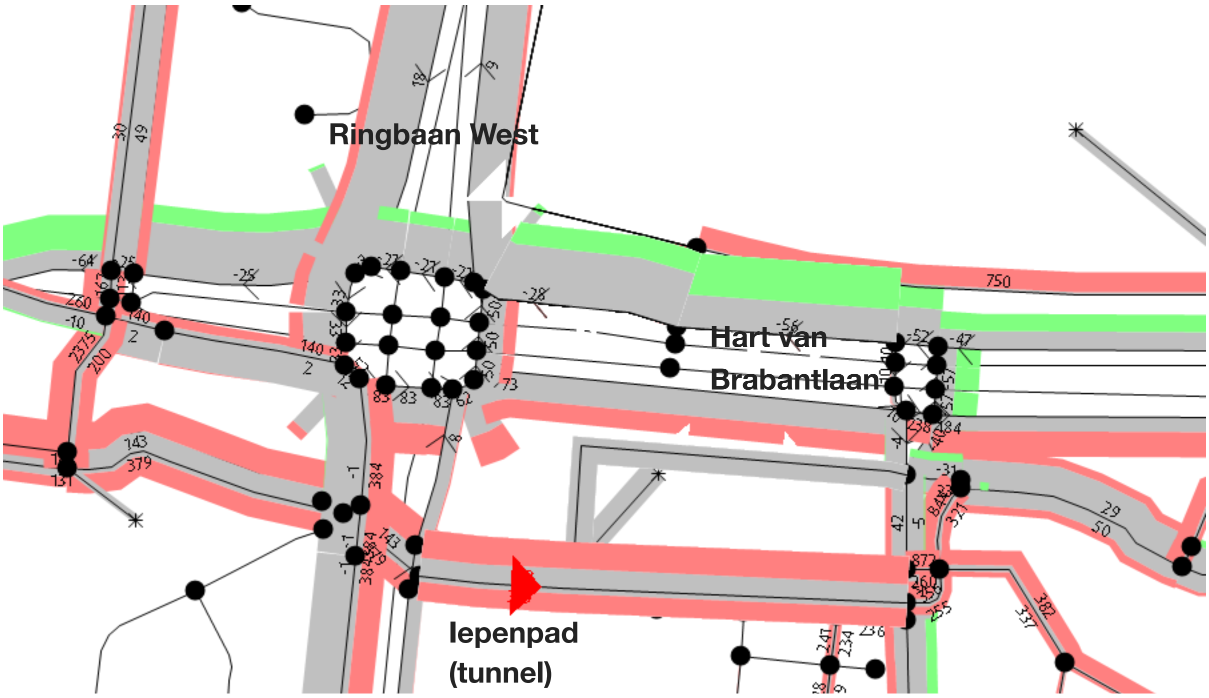


Figure 5.6. Relative difference between assignment results with settings of test 70* and the all-or-nothing test around the intersection between Ringbaan West and Hart van Brabantlaan. Green means less cyclists in test 49* compared to the all-or-nothing assignment, red means more.

Finally, the benefits of the algorithm are visible in the west of Tilburg too. Compared to the all-or-nothing assignment more cyclists take the southern route between Reeshof and the city centre (figure 5.7). The southern route has more stretches with solitary cycle paths than the northern route and does not include any normal roads as the northern route does. Furthermore, the southern route has less curves, which decreases the travel time on that route, and probably most important, the route has just 10 intersections between Reeshof shopping centre and the Statenlaan. The northern route has 30 intersections in the same section. The types of intersection are in both cases mostly equivalent intersections with both having one intersection with traffic signals (Burgemeester Baron Van Voorst tot Voorstweg).

Figure 5.7. Relative difference between assignment results with settings of test 49* and the all-or-nothing test in the western areas of Tilburg. Green means less cyclists in test 49* compared to the all-or-nothing assignment, red means more. The blue point indicates the shopping centre in Reeshof, the orange points indicate the locations where the routes cross the Statenlaan.

5.5 Comparison with other model

As described in the methodology section the results were also compared on the basis of the GEH statistic. Table 5.6 presents the GEH statistics for the Tilburg network assigned with the all-or-nothing algortihm and the final settings (test 49). These were compared to Christou (2017) who performed an user-equilibrium assignment based on perceived time in Chelmsford (UK).

Table 5.6. GEH statistic of AON and logit assignments in Tilburg and speed-based assignment in Chelmsford (Christou, 2017).

Location	Type of assignment	Number of counts with GEH < 5	percentage of counts with GEH < 5	Number of counts with GEH ≥ 5	percentage of counts with GEH ≥ 5
Tilburg (initial OD-matrix)	All-or-nothing	36	42,4%	49	57,6%
	Logit (test 49)	33	38,8%	52	61,2%
Tilburg (multiplied OD-matrix)	All-or-nothing*	34	40,0%	51	60,0%
	Logit (test 49*)	35	41,2%	50	58,8%
Chelmsford, UK	User-equilibrium (morning peak)	102	96%	4	4%
	User-equilibrium (between peaks)	106	100%	0	100%
	User-equilibrium (evening peak)	102	96%	4	4%

It is clear that the algorithm does not perform as good as the algorithm in Chelmsford, but it is important to note that these cannot be compared directly, as the area and country is totally different. What can be concluded is that it would be preferred if it was possible to obtain scores like Christou (2017) obtained. Furthermore, the algorithm does seem to score a bit better than the all-or-nothing algorithm when the multiplied OD-matrix was used. That conclusion is the same as when using the t-test. Using the t-test, however, the algorithm based on logit scored better without the multiplied matrix as well, while it scores worse using GEH. This can be explained by the different ways of measuring the closeness to the counts.

6. Validation

The calibration results were only fitted to one situation, but that does not mean that the found parameter values are valid for other locations. In this validation the best parameter combination (test 49) is applied to a part of the The Hague and Rotterdam network. The first section describes the network and the method of validation. The second section continues with the results of the validation of the route generation, followed by the results of the assignment itself. Finally, the results are compared to the results of the calibration and to other algorithms.

6.1 Network and methodology

The algorithm with the parameter set of test 49 was validated on the The Hague and Rotterdam network. Because of time limitations, it was unfortunately not possible to validate on the whole network. A cut-out was made of The Hague and surrounding towns and cities in order to include shorter-distance urban traffic and longer-distance interurban traffic. Towns and cities just within the cut-out include Monster, Wateringen, Delft, Pijnacker and Zoetermeer. The cut-out network consisted out of 2817 zones with 65617 links. In this area, 419 cyclist counts area were available. 135 of these are counted in 2014, 59 in 2015 and 229 in 2016.

The network contained extra information about cycling facilities provided by the Fietsersbond. This information was the same as in the Tilburg network, so it included the same types of intersections and road types. However, a high number of intersections did not have an associated type, which means that the algorithm could not calculate the correct utility (see section 2.3). The road types seemed to be implemented correctly in nearly all cases, because no incorrect types were found in a sample of the roads.

The validation method of the algorithms was the same as the calibration methods, but now, only the best setting is tried. In the case of the route generation, the intermediate setting, as it was called in the Tilburg network, was applied to The Hague and surrounding areas. Again, the routes of different OD-pairs were compared with Fietsersbond route planner results and a map of the area. The assignment was done with the settings of test 49, as those gave the best results in the calibration. The assignment results were compared with an all-or-nothing assignment and the scores in Tilburg using the T-test values and the GEH-statistic.

6.2 Route generation

The route generation was validated in a similar way to the calibration by inspecting the routes belonging to a selection of OD-pairs. The considered pairs were the following:

- Delft (Voorhof) - The Hague city centre
- Scheveningen - The Hague city centre
- Wateringen - The Hague (Leyenburg)
- Zoetermeer: Palenstein - town centre
- Monster - Scheveningen/Statenkwartier

The first pair is a longer-distance pair between two cities, while the second one is between a suburb of The Hague and the city centre. The third and fourth pair have a shorter distance between origin and destination. The third one is between a suburb and a neighbourhood of The Hague, the fourth between Zoetermeer's town centre and the neighbouring area of Palenstein. These were selected because of the difference in distance and type of route (between cities, between neighbourhoods and to a city centre).

In general, the generated routes seemed quite good. The longest route in the selection, between Delft and The Hague, could be improved with some extra routes in the Rijswijk area (halfway between Delft and The Hague). The *fietsbewust* route of the *Fietsersbond* route planner was for example not included (see figures 6.1 and 6.2). The *autoluw* route was not included as well, but it takes an enormous detour, so this was not regarded as a problem.

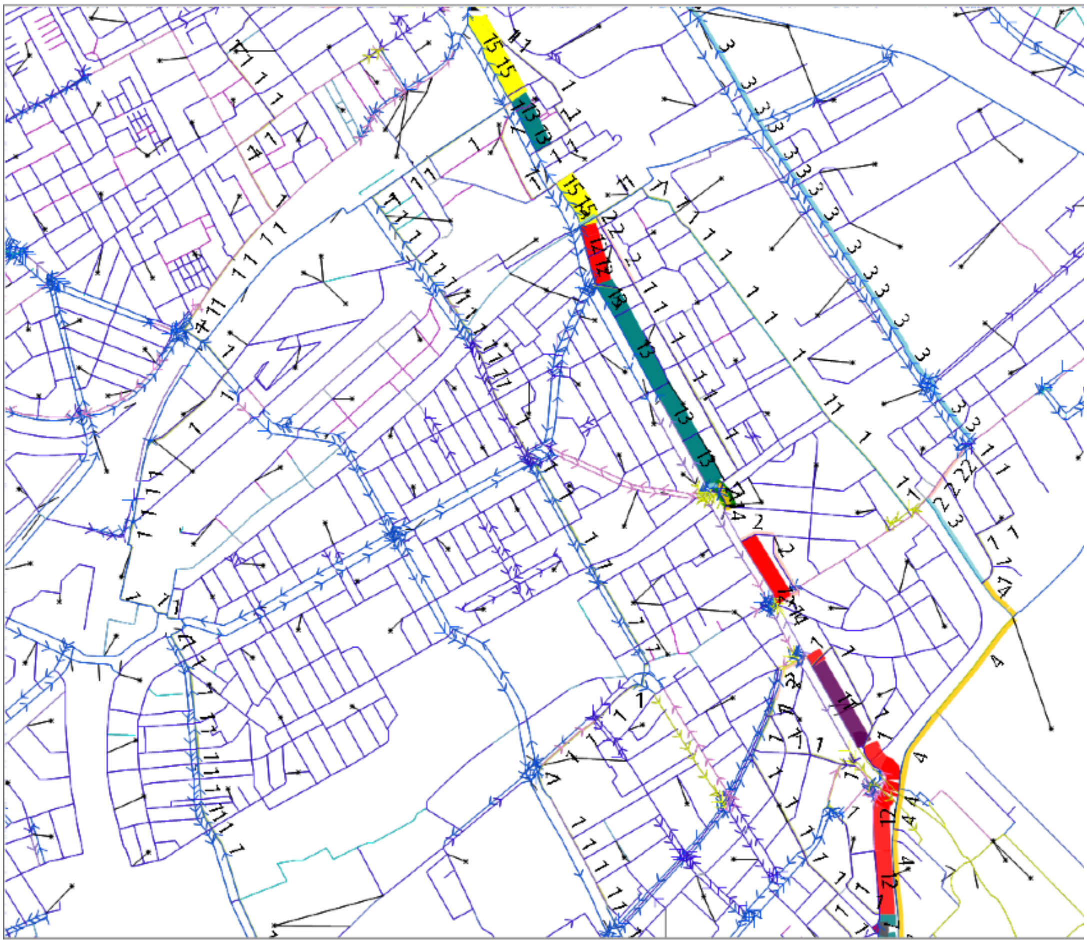


Figure 6.1. Detail of generated routes between Delft and The Hague city centre through Rijswijk.

Figure 6.2. Part of the *fietsbewust* route between Delft and The Hague in Rijswijk and The Hague south (Fietsersbond, 2018).

In the OD-pairs between Scheveningen and The Hague city centre and between Wateringen and Leyenburg, the *fietsbewust* routes were not wholly included as well. In Scheveningen, a few hundred metres are missing, possibly due to differences between the network and the real road network, but in Wateringen it did seem a reasonable route that should have been included. All pairs included the *makkelijk doorfietsen* routes and shortest routes. The *autoluw* routes were not included in the other OD-pairs as well, because they seemed quite unreasonable compared to *autoluw* routes in Tilburg or 's-Hertogenbosch (see figures 6.3 en 6.4). Table 6.1 presents an overview of the number of OD-pairs of which the planned routes were generated.

Table 6.1. Number of OD-pairs in which the *Fietsersbond* routes were included using the different settings. Half a route is indicated when just a very short part of the route is missing.

Route type in *Fietsersbond* route planner	Nr of OD-pairs
makkelijk doorfietsen route	5
shortest route	5
fietsbewust route	1,5

Route type in *Fietsersbond* route planner	Nr of OD-pairs
autoluw route	0

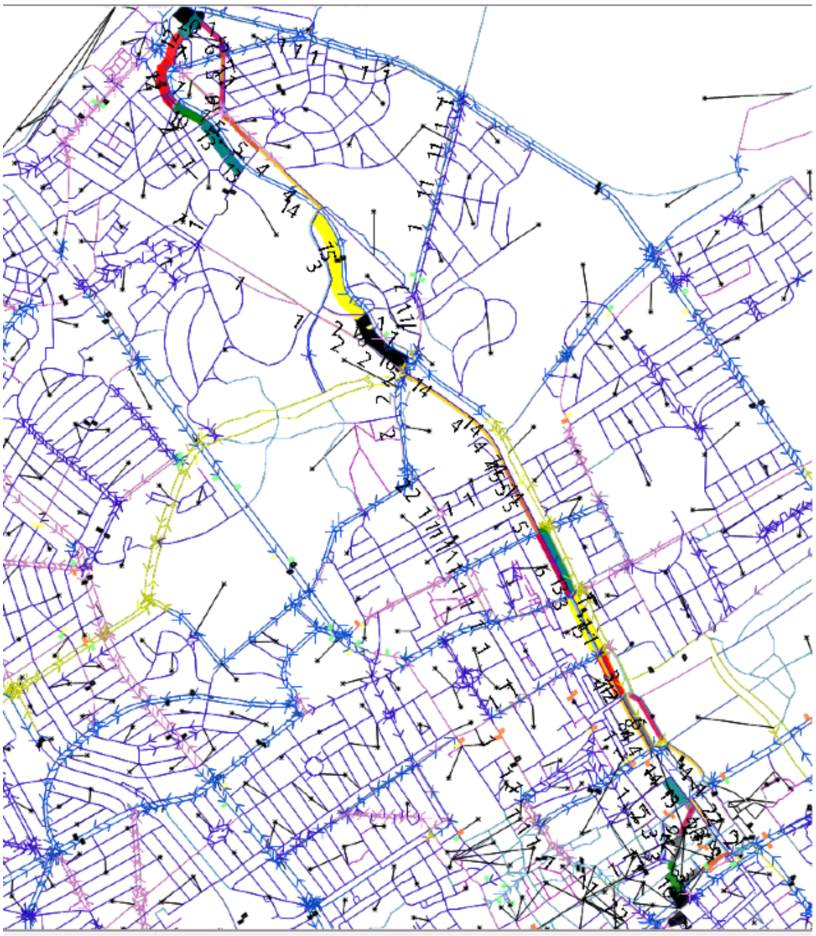


Figure 6.3. Generated routes between Scheveningen and The Hague city centre.

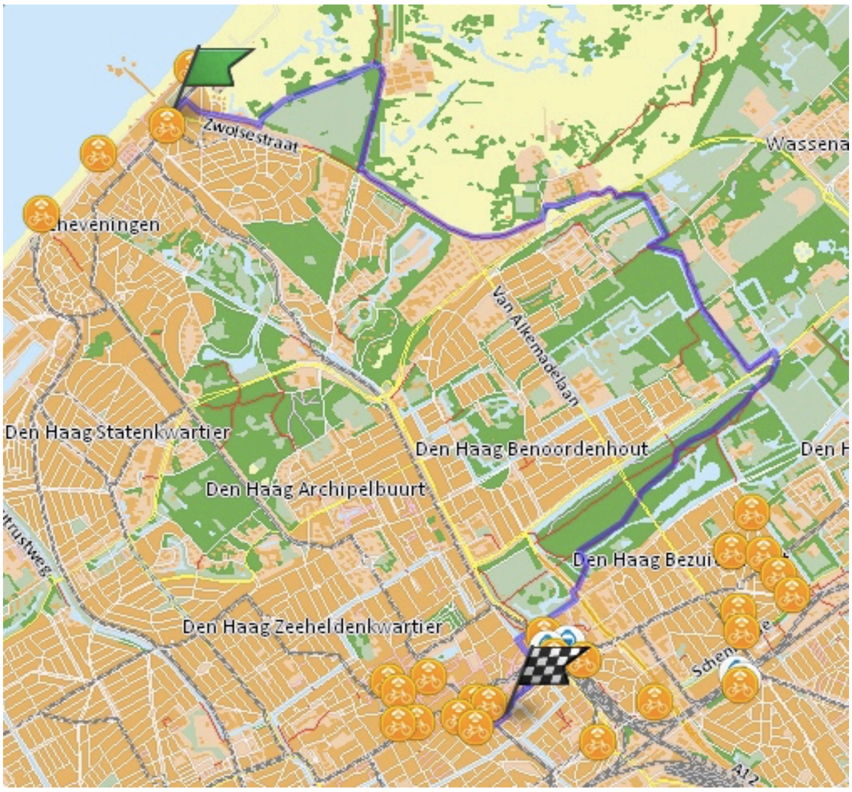


Figure 6.4. ***Autoluw*** **route between Scheveningen and The Hague city centre (Fietsersbond, 2018).**

6.3 Assignment

To validate the assignment algorithm and the parameter values, the best scoring parameter set was applied to the The Hague network. The parameter set which gave the least counts with a T-value higher than 4,5 in Tilburg were the parameters of experiment 49. Because the results are not directly comparable to Tilburg, an all-or-nothing assignment was performed as well. The assignments in The Hague city centre are visible in figures 6.5 and 6.6.

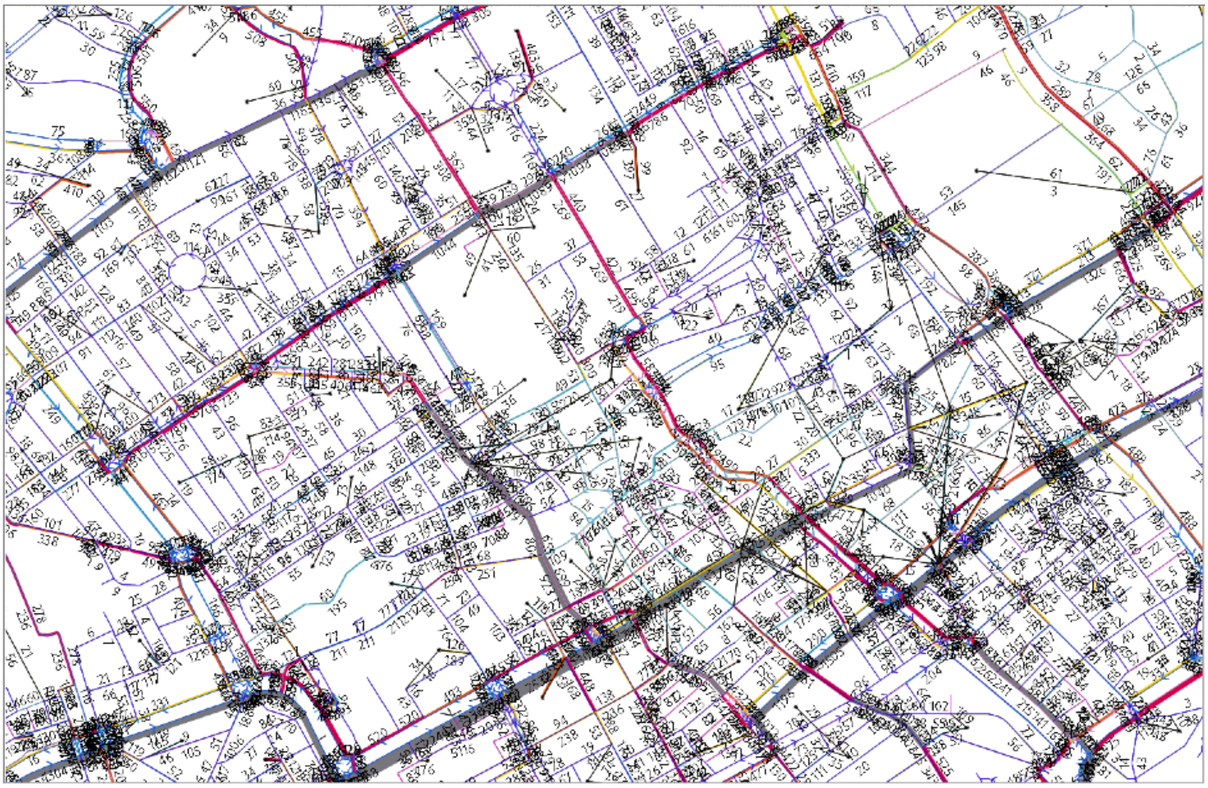


Figure 6.5. Detail of the assignment in The Hague city centre using the parameter settings of experiment 49.

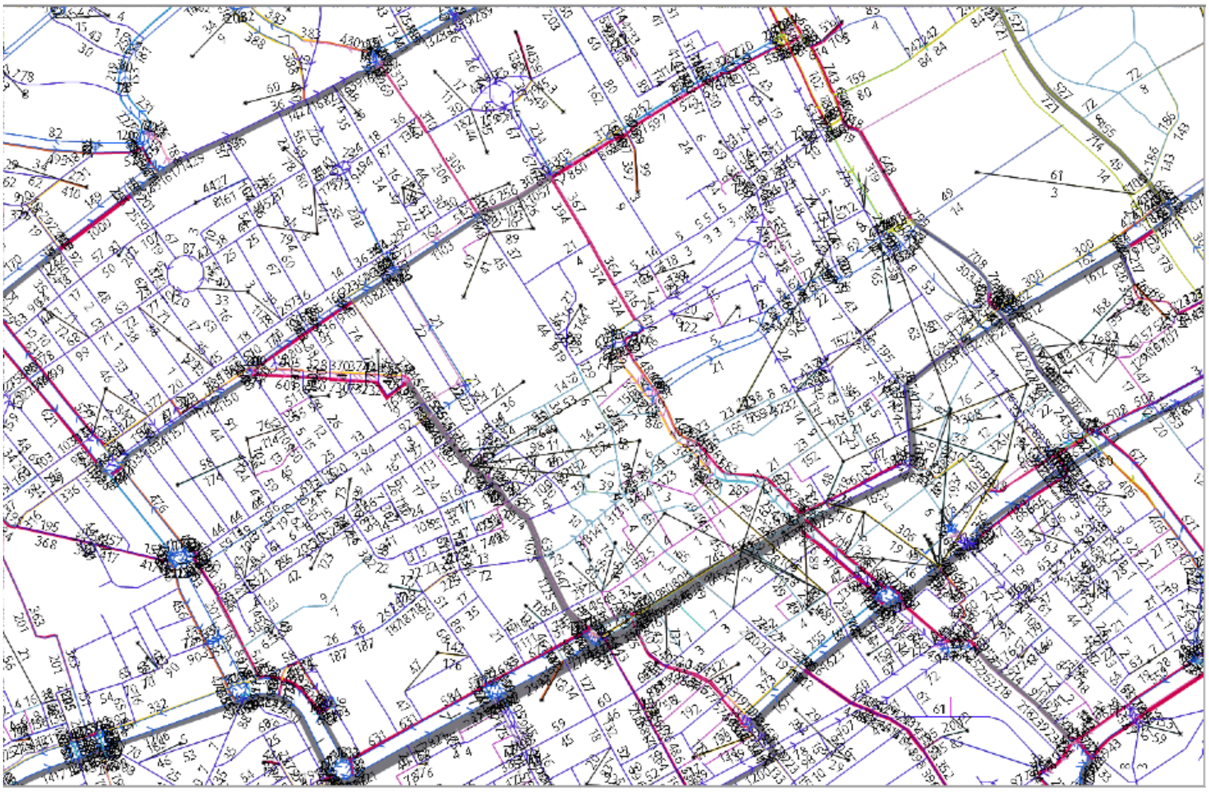


Figure 6.6. Detail of the all-or-nothing assignment in The Hague city centre.

The logit assignment again seems to distribute cyclists over more routes. In the logit assignment up to 368 more cyclists use the Spuistraat and nearly 200 the Gedempte Gracht, while the number of cyclists on the parallel Grote Marktstraat decreases with around 1000 in the same direction (figure 6.7).

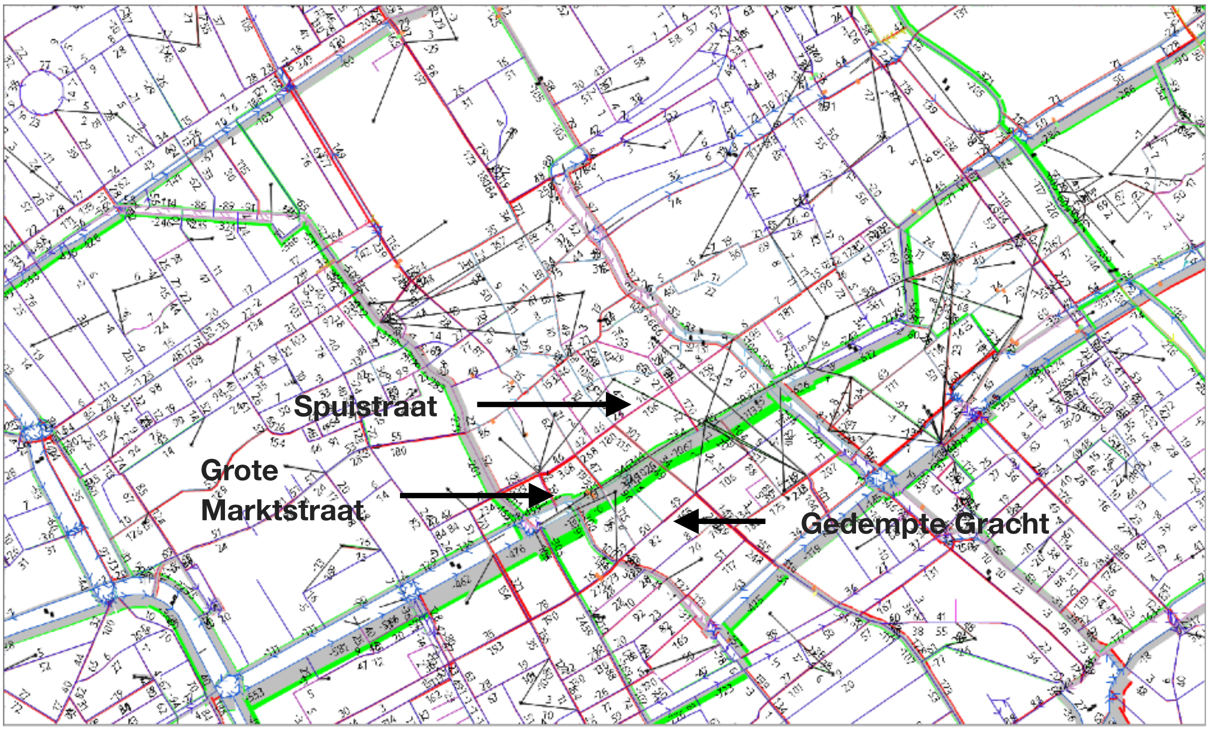


Figure 6.7. Absolute difference between logit and all-or-nothing assignment in The Hague city centre. Green indicates less cyclists in the logit assignment compared to the all-or-nothing, while red means more.

The effect of the parameters for different road types is not clearly visible in the city centre, but it is very clear on two roads between Scheveningen and the city centre (figure 6.8). The Haringkade (northeast on the map) has less cyclists than in the all-or-nothing assignment, while the Scheveningseweg (southwest) has an increase on most of its length. The Haringkade is a normal road with a cycle lane in one direction, while the Scheveningseweg has two separated cycle paths. The parameter settings implicate higher costs for roads with cycle lanes than for separate paths. This shift causes the assignment to be closer to the counts with nearly the right number of cyclists on both roads (e.g. 319 counted, 390 calculated on the Scheveningseweg).

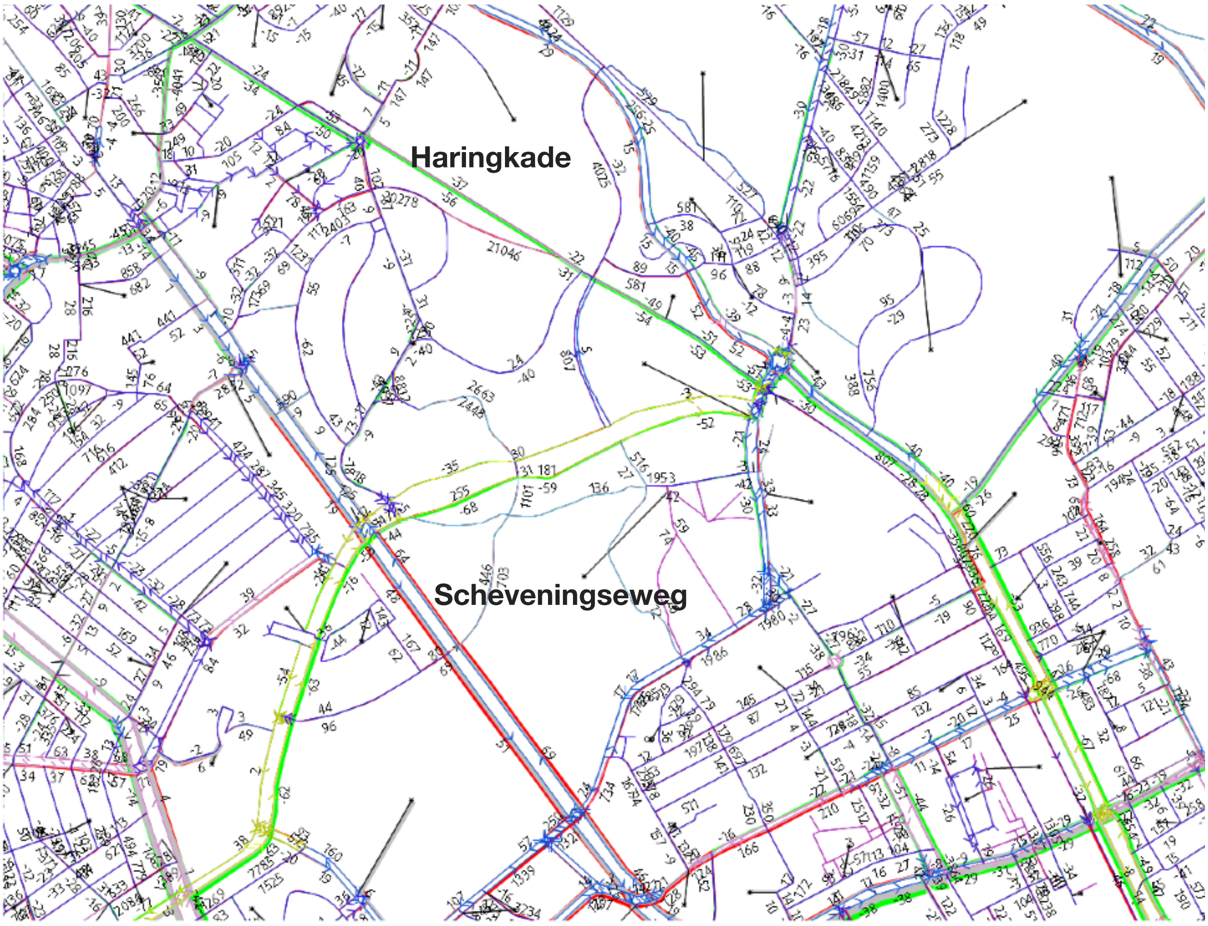


Figure 6.8. Relative difference between assignment results with settings of test 49 and the all-or-nothing test between Scheveningen and The Hague city centre. Green means less cyclists in the logit assignment compared to the all-or-nothing assignment, red means more.

Similar results emerge in the south of The Hague, where there are two roads along both sides of the HS railway station (figure 6.9). The northern road, the Parallelweg, has cycle lanes, while the southern route, the Waldorpstraat, has a separate cycle path. Again, the road with the separate path is used more in the logit assignment and the alternative is used less compared to the all-or-nothing assignment. Furthermore, part of the traffic on the Parallelweg comes from a residential street which is less used as well.

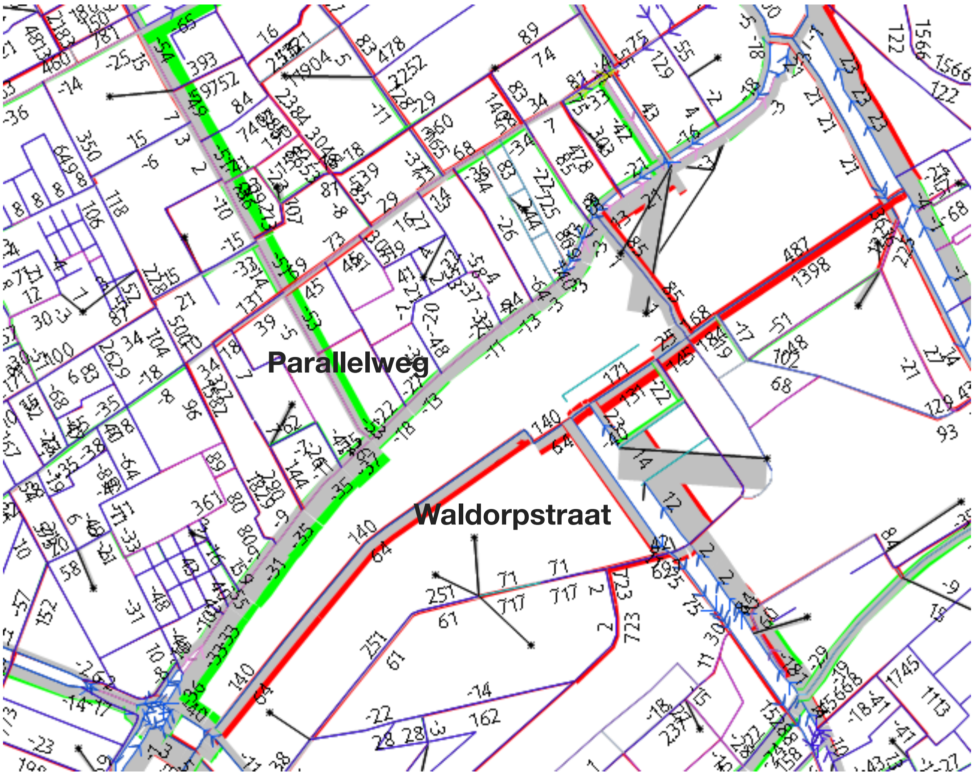


Figure 6.9. Relative difference between assignment results with settings of test 49 and the all-or-nothing test around The Hague HS railway station. Green means less cyclists in test 49 compared to the all-or-nothing assignment, red means more.

The performance of assignments was assessed with the T-test and GEH-statistic. The T-test results of the assignments in The Hague and in Tilburg are summarised in table 6.2 and the GEH-statistic results are visible in table 6.3.

Table 6.2. T-test results of the validation and calibration of the logit assignment using parameters of experiment 49 and the all-or-nothing assignment.

Location	Type	T < 3,5	3,5 < T < 4,5	T > 4,5
The Hague	**Logit (experiment 49)**	196 (46,8%)	68 (16,2%)	155 (37,0%)
	AON	181 (43,2%)	80 (19,1%)	158 (37,7%)
Tilburg (multiplied matrix)	**Logit (experiment 49)**	38 (44,7%)	16 (18,8%)	31 (36,5%)
	AON	36 (42,4%)	18 (21,2%)	31 (36,5%)

Table 6.3. GEH-statistic results of the validation and calibration of the logit assignment using parameters of test 49 and the all-or-nothing assignment.

Location	Type	Number of counts with GEH < 5	percentage of counts with GEH < 5	Number of counts with GEH ≥ 5	percentage of counts with GEH ≥ 5
The Hague	**Logit (experiment 49)**	180	43,0%	239	57,0%
	AON	154	36,8%	265	63,2%
Tilburg (multiplied matrix)	**Logit (experiment 49)**	35	41,2%	50	58,8%
	AON	34	40,0%	51	60,0%

Both the all-or-nothing assignment and the logit assignment scored more 'extremely' in the t-test compared to the performance in Tilburg. The percentage of counts in the worst category increased, but the percentage of counts in the best category increased as well. The logit assignment did score better than the all-or-nothing, but the difference was not enormous. The percentage worst counts is 0,7 percentage points lower and the percentage best counts 3,6 percentage points higher.

Similar results emerged when comparing with the GEH-statistic. The logit assignment scored 6,2 percentage points better than the all-or-nothing assignment. In Tilburg, the difference was only one count corresponding to 1,2 percentage points. All in all, the logit model with the parameter settings of experiment 49 seems to work outside Tilburg as well and scores compared to the all-or-nothing assignments better in The Hague than in Tilburg.

7. Discussion

Every research is based on assumptions or data that are not fully accurate. That of course applies as well for this research. The most important source of uncertainty in this research is the quality of the data. This applies to both the input of the model and the counts with which the output was validated. The input of the model includes the OD-matrix, the network and the calibrated parameter settings. Also, the modelling requires simplifications and assumptions which might be wrong. This chapter describes the main uncertainties and limitations of this research.

Data

The assignment depends on the number of trips in the OD-matrix. The limitations of the OD-matrix are that the values are based on a model itself. The OD-matrix was estimated with a gravity model and results of a national research in the mobility of people. This introduces uncertainty in the OD-matrix and that has consequences for the quality of the assignment. Furthermore, an all-or-nothing assignment was used to estimate the resistance against a movement between two areas. This implicates that a movement between an origin and a destination can be seen as difficult, because the distance is very long. Using the logit assignment however, that route might be attractive because it is a very cycle-friendly road and more cyclists move between those areas. Assigning cyclists based on that OD-matrix will result in too few cyclists on those pairs.

The OD-matrix contains cyclists that use their cycle as only mode in their whole trip. This means that people that cycle on their way to or from the train, are not included. It is plausible that this has caused deviations in the result as the assigned number of cyclists was lower than the counts around the railway stations of Tilburg and The Hague.

Another main limitation is the quality of the data about road and intersection properties. The used data is provided by the Fietsersbond which is an organisation run by volunteers. The data could contain inaccuracies as it is totally provided by them. The quality of the data seemed to differ strongly between the attributes. The road types are in most cases accurate, but the intersection types are missing on various locations. Overall, it is possible that the lack of data about the intersection types may have caused different results.

Furthermore, the road type and intersection data use a certain categorisation that does not necessarily class the types according to the perception of all cyclists. Some cyclists might not perceive an intersection with traffic signals that are nearly always green, in the same way as another intersection on which they have to wait a long time. In the data, these intersections are both classed as the same type. More generally, cyclists perceive the same road types differently and even cyclists from a certain area may perceive it differently from another area. The algorithm applies the same parameter values to each OD-pair, so it is assumed that everyone chooses in the same way everywhere within the area, which does not have to be true.

Besides the different perception of a factor, people find different aspects important while choosing their route. In this research a selection has been made of the factors that seemed most relevant to Dutch cyclists and of which data is available. Because there is very little research into the factors that influences route choice in the Netherlands, the factors can be different from the implemented factors in this algorithm. The selection of factors is based on research that could not be applicable to the Netherlands, because it was done abroad.

The data about the average speed causes some uncertainty as well. These are based on results of previous research and were generated by fitting observations to a formula. There is a chance that these findings are not accurate. The consequences are probably not huge, because the parameter regulating the influence of the time and speed did not seem to have a great influence on the outcomes of the assignment.

Moreover, all comparisons between model results and reality are based on cyclist counts. Most of the counts are correct on themselves, but they are not done on the same day or even type of day. Some counts are done whilst it was rainy day, others on a sunny day. It is possible that the number of cyclists on the rainy days is lower than on the sunny days. In The Hague, some counts were done using a road tubes. Those cannot register if two people pass the tube at the same time, so cause flaws in the data if two people cycle next to each other. Furthermore, visual counts have their problems too. It can be very difficult to know the exact number if large groups pass the person responsible for counting.

Finally, the network itself could create problems. Some roads might be missing in the networks and consequently, the algorithms give some cyclists a different route to what they would do in reality. More influential is the location of feeding links to centroids. These are the location where traffic starts and ends in that part of the neighbourhood. If people starting from that zone in reality do not pass that point, their route could be totally different and the assignment as well as a consequence of that. Counts will show lower use of the roads than the model, even if the assignment algorithm works correctly.

Calibration and validation process

The calibration and validation of the route generation algorithm introduces a lot of uncertainty. Because of time and storage limitations it is impossible to generate many routes, let alone all all routes. The algorithm therefore needed to be calibrated in order to find a balance between storage, computational time and quality of the route set. But the quality of the route set cannot be measured exactly. It has been assessed by comparing the map with the generated routes and planned routes of the *Fietsersbond* route planner. These are hardly exact measures. If routes are not included when they should, it can have a significant influence on the result of the assignment, because the route cannot be assigned if it is not generated.

The time limitation had its influence on the calibration of the second part of the assignment algorithm as well. If there had been more time, more parameter combinations could have been tried and therefore, a better result could have been obtained. The results could be seen as a sort of minimum possible score: what has been found can only be improved.

Another point which introduces uncertainty are the not-calibrated parameters. The path-size factor parameter was not calibrated to limit the number of variables, but in most research, this parameter is used to improve the choice model. However, the correction for overlapping routes is now better than if it was changed. There still could be some small problems with overlaps on locations with parallel roads. In the networks, some roads have a bidirectional cycle path on both sides of the main road that the algorithm regards as independent roads. People could perceive these as the same road, which introduces more uncertainty in the result.

Other parameters that were not calibrated, could not be calibrated because of the lack of roads or intersections of a certain type. For example, there were no ferries or roundabouts with traffic signals in the used networks. The ‘found’ parameters were never tested and have therefore a high uncertainty.

Finally, the algorithm was validated with the parameters that resulted from the calibration in Tilburg. It is possible that cyclists in The Hague and surrounding areas have different cycling behaviour which causes the parameters not to be transferable to the rest of the Netherlands.

8. Conclusion

The main research question of this thesis was: *How can the assignment of cyclists in a four-stage model realistically be modelled for use in the Netherlands?.* This question was answered by answering the following subquestions:

1. What are the most relevant factors in the route choice in the Netherlands?
2. What methods do exist for the assignment of cyclists in a four-stage model?
3. What is the most useful method for assignment in the Netherlands?
4. What combination of variables does describe the route choice in the Netherlands best?
5. How well does the all-or-nothing assignment perform in the Netherlands?
6. To which extent is the proposed algorithm an improvement?

Many factors seem to influence the route choice of cyclists. Previous research shows a broad range of factors, from aspects in the environment to road characteristics and personal factors. Finding the most relevant factors was not clear-cut. Nearly all literature is focussed on cities in other areas and/or is less applicable for an assignment because it is a stated-preference study. Experts at Goudappel Coffeng were asked about their opinion on the most important factors in the Netherlands and the locations where previous research took place, were compared. Nearly all sources mentioned distance as one of the most important factors together with travel time. The road type was important in most research, but in one study it was not significant. Intensities of motorised traffic were significant in most relevant studies, but no data is available about that in a bike assignment, except indirectly via the road type. The literature was less clear about for example the influence of the land use around the roads and therefore it was not selected. The selected most relevant factors were distance, time, type of intersection and road types as there seemed to be more agreement on the relevance and because of the availability of data.

The second subquestion focusses on the algorithm which should assign the cyclists using the factors found in the previous subquestion. Cyclists are currently assigned using the all-or-nothing algorithm. In literature, other options were explored, like logit and probit choice models and a user-equilibrium assignment with congestion included. Logit is most used in literature and is able to find multiple routes in contrast with all-or-nothing. A downside is the handling of overlapping routes as independent which they are not. Probit has similar abilities, but is more statistically accurate than logit. It needs Monte Carlo simulation for the calculation of the probabilities, but that takes much time and is therefore a disadvantage. The user-equilibrium assignment is based on the speed, which is on its turn based on road characteristics, car intensities and cycle intensities. To overcome some of the disadvantages of logit and probit a path-size factor can be added to the logit model to correct for the overlaps without the need for simulation.

The most relevant algorithm was selected on the basis of criteria like running speed, ability to assign different routes and the possibility to include the factors selected in the second subquestion. The logit model with path-size factor scored best on these criteria and was therefore selected as algorithm. The probit model was not chosen because of the simulation time and the user-equilibrium was not selected because of the lack of literature about the method and the difficulty in applying the factors. Logit needs routes to choose from and to do so, a Dijkstra's shortest route algorithm was used with Monte Carlo simulation to vary the costs and create multiple routes.

The fourth subquestion was answered by implementing the logit and route generation algorithm in OmniTRANS and calibrating the parameters to find the best fitting combination of values for those. First of all, the route generation algorithm was calibrated on a less detailed 's-Hertogenbosch network for testing purposes and subsequently on Tilburg. Generating 12 to 20 routes and using 0,15 as variance provided the best balance between running time, storage and quality of the generated routes. The assignment was only calibrated on the Tilburg network. Two combinations scored best: one with the least large deviations from counts and one with the most small deviations from counts. These settings both give lower costs to routes with separate cycle paths and higher costs to routes with traffic lights. Using the best setting 44,7% of the counts scored a t-value lower than 3,5, which is a small improvement over the all-or-nothing assignment.

The all-or-nothing assignment did not score well in these networks. In Tilburg, only 42,4% of the counts had a t-value lower than 3,5 and in The Hague only 43,2%. In both networks around 36% of the counts scored a t-value higher than 4,5. Furthermore, the all-or-nothing assignment does not represent cycle route choice well: it is only able to assign cyclists to one route between an origin and a destination, while cyclists choose different routes, as the literature and experts mentioned. Furthermore, in most cases more than one factor determines the route, not only time or distance.

The algorithm was compared with an all-or-nothing assignment applied to the Tilburg network to assess the improvement. The improvement seemed small, the number of good counts increased with 2,4 percentage points using the t-test and it did not improve using the GEH-statistic. The improvement was also assessed on the network of The Hague, Delft and surrounding areas. On this network the logit assignment had an increment of 3,6 percentage points of counts in the best category. The number of counts in the worst category decreased with 0,7 percentage points. The GEH-statistic improvement was bigger: an increase of 6,2 percentage points in the proportion of counts in the best category. The logit assignment can therefore be seen as a slight improvement to the all-or-nothing assignment. More positive is that cyclists were assigned to routes with more cycle facilities and less intersections, but the number are not correct in all instances.

All in all, the logit assignment seems to be a method that is an improvement of the all-or-nothing assignment for cyclists. The used parameters resulted in a small improvement, but there is potential for better scoring assignments using different and more factors. The algorithm performed even better during validation and seems therefore to work in different areas within the Netherlands, which has not been tried before. To increase the accuracy of the assignment, more research can help collecting more data about cyclists' routes in the Netherlands and optimising the algorithm by changing parameter values and adding additional factors.

9. Recommendations

As discussed in the conclusion the logit assignment has potential and could be improved by doing more research. This chapter discusses the suggestions for further research in this subject.

First of all, more research is needed into the factors that influence the route choice of cyclists in the Netherlands. Currently, very limited research has been conducted and not all research is focussed on one area. New research would preferably be based on GPS traces or similar techniques to obtain revealed-preference data instead of stated-preference which is more common. Results of these studies could be used to review the choice for factors made in this thesis and possibly make it more applicable to the Dutch situation.

Furthermore, more data should be collected about road and intersection characteristics, especially if research into the factors yields factors about which no information is present. The intersection type data should be improved by adding more intersection types as there are some areas with many unknown intersection types. An example of characteristics that need more data collection is about the environment. Some research suggests that the attractiveness of the environment is important in choosing a route. If new research confirms this and it would be added to the cost function, better data is needed than currently is available.

Another line of research could be in the way intersections are implemented. Some intersections with traffic lights are seen as an obstacle, while others are nearly always green and are not perceived as ‘bad’ as the former. It could be interesting to try the model with the perceived delay or perceived safety of the intersection instead of a standard cost for a type of intersection. This allows for differentiation within the categories that might be arbitrary. An alternative is expanding the number of categories or creating subcategories. A disadvantage of this alternative is the increasing number of variables in the calibration.

The calibration results showed that the time parameter did not have a influence on the results when using different values for both the time parameter and all other parameters. This could be explained by the fact that if people would have a preference for fast routes, it is probably also included in their preference for certain road types on which a higher velocity can be reached. If this assignment would be applied in the future, this factor could be omitted without problems.

Moreover, more parameter combinations could be tried than was the case in this thesis. Time limitations did not make it possible to try many combinations for this number of variables. A better combination of parameters could be found than in this research. Trying to apply the model on other networks could be tried as well in order to find out whether the assignment algorithm works in other areas in the Netherlands.

A more specific suggestion is further research into the route generation algorithm. The applied algorithm used Monte Carlo simulation to generate alternative routes. This takes much time and generates routes that are not relevant as well. It could be useful to review other possible algorithms that did not seem feasible in this research. Ideally, an algorithm provides a number of shortest routes in order to reduce the number of unnecessary routes and consequently, the computational time and storage.

If a new route generation algorithm will be found, but with the current algorithm as well, better data is needed about relevant routes. Ideally, information about routes that people consider and choose should be collected and compared with the generated routes. Further research could provide data about this data set.

Finally, it would be beneficial to improve the quality of the counts. Many counts are conducted on different days under different circumstances, like weather and sometimes even year. More counts conducted on the same day would provide more accurate and comparable information about the intensities. Preferably, these counts are conducted visually to improve the accuracy. More research into the route and destination of cyclists would be an improvement as well. With more information about the route and destination, the OD-matrix can be improved which results in a better assignment.

References

Bekhor, S., Ben-Akiva, M. E., & Ramming, M. S. (2006). Evaluation of choice set generation algortihms for route choice models. *Annals of Operations Research, 144*(1), 235-247. doi: 10.1007/s10479-006-0009-8

Bernardi, S., La Paix Puello, L. C., & Geurs, K. T. (2017). Evaluation of Dutch Cycling Patterns: Evidence from Smartphone Data. Retrieved from ResearchGate website: https://www.researchgate.net/publication/315892570_EVALUATION_OF_DUTCH_CYCLING_PATTERNS_EVIDENCE_FROM_SMARTPHONE_DATA

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. doi:10.1016/j.tra.2012.07.005

Centraal Bureau voor de Statistiek. (2017a, 25 October). Personenmobiliteit in Nederland; persoonskenmerken en vervoerswijzen, regio. Retrieved from https://opendata.cbs.nl/#/CBS/nl/dataset/83499NED/table?dl=B0EF

Centraal Bureau voor de Statistiek. (2017b, 25 October). Personenmobiliteit in Nederland; vervoerswijzen en reismotieven, regio's. Retrieved from http://statline.cbs.nl/Statweb/publication/?DM=SLNL&PA=83500ned&D1=a&D2=0&D3=0-2,6&D4=a&D5=0&D6=l&HDR=T&STB=G1,G4,G3,G2,G5&VW=T

Christou, G. (2017). Chelmsford Traffic and Access Strategy: Cycling Model - Technical Note. Retrieved from London: Jacobs U.K. Limited

City of Copenhagen. (2017). *Copenhagen City of Cyclists: Facts & Figures*. Retrieved from http://www.cycling-embassy.dk/wp-content/uploads/2017/07/Velo-city_handout.pdf.

Claasen, Y., & Rienstra, S. (2017). *Routekeuzegedrag van fietsers: meer dan alleen de snelste route telt*. Paper presented at the Colloquium Vervoersplanologisch Speurwerk, Gent.

Dijkstra, E. W. (1959). A Note on Two Problems in Connexion with Graphs. *Numerische Mathematik, 1*, 269-271.

Ehrgott, M., Raith, J. Y. T., & van Houtte, C. (2012). A bi-objective cyclist route choice model. *Transportation Research Part A, 46*, 652-663. doi:10.1016/j.tra.2011.11.015

Federal Highway Agency. (n.d.). Calibration of Microsimulation Models. Retrieved from https://ops.fhwa.dot.gov/trafficanalysistools/tat_vol3/sect5.htm

Fietsersbond. (2018). Plan je fietsroute of fietsrondje. Retrieved from https://routeplanner.fietsersbond.nl

Fiorenzo-Catalano, M. S. (2007). *Choice Set Generation in Multi-Modal Transportation Networks.* (PhD thesis), Delft University of Technology, Delft. Retrieved from https://repository.tudelft.nl/islandora/object/uuid:ef3b9c22-b979-4f46-9b02-110c82d67535?collection=research

Google. (2018). Google Maps. Retrieved from http://maps.google.nl

Halldórsdóttir, K., Rieser-Schüssler, N., Axhausen, K. W., Nielsen, O. A., & Prato, C. G. (2014). Efficiency of choice set generation methods for bicycle routes. *European Journal of Transport and Infrastructure Research, 14*(4), 332-348.

Han, S. (2000). *Dynamic Traffic Assignment Technoques for General Road Networks.* (PhD thesis), University College London, London, UK. Retrieved from http://discovery.ucl.ac.uk/1349430/1/343319.pdf

Hogenkamp, H. (2014). Verklaren van fietssnelheden door het gebruik van omgevingskenmerken. (MSc thesis), Wageningen UR, Wageningen.

Hood, J., Sall, E., & Charlton, B. (2011). A GPS-based bicycle route choice model for San Francisco, California. *Transportation Letters, 3*(1), 63-75. doi:10.3328/tl.2011.03.01.63-75

Hunt, J. D., & Abraham, J. E. (2007). Influences on bicycle use. *Transportation, 34*, 453-470.

Kennisinstituut voor Mobiliteitsbeleid. (2017a). De e-fiets reikt verder maar gaat nauwelijks sneller. Retrieved from https://www.kimnet.nl/mobiliteitsbeeld#personenvervoer-article7

Kennisinstituut voor Mobiliteitsbeleid. (2017b). Het fietsgebruik in binnenstedelijk woon-werkverkeer groeit, de aandelen openbaar vervoer en auto nemen af. Retrieved from https://www.kimnet.nl/mobiliteitsbeeld#personenvervoer-article12

Kennisinstituut voor Mobiliteitsbeleid. (2017c). Mobiliteitsbeeld 2017. Retrieved from https://www.kimnet.nl/mobiliteitsbeeld#personenvervoer-article12

Menghini, G., Carrasco, N., Schüssler, N., & Axhausen, K. W. (2010). Route choice of cyclists in Zurich. *Transportation Research Part A: Policy and Practice, 44*(9), 754-765. doi:10.1016/j.tra.2010.07.008

Nijmegen wil af van fietsfiles. (2016, 20 May). *De Gelderlander*. Retrieved from https://www.gelderlander.nl/default/nijmegen-wil-af-van-fietsfiles~adaed5c1/

Pixabay. (2015). Cyclists. Retrieved from https://cdn.pixabay.com/photo/2015/03/26/09/57/cyclists-690644_1280.jpg

Rieser-Schüssler, N., Balmer, M., & Axhausen, K. W. (2013). Route choice sets for very high-resolution data. *Transportmetrica A: Transport Science, 9*(9), 825-845. doi:10.1080/18128602.2012.671383

Rijksoverheid. (2017). Meer mensen op de fiets. Retrieved from https://www.rijksoverheid.nl/onderwerpen/fiets/fietsbeleid

Ryu, S., Chen, A., Su, J., & Choi, K. (2018). Two-Stage Bicycle Traffic Assignment Model. *Journal of Transportation Engineering, Part A: Systems, 144*(2). doi:10.1061/jtepbs.0000108

Sener, I. N., Eluru, N., & Bhat, C. R. (2009). An analysis of bicycle route choice preferences in Texas, US. *Transportation, 36*(5), 511-539.

Transport for London. (2012). *Cycle route choice*. London: Transport for London Retrieved from http://content.tfl.gov.uk/understanding-cycle-route-choice.pdf.

Vedel, S. E., Jacobsen, J. B., & Skov-Petersen, H. (2017). Bicyclists' preferences for route characteristics and crowding in Copenhagen - A choice experiment study of commuters. *Transportation Research Part A: Policy and Practice, 100*, 53-63.

Yen, J. Y. (1971). Finding the K Shortest Loopless Paths in a Network. *Management Science, 17*(11), 712-716. doi:10.1287/mnsc.17.11.712

Appendix A: Route generation details

This appendix provides additional information about the calibration of the route generation algorithm on the ’s-Hertogenbosch network described in section 4.3. The planned routes using the *Fietsersbond* route planner (Fietsersbond, 2018) and the generated routes are presented together with a map of the area with the location of the centroids and connectors to the centroids. It should be noted that one centroid can have multiple connectors to the road network. In order to plan a route with the route planner it is needed to plan between two real locations. Therefore in some cases multiple routes needed to be planned between different locations. Furthermore, because of the Monte Carlo simulation that varies the costs of the roads, it is possible that the generated routes in one direction are different from the opposite direction. Settings 3 and 4 resulted mainly in more combinations of routes that are generated with setting 2 as well. Sometimes an extra detour is added which is not realistic. Because of the small variation, these are not shown in this appendix.

Pair 1: Boschveld - City Centre of ’s-Hertogenbosch

The first considered OD-pair is between the area of Boschveld and the city centre of ’s-Hertogenbosch. Figure A.1 shows the locations where the feeders of the centroid connect to the road network. This pair has only one reasonable alternative through the tunnel under the railway station and should therefore result in just one route with minor variations.

Figure A.1. Centroids (stars) and connections to road network (circles) of origin Boschveld and destination city centre. (Google, 2018)

The Fietsersbond route planner gives the advices in figures A.2 and A.3 when set to easy cycling, shortest distance or as car-free as possible. Other connection points of the centroids to the road network were tested as well and resulted in the same routes with an extension to those connection points. The return trips follow the same route except for the central part where the route is at the other side of the main road because of one-way cycle paths. Only setting 4 gives alternative routes, but these are only very small detours in the city centre itself.

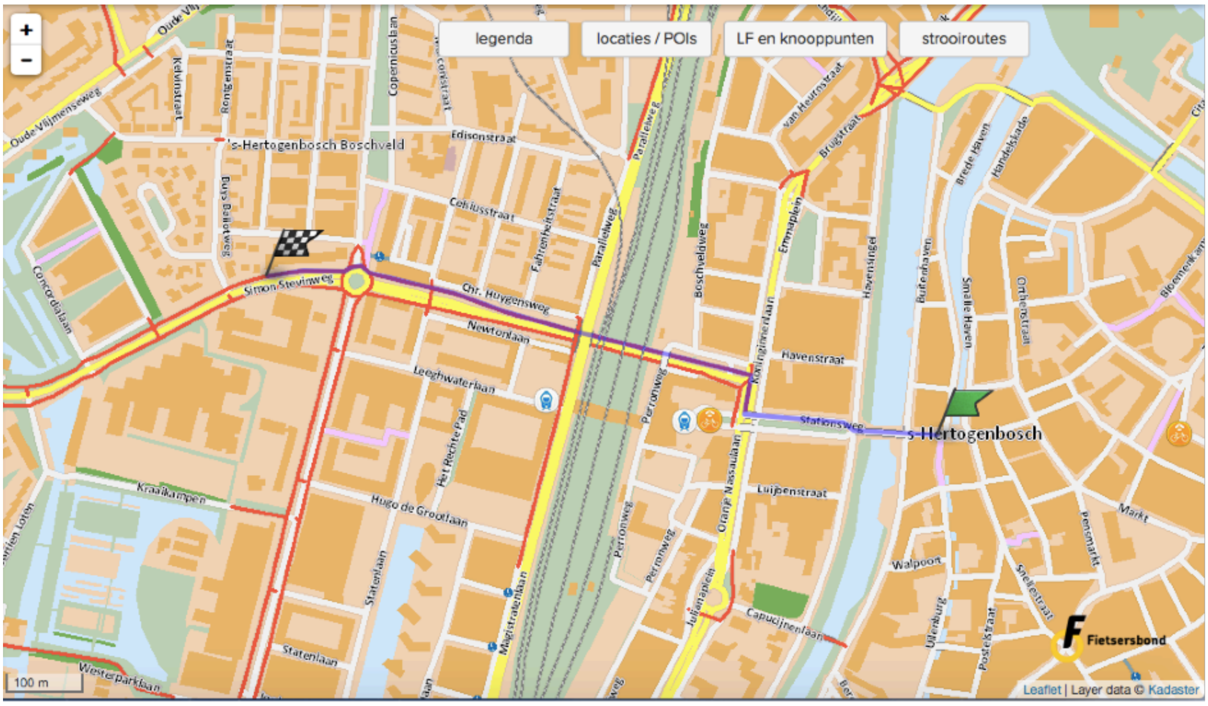


Figure A.2. Planned route (circa 1 km) between the city centre of 's-Hertogenbosch and Boschveld with setting easy cycling, shortest route or 'fietsbewust' (Fietsersbond, 2018).

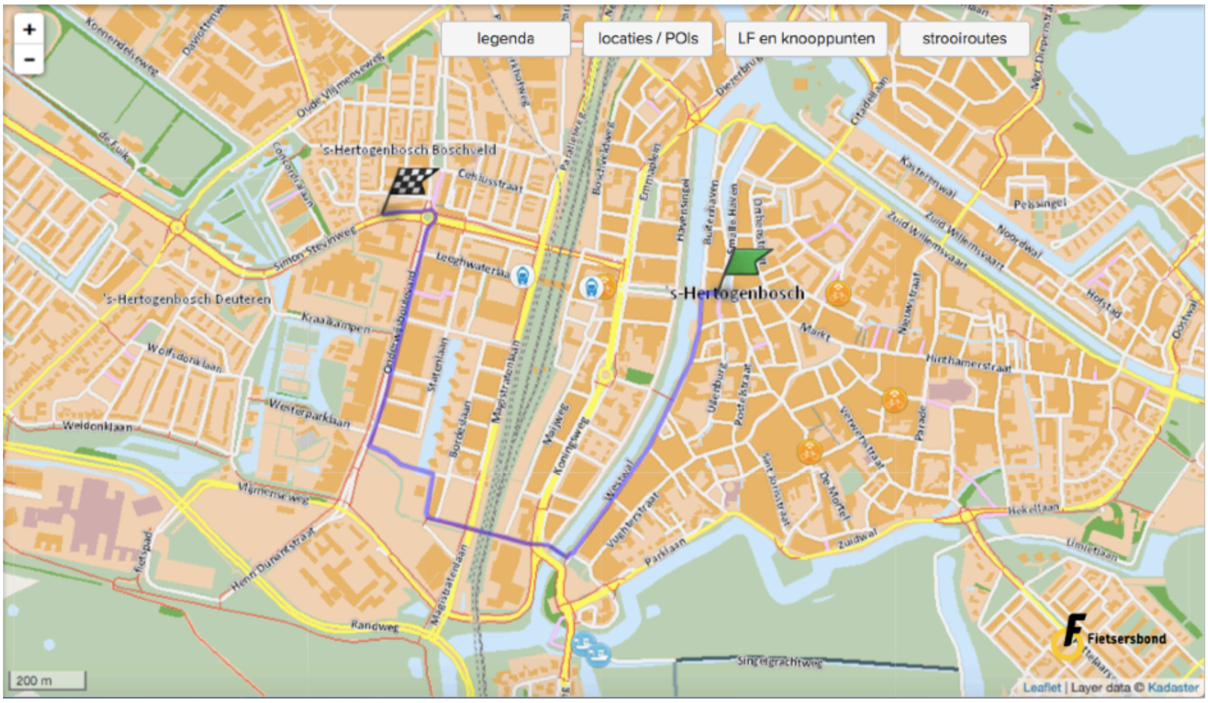


Figure A.3. Planned route (circa 2,3 km) between the city centre of 's-Hertogenbosch and Boschveld with setting low-traffic roads (Fietsersbond, 2018).

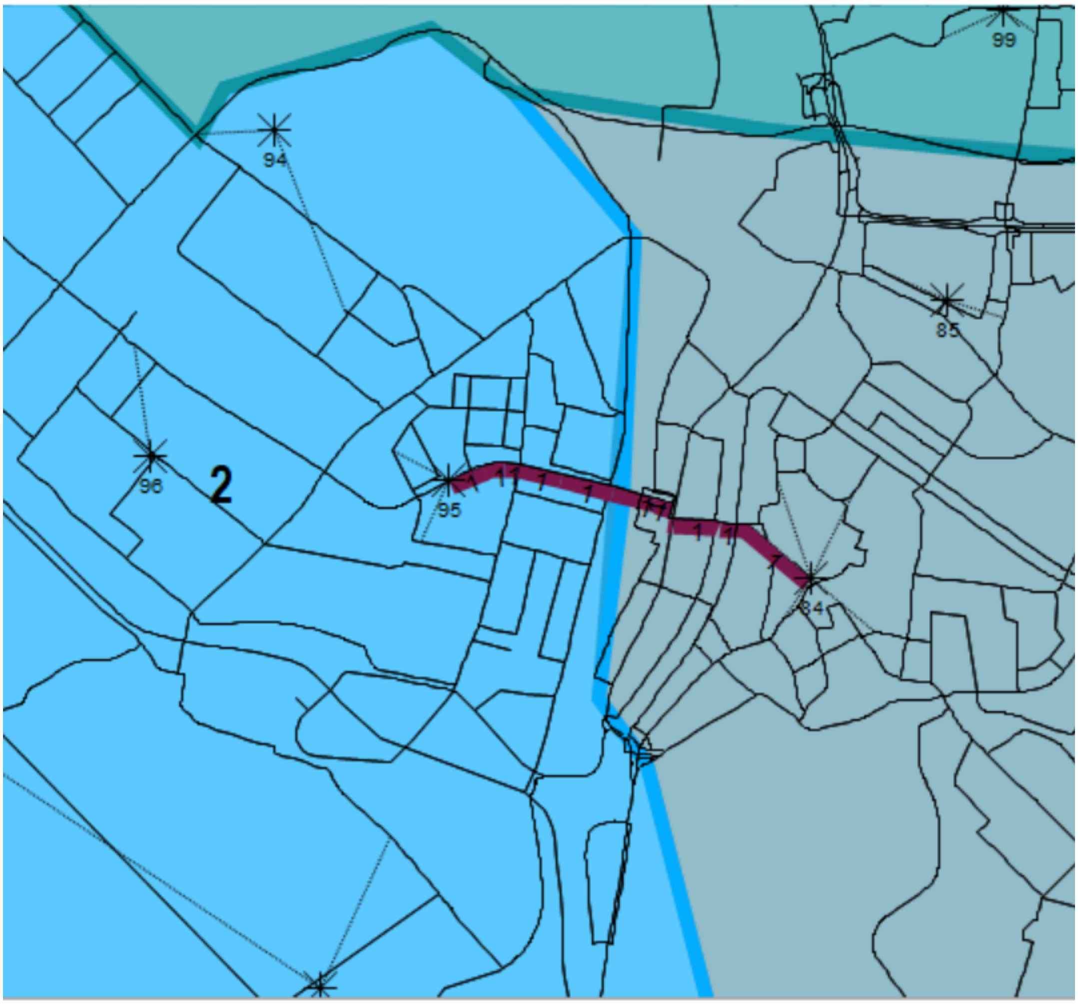


Figure A.4. Generated routes between the city centre and Boschveld with settings 1, 2 and 3.

Pair 2: De Rompert - Hintham

The second pair is between two neighbourhoods of ’s-Hertogenbosch outside the city centre. Figure A.5 shows the map of the centroids and their connection points to the road network. It shows clearly that there are lots of different routes possible between the areas.

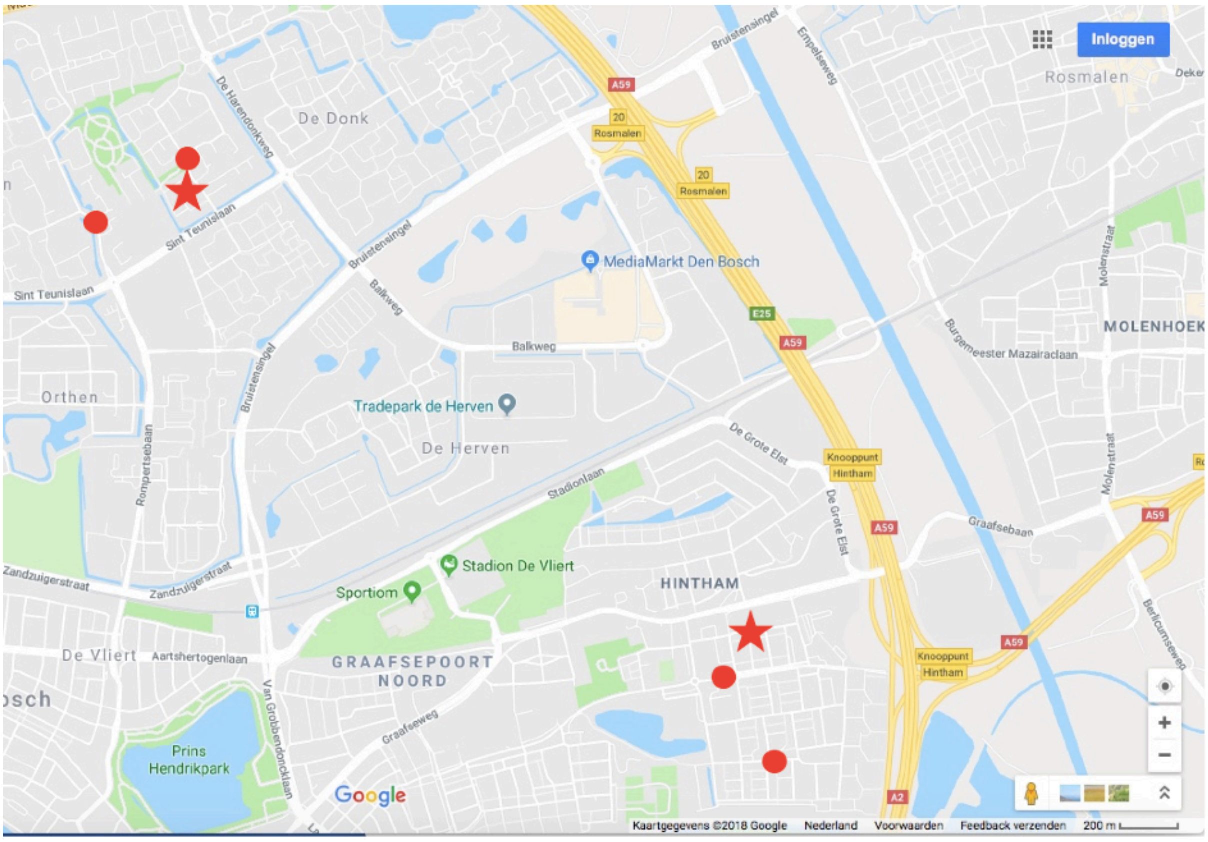


Figure A.5. Centroids (stars) and connections to road network (circles) of origin De Rompert and destination Hintham (Google, 2018).

The route planner of the Fietsersbond gave the following routes:

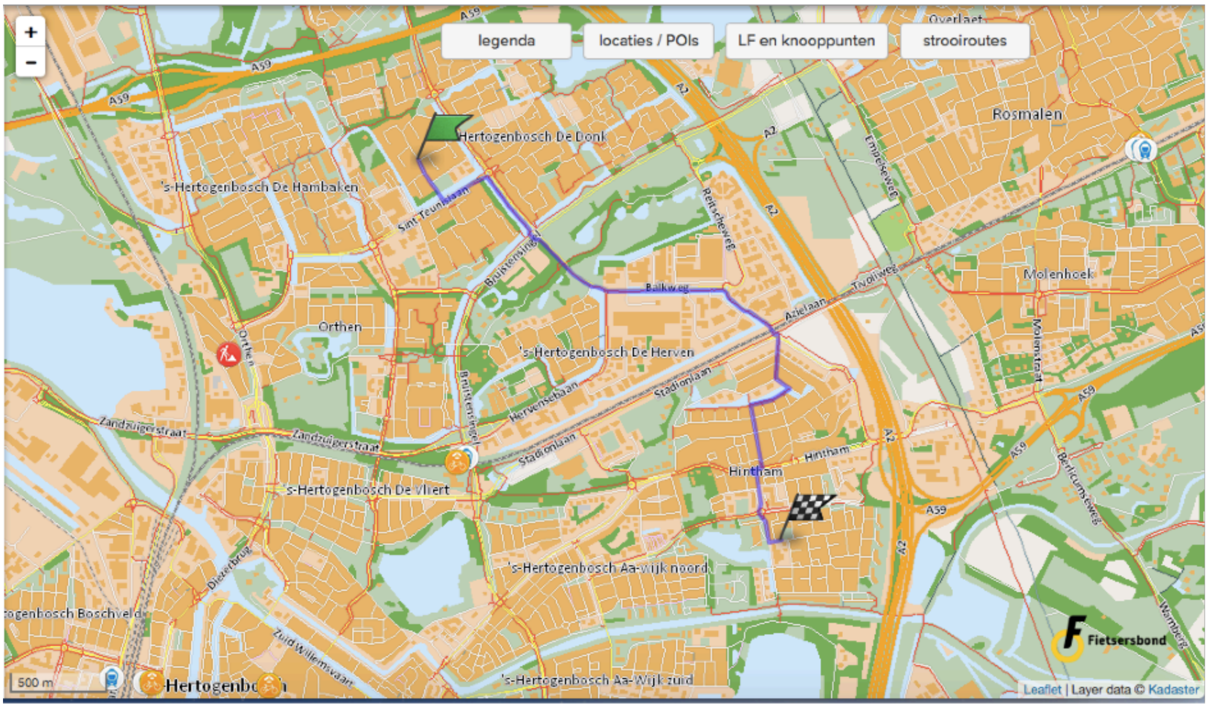


Figure A.6. Route (3,8 km) between De Rompert (eastern connection) and Hintham (northern connection) with setting easy cycling (Fietsersbond, 2018).

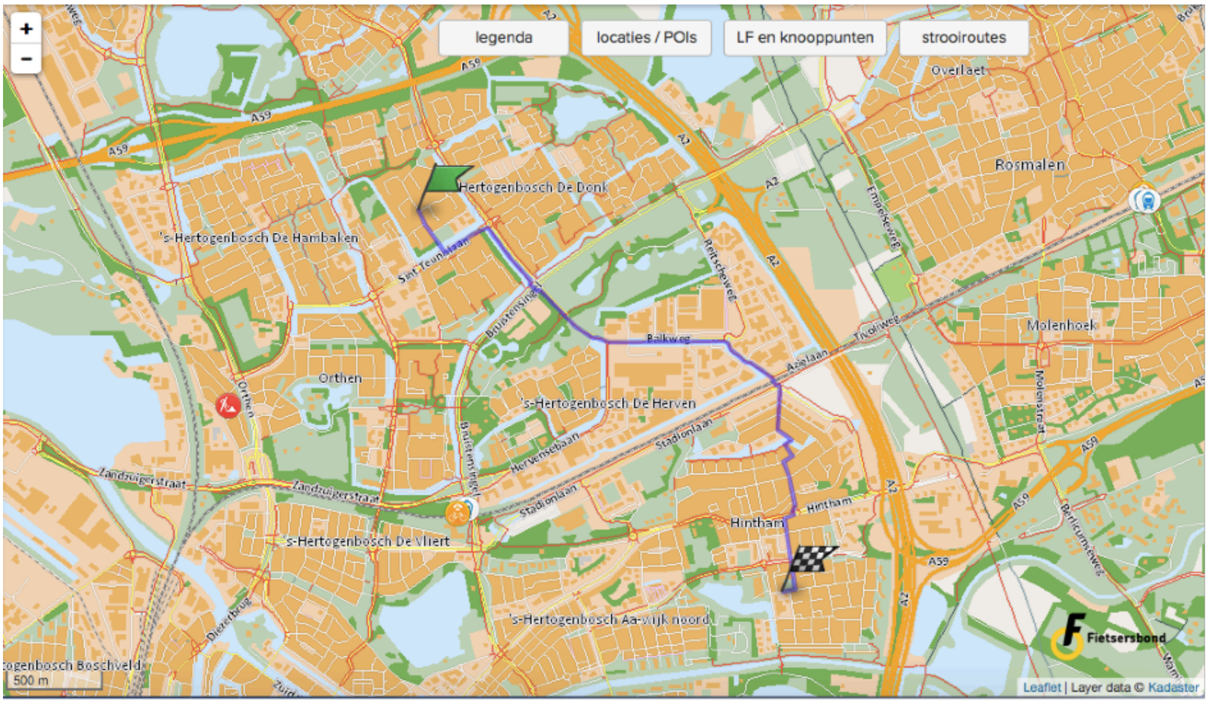


Figure A.7. Route (3,7 km) between De Rompert (eastern connection) and Hintham (northern connection) with setting shortest route (Fietsersbond, 2018).

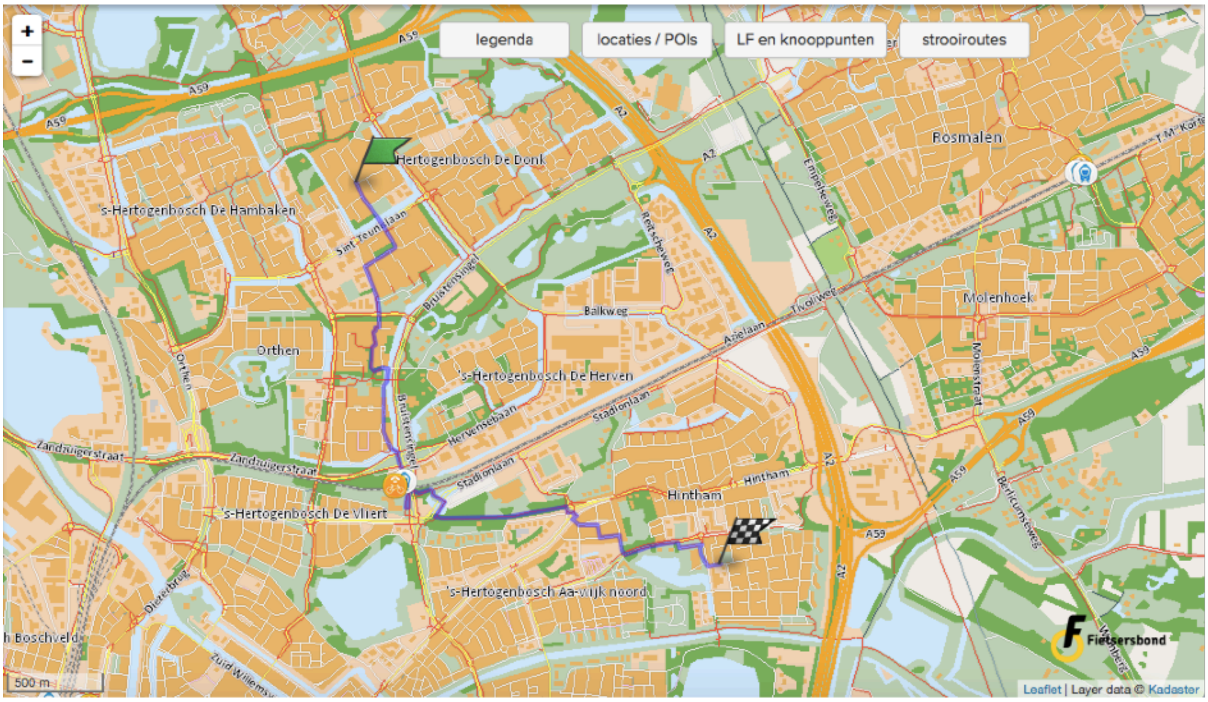


Figure A.8. Route (4,3 km) between De Rompert (eastern connection) and Hintham (northern connection) with setting low-traffic roads (Fietsersbond, 2018).

Figure A.9. Route (4,1 km) between De Rompert (eastern connection) and Hintham (northern connection) with setting 'fietsbewust' (Fietsersbond, 2018).

Routes from the westerly connection of centroid De Rompert:

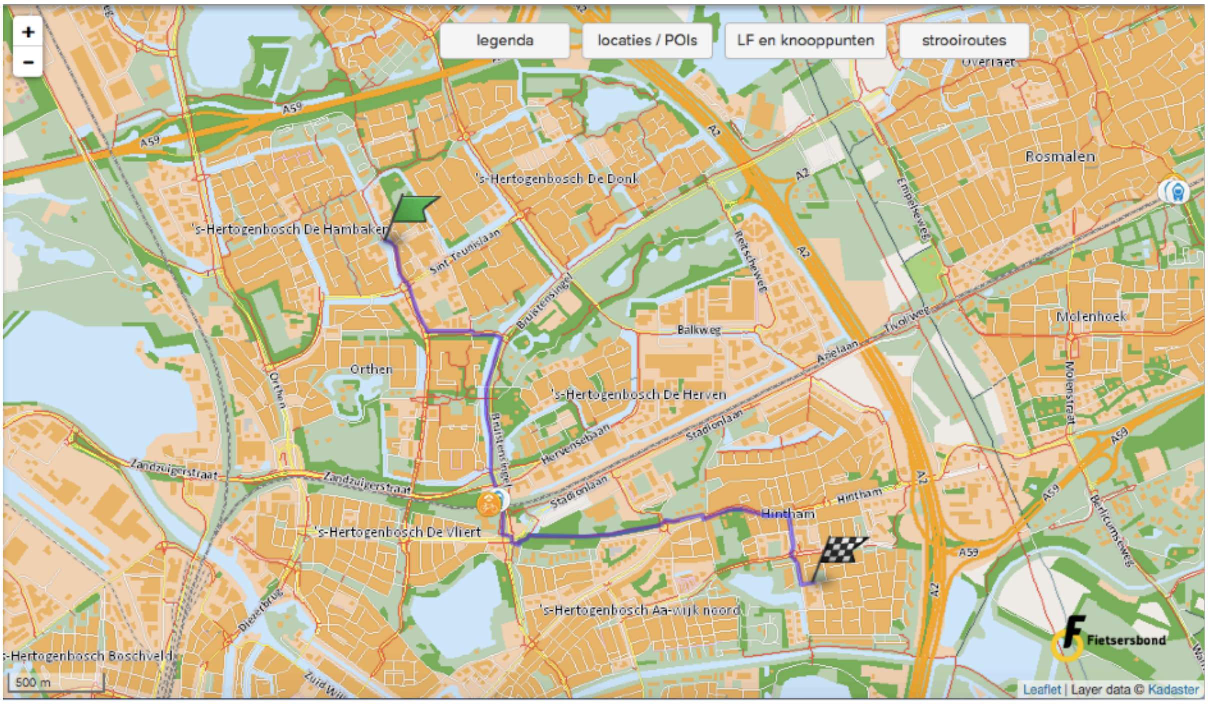


Figure A.10. Route (4,0 km) between De Rompert (western connection) and Hintham (northern connection) with setting easy cycling (Fietsersbond, 2018).

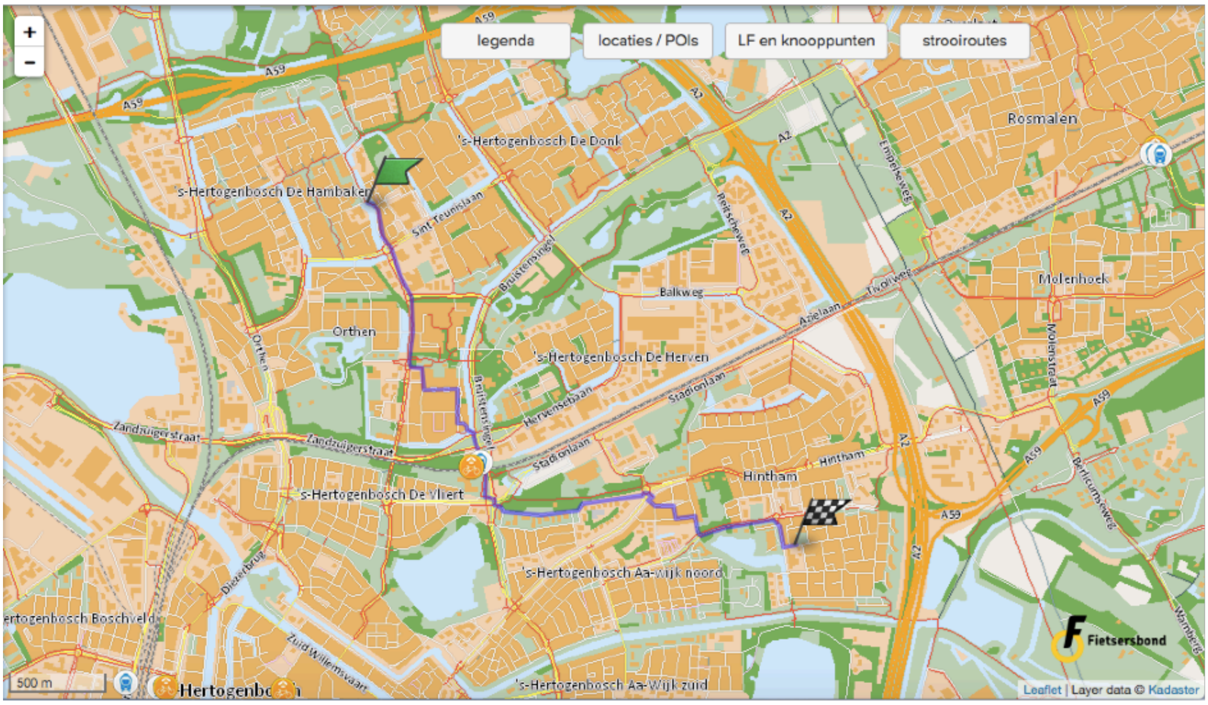


Figure A.11. Route (3,9 km) between De Rompert (western connection) and Hintham (northern connection) with setting shortest route (Fietsersbond, 2018).

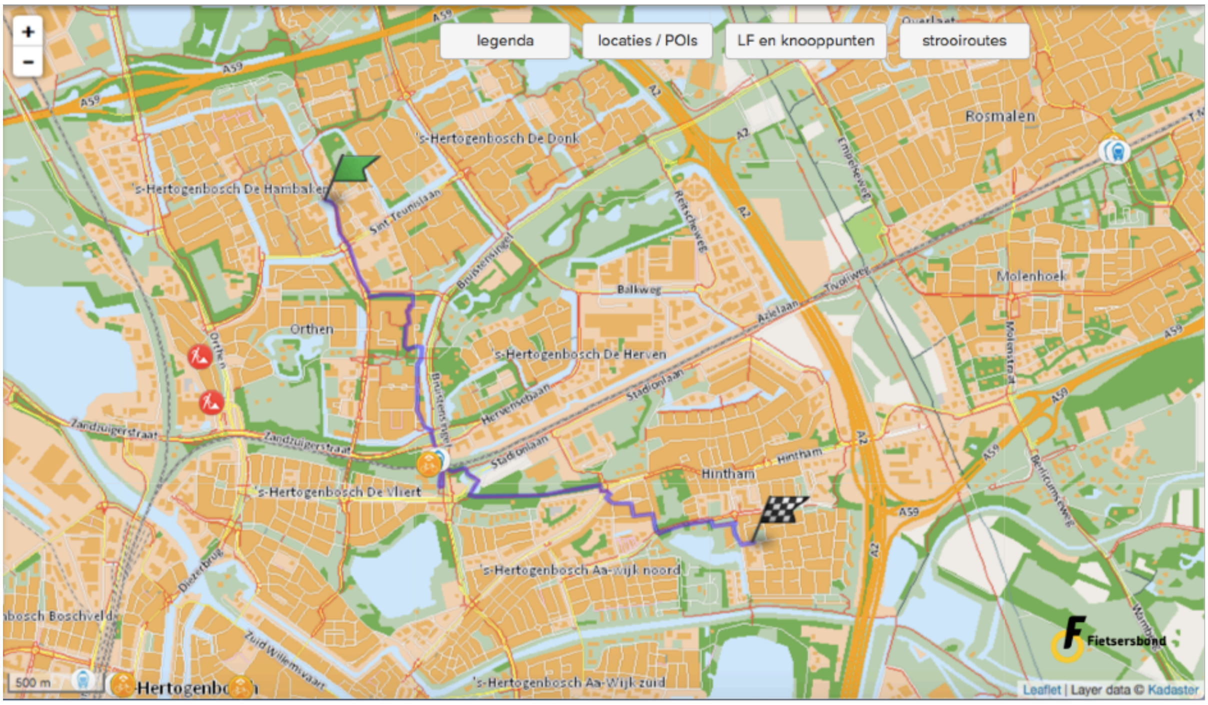


Figure A.12. Route (4,3 km) between De Rompert (western connection) and Hintham (northern connection) with setting low-traffic roads (Fietsersbond, 2018).

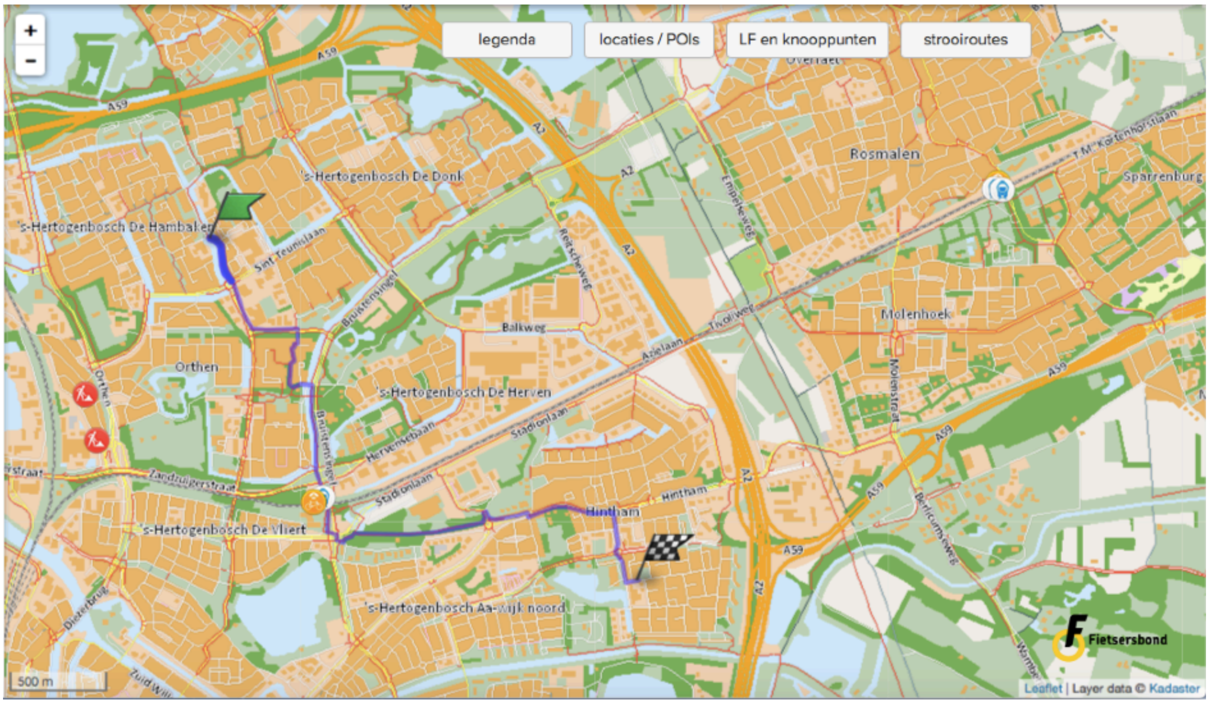


Figure A.13. Route (4,3 km) between De Rompert (western connection) and Hintham (northern connection) with setting 'fietsbewust' (Fietsersbond, 2018).

Most routes to the southern connection of the Hintham centroid were an extension of the routes to the northern connection. The shortest path between the connectors was added to the shown routes. Only the easy cycling route from the eastern De Rompert connection to the southern Hintham connection was different (figure A.14).

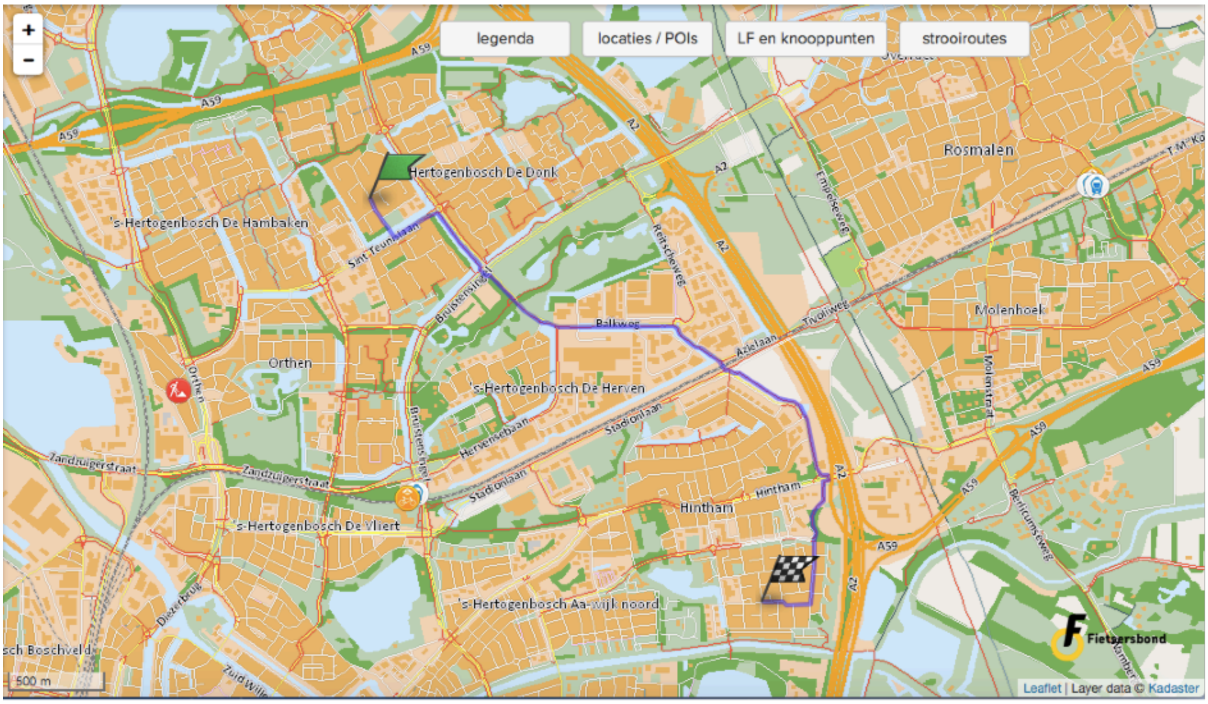


Figure A.14. Route (4,2 km) between De Rompert (eastern connection) and Hintham (southern connection) with setting easy cycling (Fietsersbond, 2018).

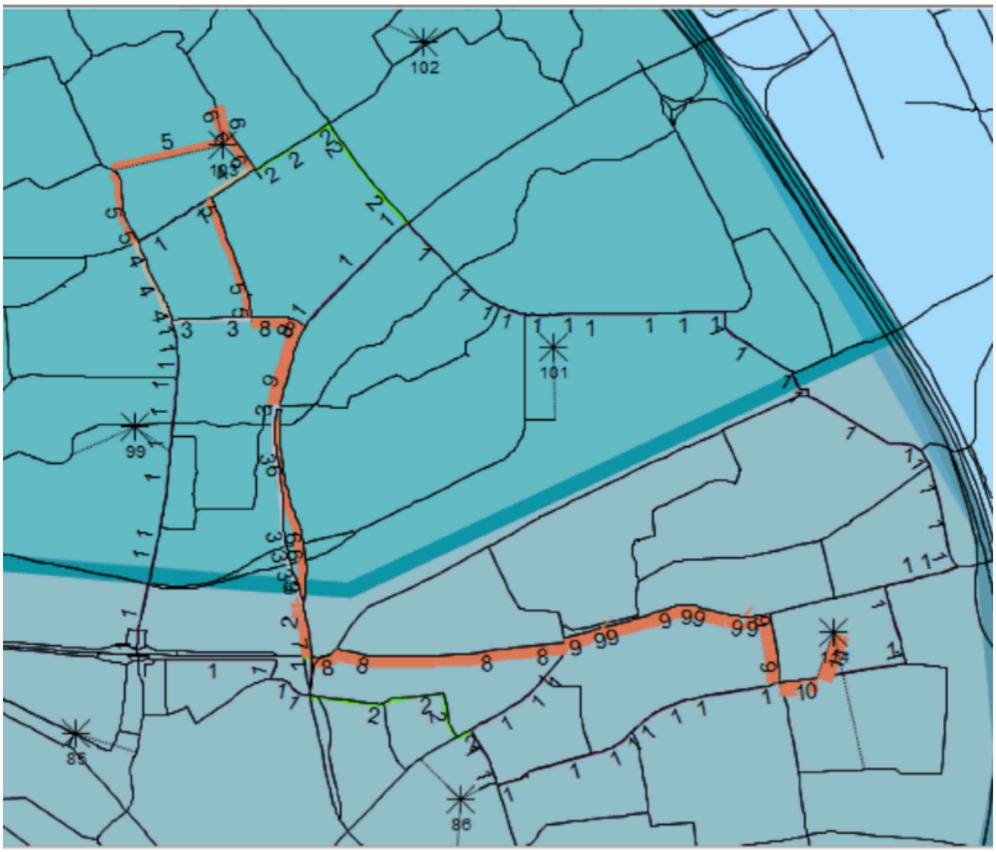


Figure A.15. Generated routes between De Rompert and Hintham with setting 1.

Figure A.16. Generated routes between Hintham and De Rompert with setting 1.

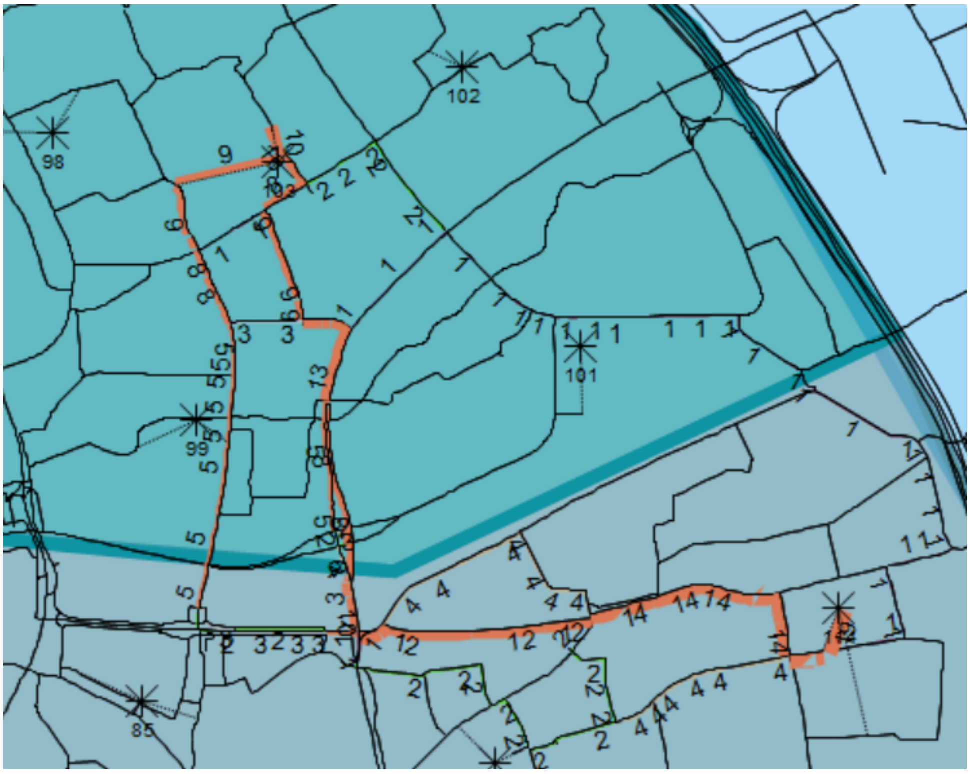


Figure A.17. Generated routes between De Rompert and Hintham with setting 2.

Figure A.18. Generated routes between Hintham and De Rompert with setting 2.

Pair 3: City Centre - De Meerendonk (84-87)

This OD-pair is between the city centre of 's-Hertogenbosch and a nearby neighbourhood. There are lots of options to choose a route from as can be seen in figure A.19.

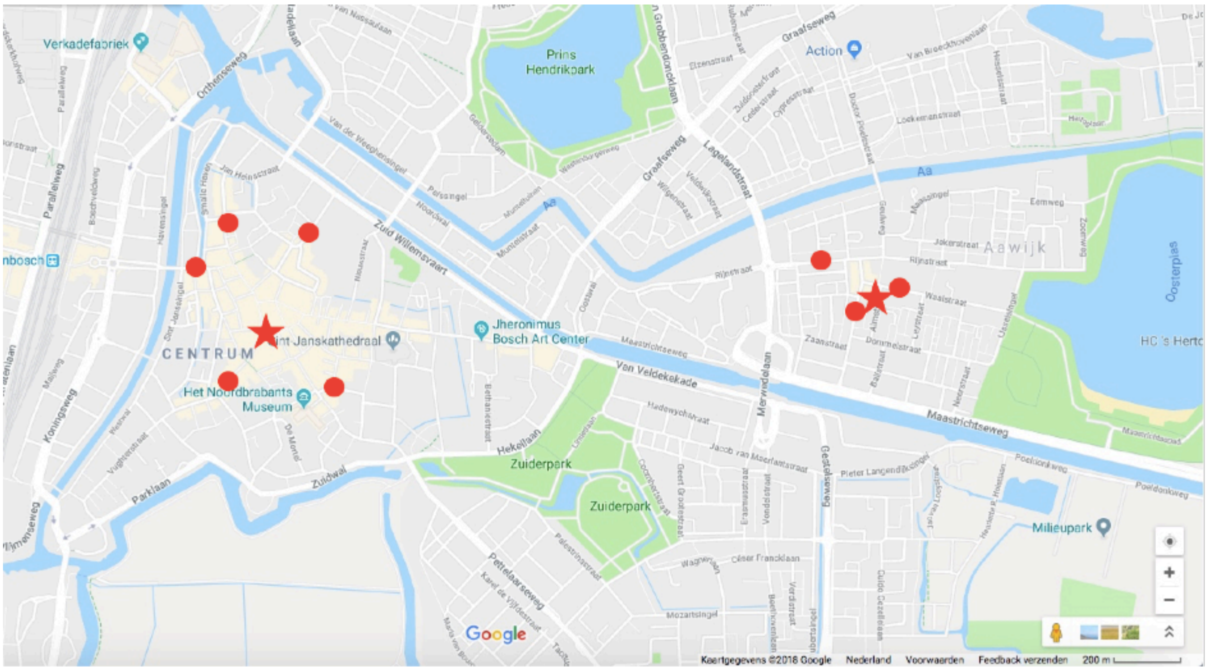


Figure A.19. Centroids (stars) and connections to road network (circles) of origin city centre and destination De Meerendonk (Google, 2018).

The routes from the two eastern connections of the city centre centroid to the two western connections of the De Meerendonk centroid are planned with the Fietsersbond routeplanner. The other connections will not be used as much, because another connection is passed on the way. Figures A.20 to A.30 show the different routes with different settings.

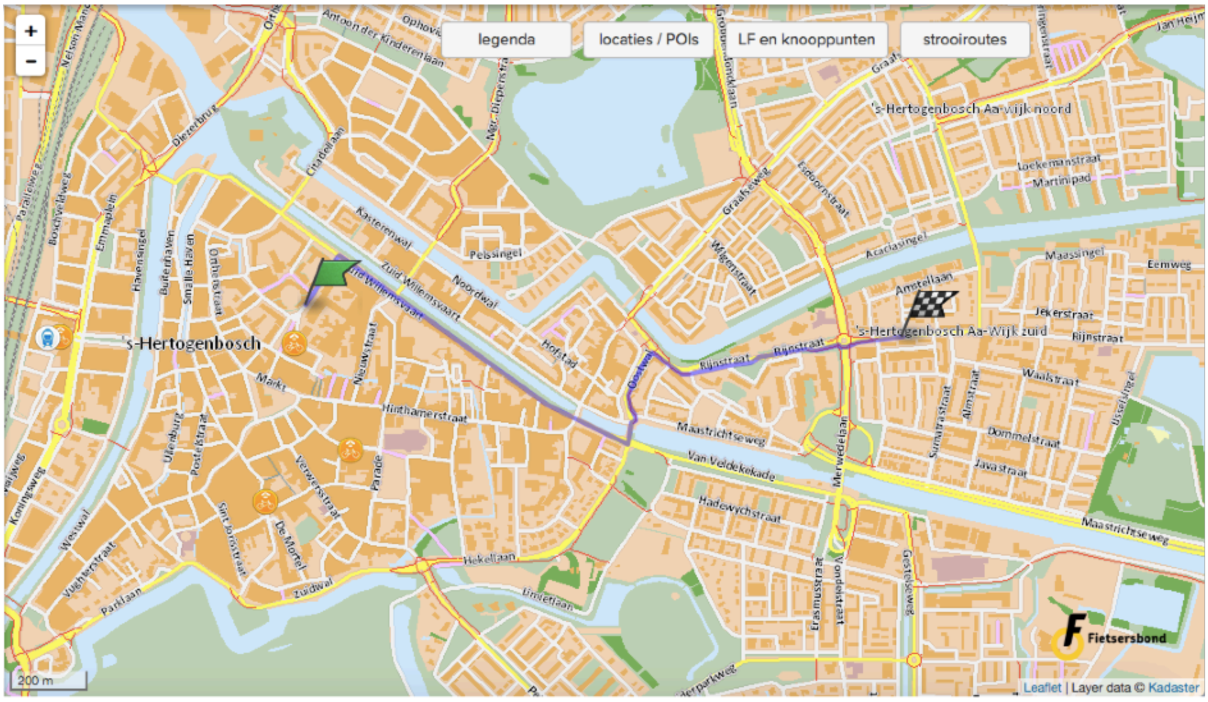


Figure A.20. Route (2,0 km) between the city centre (northeastern connection) and De Meerendonk (northwestern connection) with setting easy cycling and 'fietsbewust' (Fietsersbond, 2018).

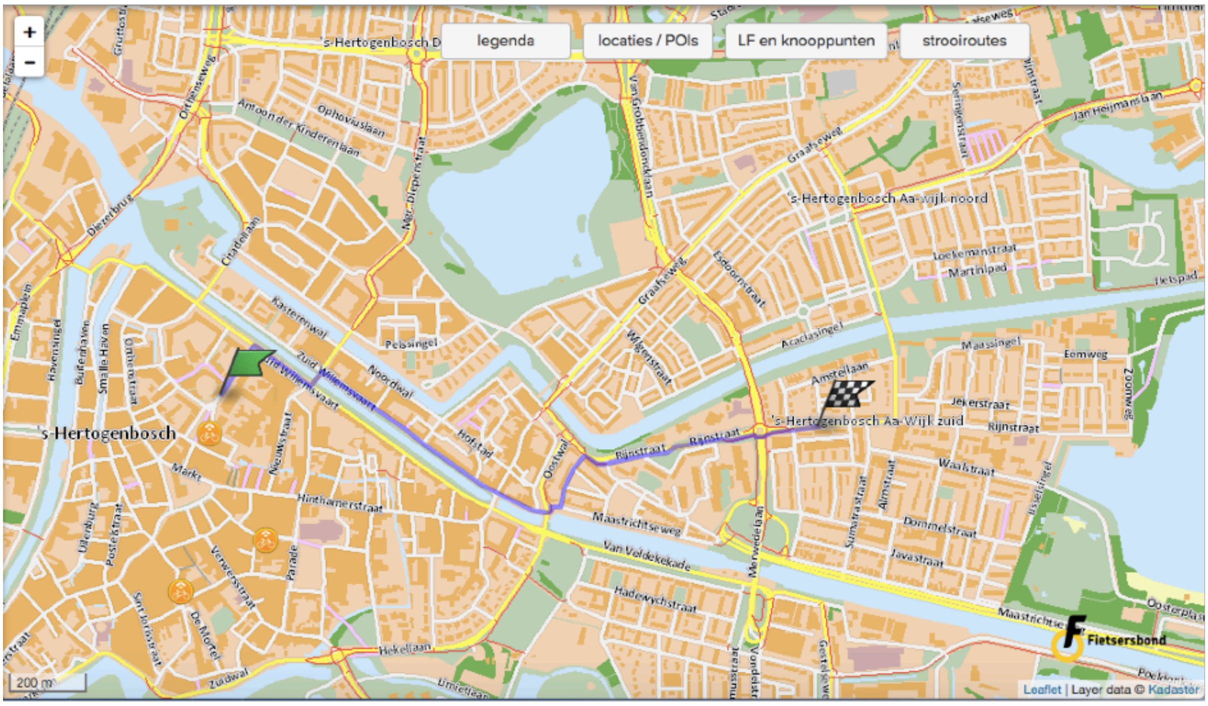


Figure A.21. Route (1,9 km) between the city centre (northeastern connection) and De Meerendonk (northwestern connection) with setting shortest route (Fietsersbond, 2018).

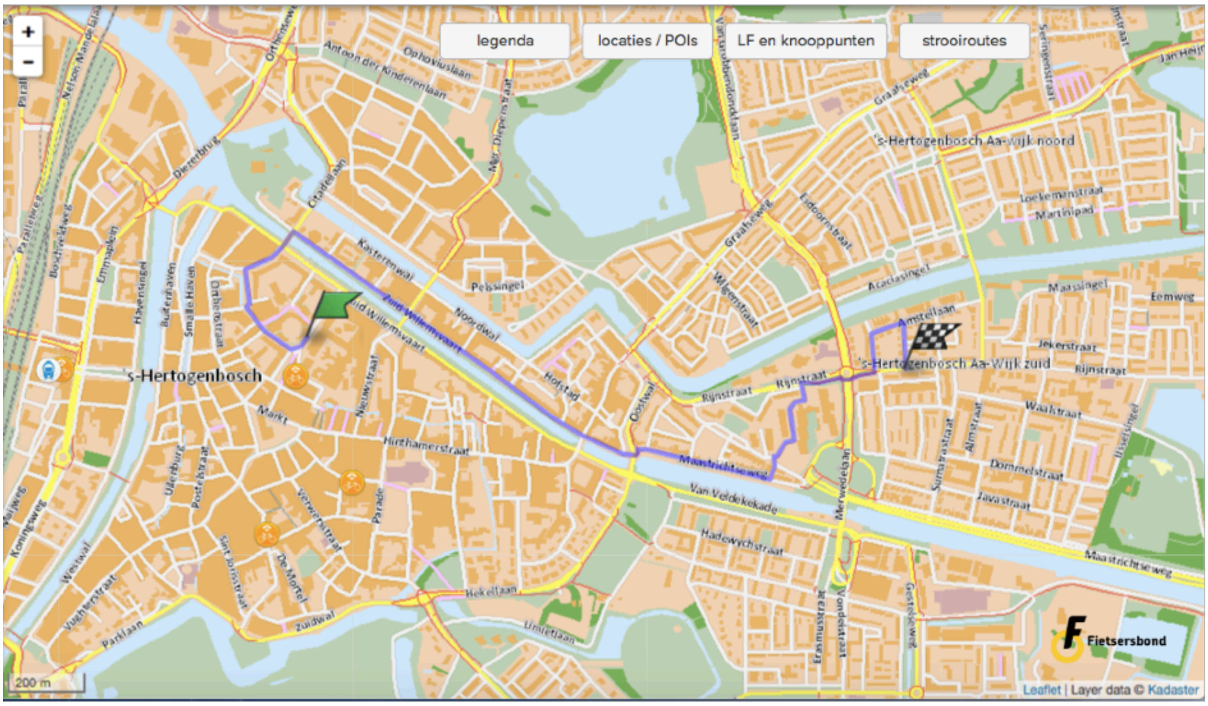


Figure A.22. Route (2,7 km) between the city centre (northeastern connection) and De Meerendonk (northwestern connection) with setting low-traffic roads (Fietsersbond, 2018).

The routes from the northeastern connection of the city centre centroid to the southwestern connection of the De Meerendonk centroid are comparable to the above shown routes. Only a small 300-metre extension is added. The shortest and low-traffic routes are different however as shown in figure A.23.

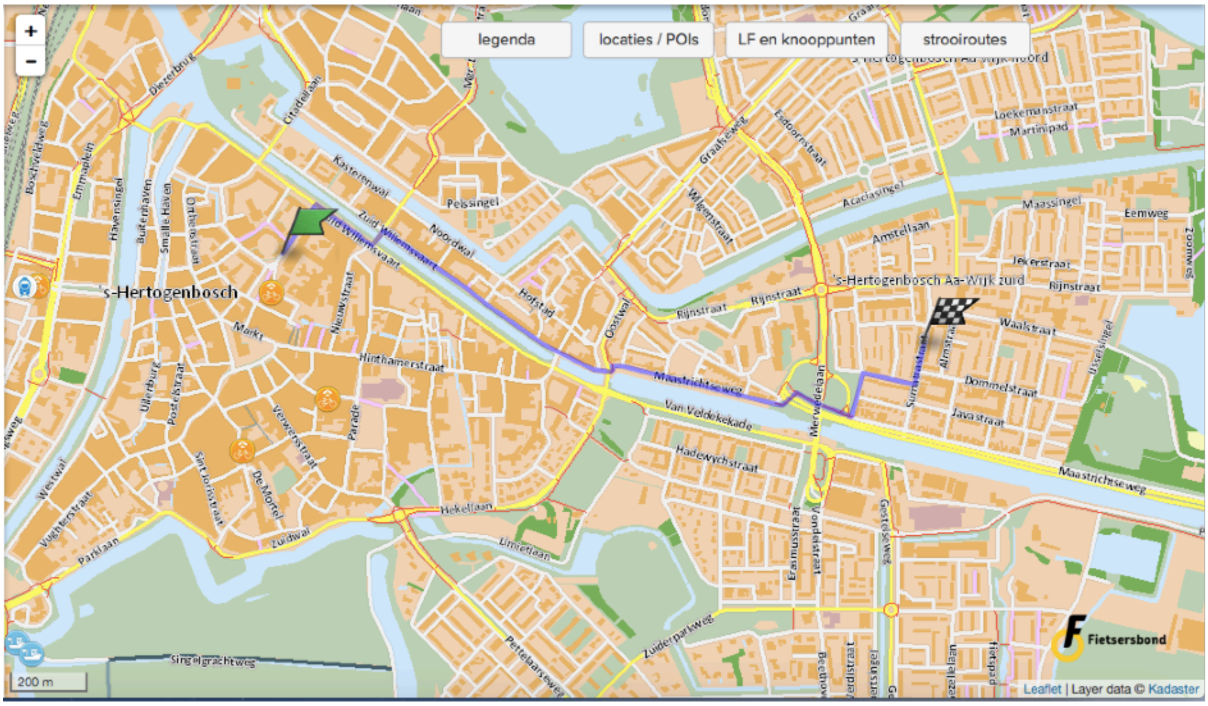


Figure A.23. Route (2,1 km) between the city centre (northeastern connection) and De Meerendonk (southwestern connection) with setting shortest route and low-traffic routes (Fietsersbond, 2018).

The routes from the southeastern connection point in the city centre to the northwestern point in De Meerendonk are shown below.

Figure A.24. Route (2,0 km) between the city centre (southeastern connection) and De Meerendonk (northwestern connection) with setting easy cycling (Fietsersbond, 2018).

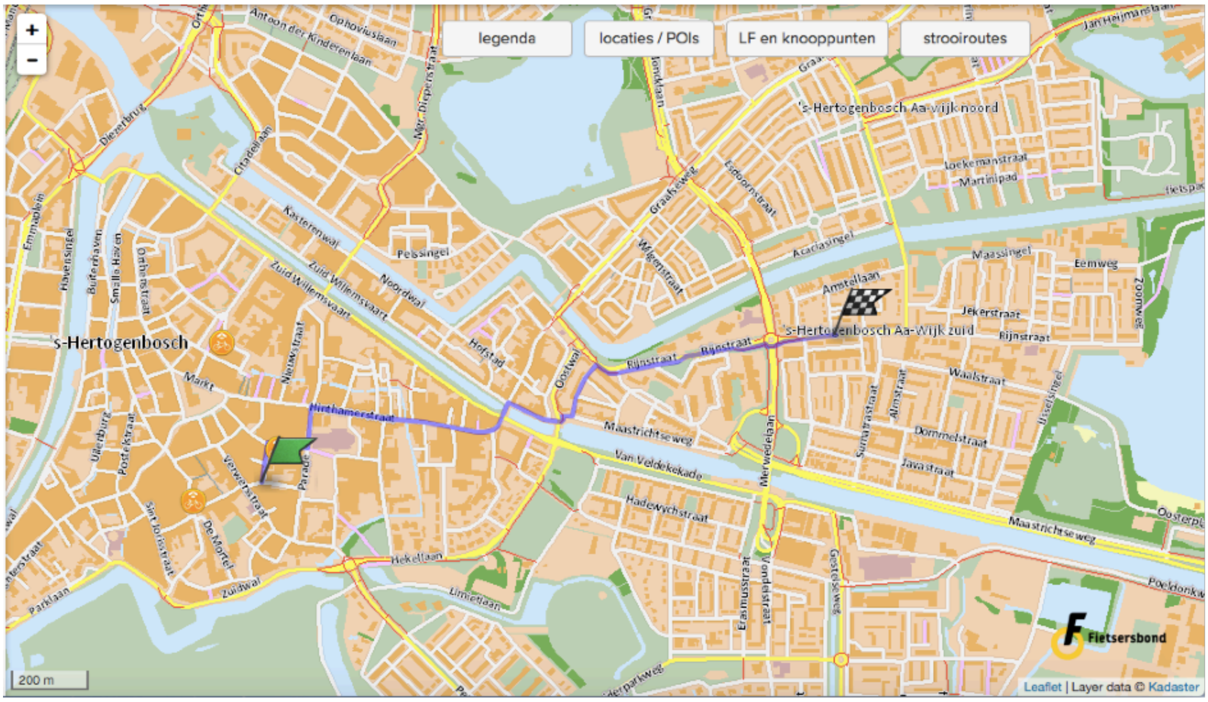


Figure A.25. Route (1,9 km) between the city centre (southeastern connection) and De Meerendonk (northwestern connection) with setting shortest route (Fietsersbond, 2018).

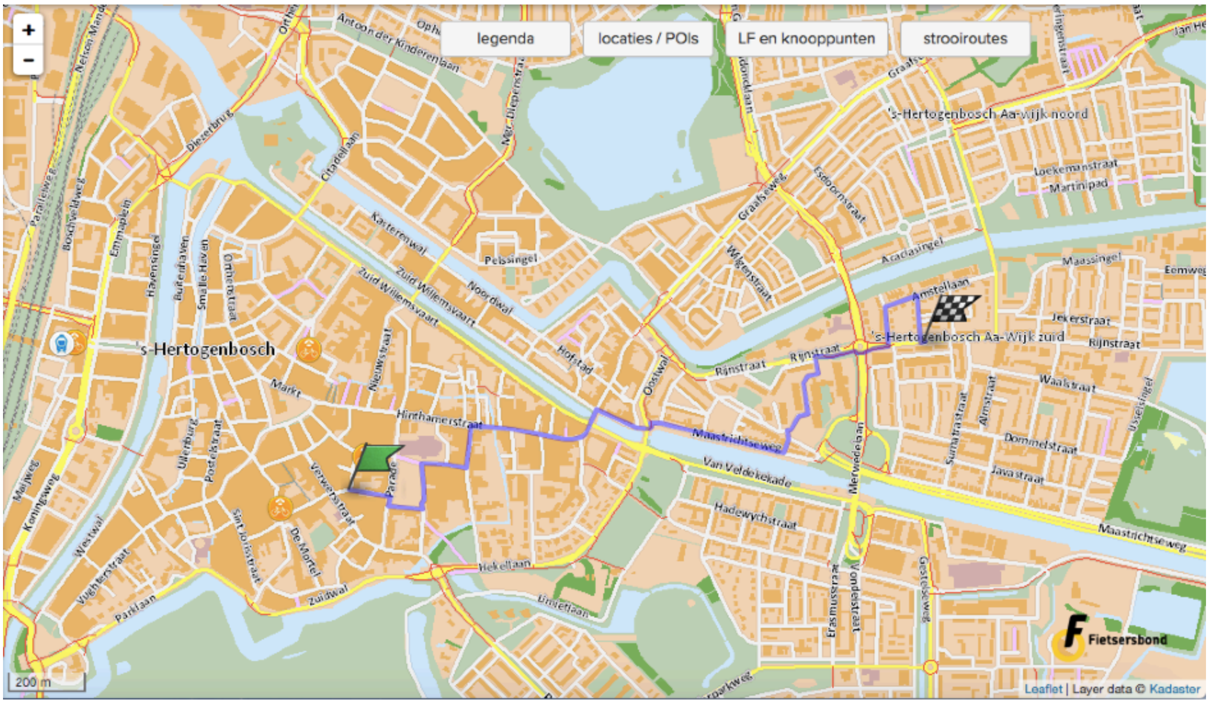


Figure A.26. Route (2,3 km) between the city centre (southeastern connection) and De Meerendonk (northwestern connection) with setting low-traffic routes (Fietsersbond, 2018).

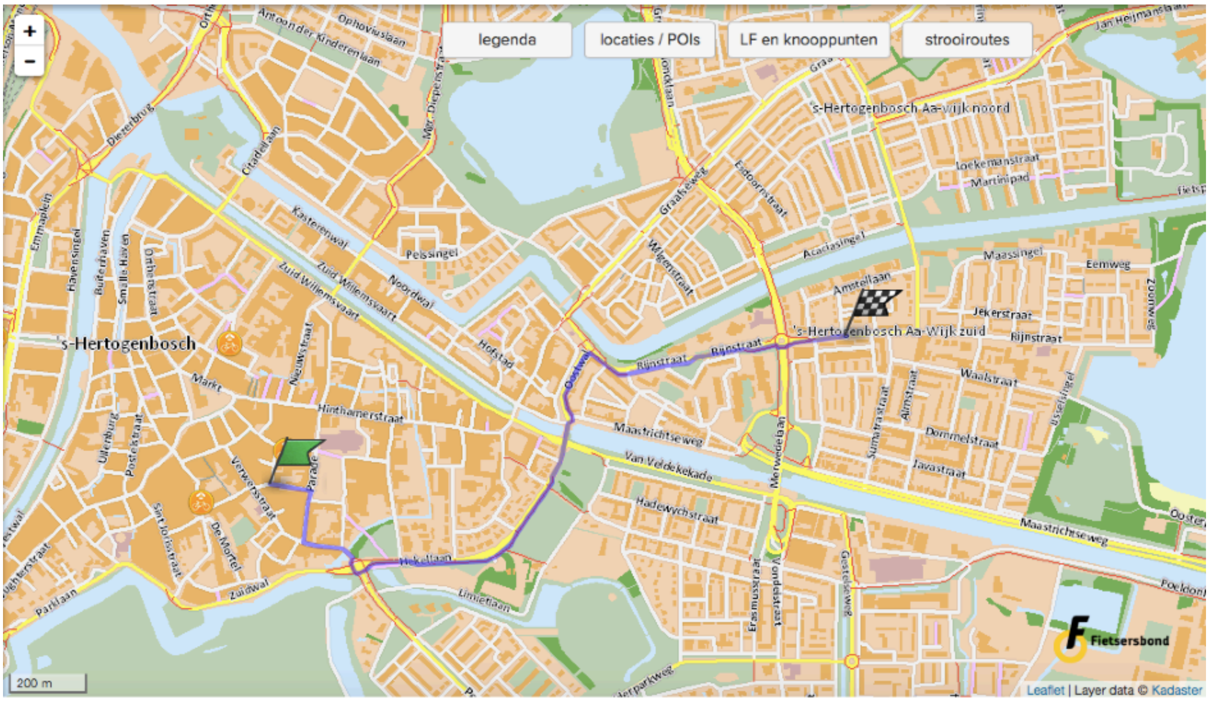


Figure A.27. Route (2,1 km) between the city centre (southeastern connection) and De Meerendonk (northwestern connection) with setting 'fietsbewust' (Fietsersbond, 2018).

Finally, the routes between the southeastern connection point in the centre and the southwestern connection point in De Meerendonk are planned. The easy cycling route is a 300-metre extension from the other connection point to the southwestern connection point.

The shortest route is shown in figure A.28, the low-traffic route in figure A.29 and the *fietsbewust* route in figure A.30.

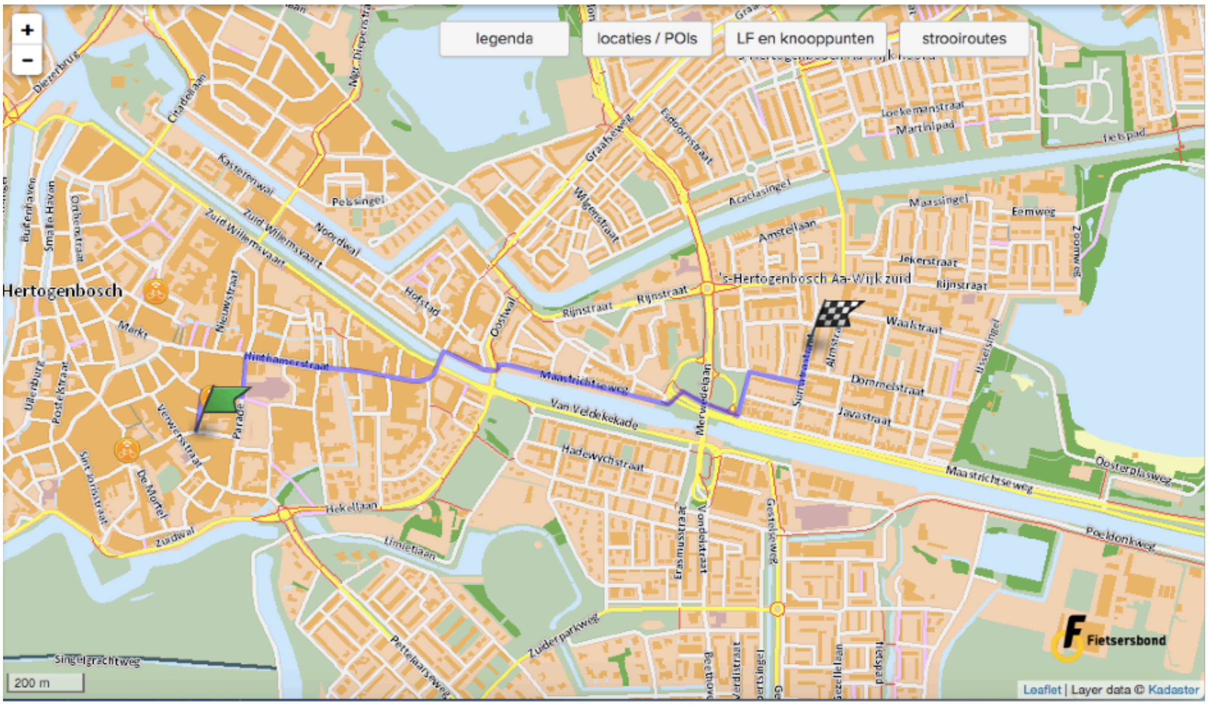


Figure A.28. Route (2,1 km) between the city centre (southeastern connection) and De Meerendonk (southwestern connection) with setting shortest route (Fietsersbond, 2018).

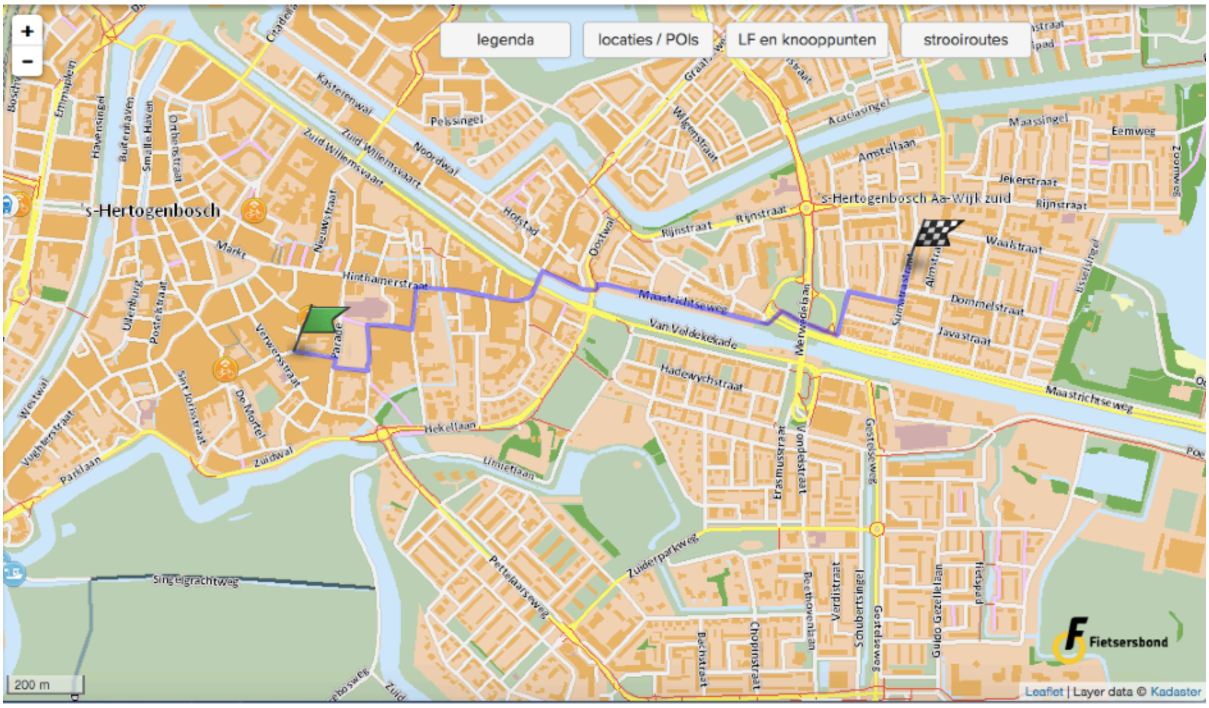


Figure A.29. Route (2,1 km) between the city centre (southeastern connection) and De Meerendonk (southwestern connection) with setting low-traffic route (Fietsersbond, 2018).

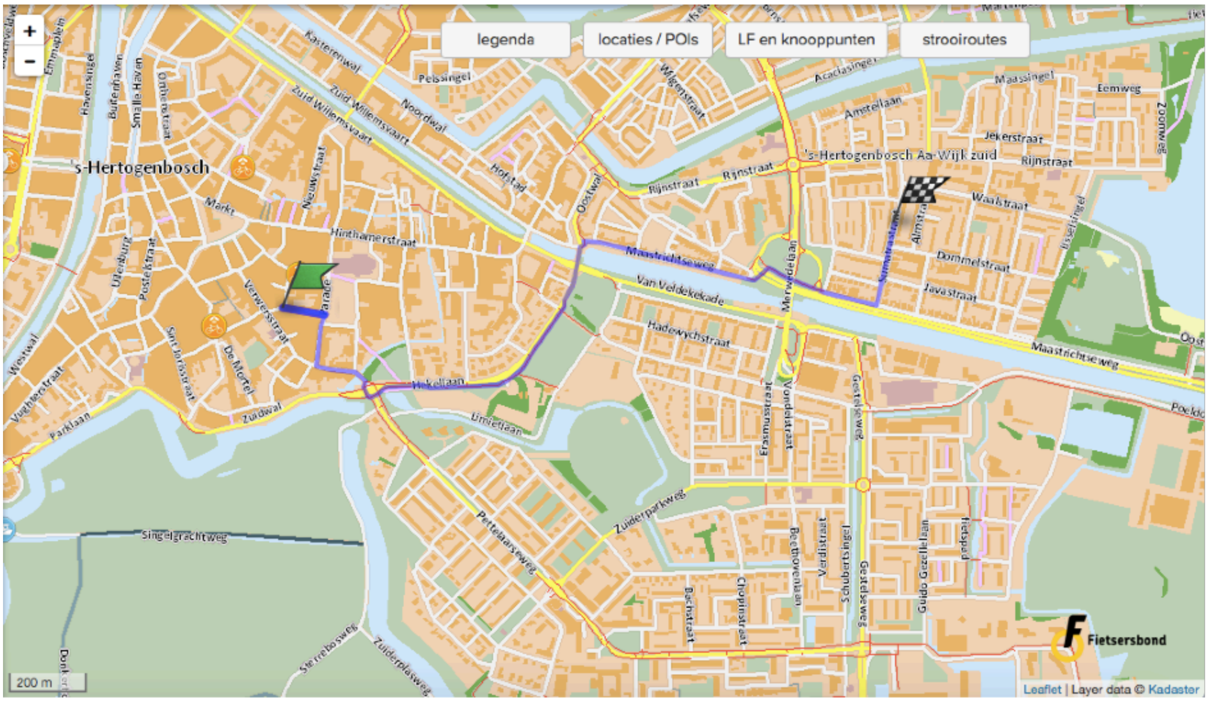


Figure A.30. Route (2,3 km) between the city centre (southeastern connection) and De Meerendonk (southwestern connection) with setting 'fietsbewust' (Fietsersbond, 2018).

Figure A.31. Generated routes between the city centre and De Meerendonk with setting 1.

Figure A.32. Generated routes between De Meerendonk and the city centre with setting 1.

Figure A.33. Generated routes between the city centre and De Meerendonk with setting 2.

Figure A.34. Generated routes between De Meerendonk and the city centre with setting 2.

Pair 4: Rosmalen - ’s-Hertogenbosch city centre (109-84)

The pair between Rosmalen and ’s-Hertogenbosch city centre has more choice and a longer distance. It seems realistic that people cycle this route from home to work. Figure A. 35 shows the locations of the centroids and figure A.36 the connections to the road network in Rosmalen.

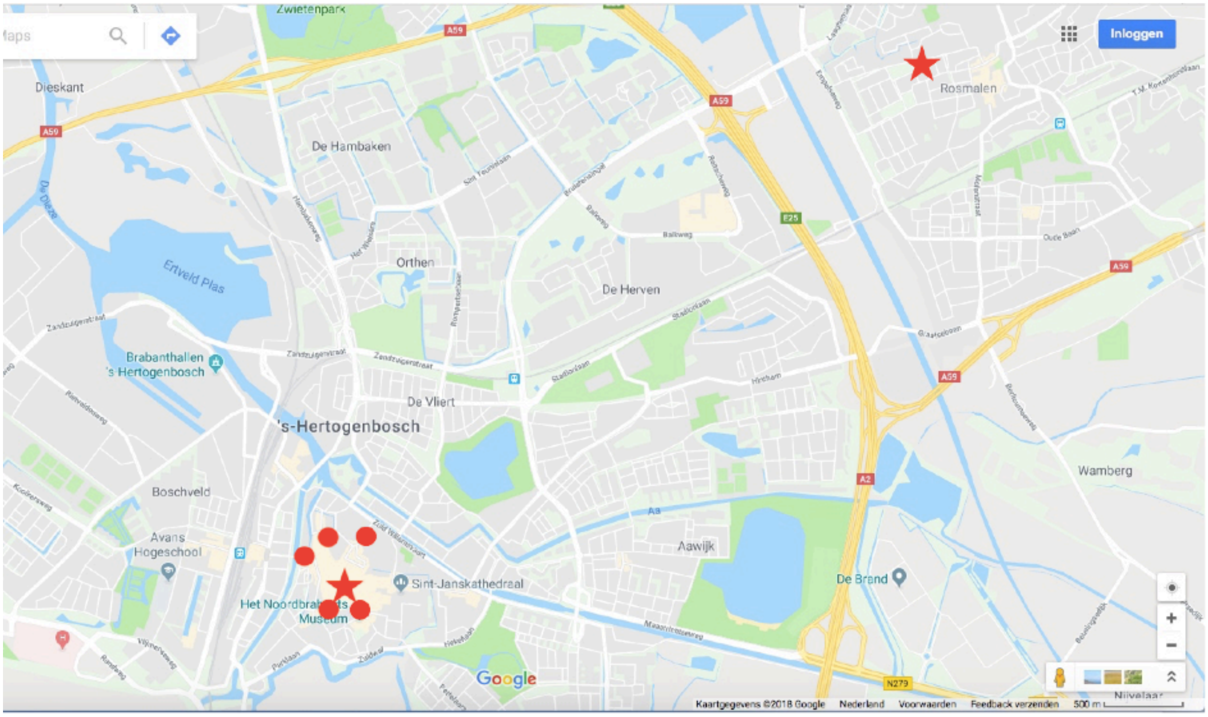


Figure A.35. Centroids (stars) and connections to road network (circles) of origin Rosmalen and destination city centre of 's-Hertogenbosch (Google, 2018).

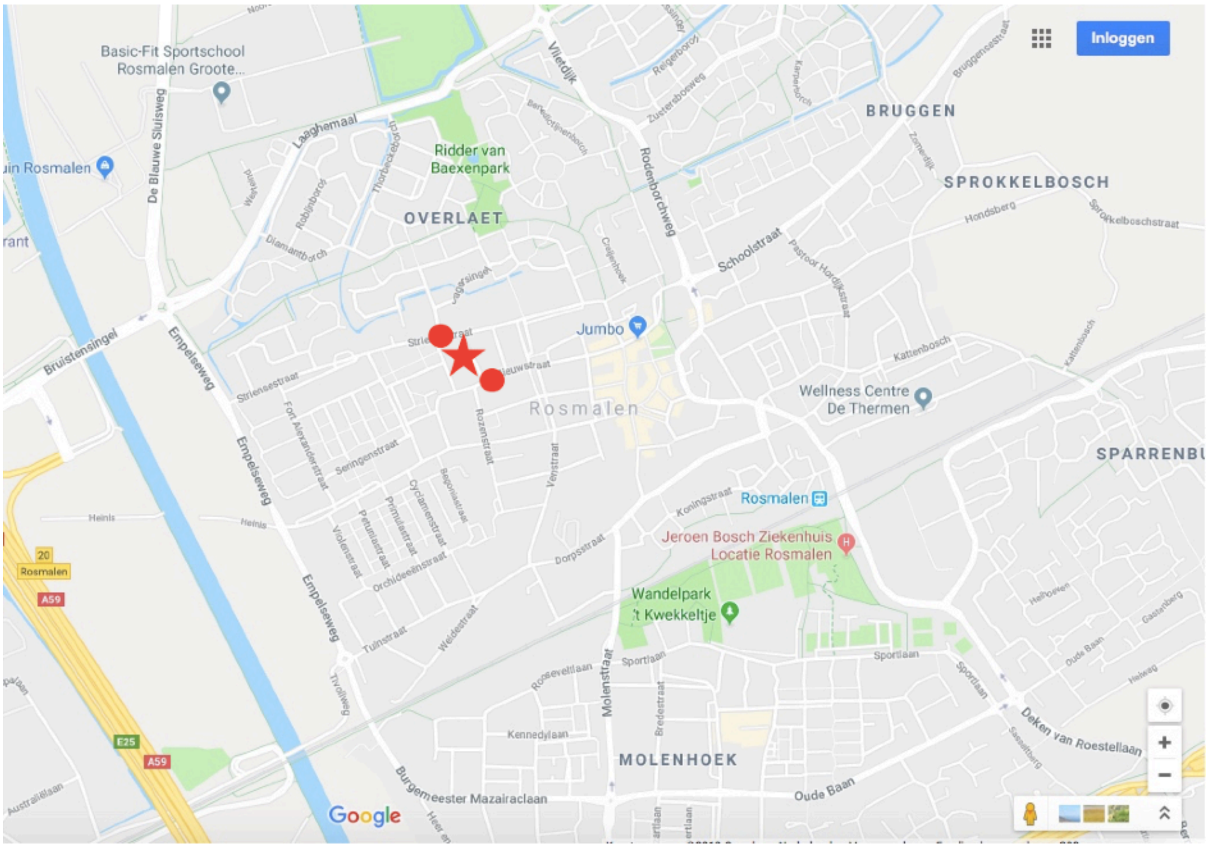


Figure A.36. Centroids (stars) and connections to road network (circles) in Rosmalen (Google, 2018).

First of all, routes were planned from the two connection points in Rosmalen to the northwestern point in the city centre in ’s-Hertogenbosch. These are shown below in figures A.37 to A.41.

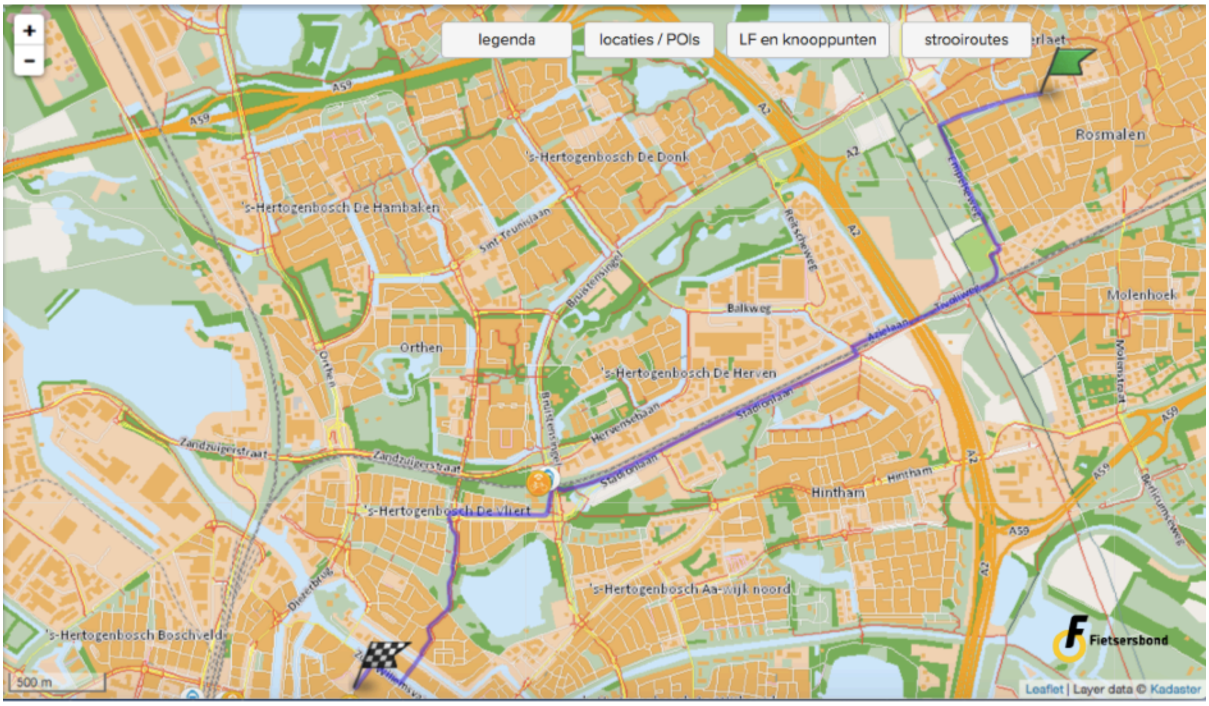


Figure A.37. Route (6,0 km) between Rosmalen (north connection) and the city centre with setting easy cycling and ‘fietsbewust’ (Fietsersbond, 2018).

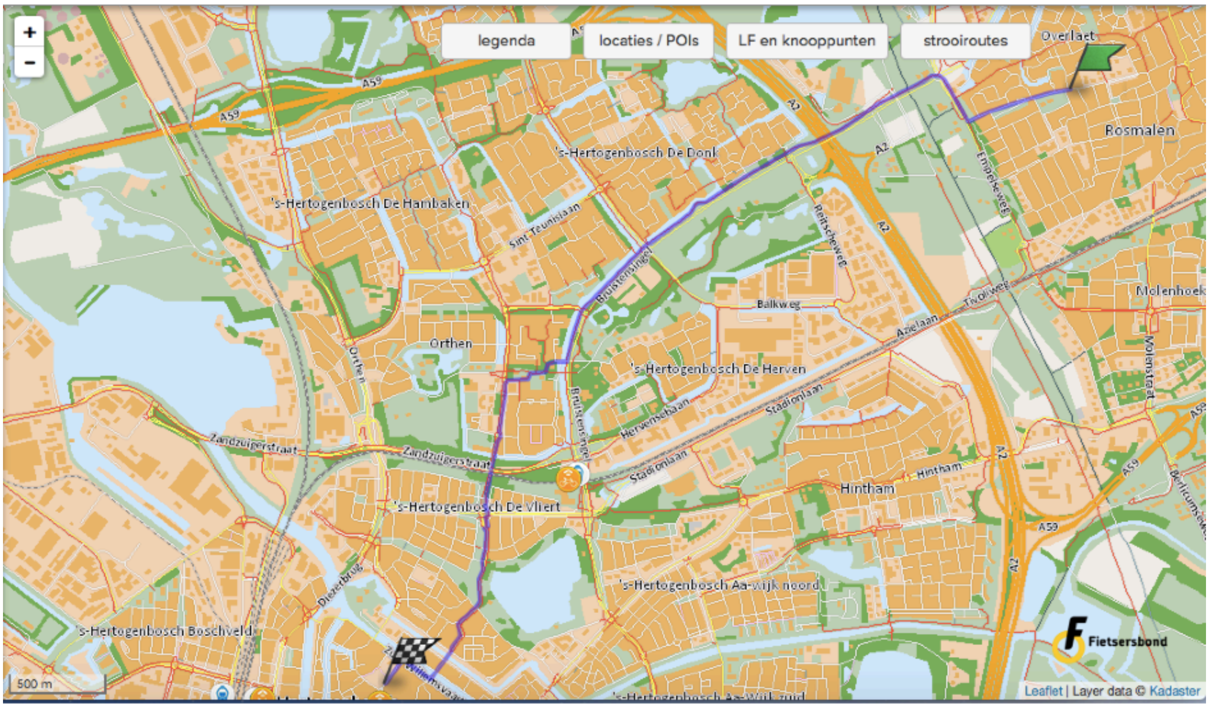


Figure A.38. Route (5,7 km) between Rosmalen (north connection) and the city centre with setting shortest route (Fietsersbond, 2018).

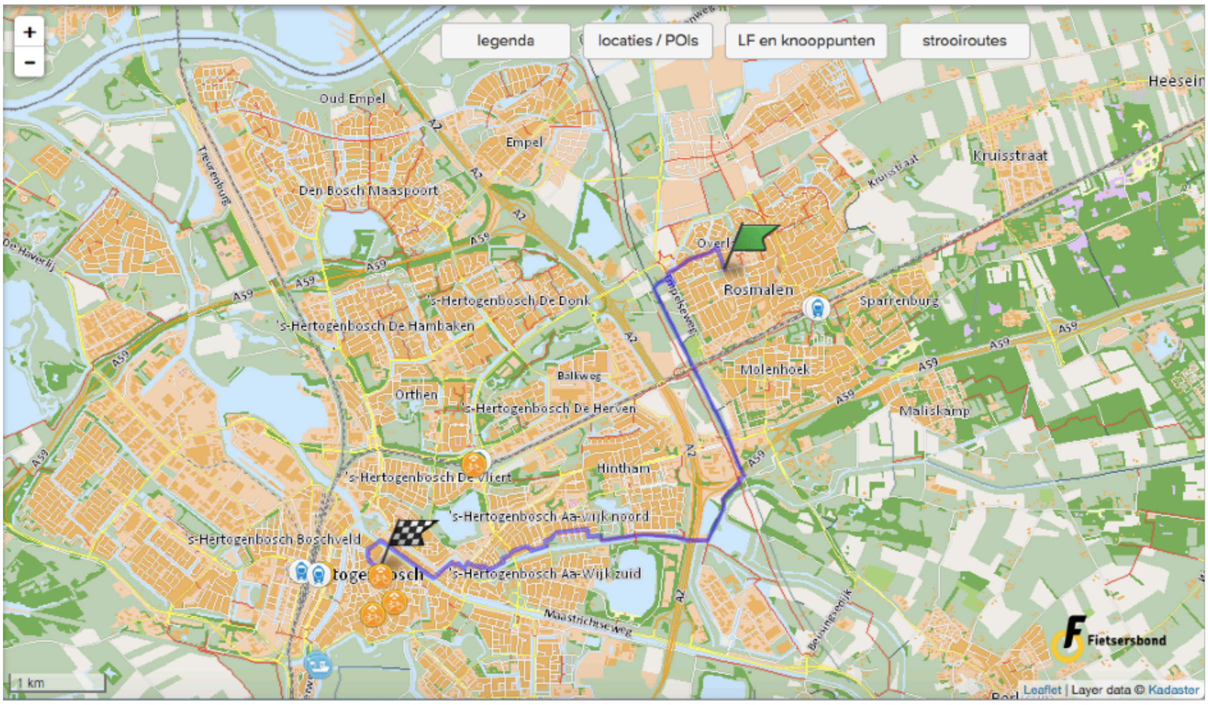


Figure A.39. Route (8,8 km) between Rosmalen (north connection) and the city centre with setting low-traffic routes (Fietsersbond, 2018).

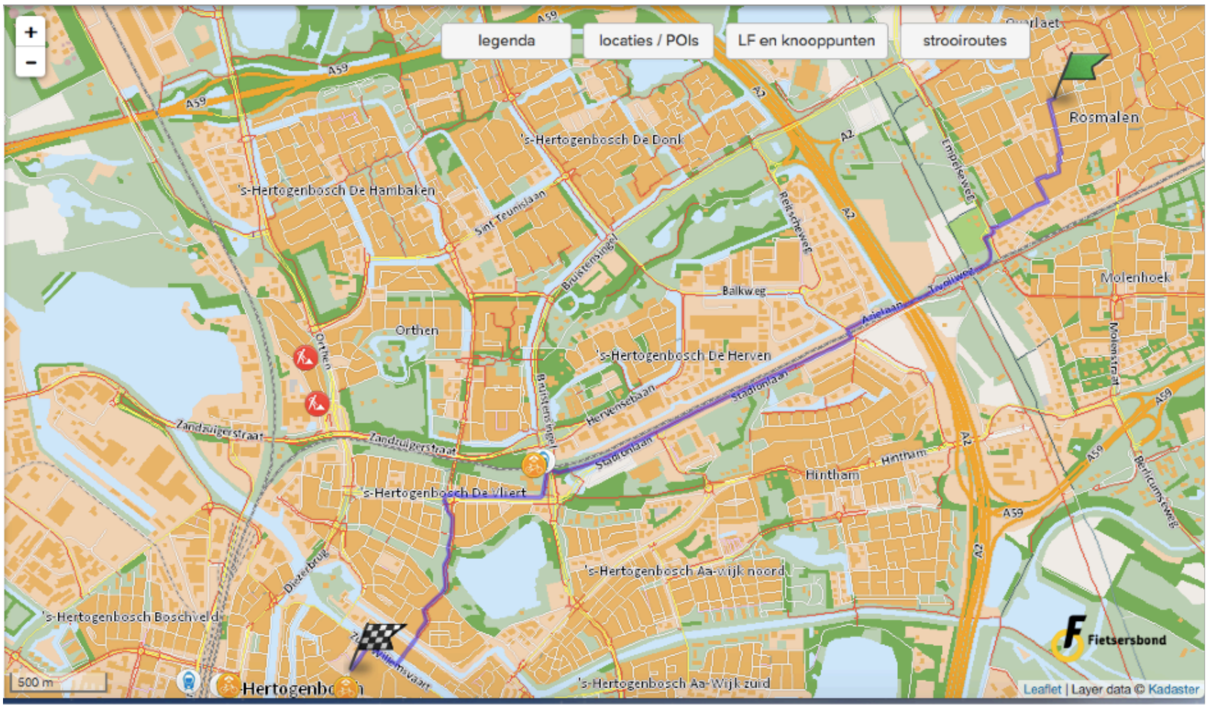


Figure A.40. Route (5,8 km) between Rosmalen (southern connection) and the city centre with setting easy cycling and 'fietsbewust' (Fietsersbond, 2018).

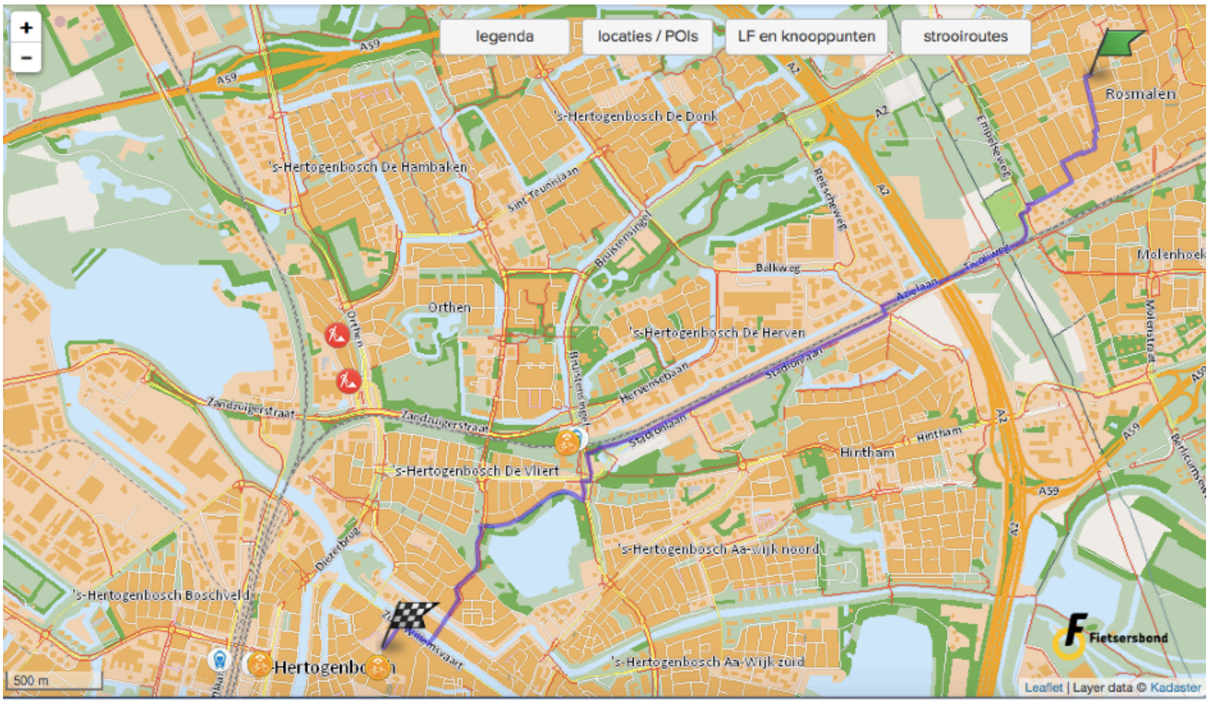


Figure A.41. Route (5,7 km) between Rosmalen (southern connection) and the city centre with setting easy cycling (Fietsersbond, 2018).

The low-traffic route from the southern connection point in Rosmalen is similar to the route from the northern point with a earlier start. Return trips are following the same route except for some one-way cycle paths. In those cases the route follows the cycle path at the other side of the main road.

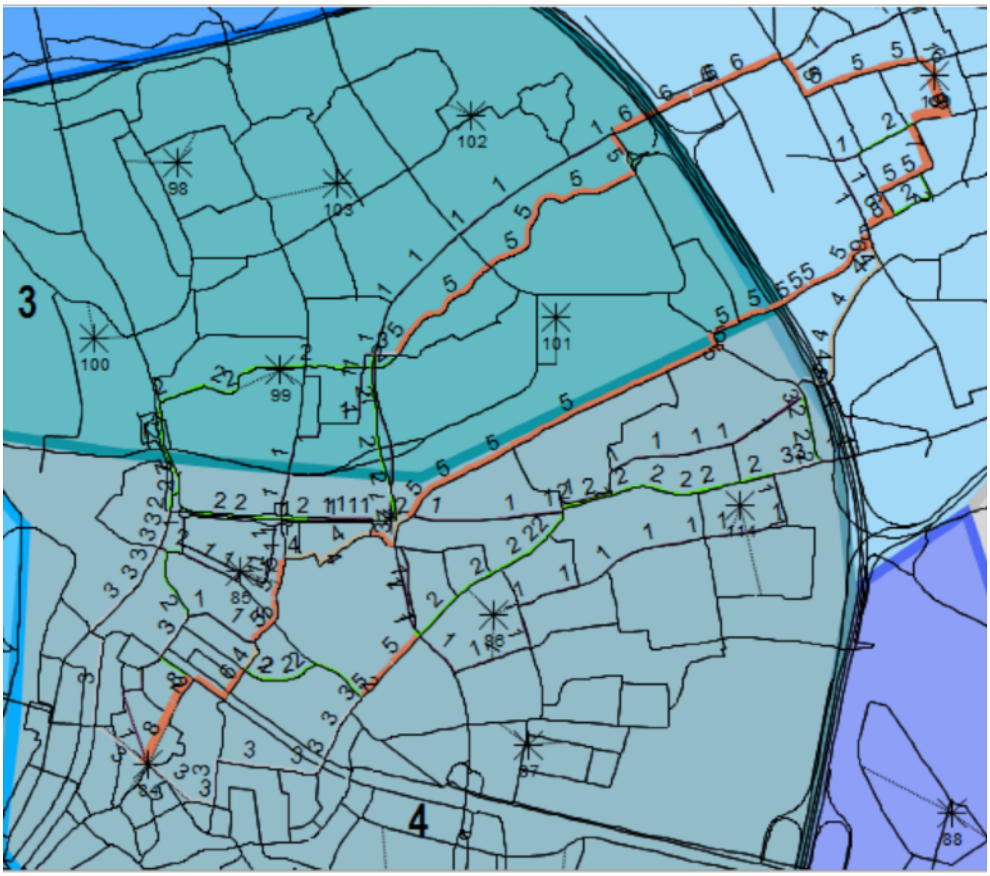


Figure A.42. Generated routes between Rosmalen and the city centre of 's-Hertogenbosch with setting 1.

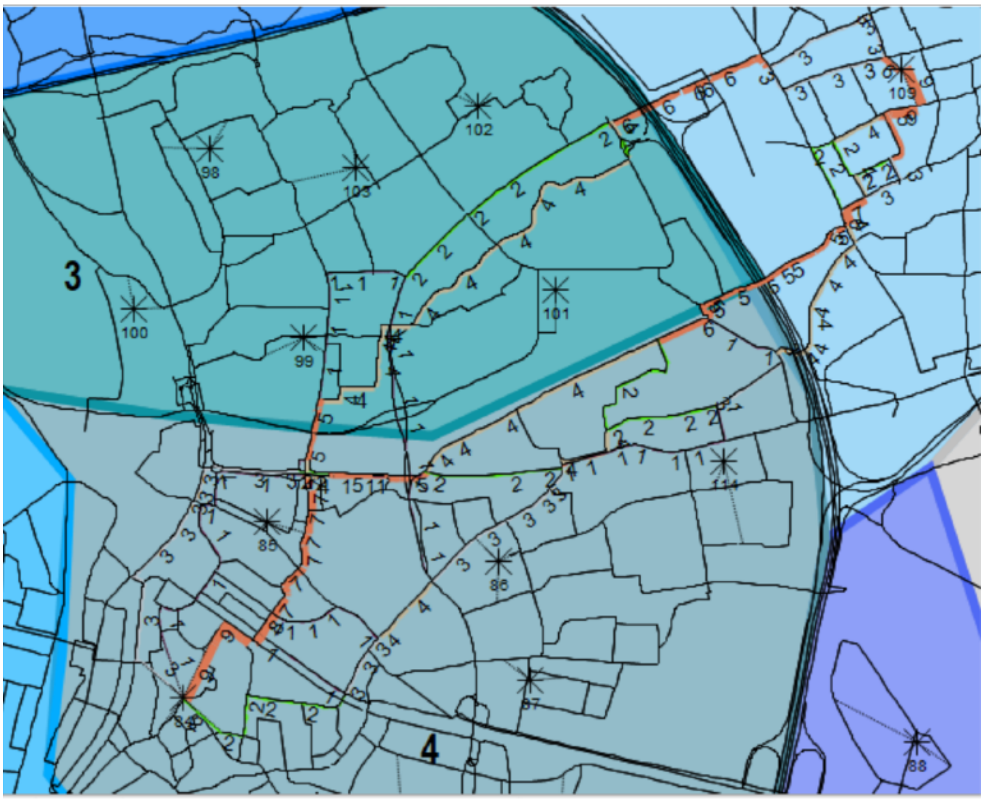


Figure A.43. Generated routes between the city centre of 's-Hertogenbosch and Rosmalen with setting 1.

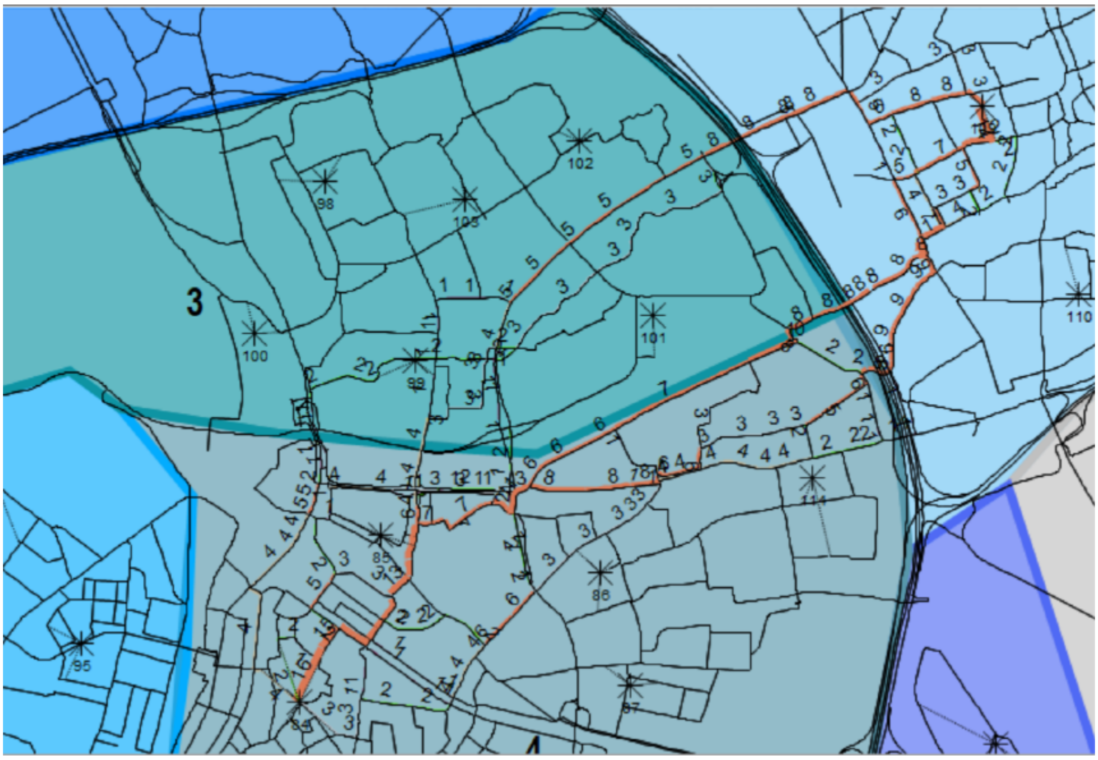


Figure A.44. Generated routes between Rosmalen and the city centre of 's-Hertogenbosch with setting 2.

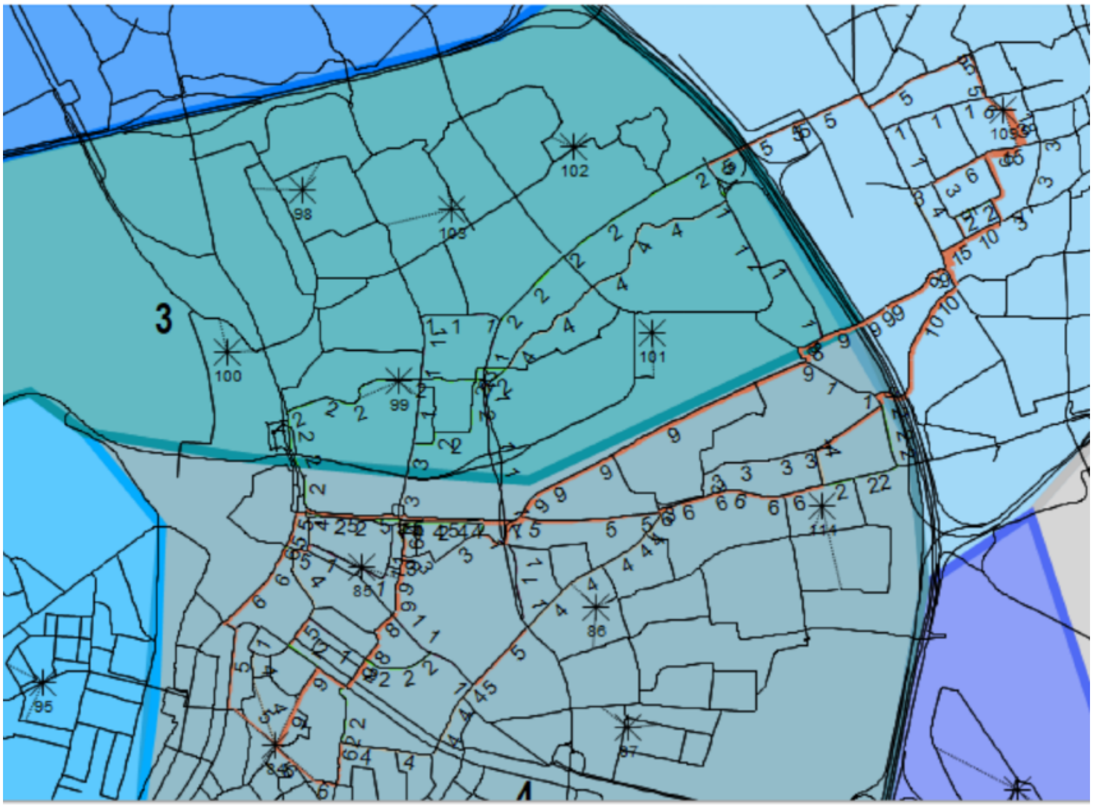


Figure A.45. Generated routes between the city centre of 's-Hertogenbosch and Rosmalen with setting 2.

Pair 5: Rosmalen - Nuland (109-128)

Finally, a rural OD-pair will be examined. This pair is between the towns of Rosmalen and Nuland. Figure A.46 shows the different parallel routes between the towns, which could mean that there is more than one possible route.

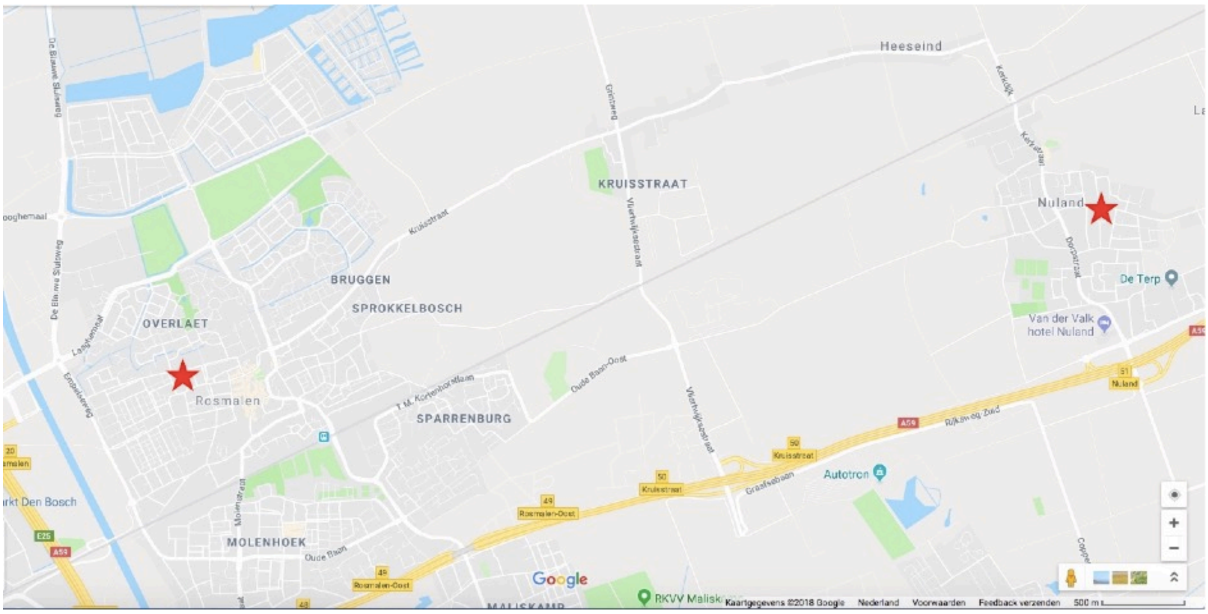


Figure A.46. Centroids in Rosmalen and Nuland (Google, 2018).

All routes are planned from the southern connection point in Rosmalen. De routes to the northern connection point have a slight variation at the start as they take the Striensestraat instead of the Nieuwstraat.

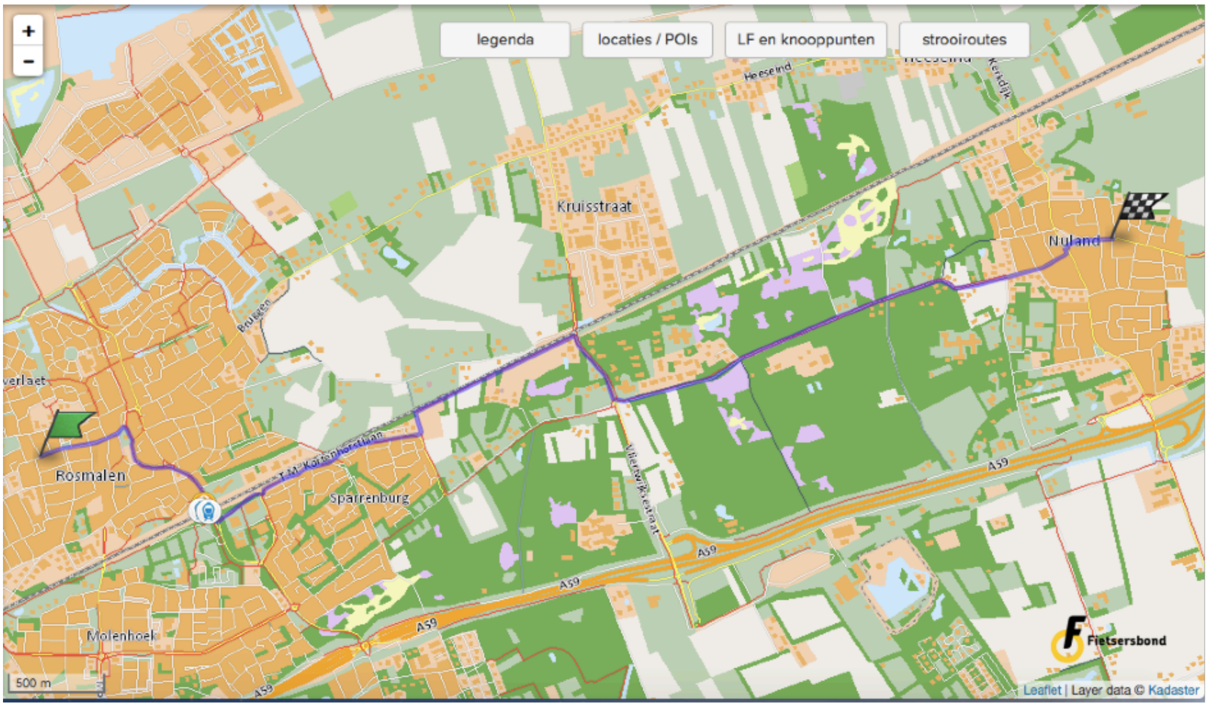


Figure A.47. Route (6,7 km) between Rosmalen (southern connection) and Nuland with settings easy cycling and 'fietsbewust' (Fietsersbond, 2018).

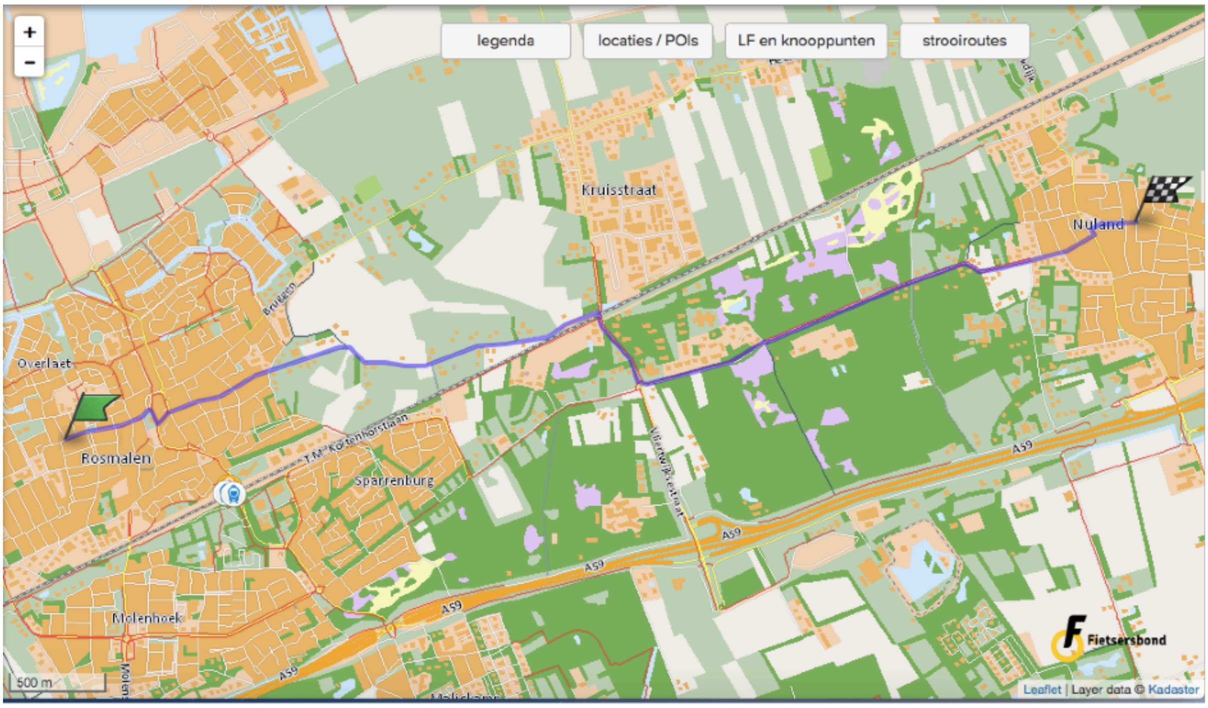


Figure A.48. Route (6,3 km) between Rosmalen (southern connection) and Nuland with setting shortest route (Fietsersbond, 2018).

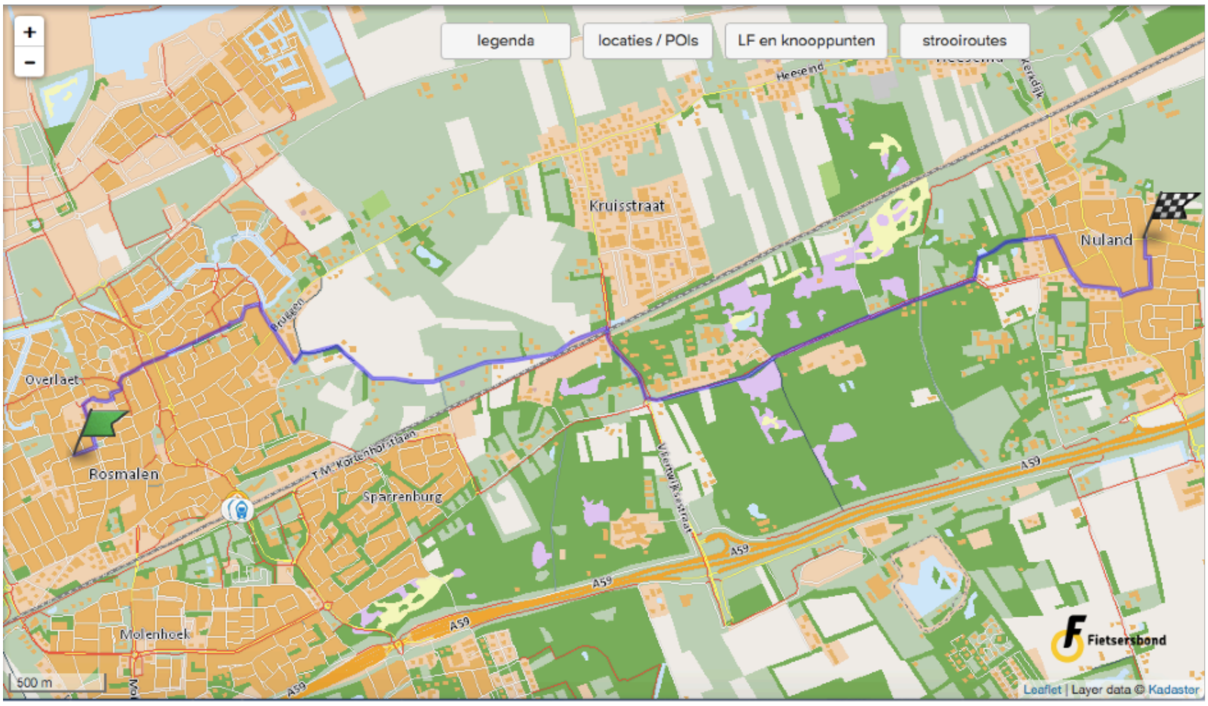


Figure A.49. Route (7,4 km) between Rosmalen (southern connection) and Nuland with setting low-traffic roads (Fietsersbond, 2018).

Figure A.50. Generated routes between Rosmalen and Nuland with setting 1.

Figure A.51. Generated routes between Nuland and Rosmalen with setting 1.

Figure A.52. Generated routes between Rosmalen and Nuland with setting 2.

Figure A.53. Generated routes between Nuland and Rosmalen with setting 2.

Appendix B: Calibration result data

This appendix provides additional information about the calibration by showing all results and more parameter combinations. Table B.1 provides the settings and results of the first tests which are also shown in section 5.3. Table B.2 shows all results of the random tests and tables B.4 and B.5 the parameter combinations of the best and worst scoring tests in table B.2. Table B.3 is a legend to the codes used in table B.4 and B.5.

Table B.1. Settings and results of tests 1 to 22.

Test number	Settings distance parameters	Settings time parameter	Settings intersection parameters	T < 3,5	3,5 < T < 4,5	T > 4,5
1	all 1,0	1,0	estimated values table 5.1	30	25	30
2	estimated values table 5.1	1,0	all 1,0	31	22	32
3	all 1,0	0,75	estimated values table 5.1	30	25	30
4	all 1,0	0,5	estimated values table 5.1	30	25	30
5	all 1,0	0,25	estimated values table 5.1	29	26	30
6	all 1,0	0,0	estimated values table 5.1	29	26	30
7	all 1,0	1,0	estimated values table 5.1 * 2	29	24	32
8	all 1,0	0,75	estimated values table 5.1 * 2	29	24	32
9	all 1,0	0,5	estimated values table 5.1 * 2	29	24	32
10	all 1,0	0,25	estimated values table 5.1 * 2	29	24	32
11	all 1,0	0,0	estimated values table 5.1 * 2	29	24	32
12	estimated values table 5.1	0,0	all 0,0	34	23	28
13	estimated values table 5.1	0,0	estimated values table 5.1	39	16	30
14	estimated values table 5.1	0,5	estimated values table 5.1	39	16	30

Test number	Settings distance parameters	Settings time parameter	Settings intersection parameters	T < 3,5	3,5 < T < 4,5	T > 4,5
15	estimated values table 5.1	0,75	estimated values table 5.1	39	16	30
16	estimated values table 5.1	1,0	estimated values table 5.1	39	16	30
17	estimated values table 5.1	0,5	estimated values table 5.1 * 0,5	37	21	27
18	estimated values table 5.1 * 0,5	0,5	estimated values table 5.1	34	20	31
19	estimated values table 5.1	0,5	estimated values table 5.1 * 2	35	20	30
20	estimated values table 5.1 * 2	0,5	estimated values table 5.1	40	18	27
21	estimated values table 5.1 * 2	0,75	estimated values table 5.1	40	18	27
22	estimated values table 5.1 * 2	1,0	estimated values table 5.1	40	18	27

Table B.2. Number of counts where the assignment has a t-value under 3,5, between 3,5 and 4,5 and more than 4,5 with different random parameter combinations.

Test	T < 3,5	3,5 < T < 4,5	T > 4,5	Test	T < 3,5	3,5 < T < 4,5	T > 4,5
24	38	14	33	**44**	37	12	36
25	32	18	35	**45**	33	18	34
26	38	18	29	**46**	37	16	32
27	36	23	26	**47**	31	20	34
28	36	16	33	**48**	30	20	35
29	29	27	29	**49**	38	21	26
30	37	17	31	**50**	33	21	31
31	39	16	30	**51**	35	19	31
32	30	19	36	**52**	30	18	37
33	36	22	27	**53**	39	17	29
34	36	21	28	**54**	33	19	33
35	35	21	29	**55**	38	21	26
36	33	17	35	**56**	34	19	32
37	30	25	30	**57**	34	18	33

Test	T < 3,5	3,5 < T < 4,5	T > 4,5	Test	T < 3,5	3,5 < T < 4,5	T > 4,5
38	31	20	34	**58**	35	18	32
39	36	16	33	**59**	34	19	32
40	32	22	31	**60**	31	20	34
41	31	18	36	**61**	40	14	31
42	31	20	34	**62**	33	17	35
43	34	20	31	**63**	31	20	34

Table B.3. Parameter number and associated type of road or intersection.

Parameter number	Distance factors	Parameter number	Intersection factors
1 (and 2)	Separated cycle (and moped) path along road	**1**	Unknown type
3	Bicycle boulevard	**2**	Equivalent
4	Normal road	**3**	Give way
5 (and 6)	Solitary cycle (and moped) path	**4**	Priority
7	Ferry	**5**	Traffic signals
8	Service road	**6**	Roundabout
9	Pedestrian short cut	**7**	Roundabout with traffic signals
10	Pedestrian area		
11	Cycle lane		

Table B.4. Distance parameter combinations of the best and worst results.

Test / para meter	1	2	3	4	5	6	7	8	9	10	11
24	0,4593	0,3975	0,2940	0,6457	0,0346	0,2936	0,5292	0,3348	0,9202	0,8706	1,2803
26	0,5291	0,9088	0,3260	1,4879	1,0037	1,0395	0,2632	1,2456	0,3692	0,9062	1,2277
32	1,1251	1,2513	1,4171	1,1233	0,2664	1,3758	1,0966	0,3016	1,1582	0,7532	1,1447
33	1,3905	0,2252	1,3357	0,9230	1,2802	0,1543	0,1093	0,5161	0,4776	0,5819	1,1747
37	0,0440	0,7853	1,4325	0,8383	0,3314	0,5429	1,2128	0,8087	0,2036	1,0500	0,1253
46	1,0377	0,7509	0,2528	1,3400	1,0465	0,0440	0,8065	0,3464	0,2974	0,8274	0,1883
48	1,1627	1,3508	1,4830	0,4713	0,9622	1,3231	0,9051	0,7134	0,7003	0,4702	1,0052

Test / para meter	1	2	3	4	5	6	7	8	9	10	11
49	1,4653	0,1281	1,0864	0,8924	1,4449	0,9171	0,5597	0,5176	1,0957	1,2432	0,1558
52	1,1997	0,6466	1,0643	0,3303	0,9276	1,2886	1,2647	0,5283	0,8752	0,6789	0,6626
53	0,2663	0,2779	0,5925	0,8617	0,9885	0,6681	1,1483	0,4008	0,4368	0,7037	1,2607
55	0,9114	0,0954	0,4091	1,1777	0,0666	0,3772	1,0909	0,8027	0,1621	0,8480	0,3744
61	0,8241	0,7859	0,8163	1,3216	0,2477	0,5072	0,9049	0,6706	0,8773	0,6818	1,3404
63	1,2434	1,2354	1,4234	0,7252	0,0839	0,9089	1,0292	0,9782	0,5973	0,3368	1,3617

Table B.5. Time and intersection (= is) parameter combinations of the best and worst results.

Test / parameter	time	is 1	is 2	is 3	is 4	is 5	is 6	is 7
24	0,98038	0,012899	0,046116	0,041615	0,060272	0,018307	0,041791	0,013409
26	1,42542	0,004526	0,035255	0,004189	0,015048	0,033833	0,016328	0,076097
32	0,34969	0,088180	0,073632	0,081057	0,024773	0,036489	0,074637	0,082153
33	1,26261	0,035497	0,012392	0,087946	0,050592	0,040802	0,088415	0,088212
37	0,27338	0,017925	0,061745	0,082943	0,026219	0,063455	0,096775	0,021143
46	0,50945	0,012772	0,039835	0,088563	0,014804	0,099945	0,002618	0,078304
48	0,54281	0,088080	0,076898	0,060897	0,004459	0,076562	0,049073	0,041909
49	0,05265	0,039980	0,012869	0,032437	0,002276	0,046445	0,009120	0,067752
52	1,14852	0,050802	0,051579	0,044392	0,036610	0,087547	0,041225	0,047198
53	0,83113	0,009003	0,057238	0,086486	0,029200	0,085415	0,026889	0,042638
55	1,35629	0,083603	0,090309	0,007569	0,033566	0,074193	0,056082	0,021806
61	0,59730	0,023824	0,020612	0,004129	0,031124	0,043145	0,027948	0,009082
63	0,27151	0,062154	0,009124	0,057598	0,014394	0,021597	0,092584	0,078900

Figures B.1 and B.2 provide an overview of the assignment results of test 49 and 70.

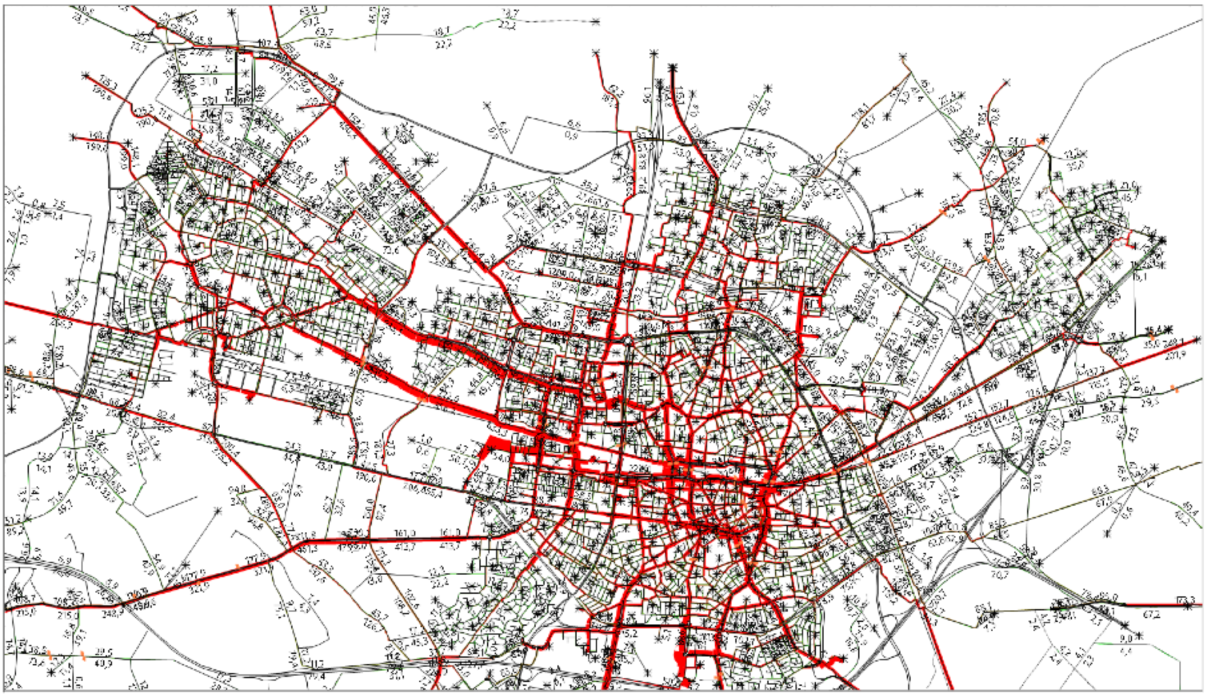


Figure B.1. Assignment result of parameter settings belonging to test 49* with an increased matrix.

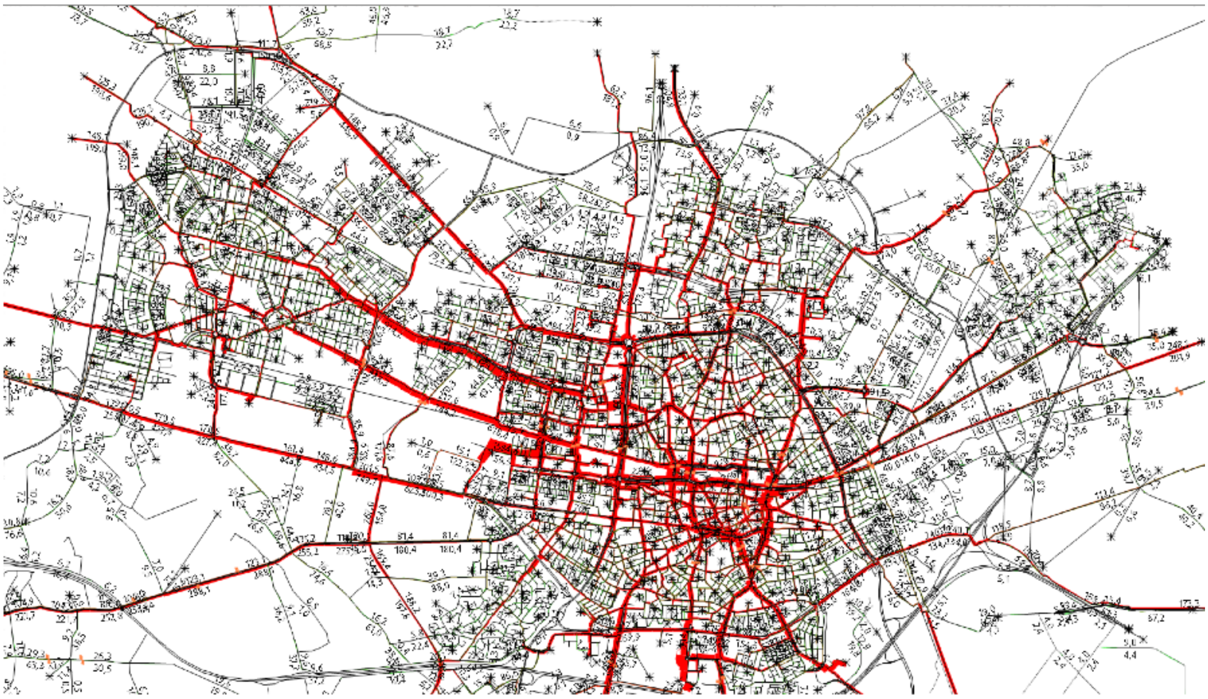


Figure B.2. Assignment result of parameter settings belonging to test 70* with an increased matrix.

The results of the random parameter setting tests were used to create a set of best values. In every test, as described in chapter 5, one value is changed to a higher or lower value as shown in table B.6. Table B.7 shows the changes made in each test and the results of the tests.

Table B.6. Parameter values derived from previous tests with random settings.

Distance factors	Low value	Middle value	High value	Intersection factors	Low value	Middle value	High value
Separated cycle (and moped) path along road	0,1	0,8	1,5	Unknown type	0,01	0,02	0,04
Bicycle boulevard	0,3	0,8	1,1	Equivalent	0,01	0,02	0,04
Normal road	0,9	1,3	1,5	Give way	0,0	0,0	0,03
Solitary cycle (and moped) path	0,2	0,2	1,2	Priority	0,02	0,03	0,05
Ferry	0,9	0,9	0,9	Traffic signals	0,03	0,04	0,08
Service road	0,4	0,7	0,8	Roundabout	0,02	0,03	0,06
Pedestrian short cut	0,4	0,9	1,0	Roundabout with traffic signals	0,01	0,01	0,01
Pedestrian area	0,6	0,7	0,7				
Cycle lane	0,4	1,3	1,3				

Table B.7. Parameter settings and results of tests 64 to 97.

Test	Parameter setting	T < 3,5	3,5 < T < 4,5	T > 4,5
64 to 69	time parameter from 0 to 1 with steps of 0,2	37	18	30
70	separated cycle (and moped) path parameter low value	40	16	29
71	bicycle boulevard parameter low value	40	14	31
72	normal road parameter low value	33	20	32
73	solitary cycle (and moped) path parameter low value	39	17	29
74	service road parameter low value	40	14	31
75	pedestrian short-cut parameter low value	39	15	31
76	pedestrian area parameter low value	39	15	31
77	cycle lane parameter low value	37	19	29
78	separated cycle (and moped) path parameter high value	32	23	30
79	bicycle boulevard parameter high value	39	17	29
80	normal road parameter high value	37	19	29
81	solitary cycle (and moped) path parameter high value	37	17	31

Test	Parameter setting	T < 3,5	3,5 < T < 4,5	T > 4,5
82	service road parameter high value	39	15	31
83	pedestrian short-cut parameter high value	39	15	31
84	pedestrian area parameter high value	39	15	31
85	cycle lane parameter high value	39	15	31
86	unknown intersection parameter low value	39	16	30
87	equivalent intersection parameter low value	39	15	31
88	give way on intersection parameter low value	39	15	31
89	priority intersection parameter low value	38	16	31
90	traffic signals parameter low value	40	14	31
91	roundabout parameter low value	39	15	31
92	unknown intersection parameter high value	37	18	30
93	equivalent intersection parameter high value	39	16	30
94	give way on intersection parameter high value	39	14	32
95	priority intersection parameter high value	39	15	31
96	traffic signals parameter high value	40	14	31
97	roundabout parameter high value	39	15	31

Table B.8. Number of counts where the assignment has a T-value under 3,5, between 3,5 and 4,5 and more than 4,5 with different random parameter combinations in between those of test 49 and 70. Note that all results are obtained using the increased OD-matrix.

Test	T < 3,5	3,5 < T < 4,5	T > 4,5	Test	T < 3,5	3,5 < T < 4,5	T > 4,5
100	36	13	36	113	33	13	39
101	33	13	39	114	36	11	38
102	34	14	37	115	35	13	37
103	38	9	38	116	35	11	39
104	35	13	37	117	36	10	39
105	35	13	37	118	38	12	35
106	31	17	37	119	37	11	37
107	38	12	35	120	37	12	36
108	34	12	39	121	34	18	33
109	38	13	34	122	36	12	37

Test	T < 3,5	3,5 < T < 4,5	T > 4,5	Test	T < 3,5	3,5 < T < 4,5	T > 4,5
110	36	15	34	**123**	32	14	39
111	36	13	36	**124**	31	15	39
112	34	13	38				