Cyclist Route Choice Modeling

A research on influence of Openness & Monotony on cyclist route choice

Technical Report

Anastasia Anastasiadou, Meylin Herrera Herrera,
Davey Oldenburg, Ioanna Tsakalakidou & Erik van der Wal

TU Delft

Provincie Noord-Brabant

[This page was intentionally left blank]

Cyclist Route Choice Modeling

A research on influence of Openness & Monotony on cyclist route choice

Authors:

Anastasia Anastasiadou
Meylin Herrera Herrera
Davey Oldenburg
Ioanna Tsakalakidou
Erik van der Wal

Supervisors:

ir. Edward Verbree
dr. Kees Maat

Clients:

B-riders project
Professorship of Urban Intelligence
Province of Noord-Brabant

Preface

An important part of the MSc Geomatics program at Delft University of Technology is the synthesis project. This is a group project that connects all aspects of the first year of Geomatics program. The span of the project is nine weeks and takes place at the end of the first year. As there were five possible selections for a project subject, this project aims to research the effect of the built environment on cyclist route choice. The novelty of this research is the establishment of models that describe openness and monotony of the built environment, two concepts that have never been related to cyclist travel behavior before.

A. Anastasiadou, M. Herrera Herrera, D. Oldenburg, I. Tsakalakidou & E. van der Wal

Delft, June 2017

Acknowledgements

During this research, a few people and organisations have been of great support. First and foremost, we would like to thank Drs. Ing. Paul van de Coevering, representing the Professorship of Urban Intelligence and the Province of Noord-Brabant for the provision of data and input during the course of this project. Furthermore, we owe our thanks to dr. Kees Maat and ir. Edward Verbree for their supervision. Their views on the research, and help during the process, brought us to the point where we are now. Finally, we would like to thank dr. ir. Mathias Lemmens, in cooperation with Edward Verbree, for the excellent organization of the project in general.

Table of contents

Cyclist Route Choice Modeling
A research on influence of Openness & Monotony on cyclist route choice

Anastasia Anastasiadou, Meylin Herrera, Davey Oldenburg, Ioanna Tsakalakidou, Erik van der Wal

Supervisors: ir. Edward Verbree, dr. Kees Maat

Clients: B-riders project, Professorship of Urban Intelligence, Province of Noord-Brabant

Keywords: built environment, cyclists, openness, monotony, bike route choice

Abstract

In previous studies on cyclist route choice, many influencing factors have been identified. Openness of the built environment, which can be described as the extent of open space above and around a specific point, has no yet been related to cyclist travel behaviour. Monotony of the built environment, described as the extent of visual variation of elements that form the built environment for a sequence of locations, has been related to transportation problems but not to cyclist route choice in particular. This research seeks to bridge this gap in existing literature by determining the effect of the openness and monotony of the built environment on cyclist route choice in the Province of Noord-Brabant. The openness value for a specific route has been modeled by accounting for the building heights and distance to a building for a sequence of locations. Applying a linear regression analysis with the openness and monotony model as input shows that openness of the built environment has a negative influence on the amount of distance people are willing to diverge from the shortest path, while on the other hand, cyclists prefer to use roads with higher variation in the built environment. However, the results show that the proportional influence of both factors can be considered as low.

1. Introduction

Since The Netherlands covers a relatively small area, with a high density of cities, towns and villages, bicycles and e-bicycles have been recognized as an appropriate means of inter-city transportation by governments on different scale levels within the country (Ministry of Transport, Public Works and Water Management, 2009). To strengthen this recognition, the Province of Noord-Brabant has established a policy to promote and increase the use of bicycles by, among other measures, the development of high speed bike lanes between the larger city in the province. (Province of Noord-Brabant, 2009). In order for the high speed bike lanes to be a stimulant for bicycle use, the Province of Noord-Brabant wishes to ensure the potential use of these lanes, by taking into account factors that affect the cyclist' route choice. Over the years many studies have been performed on cyclist' travel behaviour and the factors that influence the route choice. An extensive set of influencing factors has been identified. These factors can be related to road facilities, safety, cyclist characteristics and the built environment. Despite existing research on the influence of the built environment on cyclist travel behaviour, the potential influence of openness and monotony of the built environment has not been examined yet. Openness of the built environment can be defined as *the extent of open space above and around a specific point,* whereas monotony of the built environment is described as *the extent of visual variation of elements that form the built environment for a sequence of locations.* This research aims to further examine how the configuration of the built environment, described in terms of openness and monotony, affects cyclist route choice in the province of Noord-Brabant, by answering the following research question:

How do openness and monotony of the built environment affect cyclist' route choice in the Province of Noord-Brabant?

Identifying the influence of openness and monotony of the built environment on cyclist travel behaviour will contribute to the investigation of how high speed lanes in the province of Noord-Brabant can meet the user requirements, as well as to the implementation of strategic analysis of cycling facility improvement schemes in the whole province.

2. Theoretical framework and definitions

Former studies have identified a set of characteristics that influence the use of bicycles and the choice of bicycle routes. However, both openness and monotony of the built environment in relation to cyclist route choice have not been covered yet. An objective of this research is to bridge the gap in existing literature, by establishing definitions for both concepts based on theory. These definitions will be the base for the spatial and mathematical modeling of openness and monotony of the built environment.

2.1 Openness of the built environment

The term of openness has been approached throughout the literature in various ways and research comes from different disciplines. Benedikt (1979) has developed the term *Isovist*, which is “the set of all visible points from a given vantage point in space and with respect to an environment”. Oke (1981) approaches openness by means of calculating the proportion of the sky that is visible from a certain point: the *sky view factor.* Finally, Fisher-Gewirtzman & Wagner (2003) introduced the *spatial openness index*, which measures the volume of open space that can potentially be seen from a certain point. A common aspect in the approaches of the three studies is that openness is related to the visual scene of a person on a certain location. Accounting for the results of previous studies, openness of the built environment can be defined as follows: *the extent of open scene above and around a specific point. In the built environment an environment is considered as open when no obstacles of the built environment are interfering with the visual scene of a person, or the interference can considered to be low.*

2.2 Monotony in the built environment

Like openness, monotony in the built environment is related to the visual experience of cyclists while cycling. Previous studies have covered monotonous road environments, and the effect on drivers’ fatigue (Thiffault & Bergeron, 2003; Zhao & Rong, 2013). Based on these studies, and taking into account the definition of the built environment, a definition of monotony in the built environment can be established. Since this research focuses on cyclists, and therefore encompasses a transportation problem, the emphasis is on how cyclists experience the built environment during their trip, and not on one particular location. Therefore, monotony in the built environment can be defined as *the extent of visual variation in elements that form the built environment for a sequence of locations.*

3. Data description and management

The province of Noord-Brabant is located in the south of the Netherlands, and consists of sixty six municipalities that form an area covered by built environment as well as rural areas. For the purpose of data manageability the decision has been made to focus on a smaller sample area, instead of applying the methodology on the entire province of Noord-Brabant. The sample area has been selected around Breda and Tilburg, two cities that are located in each others proximity, and their near surroundings. The selection of the area is based on the fact that the two municipalities are among the biggest of the province, meaning that a considerable number of daily travels is generated. Additionally, considering the fact that they are neighboring municipalities, this selection allows us to research both urban environment and the rural areas among and around the municipalities.

The development of the models describing openness and monotony of the built environment in a spatial and mathematical way has been based GPS tracking data of a set of commuters in the province of Noord-Brabant. The tracking data has been linked to datasets on existing bicycle road networks, to enable assignment of travel information to the road network. The connection between the different datasets is depicted in Figure 1.

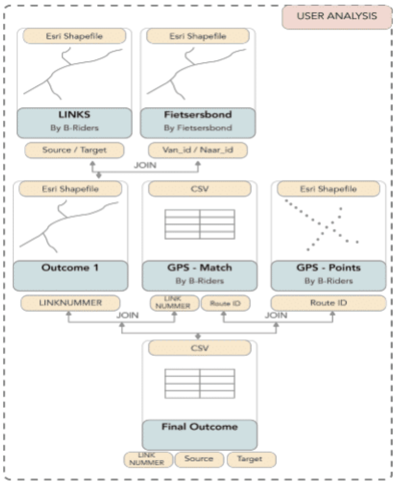


Figure 1: *Datasets connections*

4. Methodology

The methodology consists on the development of a set of differents models: a base route model, an openness model and multiple monotony models. Finally, a linear regression analysis provides the output on which conclusions can be drawn to answer the main research question.

4.1 Base route model

The base route model is consisting of two main stages: the shortest path analysis and the routes analysis based on GPS tracking data. A routing topology network was created based on a graph containing edges, vertices and distances that later served as the input for applying the A* algorithm and finding the shortest trajectory for source (origin) - target (destination) points of every route in the sample area. In addition, spatial patterns of cyclists were analyzed by matching the GPS tracking data with the bicycle road network. The number of travels made by cyclists over the road segments were calculated and the preferences of the user group on the bicycle road network were visualized on the intensity map. Theoretical travel times per road segments were calculated based on the average speed provided within the GPS measurements and the edge lengths generated from the routing topology network. This average travel time was the base for calculating the travel times per shortest paths and observed routes.

4.2 Openness model

The definition of openness of the built environment that has been used as the base for the development of the openness model takes into account the (interference of the) visual scene of a cyclist on a certain location. Many elements of the built environment have been found to affect the value of openness in previous studies. However, the scope of this research only considers buildings as obstacles of a cyclist's view, due to time limitations and a desired simplification of the input variables.

The openness on a certain location is therefore affected by the configuration of the buildings in the neighborhood of that location. First of all, the visual field of a cyclist is affected by the distance to a building, as well as the height of a building. The main idea is that the perception of openness of the built environment is negatively influenced by the height of a building, but positively by the distance to a building. In other words, higher buildings will lead to a less open feeling, while a larger distance to buildings will create a more open experience (Figure 2).

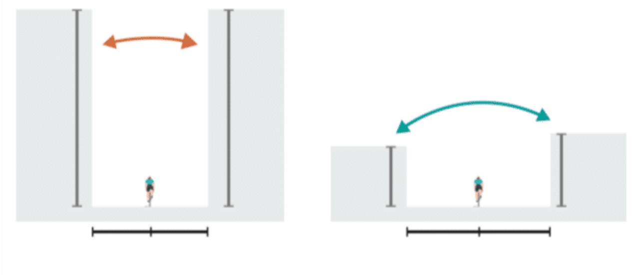


Figure 2: *A less open visual scene (left) versus a more open visual scene (right).*

The applied methodology to compute the openness values for a certain location takes into account the possible visual scene of a cyclist while cycling (Figure 3). A vertical angle of 15°, starting from the horizontal line of sight, is used as this is the angle of view of a person while driving and looking straight (Cichański, A & Wirwicki ,2010). The next step is to estimate how far a building must be in order to be visually evident within the visual field created by the 15° angle. Classifying the heights of buildings (H) into four classes (Table 1), four distances were calculated using the formula $distance = \frac{H}{tan15°}$, resulting in four zones.

Building height classes (m)*	Resulting zones (m)
[-0.8-12.3]	45.52
(12.3-18.4]	68.28
(18.4-55.2]	205.59
(55.2-125.25]	568

Table 1: *Building height classes and resulting zones*

Each zone influences the cyclist according to the height of the buildings existing in the respective distance. This approach indicates that the variables of distance and building height are interrelated in a way that whether a building is included in a person's view depends on the distance of the building to that person, but also on the building height (Figure 3).

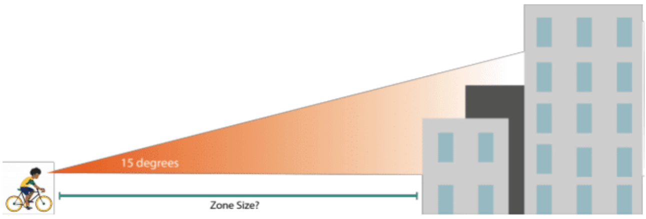


Figure 3: *Visualization of the openness calculation.*

The implementation process for the computation of openness values involved three main stages: finding point neighbours to every sample point, computing point intersections towards the buildings and finally calculating and visualizing openness values for the whole sample area.

Buildings neighbours were found by applying the nearest neighbours algorithm given a certain distance to the sample point (FDN algorithm). The point intersections were calculated based on the *sight-and-light* algorithm, adjusted from NCase, 2018.The implementation starts by casting rays from the point towards the buildings within its vicinity. With a suitable angle of 10°, a total of 36 rays were casted for every sample point on the road network (Figure 4).Since the algorithm is based on operations with parametric equations, the rays and all the line segments were combined in a linear system of equation the allowed to solve the independent parameters *T1* and *T2* and find the intersecting points.

$$rpx+rdx*T1 = spx+sdx*T2$$
$$rpy+rdy*T1 = spy+sdy*T2$$

Where:

rpx= ray a.x	,	*rpy= ray a.y*
spy= segment a.y	,	*spx= segment a.y*
rdx= ray b.x - ray.a.x	,	*dy= ray b.y - ray.a.y*
sdx= seg b.x - seg.a.x	,	*sdy= seg b.y - seg.a.y*

At last stage of the algorithm, intersected points were filtered so that that buildings that affect the value of openness are the buildings with heights within the height range of the previously defined zones (Table 1).The openness on each intersection point is represented by the ratio of the distance and height and can be calculated as follows:

$$Intersection\ Point\ Openness = \frac{D\ (m)}{D(m) + H\ (m)}$$

Where:

D = distance of the building to the sampling point
H = the building height.

Figure 4: *A less open visual scene (left) versus a more open visual scene (right).*

The final openness values are calculated as the average of all the intersection points openness values by the formula:

$$Sample\ Point\ Openness = \frac{\sum_{i=1}^{n=36}(\frac{D(m)}{D(m)+H(m)})}{36}$$

4.3 Monotony model

Monotony of the built environment was modelled by the amount of land use changes around a particular road segment. The land use can be described as the intended use purpose of a particular piece of land, including residential, industrial or forest area. Based on the input dataset containing polygons with land uses, a buffer was created around each road segment, and an spatial analysis was performed in order to assess whether a polygon overlaps, crosses, intersects or falls completely within this buffer (Figure 5).This approach enables the computation of the amount of land use changes per meter, as well as the amount of distinct land uses per meter for each road segment.

For a more complete picture of the variation in road environment, a more detailed analysis take into account the amount of distinct functions of the buildings that are closest to a road segment. To obtain the desired values, a similar procedure was applied as for computing the amount of distinct land uses per meter (Figure 5). A buffer is created around every road segment, and for each building that intersects or falls within this buffer, the function is evaluated and the amount of distinct functions is counted. The size of this buffer was set to 20 meters, as this ensures that only the first line of buildings is considered. If inside the buffer exist multiple buildings that are placed behind each other, it was not included in the analysis, due to these buildings do not contribute to the experience of monotony of the surrounding environment as they are blocked by the buildings in front.

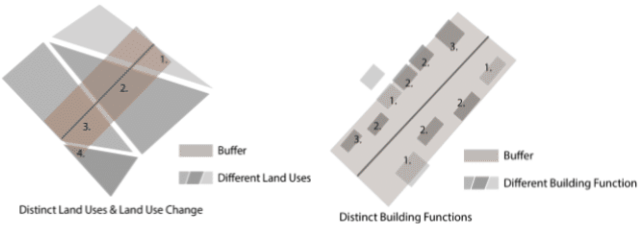


Figure 5 : *Modeling the amount of and use changes (left), and the amount of distinct building functions (right).*

4.4 Statistical analysis

As this research seeks to determine how and to what extent openness and monotony of the built environment affect cyclist route choice, linear regression analyses have been performed. To allow for conclusions on the extent of the influence, the dependent variable for these analyses has been modeled as the divergence from the shortest path, as a percentage of the length of the shortest path. The independent variables describe openness and monotony of the built environment. Finally, the travel time per route, the number of crossings per route, and the percentage of separated bicycle lanes per route have been included in the statistical model as control variables. This allows for a conclusion on the independent effect of openness and monotony, controlled for other potentially influential factors.

To take into account the values of the independent variables for both the shortest path and the observed routes, two separate linear regression analysis have been performed. The first model takes the values for the independent variables that correspond to the shortest paths into account, while the second model includes the differences between the values of the observed route and values of the accompanying shortest path.

5. Results

5.1 Results of modeling Openness and Monotony

The outcome of the openness model results in lower openness values in the builted areas, with the lowest values in the two city centers, Tilburg and Breda. Openness values moderately increase when moving towards the suburbs and the rural areas, getting higher values on the big highways .

About monotony, the variables that are considered are the number of land uses, the unique land uses and the unique building functions,measured per meter for each road segment. The variation in uses is greater within the city centers, expressing lower values for monotony. On the contrary with regions that are distant from the urban region, that have lower variation and consequently higher monotony.

5.2 Results of the statistical analysis

The output of the statistical analysis that stands out most is the insignificant predictive value of the number of distinct land uses per meter (Table 2). Since this coefficient is insignificant, it is statistically not possible to assign a conclusion to the predictive value of the number of distinct land uses per meter. The statistical insignificance of this predictor can be explained by the correlation with the other predictors. For the model with shortest path values and the model with the differences between the observed routes and corresponding shortest paths, the correlation with the variable describing the number of land use changes per meter is 0.975 and 0.982 respectively. Therefore, the values for the one variable almost completely explain the values for the other.

With both models displaying similar results, it is found that openness of the built environment influence the willingness of cyclists to diverge from the shortest path negatively. According to the first model, an increase of the openness by 1% will result in a decrease of 0.670% in the divergence from the shortest path. For the second model, the results can be interpreted in a similar fashion: In this case the coefficient of -3.102 resembles a decrease of 3.102% in the divergence from the shortest path when the difference in openness between the observed route and a shortest path increases by 1%.

Variable	Shortest path	Difference shortest path - observed route
Constant	118.836 (0.000)	23.648 (0.000)
Number of land use changes per meter [n/m]	466.591 (0.000)	-597.444 (0.000)
Number of distinct land uses per meter [n/m]	-102.843 (0.384)	194.140 (0.236)
Number of distinct building functions per meter [n/m]	210.907 (0.021)	-939.023 (0.000)
Openness [%]	-0.670 (0.000)	-3.102 (0.000)
Travel time [seconds]	-0.016 (0.000)	0.156 (0.000)
Number of cross roads per route [n/route]	-0.763 (0.000)	0.149 (0.050)
Percentage of separate bicycle lanes [%]	-4.27 (0.000)	0.438 (0.000)

Table 2: *Output of the linear regression analyses*

Unlike for the openness, the two models show different results for the statistically significant variables describing monotony of the built environment. The output of the first analysis indicates that a higher number of land use changes and a higher number of distinct building functions increase the willingness from cyclists to diverge from the shortest path. However, the analysis that takes the difference between observed routes and shortest paths as input suggests a negative influence of those variables. This can be explained by the fact that for both variables, there are many negative differences.

The question is how representative the established regression models are when it comes to their total predictive value of the divergence, in distance, from the shortest path by cyclists. The explained variance for the regression model based on the values of the shortest path, and for the regression model based on the differences between observed route and shortest path, are 5.8% and 3.6% respectively. This indicates that the established regression models explain 5.6% and 3.6% of the dispersion of the input data. The fact that both models have a relatively low explanatory strength on the amount of divergence from the shortest path is caused by the limited amount of variables that are included in the model. In previous work, many factors have been found influential on cyclist route choice. However, due to time and data limitations only seven predictors could be included in the final regression models. Furthermore, there is many factors that influence human behaviour in general, while people do not always make rational decisions either. Therefore, it can be considered as reasonable that the explanatory strength of both regression models is relatively low.

Interesting from the perspective of this research is the proportion of the effect of the independent variables describing openness and monotony on the explained variance for each of the regression models. Entering openness of the built environment to the model based on the values for the shortest path increases the explained variance by 0.3%, while the explained variance increases by 0.7% when entering openness in the model based on differences between observed routes and shortest paths. Entering the number of land use changes per meter increases the explained by 0.4% and 0.3% respectively, while the number of distinct building functions per meter does not result in extra explained variance for the model based on shortest path values and an increase of 0.3% for the model with differences between observed routes. From these statistics it can be concluded that the predictors describing openness and monotony of the built environment have a relatively small influence on the predictive strength of the regression models.

6. Conclusions

The main objective of this research has been to bridge a gap in existing literature on cyclist travel behaviour, by examining the effect of openness and monotony of the built environment on cyclist route choice. From the results of multiple regression analyses it can be concluded that the openness of the built environment has a negative influence on the divergence from the shortest path. This indicates that cyclists in the sample area experience more open routes as a negative factor and thus prefer ‘less open’ roads. On the contrary, cyclists do prefer routes that have more variation on the surrounding built environment. However, the proportional influence of openness and monotony of the built environment can be considered low in relation to other factors.

7. Recommendations

In this research the effect of openness and monotony of the built environment on the bike route choice has been examined. The influence of these aspects on the bicycle route choice has not been extensively reviewed before. For this purpose, new models describing openness and one the monotony have been developed, and with this development several limitations were discovered along the way. The following recommendations are proposed with respect to the limitations:

- Consideration of the bicycle lane direction.
- Incorporation of more elements of the built environment as vegetation and other possible artefacts.
- Consideration of the real visible angle. The current approach of this research considers all the buildings around a specific point within an angle of 360°, when in reality, only building interfering with the cyclist' visual field should affect the result of openness of a sampling point.
- Aggregation of values about openness and monotony per each route.
- Combination of extra control variables as weather conditions, traffic volumes or facilities on the roads in the applied formulas.
- Result validation by applying the models and the statistical analysis in another sample area.

References

Benedikt, M. L. (1979). To take hold of space: isovists and isovist fields. Environment and Planning B: Planning and design, 6(1), 47-65.

Cichański, A., & Wirwicki, M. (2010). Ergonomics analysis of anthropo-technical system In the environment of CATIA program.". Journal of Polish Cimac, 5(3), 19-25.

Fisher-Gewirtzman, D., & Wagner, I. A. (2003). Spatial openness as a practical metric for evaluating built-up environments. Environment and Planning B: Planning and Design, 30(1), 37-49.

Ministry of Transport, Public Works and Water Management. (2009). Cycling in the Netherlands. Den Haag: Ministry of Transport, Public Works and Water Management.

Oke, T. R. (1981). Canyon geometry and the nocturnal urban heat island: comparison of scale model and field observations. International Journal of Climatology, 1(3), 237-254.

Province of Noord-Brabant (2009). Fiets in de versnelling. Provincie Noord-Brabant.

Thiffault, P. & Bergeron, J. (2003). Monotony of road environment and driver fatigue: a simulation study. Accident Analysis and Prevention, vol. 35, pp. 381-391.

Zhao, X. & Rong, J. (2013). The relationship between Driver Fatigue and Monotonous Road Environment, in: W. Wang and G. Wets, Computational Intelligence for Traffic and Mobility, Atlantis Computational Intelligence Systems, vol. 8, pp. 19-36.

[This page was intentionally left blank]

1

Introduction

With an increasing awareness of sustainability and climate change under the great majority of society worldwide, cycling is a means of transport of increasing relevance (Strauss et al., 2015). Apart from the environmental issues, the shift towards soft mobility, i.e., pedestrian, bicycle transportation, can evidently have a benefit on health (Heinen et al., 2011; Unwin, 1995). Governments and other organizations all over the globe are taking measures to promote the use of bicycles and electric bicycles (Strauss et al., 2015). The Netherlands is a leading country when it comes to the ownership and use of bicycles compared to the size of the population (Ministry of Transport, Public Works and Water Management, 2009). Since The Netherlands covers a relatively small area, with a high density of cities, towns and villages, bicycles and e-bicycles have been recognized as an appropriate means of inter-city transportation by governments on different scale levels within the country (Ministry of Transport, Public Works and Water Management, 2009).

To strengthen this recognition, the Province of Noord-Brabant has established a policy to promote and increase the use of bicycles by, among other measures, the development of high speed bike lanes (Province of Noord-Brabant, 2009). These high speed bike lanes form a connection between the larger cities in Noord-Brabant and should enable fast traveling between these cities. In order for the high speed bike lanes to be a stimulant for bicycle use, the Province of Noord-Brabant wishes to ensure the potential use these lanes, by taking into account factors that affect the cyclist' route choice.

Over the years many studies have been performed on cyclists' travel behaviour and the factors that influence the route choice, pointing out factors like the length of the route, the amount of motorized traffic, the waiting time and the stops (Dill & Gliebe, 2008). In other studies, factors related to road facility and infrastructure are found to be more important, like the accessibility, turn frequency or road intersections, connectivity and slope (Zhao, P. 2014; Handy, S. L., & Xing, Y. 2011; Cervero et al., 2009; Broach et al., 2012) . At the same time it is unclear if the form of the surrounding environment has an influence on bicycle route choice as there are studies with conflicting conclusions on the topic (Moudon et al., 2005; Handy, S. L., & Xing, Y. 2011). Despite the fact that studies give a broad sense on how the road facility and design affect the route choice of cyclists, the contradiction on the possible influence of the configuration of the built environment needs to be more explored.
Prompted from that, this research aims to examine how the configuration of the built environment affect the cyclists' route choice. Existing studies support that the perception of an environment while moving within it is affected by its geometrical properties (Benedikt 1979; Batty 2001). Based on this aspect, the geometry of the built environment is approached by researching the *openness* of it.

The concept of *openness* of the built environment has been treated in previous studies stemming from various fields. The sky view factor is an approach, linked to the field of street climate design, that measures the effects of building geometry on sun radiation in urban canyon environments (Oke, 1981). In the field of architecture, Benedikt (1979) introduced the isovist model, which defines the extent of visible surfaces from a given point. Fisher-Gewirtzman & Wagner (2003) analyse the spatial openness concept and introduce the spatial openness index (SOI). These different concepts are approaching the term openness each one from its own scope. However, the potential influence of openness on cyclist route choice behaviour has remained underexposed so far.

Another aspect that will be examined in this research is that of *monotony* of the built environment, which is a feature that can be affected by the configuration of the environment. Road environment is considered monotonous when the environment remains unchanged or will change in a predictable pattern (Zhao & Rong, 2013). A correlation of the monotony of an environment and the movement within it is already evident in previous research (Thiffault & Bergeron, 2003; Zhao & Rong, 2013). However, its possible influence on cyclist route choice is not recognised yet.

Identifying the influence of the configuration of the built environment on openness and monotony and in extent on cyclist travel choices will contribute to the investigation of suitability of high speed lanes in the province of Noord-Brabant, as well as to the implementation of strategic analysis of cycling facility improvement schemes in the whole province. In an attempt to disclose on how the high speed bike lanes can meet the user requirements, this research seeks to determine the effect of the openness and monotony of the built environment on cyclist route choice in the Province of Noord-Brabant, by answering the following research question:

How do openness and monotony of the built environment affect cyclists' route choice in the Province of Noord-Brabant?

The focus of this research will be on the analysis of secondary data provided by the B-riders,the Province of Noord-Brabant and the Ministry of Infrastructure and Water Management, and is a project of Noord-Brabant citizens that commute with bicycles. The GPS-based data were generated by a group of commuters that travel by electric bicycles. Through literature study, a theoretical framework on openness and monotony of the built environment will be established on which the applied methodology will be based, a necessary step for the establishment of the parameters within the models. The final results will be obtained through a regression analysis, which allows for a prediction of the influence of openness and monotony of the built environment on cyclist travel behaviour.

Section two of this paper provides a theoretical framework that forms the base of this research. Section three elaborates on the study area and gives insight on the data that has been used and the way they have been processed. The applied methodology is discussed in section four, where later the results are presented and discussed in section five. Finally, in section six conclusions are drawn and recommendations for further research are made in section 7.

2

Theoretical framework

This section provides a theoretical framework for this research, by examining previous studies in order to answer a set of subquestions. The subquestions cover diverging aspects that lead to an answer on the main research question. Therefore, the aim of this section is to clarify the relevance of this research, as well as to make an attempt on providing a hypothetical answer to the sub questions and main research question, based on existing literature.

2.1 Importance of openness and monotony of the built environment

Openness and monotony are identified as aspects of the built environment whose influence on cycling activity is still unexplored. Within this section, the relevance of researching the effect of openness and monotony, among other factors, will be discussed by answering the following subquestion:

Why are openness and monotony of the built environment important for the province of Noord-Brabant, in particular with respect to fast bike lanes?

Former studies have identified a set of characteristics that influence the use of bicycles and the choice of bicycle routes. Ipek et al. (2016) found that the travel time and motorized traffic volume are the most important factors in bicycle route choices. Other route attributes with a high impact include the number of stop signs, traffic lights, cross-streets, speed limits, and the existence of continuous bicycle facility on the route. Studies by Akar & Clifton (2009), Hopkinson & Wardman (1996), and Winters et al. (2011) concluded that the main factors influencing route selection are safety-related factors as for instance the proximity of the bicycle-lanes to other means of traffic, separation of bicycle lanes from vehicle roads, risk of injuries through collision with cars or how slippery is the route during the presence of rain or ice. Qing Shen et al. (2014) also relate traffic volume to safety, in the sense that a large number of cyclists or other vehicles may increase the possibilities of accidents. Zimmerman et al. (2016) state that traffic volume (this can be cyclists or other means of transport on shared roads) has a negative correlation with the choice of a certain bicycle route, which is supported by the findings in the other studies. Especially commuters are sensitive to traffic volume and the consistency of traffic (Broach et al., 2012), since they are usually driven by a tighter time schedule than non-commuters. Also, the amount of air pollution is a factor that has been recognized as a major influence on route selection.

Nonetheless, factors that are more closely related with the built environment also play an important role in cyclist route choice. These factors include safety-related variables such as the proximity to a traffic area, presence of bicycle-vehicles shared roads, lighting availability during the entire trip, quality pavement of the routes, bicycle facilities and motorcycle volumes among others. Traffic lights and cross-intersection, indirectly related with travel time, also seem to have an important value for commuters when deciding the preferred route.

Directness and connectivity are also considered as a major influence parameter in the Netherlands. As investigated by Balci (2017), commuters in Utrecht prefer to ride along direct routes avoiding deviation of a road between the end and the start point of the travel. Additionally, the authors state that major roads (more connected roads) affect cyclists route choice preferences in a positive way, as well as they conclude that cyclists do not prefer motorized vehicles shared roads.

As becomes clear from review of the existing literature, many factors that affect the attractiveness of routes for cyclists have been identified. However, both openness and monotony of the built environment in relation to cyclist route choice have not been covered yet. The aim of this research is to bridge the gap in existing literature, by providing the province of Noord-Brabant with predictions on the influence of openness and monotony on the experienced attractiveness of routes by cyclists.

2.2 The built environment

As this research partly focuses on the establishment of a definition of openness and monotony with respect to the built environment, it is in the first place necessary to define the built environment by itself. For the definition of the built environment, a distinction is made between urban and rural areas, as the characteristics of both types of areas differ significantly. Therefore, the following subquestion will be answered in this section:

What is the built environment in urban and rural areas?

The concept of the built environment refers to the design, construction, management and use of all the infrastructure that is man-made, as long as it has relationship to human activities over time. Generally, it encompasses places and spaces that are created by people and in a more abstract view to any physical adjustment of the natural environment through human activities (Lawrence, 1990). Regardless of the scale, all features that constitute the built environment covers elements or combinations of density, diversity and design, that are the three most significant dimensions of it (Cervero, 1997).

The built environment also includes spaces that have no boundaries and are not enclosed such as the uncovered areas in public squares, parks or streets. Recently, the first definition is expanded by including also different factors that are significant elements of the term as walkability, bikeability, mental health, community gardens.

2.3 Openness of the built environment

Throughout the literature about cyclist route choice, many factors have been reported that affect the decision of a cyclist to select one route over another. However, spatial openness has not yet been related to cyclist route choice in such a manner. As this research attempts to bridge that gap in existing literature, the following subquestion has been developed to define the concept of openness with respect to the built environment:

How is openness of the built environment defined in existing literature?

In a simplistic manner it is possible to say that openness implies the extent of open scene above and around a specific point, the sense of an open-free or close-narrow urban environment. The term of openness has been approached throughout the literature in various ways and research comes from different disciplines. It can be related to what Benedikt (1979) defines as *isovist.* In order to define the isovist, Benedikt considers all the surfaces that are included in a connected region bounded by a smooth convex boundary. In this region, he considers the spatial arrangement of the surfaces and any change in their position defines a new environment. “The isovist is the set of all visible points from a given vantage point in space and with respect to an environment”. The shape and size of the isovist depends on the selection of the vantage point, as its position defines the visible region of this point. As a next step, Benedikt defines the *isovist fields* which are all the isovists that belong to a given path. The isovist approach can be used both for 2D and 3D space. Benedict takes this approach a step further and considers straight lines from the vantage point that radiate towards the boundary of the surfaces inside the isovist environment. The length of these line segments is specified by the coordinates of the vantage point and the point on the boundary. Though, Benedikt does not define how far or close a surface should be in order to be inside his environment.

In a different approach, stemming from the field of street climate design, Oke (1981) used the notion of *sky view factor* to study the effects of building geometry on sun radiation in urban canyon environments. The sky view factor is used to measure the amount of sky that is visible from a certain point in the middle of the canyon. In order to measure this factor Oke uses the ratio of the building height and the street width at a specific point. Based on that, Oke (1988) recognises two structures of urban geometries, the *shelter* (narrow streets and compact built environment) and *dispersion* (low building density and separation), where shelter covers from wind and cold while dispersion protects from pollutants and sunlight access.

Later in their research, Fisher-Gewirtzman & Wagner (2003) analyse the spatial openness concept and introduce the *spatial openness index (SOI)* to measure the volume of open space potentially seen from a given point. To compute the SOI, Fisher-Gewirtzman & Wagner consider the world as a part of a 3D integer grid. The SOI inputs are: the open space *S*, as a subset of 3D space, a set of built volumes *B={b1,...,bn}* and a function that defines for every cube of the grid the openness value. The result is given by the number of grid points in the open space *S* that are visible to the center of the cube *c*. The visibility of a grid point is defined by a boolean function of 0 or 1 where one is the value if a grid point is visible from the cube *c* and 0 if it is not. It resembles with Benedikt’s isovist, but there is a possibility to introduce correction factors, such as weights, through which the openness is expressed not only as a visual factor but also as a factor of qualitative attributes such as natural light, air and near or distant views.

The visual environment is connected with the satisfaction of people with regard to their surroundings. This is supported by the findings of Hur et al. (2010) which among others support that the satisfaction of residents is associated with the perceived openness and the physical measures. The term of openness of the built environment has not been correlated with the effect it may imply on the route selection. Aiming to explore the effect of this factor to the commuters' routes choice, the term is further analysed and some measures for applying openness in the study area are defined.

The openness of a cyclist's view is affected by the existence of surroundings of the built environment. Component of the built environment is everything that is man-made, like the buildings, the road and transportation networks and the infrastructures. Nevertheless, not all of these components of the built environment can affect the openness factor. Thus, for this research, the built environment was considered as "*everything that is a more-than-2D object that is higher than the eye level and can interfere with the cyclists' line of sight"*. Components of our built environment could be buildings, bridges, large vehicles (i.e. buses, trucks). At this point it is important to note that even though trees are clearly objects that affect openness, they were considered as part of natural environment and not of the built environment. Accounting for the results of previous studies, openness of the built environment was defined as follows: *the extent of open scene above and around a specific point. In the built environment an environment is considered as open when no obstacles of the built environment are interfering with the visual scene of a person, or the interference can considered to be low.*

2.4 Monotony in the built environment

Like openness, monotony in the built environment is related to the visual experience of cyclists while cycling. Previous studies have covered monotonous road environments, and the effect on drivers' fatigue (Thiffault & Bergeron, 2003; Zhao & Rong, 2013). Accounting for the findings within these studies, a definition of monotony in the built environment can be formed by answering the following subquestion:

How is monotony in the built environment defined in existing literature?

Thiffault & Bergeron (2003) approach monotony in general as a set of sensory stimuli for the human brain. They describe a situation as monotonous when the stimuli remain unchanged or change in a predictable way. With respect to the built environment, this could be understood as predictable changes in the elements that form the built environment, or no change at all. In a similar fashion, monotony of road environments is treated by Zhao & Rong (2013). According to their study, a road environment is considered as monotonous when the environment remains unchanged or will change in a predictable pattern.

Based on the studies by Thiffault & Bergeron (2003) and Zhao & Rong (2013), and taking into account the definition of the built environment, a definition of monotony in the built environment can be established. Since this research focuses on cyclists, and therefore encompasses a transportation problem, the emphasis is on how cyclists experience the built environment during their trip, and not on one particular location. Therefore, monotony in the built environment can be defined as *the extent of visual variation in elements that form the built environment for a sequence of locations.*

2.5 Hypotheses

By answering the subquestion, a theoretical background has been established in order to fulfil the main objective of this section: providing a hypothetical answer to the main research question. Although openness of the built environment has been researched in previous studies, a direct link to cyclist route choice is still missing. Therefore, the hypothesis on the effect of openness on cyclist route choice will be based on the perception of openness of surroundings by human beings in general. Based on previous researches, it is expected that more open roads will be more preferred by cyclists, confirming with the findings of Hur et al. (2010) on the satisfaction of residents in a certain area. However, forming a hypothetical conclusion on the magnitude of the effect of openness on route choice cannot be drawn based on existing studies.

For the monotony of the built environment, findings in existing literature are more relatable to transportation problems in general. As previous studies state that monotonous road environments increase the fatigue of drivers while being on the road (Thiffault & Bergeron, 2003; Zhao & Rong, 2013), it is assumed that more variation will have a higher preference among cyclists and that monotony is therefore can be a significant explanatory factor for the selection of a route.

However, as can be derived from existing literature, the experience of satisfaction by both openness and monotony of the built environment depends on personal preferences of individuals. Therefore, the expectation is that it will be complex to develop strong predicting coefficients for both concepts.

3

Data description and preparation

This section will introduce the study area and the further selection of the sample area. Additionally, the data that has been used to develop the desired models will be presented. Furthermore, the storage and pre-processing of the data will be discussed in this section, as well as the selection of the sample area.

3.1 Study area

Traditionally, The Netherlands has been a country where cycling is a main means of transportation. Over the years, a well-structured and elaborate infrastructure of bicycle lanes has been developed, and improvements are being made every day (Province of Noord-Brabant, 2009). The province of Noord-Brabant has recognized cycling not only as a means of transport for short distances, but also for regular intercity transportation. In their policy on bicycle use, 'fiets in de versnelling' (Province of Noord-Brabant, 2009), the province of Noord-Brabant provides a vision and action program until 2020. This document is clearing in which ways the province wants stimulate bicycle use, to participate in the improvement of the accessibility, quality of life and health. The main actions are three topics: increase comfort and ease, seduce travellers to take the bicycle, and reinforce each. To achieve increase in comfort and ease for cyclists during intercity travel, the ambition for 2020 is to realise multiple high speed bike lanes and research other possibilities for a faster network. The idea behind the high speed bike lanes is to improve the connections between the larger cities in the province, and to make it more interesting for travellers to take the bicycle (Province of Noord-Brabant), as displayed in Figure 3.1.

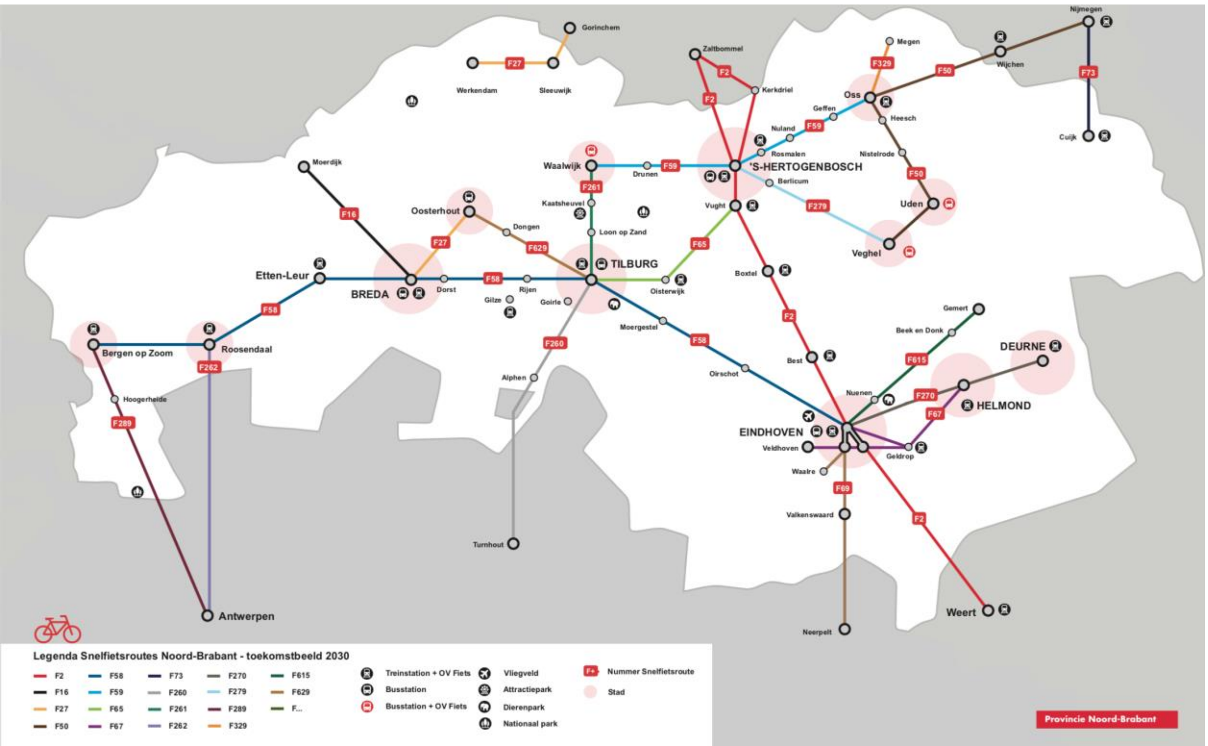


Figure 3.1. *Fast bike lanes Noord-Brabant future vision 2030 (Noord-Brabant, 2017)*

The province of Noord-Brabant is located in the south of the Netherlands, and consists of sixty-six municipalities that form an area covered by built environment as well as rural areas. For the purpose of data manageability the decision has been made to focus on a smaller sample area, instead of applying the methodology on the entire province of Noord-Brabant. The sample area has been selected around Breda and Tilburg, two cities that are located in each others proximity, and their near surroundings. The selection of the area is based on the fact that the two municipalities are among the biggest of the province, meaning that they generate a considerable number of daily travels. Additionally, considering the fact that they are neighboring municipalities, this selection allows us to research both urban environment and the rural areas among and around the municipalities.

3.2 Data

The datasets used for this project are secondary data about cyclists' routes and network infrastructure provided by B-riders and Fietsersbond. The main dataset about cyclists *GPS_match* was provided through a csv file with distinction between the different users (userid), and the routes (routeid) of the citizens of Noord Brabant for the year 2014. Additional dataset, *GPS_points* a point shapefile with raw data of multiple GPS tracking points (approximately 45.5 million points) representing the gps location of each cyclist every few seconds for the whole extent of the Netherlands.

The data for the bicycle lane network consists of two shapefiles (*links* and *fietsersbond*) of linear geometry covering the whole extent of the Netherlands. In order to connect the users' trips with the bicycle lanes two joins on the datasets were performed. First, the linear shapefile *links* representing lines from B-riders was connected with the *fietsersbond* using as common keys the *source* and *target* point of the cyclist. This join had as purpose to keep the union of the geometries of the two shapefiles and ad the same time to collect all the information of the attributes in one dataset. Second join was performed on the output shapefile of the first join with the csv file *GPS_match* using as common attribute the distinct *linknummer* from both files. The shapefile with the GPS tracking points was joined with the rest of the files based on routeid, another unique parameter of the datasets (Figure 3.2).

The datasets were clipped with the bounding box of the sample area resulting to 53.135 routes of a total amount of 593 total users. The average number of routes per user in the sample area is 90 for a time period of one year.

The calculations of the built environment parameters are based on spatial data acquired from the Dutch Kadaster. The shapefile used contains polygons of all the buildings with the required information of the heights. The Dutch Kadaster combined their buildings data with the data from the Dutch AHN. The AHN (Actueel Hoogtebestand Nederland) is the country wide lidar laser scan of the Netherlands. By merging them with the buildings polygons, all buildings contain height information such as median height, maximum height and minimum height. The result is the merged 3D buildings dataset, used in this research to derive the building height.

Finally, a dataset containing the land use for each parcel within The Netherlands, provided by Dutch Statistics, has been used for the computation of values to express the monotony of the built environment.

Based on these datasets, new data has been generated in order to perform the analysis: the shortest distance from the road segment[1] to the buildings, the average height of each building and the number of different land uses in a specified area. These variables are used to develop the models about openness and monotony.

---

[1] *Road segment represents part of the geometry of a line on the road network and more specifically, refers to a part of a line that is bounded by two distinct endpoints.*

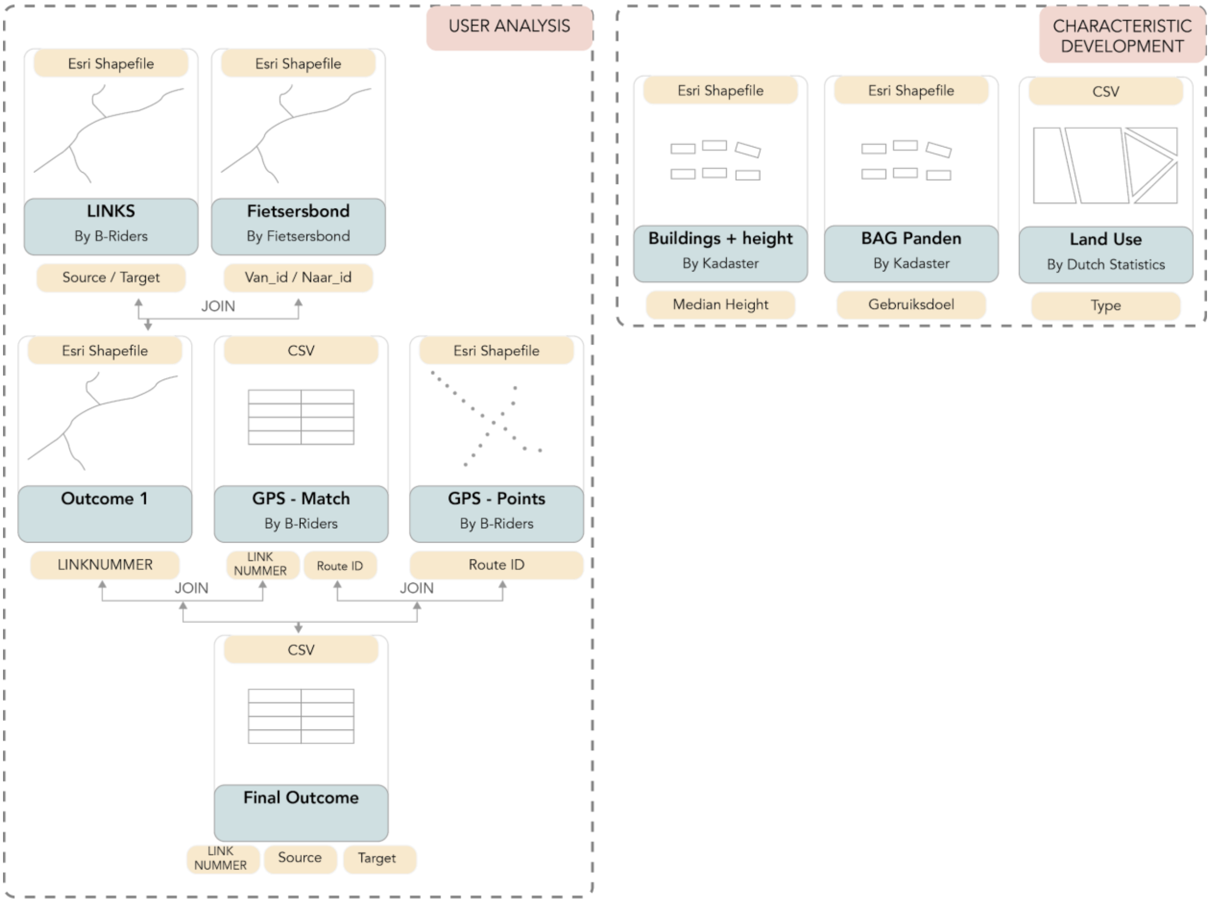


Figure 3.2. *Dataset connections*

3.3 Software Specification

For the implementation of the project multiple tools were used for analysis, computation and visualization including QGIS, FME, Python, PostgreSQL and SPSS.

More specifically, *Python* is an open source programming language, used to clean the raw data from outliers and to perform most of the calculations involved in the base route and openness model. For the coding requirements of this project the external libraries were installed in python. *Pandas* for data manipulation and analysis, *shapefile* and *shapely* for shapefile handling and *math* for mathematical operations. Additionally, the program was used as a tool for connection to the DBMS server by using the module *Psycopg2.*

PostgreSQL is an open source object-relational database system. It was used as the database management system of the project, allowing data handling between several datasets and multiple non-spatial and spatial operations using the PostGIS extension.
It was useful for generation of relational outputs like csv files, and spatial outputs that included geometry properties. Additionally, it served as the main tool for connections of datasets through different systems including: QGIS - PostgreSQL and PostgreSQL- Python (Figure 3.3).

PGRouting is an open source library, that works as an extension of PostgreSQL to provide geospatial routing functionality. In conjunction with python, it was used for the route modeling of the network: shortest paths, crossings, path distances and travel times.

QGIS is a geographic information system, open source tool, utilised for analysing and editing the datasets and was the only tool used for visualization. The capability of interconnecting PostgreSQL with QGIS, made all the procedures are less time-consuming while less storage was needed.

Feature Manipulation Engine mostly known as *FME* is a platform for translation of spatial data between geometric and digital formats. FME was used complementary to GIS software to perform processes like the reprojection of the datasets and for some spatial relation of data for the base route model.

SPSS is a software package operated for statistical analysis. The provided possibilities of the software were used in the final step of the project, to perform a statistical analysis on the output variables and to extract the final results.

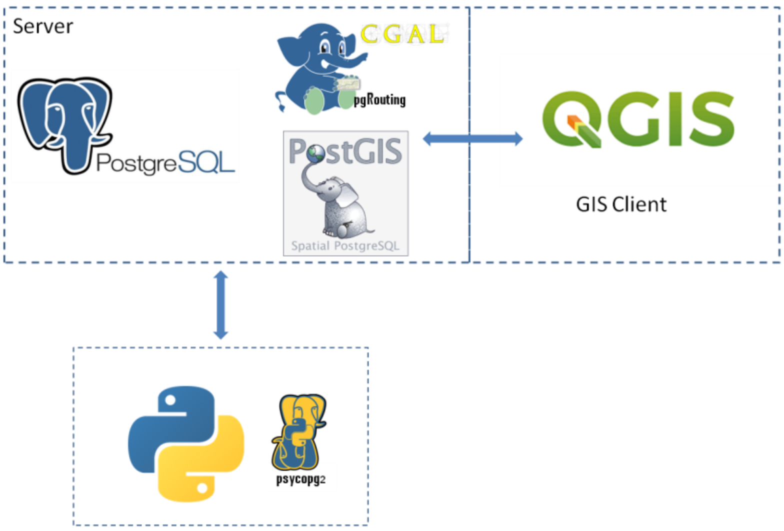


Figure 3.3. *Connections between different storage and processing tools*

3.4 Data storage

Since most of the dataset contains geometry information as GPS tracking points, or more complex structures as the linear networks of the roads, spatial operations need to be carried out for the data processing phase (the storage part in Figure 3.4). To deal with these operations, the data was stored in the open source database system PostgreSQL, using the spatial extension PostGIS in order to add support to geographic objects.

The complexity of the storage process is mainly based on the big amount of data to be handled (14GB). Such databases require additional functionality to process spatial data types efficiently. Different strategies were applied in order to reduce cost of processing time and improve the computations performance:

- *Standard indexing with data structure B-tree*: the most common default access method in database systems for indexing standard types as numbers, strings and dates. This indexing strategy was implemented in order to speed up the searching by organizing the data into a search tree which can be quickly traversed to find a particular record. It created a hierarchical tree based on the values of the column being indexed (*userid, routeid*).

- *Spatial indexing with data structure R-tree:* a common data structure widely implemented in various database systems including postgreSQL and Oracle spatial. Once the data was loaded into spatial tables, the R-tree spatial index allowed efficient access to the data avoiding *sequential scan* of every record in the database. As spatial indexes are unable to index the geometric features themselves, it indexes the bounding boxes of the features. This approach was implemented in all the datasets containing geometries, reducing the query processing time up to 60%, especially when determining relationship between geometries (i.e spatial function as `ST_Intersects, ST_DWithin, ST_Contain)`.

- *Optimizing joining operations*: a variety of joining operations could be performed within the database. However, one strategy for improving query performance was to to avoid as much as possible the use of large amount of information within the same query. One simple way to deal with this issue was to create materialized views (snapshot of a query saved into a table) avoiding long nested queries operations; as this view actually is a table, it was also possible to create indexes to speed up the searching operations. Additionally, another approach to reduce the number of processing elements was to generate logically equivalent expressions that returns same results at a less possible cost; for instance, performing queries using the indexed routeid instead of GPS tracking point (unique id), reduced the query processing time by more than 1000%.

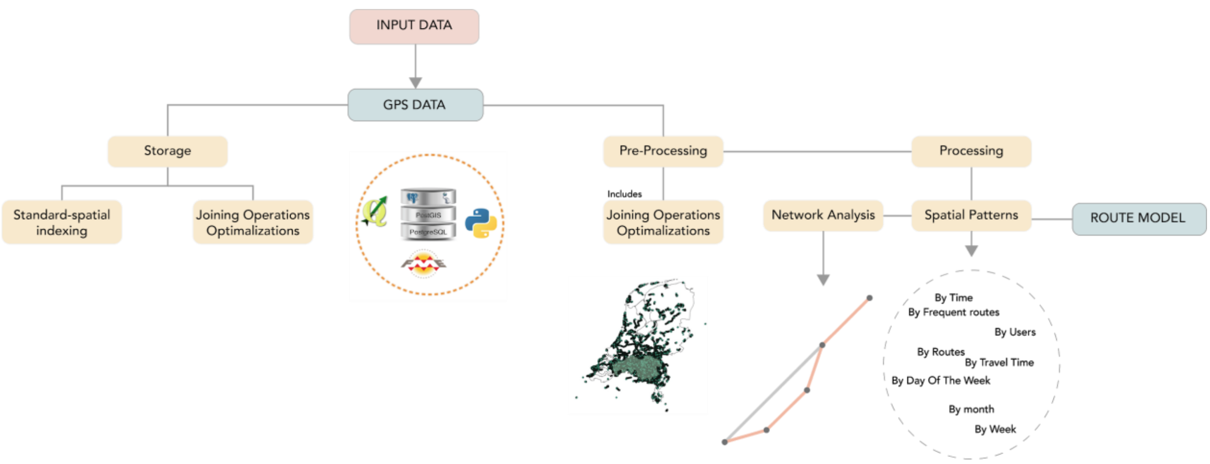


Figure 3.4. *Schematic diagram for the data storage, pre-processing and processing phase.*

3.5 Pre-processing of the data

The preprocessing of the data was necessary in order to fit with each other, decrease the computation time, and be able to extract information out of them (the pre-processing part of Figure 3.4). Part of the provided datasets was processed and ready to use. Though, actions like clipping with the border of the study area, setting a common coordinate reference system or cleaning out unusable information have been performed.

3.5.1 Determining a suitable Coordinate Reference system

The different datasets were provided in two different coordinates systems: Amersfoort RD/New, EPSG 28992 and WGS84/ EPSG 4326. As multiple datasets have to be combined to obtain the desired information, the data should be stored in a common coordinate reference system (CRS) to enable spatial analysis. When defining the most suitable CRS, two main parameters were considered:

- *The scale and extent of the data and stakeholders' needs:* since the data concerns to the Province of Noord-Brabant, the official coordinate system of the Netherlands (Amersfoort RD/New, EPSG 28992) will be the most suitable for the project; however, depending on the client's needs, others CRS could be used.
- *Running spatial analysis:* as postGIS spatial operations were used for calculating distances, it was necessary to use WGS84, EPSG 4326. The spatial function st_length_spheroid, included within pgRouting, calculates the 2D length of a geometry on an ellipsoid, thus coordinates of the geometry should be in longitude/latitude.

As these two considerations derived in discrepancies when trying to determine a common reference system, it was decided to use both, RD Amersfoort and WGS84.The first, mostly used for the analysis phase including the openness and monotony models, as well as for the output files. The second one, more suitable for the base route model, meaning that back and forth transformations should be performed. For reprojecting the input datasets, the *Reprojector* transformer from FME was used to change feature coordinates of input layers from the source to the target CRS. After performing all the spatial operations, all the output geometry layers were converted into Amersfoort RD New, EPSG 28992, using the same transformer.

4

Methodology

This section outlines the methodology that has been applied for modeling purposes within this research. Generally, the methodology consists of the development of four different models (Figure 4.1): a base route model, models for openness and monotony of the built environment and a statistical model that will provide the output on which conclusions can be drawn to answer the main research question.

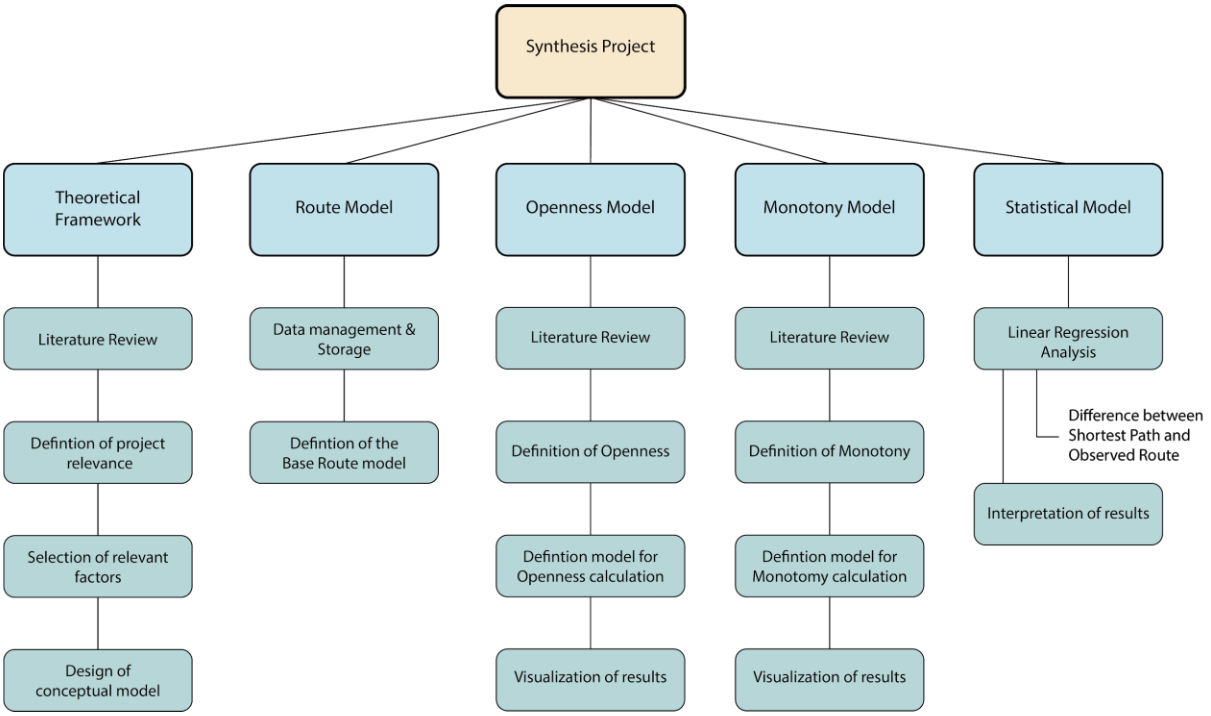


Figure 4.1. *Schematic overview of the research steps.*

4.1 Base route model

The processing of the data is represented by the base route model and it basically includes two main parts: The *shortest path analysis* of the bicycle lanes and the *routes analysis* for understanding the spatial pattern of cyclists based on the input GPS tracking data (in the processing part of Figure 3.4).

4.1.1 Shortest-Path Analysis

As part of the base route model, the shortest path model from origin to destinations was generated based on the A*algorithm. In overall terms, the development of the model included the following spatial and mathematical operations:

- *Routing topology network:* a graph based on edges and vertices was built from the bicycle road network. For any given edge in the road network, the ends of that edge were connected to a unique node which at the same time is connected to other edges of the network (Figure 4.2). This operation generates two main tables: an edge-table with source and target attributes with the ids of the vertices of the segments as well as the associated vertices table containing detailed information as potential gap problems and dead ends.

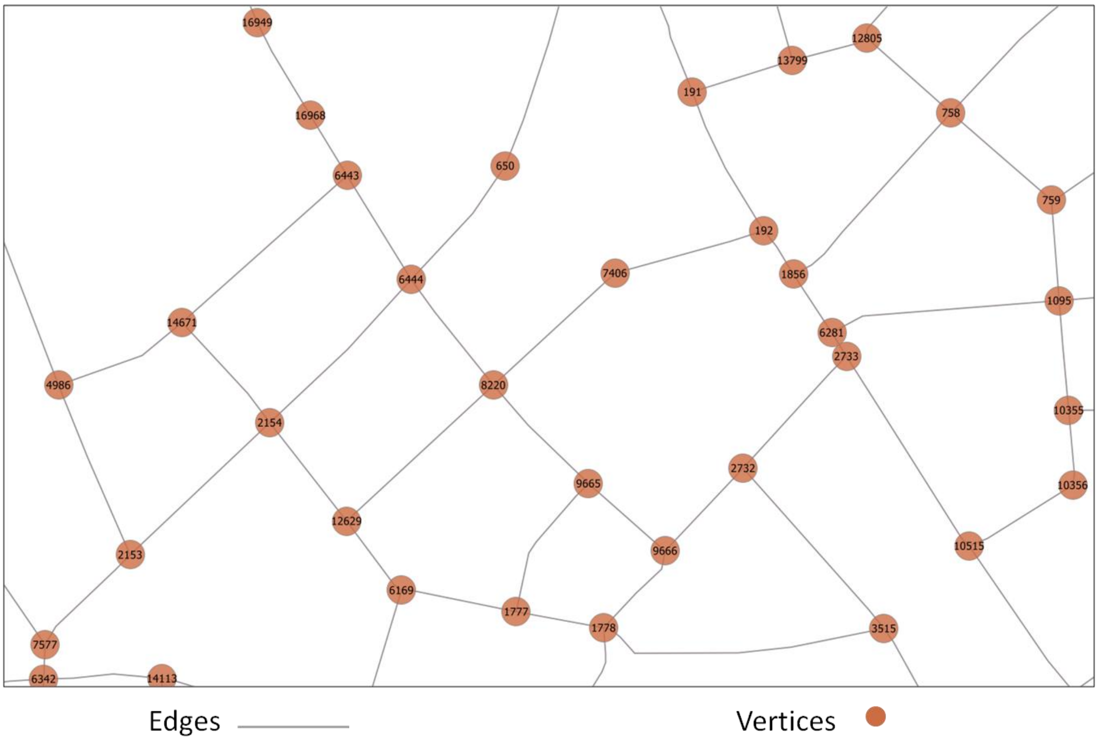


Figure 4.2. *Topology network.*

- *Topology checking:* validation of the geodatabase topology was carried out using pgrouting functions and queries to detect, repair and eliminate errors generated during the routing operation as road segments without starting or ending nodes, not connected nodes, crossing edges, isolated roads and dead-ends (Figure 4.3).

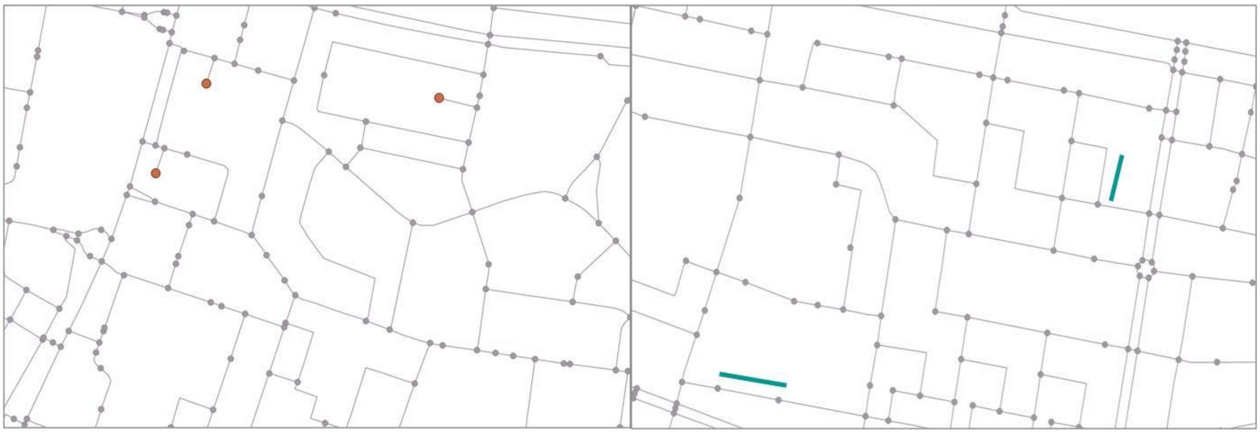


Figure 4.3. *Topological errors: dead-ends (left), and isolated edges (right).*

- *Edge length computation:* distances between vertices were calculated using PostGIS spatial analysis functions that consider the curvature of the earth by including the spheroid associated with the used CRS. Since the input coordinates of the function should be longitude/latitude, in this case the geometry was stored and operated in EPSG 4326, for which the WGS84 spheroid was included in the calculations.

- *Shortest path*: After obtaining a cleaned topology network with the edges lengths, the path finding A* algorithm (with cost distances) from Pgrouting was used to find the shortest path from starting to ending vertices in the bicycle network (Figure 4.4, appendix F). A* was a better approach than Dijkstra algorithm as it doesn't have to visit all the vertices in the network, thus avoiding unnecessary processing. Travels made by cyclists were obtained as routes between origin (source) and destination (target) locations derived from the datestamps of the measurements (appendix A). Using a python script to iterate per every route and invoke A* algorithm, the shortest path was calculated for every pair source – target, and list of road segments per paths were generated and later compared with the road segments of the observed routes.

- *Cross-roads:* from the topology network, road intersections were found based on the number of edges connected with every vertex, thus if a vertex has more than three edges it was considered as a crossing point. Number of intersections were calculated per every shortest path and *routeid,* being later included as control variables for the statistical model.

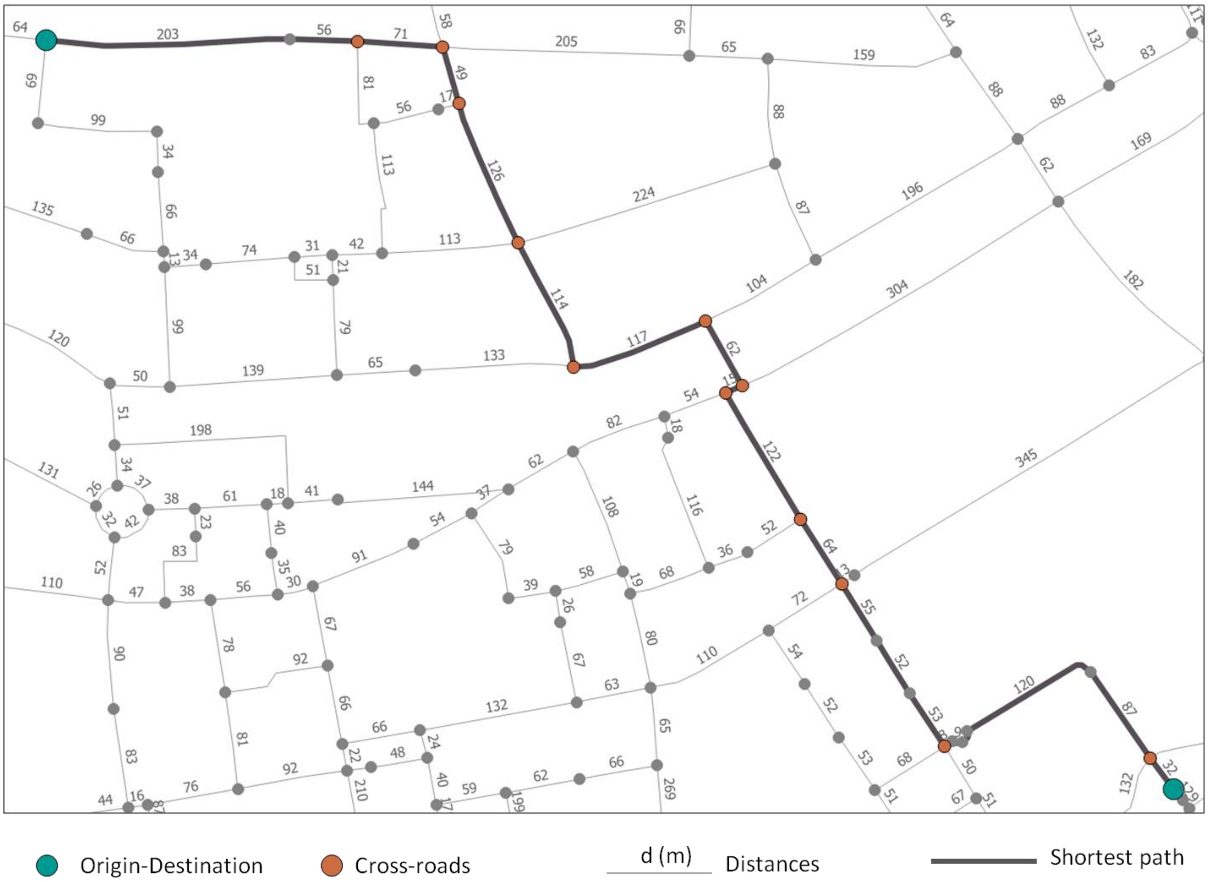


Figure 4.4. *A* shortest path example in the road network.*

4.1.2 Routes intensity and travel times

In order to understand the cyclists' preferences on the bicycle lane network, it was crucial to analyze the spatial patterns movements based on the GPS tracking data. First of all, the GPS dataset was matched with the bicycle roads network. By joining the *routeid* with the roads id's *linknummer*, the number of travels made by cyclists over the bicycle lanes could be calculated. This query operation generated a GPS intensity table that contains the *linknummer* column and the frequency of travels for every road segment. The output map was visualized using a QGIS graduated styling by size and color; the thickness of the lines shows road trip frequency during one year (thicker: higher frequency) and the color lines categorize three main classes: light green for low intensity (up to 300 trips), grey for medium intensity (301 to 1200 trips), and orange for high intensity (up to 4700 trips). The map clearly highlights the preferences of the user group on the bicycle road network in the sample area (Figure 4.5, appendix E).

Figure 4.5. Visualisation of Intensity of travels within a year. A more detailed version of this map is to be found at appendix E.

The theoretical travel times per road segments were calculated based on the average speed provided within the GPS measurements and the edge lengths generated from the routing topology network. This average travel time per link was the base for calculating the travel times per shortest paths and observed route (appendix A).

4.2 Modeling openness of the built environment

Openness is one of the two main subjects that are considered part of attractiveness within this research. Openness for the built environment is defined as: *the extent of open scene above and around a specific point. In the built environment an environment is considered open when no obstacles of the built environment are interfering with the visual scene of a person, or the interference can considered to be low.*

The openness on a specific location is affected by the configuration of the buildings in the neighborhood of that location. More specifically, the openness of a point is affected by the distance of the building to that point, the height of the building and the width of the building. When talking about a cyclist, the openness of a location is translated into the visual field of the cyclist to that location. The distance of the buildings from a cyclist, affects the visual field, depending on how far or close a building is from this person. This distance is defined by the proximity of the buildings of the area to this road. The main idea is that the the environment is considered more open when the width of the street is larger and the building height is lower (Figure 4.6).

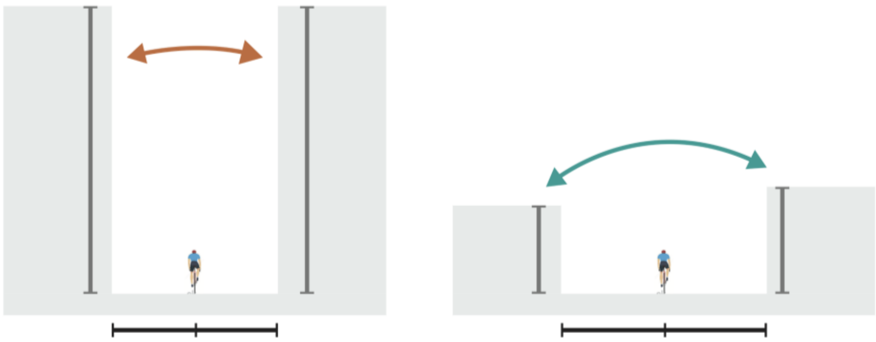


Figure 4.6. *Lower openness value (left), higher openness value (right).*

Besides this qualitative description of openness, a quantitative measure needs to be established by answering the following research subquestion:

How can the openness of the built environment be quantitatively measured using built environment data?

4.2.1 Background

Many elements of the built environment affect the value of openness of a specific point. In the scope of this research, in order to simplify the input variables, as obstacles of a cyclist's view only the surrounding buildings are considered and the trees or other possible attributes of the built environment are excluded. In this context, for the openness calculation, the way that the variables of height and width of a building as well as its distance to a point affect the visual field of a person and in extension the value of openness must be defined.

The higher the building and the shortest its distance from a point, the lower the openness value. For the definition of the maximum distance that a building should be in order to be considered inside a person's visual field, anthropometric data are used (appendix B, Figure B1) for the vertical field of view. More specifically, the methodology applied uses the angle of 15° starting from the horizontal line of sight, to the vertical direction. This is the angle of view of a person driving looking straight. The next step is to find how far a building must be in order to be visually evident using the 15° angle (appendix B, Figure B2). To specify this distance, the building heights of the area were considered as reference value. The building heights were classified into four classes and the right limit of each class was used as the reference building height. Using the tan15° the formula $distance = \frac{H}{tan15°}$ results in four distances that represent the size of the zones (Figure 4.7). The classes of building heights and the respective zones are shown in Table 4.1 and their visual representation in Figure 4.8. The resulting distances form the zones of influence of a building to the cyclist according to its height. This approach indicates that the variables of distance and building height are interrelated in a way that if a building is included in a person's sight view depends on the distance of the building to that person but also from the building height.

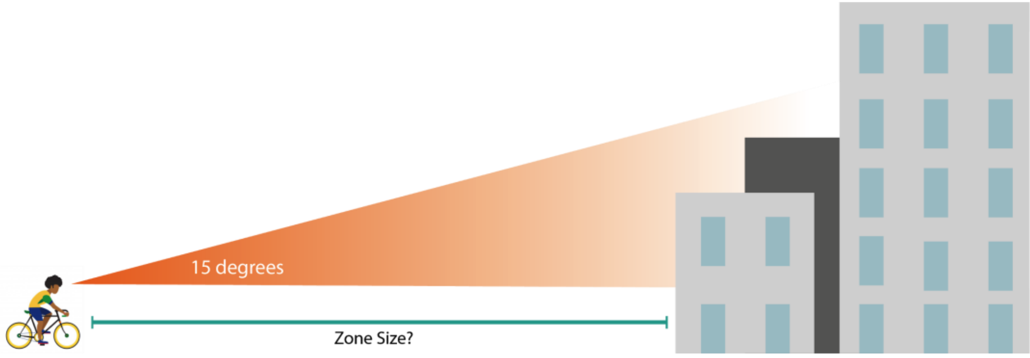


Figure 4.7. *Distance corresponding to zone size*

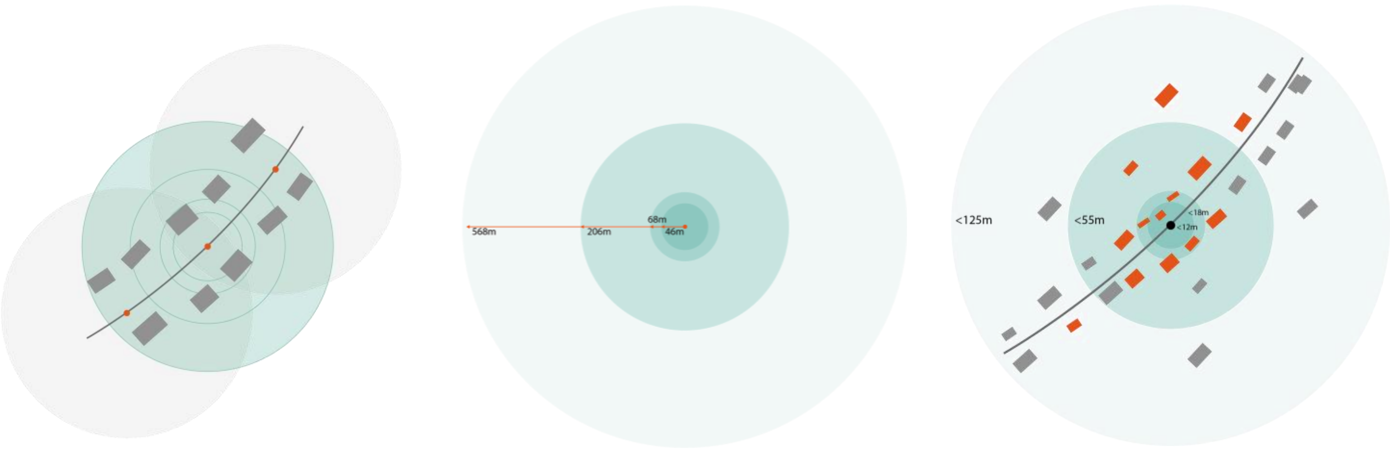


Figure 4.8. *Resulting zones. Left: zones are considered for every sample point. Middle: radius of each zone. Right: Maximum building height in zone.*
The definition of the zones indicates that buildings with height be visually evident, while objects in larger distance may be potentially noticeable or insignificant.

Building height classes[2] (m)	Resulting zones (m)
[-0.8-12.3]	45.52
(12.3-18.4]	68.28
(18.4-55.2]	205.59
(55.2-125.25]	568

Table 4.1. *Building height classes and resulting zones*

[2] *classification with Natural Breaks (Jenks): the variance within each class is minimal while the variance between classes is maximal (QGIS documentation)*

The variable of building width is implicitly incorporated in the processing algorithm of openness by considering lines towards the buildings all around the cyclist, with interval angle of 10°. The algorithm considers only the visible face of a building from a specific point.

4.2.2 Towards the quantification of openness

The quantification of openness is based on combination of methods found on literature. From Benedikt's (1979) isovists, the logic of finding all visible points from a given vantage point was borrowed. To enhance this, the zones defined in the previous section were used to introduce the human visual capability and to have a restriction on the buildings that should be considered as affecting the openness of a point. Additionally, the ratio of the building height to the street width was used for in the *sky view factor* but re-adjusted to fit to this research. The re-adjusted formula differentiates from the sky view factor in the part that instead of using the width street as distance, the actual distance of a building to a specific point was used. For the quantification of openness, the main procedures are:

1. Sample points

The measurement of openness is conducted on each point of a set of sample points distributed every 20 meters along the bicycle lane network of the study area (Figure 4.9). The distance of 20 meters interval is selected so that the area is covered sufficiently and at the same time the computation time is efficient. The total 113.688 points are created on every road segment, based on distinct *linknummer* of the bike lane network dataset.

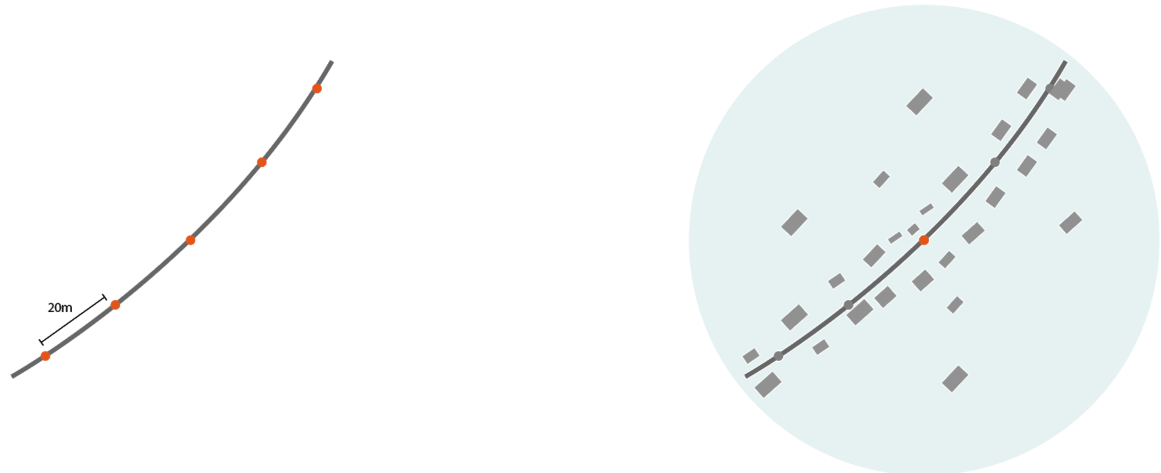


Figure 4.9. *Left: Sample points on the road. Right: Building neighbours.*

2. *Implementation process*

The implementation process for the computation of openness values involved three main python programs: first one, for finding point neighbours to every sample point, the second for retrieving the point intersections towards the buildings, and the final one for applying the openness formula and calculating the openness values for every point on the road segment.

Figure 4.10. *Implementation of the openness calculation. The figure is a caption from QGIS working space showing the sample points along the bicycle lane network. It demonstrates for one sample point, the resulting casted rays and intersection points as produced from the code in python. The openness value of the example point is only affected by the buildings within the three out of four total zones. The whole process is presented with step by step illustrations (Figure 4.11 until Figure 4.13)*

Building Neighbours

For optimization purposes, as a first stage of the openness calculation, buildings neighbours were found based on the nearest neighbours algorithm given a certain distance (fixed distance neighbourhood algorithm, appendix G). Two search criteria were used for performing this computation: first, based on the distance from the sample point to the building (250 m) and second, based on a further distance (600m) relative to the heights of the neighbours buildings (> 50m height).

Points intersections

The methodology followed to calculate the point intersections was based on the *sight-and-light* algorithm (NCase, 2018). To start the computation in python, the input sample points *(point_id, linknummer*) and the buildings polygons (*building_id, heights,centroids*) were retrieved from the DMBS server using SQL language (appendix H). Then the following steps are followed.

- *Sample points* coordinates file was read and parsed into a list of tuples [(x0, y0),(x1, y1), (x2, y2), …, (xn-1, yn-1)] with each tuple containing one sample point's x and y coordinates. Additionally to points coordinates, the point ID and the associated *linknumme*r were also included as part of the information retrieved from the points.

- *Buildings polygons* were parsed as list of lists, with each list containing a list of tuples with coordinates vertices of the polygons, height values (m) and the building IDs:

  [[(x0, y0), (x1, y1), (x2, y2), …, (xi-1,yi-1) ,*h1,id1*] , [(x0, y0), (x1, y1), (x2, y2), …, (xi-1, yi-1), *h2,id2*], ……., [(x0, y0), (x1, y1), (x2, y2), …, (xi-1, yi-1),*hn*,*idn*]]

- In addition, the *neighbours* to each sample points were read and parsed from the building neighbours python script as a list of points IDs with their neighbours buildings IDs.

The implementation starts by casting rays from the point towards the buildings within its vicinity (Figure 4.8, Figure 4.11). Since the the length of an arc depends on the radius of a circle and the central angle, for an angle equal to 360 degrees (2π), the arc length is equal to circumference so that it could be calculated as:

$$L = r \cdot \alpha$$

Where:
r = *radius*
α = *angle*

That means that the arcle angle (angle coverage) between the lines will increase progressively as the distance increase. In order to have a reasonable balance between coverage and computational cost, 10° angle was chosen as a suitable value for getting accurate results, having in mind that the probability of incidence will be lower at higher distances, as in the case of 100 m height buildings located 500 m away from the cyclist position.Thus, with a chosen angles of 10° and 360° of incidence, a total of 36 rays were casted for every sample point on the road network.

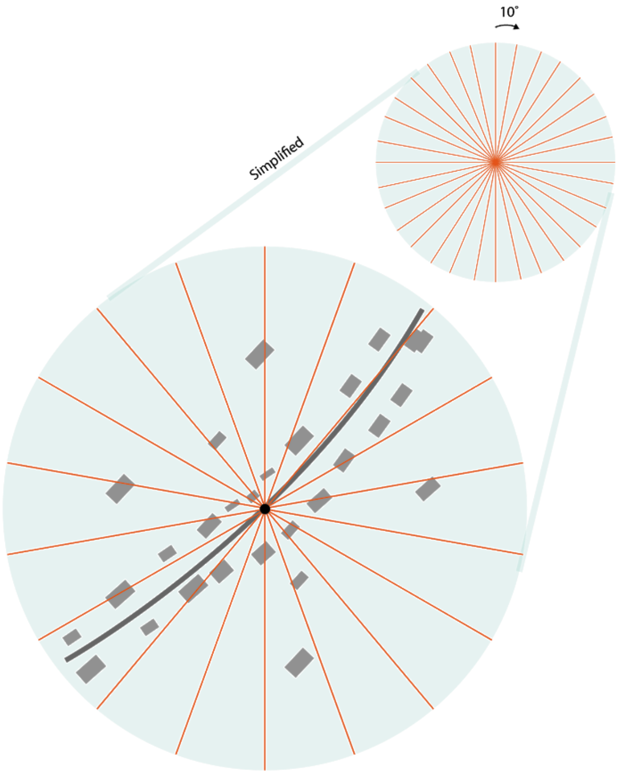


Figure 4.11. *Rays casted from one sampling point*

The implemented algorithm is based on operations with parametric equations, thus to find intersection between the ray and all the line segments, both rays and building segments were written as parametric form of the line (Wikipedia ,2018):

$$x = x_0 + \alpha \cdot t$$
$$y = y_0 + \beta \cdot t$$

Being the two coordinates x,y represented as functions of the same independent variable *t*. After setting the parametric equation for rays and line segments, the following equations were found:

*Ray X = r_px+r_dx*T1*
*Ray Y = r_py+r_dy*T1*

*Segment X = s_px+s_dx*T2*
*Segment Y = s_py+s_dy*T2*

3

Ray(sample points)

```
r_px = ray.a.x
r_py = ray.a.y
r_dx = ray.b.x-ray.a.x
r_dy = ray.b.y-ray.a.y
```

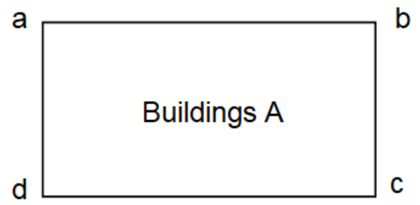
Segments (building segments)

```
s_px = segment.a.x
s_py = segment.a.y
s_dx = segment.b.x-segment.a.x
s_dy = segment.b.y-segment.a.y
```

3

The list of segments were created based on the following data structure:

{a:{x:100,y:150} b:{x:200,y:150}},
{b:{x:200,y:150},c:{x:200,y:50}},
{c:{x:200,y:50},d:{x:100,y:50}},
{a:{x:100,y:50}, b:{x:100,y:150}}



Every building is composed of an *x* number of segments; for the example above, 4 segments ab,bc,cd,da were constructed considering that the end point of every segment, is the starting point of next one. Besides that, the starting point and ending point of each polygon should be the same.

Having the two pairs of equations (rays + segments), they were converted to a linear system by setting the two X equations equal, as well as the two Y equations as follows:

*r_px+r_dx*T1 = s_px+s_dx*T2*

*r_py+r_dy*T1 = s_py+s_dy*T2*

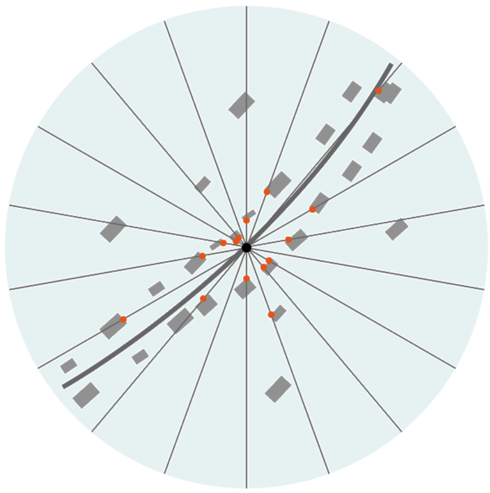


Figure 4.12. *Point intersections from building neighbours.*

The linear system of equations were then arranged to solve the parameters T1 and T2. The values of those parameters, give the percent distance that the intersection is between the endpoints on each line. So that if the values are between 0 and 1, then the intersection point lies internal to the two line segments. Otherwise, if the values are greater than 1 or less than 0, the lines intersect but at some external point. Additionally, intersections will be retrieved if and only if the rays and the lines segments are not parallel, otherwise there is no intersection. The rays meet the buildings in the vicinity of the point where only the first intersected building is considered (Figure 4.12). The rays meet a building more than once, considering in this way also the horizontal dimension of it. The algorithm that detects point intersections is depicted in the pseudo code below.

Pseudo Code – Point Intersections

Input: *- Sample points neighbourhood object (List of points with segments neighbours)*
Output: *Intersection points list*
Initialize algorithm:
Point_Intersections = [ ]
Find point intersections:
For each sample point
    Calculate ray angles
    Iterate over ray angles
        Calculate the ray
        Find ray- segment intersection
        Check for closest intersection
        Append points based on building height and length
Return Point_Intersections

At last stage of the algorithm, intersected points were filtered based on the buffer zones defined in section 4.2.1, so that the buildings that affect the value of openness are those buildings with heights within the height range of the corresponding zone (Table 4.1, Figure 4.13).

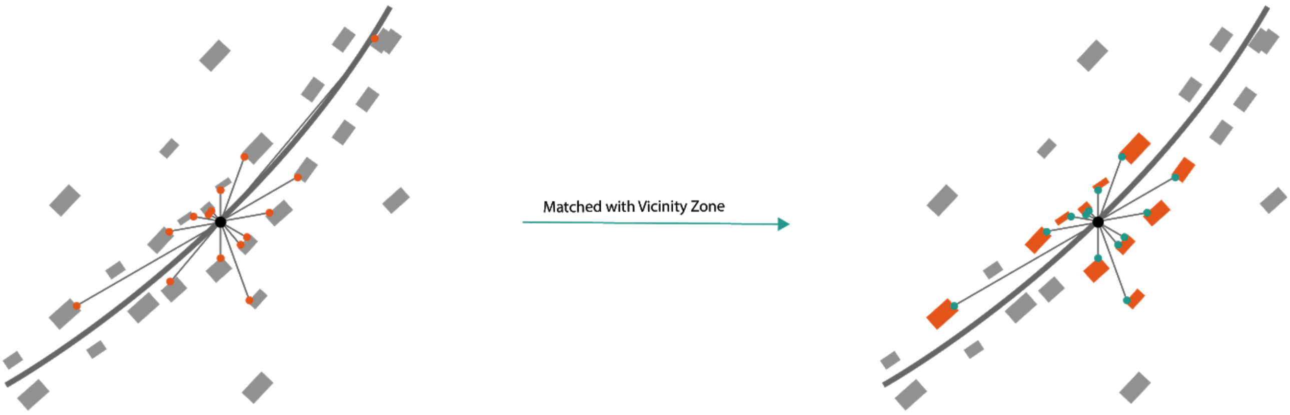


Figure 4.13. *Buildings contributing to the openness value*

3. Openness on each sample point

The openness is calculated on each intersection point where the height of the building that the ray hits and the distance from the sample point to the intersection point are considered. To get the most accurate result of the building height, the median value of each building was used. By taking the average, the errors would also be considered and this affects the trustability of the calculation. The openness on each intersection point is represented by the ratio of distance and height and is calculated with the formula:

$$Intersection\ Point\ Openness = \frac{D\ (m)}{D(m)\ +\ H\ (m)}$$

where D is the distance of the building to the sampling point and H is the building height. In the denominator the distance is added so that the resulting value is between 0 and 1. Then the final openness value is calculated by the average of all the intersection points openness values by the formula (full algorithm to be found in appendix I)

$$Sample\ Point\ Openness\ = \frac{\sum_{i=1}^{n=36}(\frac{D(m)}{D(m)+H(m)})}{36}$$

4.3 Modeling monotony of the built environment

Besides accounting for openness of the built environment, this research attempts to evaluate the effect of monotony of the built environment on cyclist route choice. Based on findings in existing literature, the following definition of monotony of the built environment has been formed: *the extent of visual variation in elements that form the built environment for a sequence of locations.* For the purpose of creating a model that can spatially represent the monotony of the built environment, surrounding a certain road segment, a definition of the concept of monotony has to be established based on actual built environment data:

How can monotony of the built environment be quantitatively measured using data of built environment?

Unlike the openness of the built environment, monotony will not expressed in a single value. Initially, monotony will be modelled by the amount of land use changes around a particular road segment. The land use can be described as the intended use purpose of a particular piece of land, like residential, industrial or forest area. Data on the land use in the area of interest of this research has been retrieved from the dataset ‘Bodemgebruik 2012’, provided by Dutch Statistics. This dataset contains polygons with one land use assigned to it, and the spatial analysis to obtain the amount of land use changes around a road segment is fairly straightforward. As depicted in Figure 4.14, a buffer will be created around each road segment. For each polygon representing a land use, a spatial check is performed to assess whether a polygon overlaps, crosses, intersects or falls completely within this buffer. Through this approach, the number of land use polygons around each road segment can be computed. On top of that, it provides enough information to compute the amount of land use changes per unit of distance, as the length of each road segment is known.

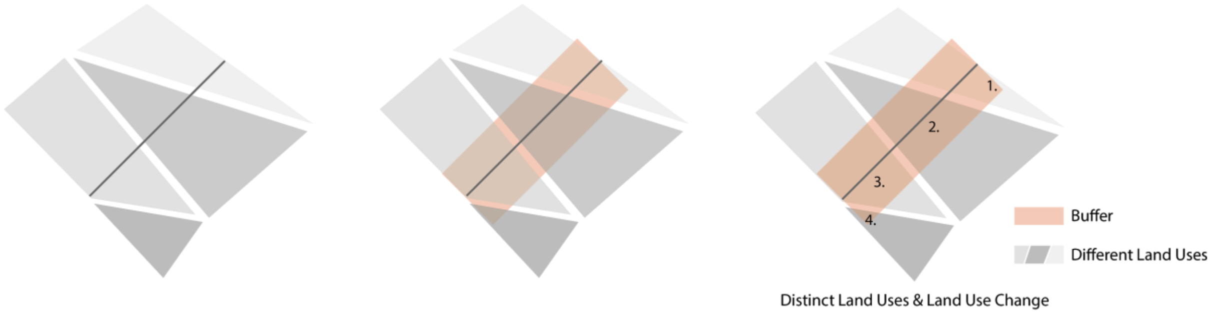


Figure 4.14. *Steps of calculating the amount of land use changes.*

However, as it has been found in previous studies, the predictability of the pattern of land use changes is an important factor within the concept of monotony (Thiffault & Bergeron, 2003; Zhao & Rong, 2013). Therefore, a value that describes the amount of distinct land uses surrounding a certain road segment is also used to express monotony. The process of obtaining this value is very similar to the process for obtaining the amount of land use changes. Within the buffer that has been created around a road segment, the amount of distinct land use polygons that overlap, cross, intersect or fall completely within this buffer is counted. Similarly to the computation land use changes per unit of distance, a value can be computed to express the amount of distinct land uses per unit of distance.

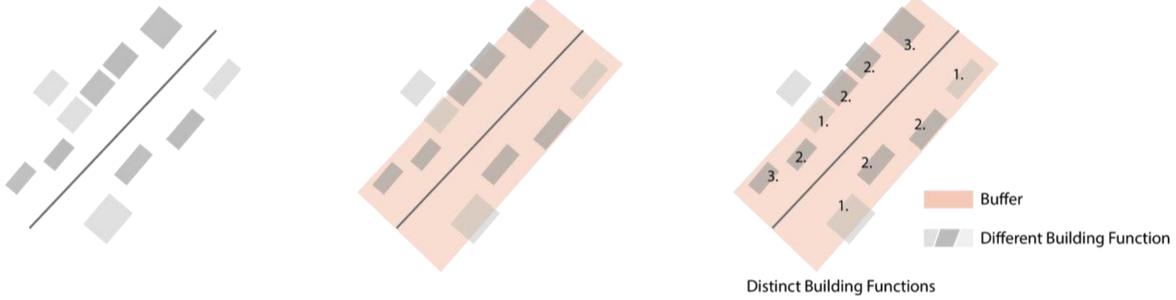


Figure 4.15. *Steps of calculating the amount of distinct building functions.*

For a more complete picture of the variation in road environment, a more detailed analysis will take into account the amount of distinct functions of the buildings that are closest to a road segment. The function of a building can be described as the intended use of a building, like residential, commercial or industrial. To obtain the desired values, a similar procedure is applied as for computing the amount of distinct land uses per unit of distance (figure 4.15). A buffer is created around every road segment, and for each building (retrieved from the 'Basisregistratie Adressen en Gebouwen' dataset by the Dutch Kadaster) that intersects or falls within this buffer, the function is evaluated and the amount of distinct functions is counted. The size of this buffer has been set to 20 meters, as this ensures that only the first line of buildings is taken into account. If inside the buffer exist multiple buildings,that are placed behind each other, will be not included in the analysis, due to these buildings do not contribute to the experience of monotony of the surrounding environment as they are blocked by the buildings in front.

4.4 Statistical analysis

4.4.1 Linear regression model

With the main aim of this research being to predict the pure effect of openness and monotony of the built environment on route selection, a statistical analysis that accounts for these potential effects, controlled for the effect of other factors, has to be performed. An often applied analysis in previous research is the multivariate logistic regression analysis, which allows for the formation of a multivariate regression relation between a dependent variable and a number of independent variables (Pradhan, 2010). Multivariate logistic regression assigns each independent variable that is included in the model a coefficient that measures the variables' independent contribution to variations in the dependent variable (Pradhan, 2010). However, this analysis would only provide information on the odds of choosing a different route over the shortest path, for an increase or decrease of the independent variables. Therefore, the difference between the shortest path and the observed route is not being represented as a dichotomous choice between either the one or the other, but as a numerical difference to enable a linear regression analysis. A linear regression analysis allows for the prediction of the value of the dependent variable, based on the values of the independent variables, through a linear relation (Wikipedia, 2018b). As the computation of the shortest path is based on the distance between a start and end point of a route, the dependent variable is defined as the divergence in distance from the shortest path:

$$Divergence\ (\%) = \frac{Length\ observed\ route\ -\ length\ shortest\ path}{length\ shortest\ path} \cdot 100$$

Included in two distinct statistical models are values for openness, the amount of land use changes (per meter), the amount of distinct land uses (per meter), and the amount of distinct building functions are included, as well as values for the control variables for both the shortest path and the observed route. The first model takes the values for the independent variables that correspond to the shortest paths into account, while the second model includes the differences between the values of the observed route and values of the accompanying shortest path. This approach delivers comparative regression coefficients for the variables for openness and monotony of the built environment, which provides sufficient information to predict the effect of both concepts on the divergence from the shortest path. The predictions can be used as the base for the formation of conclusions on the effect of openness and monotony of the built environment on cyclist route choice, controlled for other independent variables.

4.4.2 Control variables

In order for the final statistical model to provide output that can be used for a valuable conclusion on the independent effect of openness and monotony of the built environment on cyclist route choice, other potential influencing factors should be included in the statistical model as control variables. The selection of the control variables is based on existing literature, as well as the availability of necessary data. Factors that have been recognized as influential on cyclist route choice in multiple previous studies have been selected as control variables. The following factors have been included in the statistical model, with the values being representative for an entire (shortest or observed) route: *travel time, the number of crossings, and the percentage of separated bicycle lanes.*

4.4.3 Paired-samples t-test

Prior to the linear regression analysis, a comparison between the values for the independent variables for the observed routes and accompanying shortest paths has been made by means of paired-samples t-tests. These tests declare whether the differences between the observed routes and the shortest paths, for every independent variable, are statistically significant (Papinski & Scott, 2011). In this way, information is provided about the magnitude of the differences between shortest path and observed route, as well as the relevance of applying a linear regression analysis on the differences.

5

Results

This section will treat the application of the statistical analysis and its output. First of all, the results of modeling the openness and monotony of the built environment will be discussed and visualized in section 5.1. The results of the different statistical analyses are covered in section 5.2, while the interpretation and value of the statistical analyses is discussed in section 5.3.

5.1 Results of modeling Openness and Monotony

Openness

The openness model calculates the openness value for each sample point on the road network. As a next step, the average openness per road segment was calculated from the values of all the points on the segment. The result is shown in the Figure 5.1. As can be observed, the openness values are lower in the builted areas, with the lowest values in the center of Breda and Tilburg. Openness gradually raises when moving towards the rural streets and gets the higher value on the big highways.

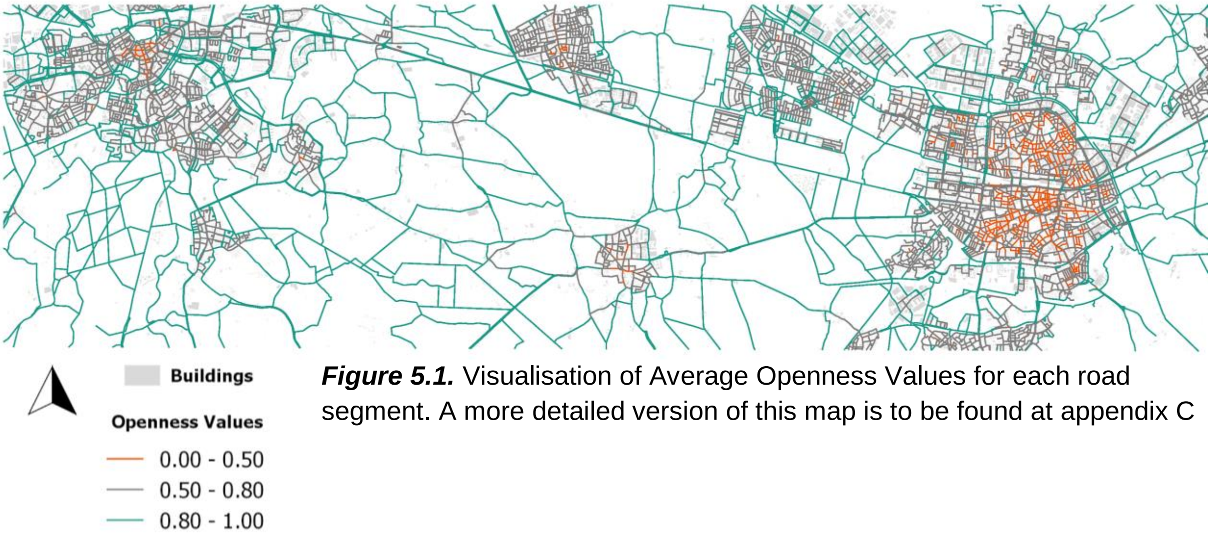


Figure 5.1. Visualisation of Average Openness Values for each road segment. A more detailed version of this map is to be found at appendix C

Monotony
The model for monotony considers three variables: the number of land uses per meter for each road segment, the distinct (unique) land uses per meter for each road segment and the distinct (unique) building functions per meter of a road segment. The visualisation of the monotony result for each road segment (Figure 5.2) highlights that the within the built areas the variation is bigger, with 2 to 8 buildings per meter of line segment. The higher the divergence from the urban fabric, the lower the variation of the land uses and the building functions.

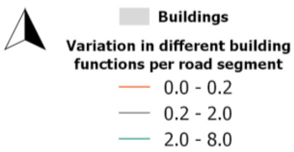


Figure 5.2. Visualisation of Variation in different building functions per road segment. The values represent the number of buildings per meter or road segment. A more detailed version of this map is to be found at appendix D.

5.2 Results of the statistical analysis

Initially, a comparison between the observed routes and the accompanying shortest paths has been made based on the differences between the values of the independent variables for both types of route. The results of this comparison, obtained through paired-samples t-tests, are displayed in table 5.1 with the mean difference between both types of route and the significance values.

Variable	Mean difference	Significance
Number of land use changes per meter [n/m]	0.002	0.000
Number of distinct land uses per meter [n/m]	-9.922*10^-4	0.000
Number of distinct building functions per meter [n/m	0.002	0.000
Openness [%]	2.893	0.000
Travel time [seconds]	113.104	0.000
Number of cross roads per route [n/route]	3.618	0.000
Percentage of separate bicycle lanes [%]	0.39	0.000

Table 5.1. *Output of the paired-samples t-test for every independent variable.*

The results of the paired-samples t-tests show significant differences between the values of all independent variables between the observed routes and their accompanying shortest paths at a 0.05 significance level. On average, observed routes have less distinct building functions per meter, but have a slightly higher variation of road environment in terms of distinct land uses and land use changes per meter. Similarly, observed routes can be considered more open than the corresponding shortest paths.

Also, for the control variables, there are statistically significant differences. As expected, the observed routes are significantly longer than the shortest paths in terms of travel time. This may suggest that cyclists choose for a certain route based on other factors, since they are not optimized for travel time. Finally, the observed routes have on average 3.6 more crossings per route than the corresponding shortest paths, as well as a higher percentage of separate bicycle lanes. However, this is only a slight positive difference. The complete results of all paired-samples t-tests have been included in appendix L.

Since from the paired-samples t-test can be concluded that there are significant differences between the observed routes and the shortest paths, a reasonable continuation is to predict the influence of the independent variables on the choice for a route other than the shortest path. As mentioned in section 4.4.1, a linear regression has been performed with the dependent variable being the divergence in distance from the shortest path by an observed route. In the first statistical model, the values of the independent variables for every observed route correspond to the values of the related shortest paths. The aim of this analysis is to predict the influence of the values for the shortest path on the magnitude of the divergence in distance. The results of the first regression analysis are included in Table 5.2.

Variable	B coefficient	Significance
Constant	118.836	0.000
Number of land use changes per meter [n/m]	466.591	0.000
Number of distinct land uses per meter [n/m]	-102.843	0.384
Number of distinct building functions per meter [n/m]	210.907	0.021
Openness [%]	-0.670	0.000
Travel time [seconds]	-0.016	0.000
Number of cross roads per route [n/route]	-0.763	0.000
Percentage of separate bicycle lanes [%]	-4.27	0.000

Table 5.2. *Output of the linear regression analysis with the values for the shortest paths as input.*

The result that stands out most is the insignificant predictive value of the number of distinct land uses per meter. Since this coefficient is insignificant, it is statistically not possible to assign a conclusion to the predictive value of the number of distinct land uses per meter. The statistical insignificance of this predictor can be explained by the correlation with the other predictors. As is shown in table 5.3, the variable describing the number of distinct land uses per meter correlates heavily with the variable describing the number of land use changes per meter. This resembles that the values for the number of distinct land uses per meter can be almost entirely explained by the values for the number of land use changes per meter. The full correlation matrix is included in appendix N.

Variable	Correlation
Number of land use changes per meter [n/m]	0.975
Number of distinct building functions per meter [n/m]	0.136
Openness [%]	-0.045
Travel time [seconds]	0.124
Number of cross roads per route [n/route]	0.226
Percentage of separate bicycle lanes [%]	0.415

Table 5.3. *Correlation between the number of distinct land uses per meter and the other predictors.*

However, the other two variables describing monotony do have significant predictive influence on the divergence from the shortest path. An increase of one distinct land use or distinct building function, would result in an increase of the divergence of 467% and 210% respectively. As these numbers seem extremely high, the statistics describing these variables (appendix K) show that the values for both variables are relatively small compared to the other variables, with mean values of 0.0753 land use changes per meter and 0.0198 distinct building functions per meter. Where an increase in the two significant variables describing monotony have a positive influence on the magnitude of the divergence from the shortest path, it is found that openness has a negative predictive value. According to the model, an increase of the openness by 1% will result in a decrease of 0.670% in the divergence from the shortest path.

For the second analysis, the differences in the values of the independent variables for the observed routes and corresponding shortest paths are taken is input. The output of the linear regression analysis, shown in table 5.4, takes a slightly different form than the output of the previous analysis. As it was the case for the analysis with the shortest path values as input, the coefficient of the number of distinct land uses per meter is found statistically insignificant. Again this is caused by high correlation with other predictors. However, the coefficients of the other two variables describing monotony are negative. This can be explained by the fact that the input values can be both positive and negative, as they are differences, which results in negative mean values for both variables (appendix M). The output of this analysis suggests that the number of distinct building functions per meter has a higher impact on the divergence from the shortest path than the number of land use changes per meter. This is in contrast with the results in table 5.2, where the impact of the number of land use changes per meter is higher than the impact of the number of distinct building functions per meter.

Similar to the results in table 5.2, the predictive value of openness is found to be negative. In this case the coefficient of -3.102 resembles a decrease of 3.102% in the divergence from the shortest path when the difference in openness between the observed route and a shortest path increases by 1%. A final result that stands out is that the predictive value of the difference in number of cross roads per route is found insignificant, at a significance level of 0.05.

Variable	**B coefficient**	**Significance**
Constant	23.648	0.000
Number of land use changes per meter [n/m]	-597.444	0.000
Number of distinct land uses per meter [n/m]	194.140	0.236
Number of distinct building functions per meter [n/m]	-939.023	0.000
Openness [%]	-3.102	0.000
Travel time [seconds]	0.156	0.000
Number of cross roads per route [n/route]	0.149	0.050
Percentage of separate bicycle lanes [%]	0.438	0.000

Table 5.4. *Output of the linear regression analysis with the differences between the observed routes and corresponding shortest paths as input.*

5.3 Discussion on the statistical analysis

From the two linear regression models that have been established it can be concluded that monotony, represented by the number of land use changes per meter and the number of distinct building functions per meter, and openness of the built environment have an influence on the amount of distance cyclists are willing to diverge from the shortest path. However, the question is how representative the established regression models are when it comes to their total predictive value of the divergence, in distance, from the shortest path by cyclists.

The predictive value of the model is determined by the explained variance, which measures the proportion to which the regression model takes the dispersion of a dataset into account (Wikipedia, 2018b). It is described by the R Square statistic, where a higher value for the R Square means a higher predictive value of the regression model. Table 5.5 shows the R Square values for the regression model based on the values of the shortest path, and for the regression model based on the differences between observed route and shortest path, which are 5.8% and 3.6% respectively. This indicates that the established regression models explain 5.6% and 3.6% of the dispersion of the input data.

	Input: shortest path	**Input: observed route - shortest path**
R Square	0.058 (0.000)	0.036 (0.000)

Table 5.5. *R Square values for the two regression models, with the corresponding significance value.*

The fact that the values for the explained variance are low for both regression models can be explained relatively easy. As has been discussed in section 2.1, many factors have been found in previous studies that influence cyclist route choice. These are factors with respect to the built environment, safety and cyclist characteristics. However, due to time limitations and limitations on available data, only seven predictors could be included in the final regression models and many potential influencing factors are not being accounted for. But, even without limitations in time and data availability, it would still be fairly impossible to take all potential influences on human decision into account as there is many factors that potentially influence human behaviour. Additionally, predicting the influence of certain characteristics on human decision making is a complex process in general, since humans do not always make rational choices. Therefore, it can be considered as reasonable that the explanatory strength of both regression models is relatively low.

Interesting from the perspective of this research is the proportion of the effect of the independent variables describing openness and monotony on the R Square statistic for each of the regression models. By entering the independent variables stepwise into the regression model, statistics can be obtained on the change of the R Square value for every step (table 5.6).

Variable	R Square change: shortest	R Square change: difference
Number of land use changes per meter [n/m]	0.4%	0.3%
Number of distinct building functions per meter [n/m]	0.0%	0.4%
Openness [%]	0.3%	0.7%
Travel time [seconds]	4.5%	2.0%
Number of cross roads per route [n/route]	0.2%	0.2%
Percentage of separate bicycle lanes [%]	0.4%	0.0%

Table 5.6. *R Square changes for the predictors in the two regression models*

The results in table 5.6 clarify that the travel time accounts for a large share of the explained variance of both regression models. Entering openness of the built environment to the model based on the values for the shortest path increases the explained variance by 0.3%, while the explained variance increases by 0.7% when entering openness in the model based on differences between observed routes and shortest paths. Entering the number of land use changes per meter increases the explained by 0.4% and 0.3% respectively, while the number of distinct building functions per meter does not result in extra explained variance for the model based on shortest path values and an increase of 0.3% for the model with differences between observed routes. The number of distinct land uses has been left out of the stepwise linear regression analysis because it is not found statistically significant. From these statistics it can be concluded that the predictors describing openness and monotony of the built environment have a relatively small influence on the predictive strength of the regression models.

6

Conclusions

The main objective of this research has been to bridge a gap in existing literature on cyclist travel behaviour, by examining the effect of openness and monotony of the built environment on cyclist route choice. With the spatial focus of the research being on the province of Noord-Brabant, this research seeks to answer the following research question by means of literature review, spatial modeling and statistical analysis:

How do openness and monotony of the built environment affect cyclist' route choice in the Province of Noord-Brabant?

Initially, the main challenge of this research has been to define the concepts of openness and monotony of the built environment. Both concepts have been treated in existing literature, but the link to cyclist route choice has remained underexposed so far. Based on multiple definitions and methods used in previous studies, the openness of the built environment could be described as *the extent of open scene above and around a specific point. In the built environment an environment is considered as open when no obstacles of the built environment are interfering with the visual scene of a person, or the interference can considered to be low.* To obtain openness values on a certain point that comply with this definition, the height of surrounding buildings and the distance to those buildings are considered. By averaging the openness values over a sequence of locations, an openness value could be assigned to an entire route.

In a similar fashion, the monotony of the built environment can be described as follows: *the extent of visual variation in elements that form the built environment for a sequence of locations.* Taking into account the findings from previous studies, this definition allows for multiple variables to describe monotony of the built environment. Visual variation can be described by the land uses in the built environment surrounding a route, as well as the by the buildings surrounding a route.

By means of analysing GPS measurements of cyclists, a set of distinctive observed routes could be identified. However, due to data storage and processing limitations, the spatial scope of the research has been limited to a sample area in Noord-Brabant that includes the cities of Breda and Tilburg. This sample area is considered as representative for the entire province as it includes urban and rural areas, and inter-city connections between both urban areas.

Based on applied methods and findings in previous studies, the assumption has been made that cyclists in the first place opt for the shortest path when selecting a route, measured in distance. For every distinct combination of starting and ending points in the observed routes, a shortest path has been calculated. To examine whether the openness and monotony of the built environment affect the route choice by cyclists, the observed routes were compared with the shortest paths in terms of distance, with the difference being expressed as the percentual divergence from the shortest path. For every observed route and corresponding shortest path the values for the variables describing openness and monotony of the built environment were computed. Since many factors have been found to influence cyclist route choice, the values for multiple control variables have been calculated as well to obtain a more valuable effect of openness and monotony of the built environment. These variables are the *travel time*, the *number of crossings*, and the *percentage of separated bicycle lanes*.

With the main objective of this research being to estimate the effect of openness and monotony of the built environment on divergence from the shortest path, it has been tested whether there are significant differences between the independent and control variables between the observed route model and the shortest path model. From these tests it can be concluded that for every variable that is included in the analysis, there are statistical differences between both models. In the first place, this suggests that people do diverge from the shortest path, and it assures that the observed routes are significantly different from the shortest paths in terms of the independent and control variables.

By means of linear regression analyses, a conclusion could be drawn on how and how much the openness and monotony, described by the number of land use changes per meter, the number of distinct land uses per meter and the number of distinct building functions per meter, of the built environment affect the divergence from the shortest by cyclists in the sample area. The first regression model takes the values for the corresponding shortest path as input for every observed route. From the output of this analysis it can be concluded that the openness of the built environment has a negative influence on the divergence from the shortest path. This indicates that cyclists in the sample area experience more open routes as a negative factor and therefore prefer 'less open' roads. This does not comply with findings from previous research (Hur et al., 2010), where the experience of more openness on a certain location is considered as something positive. However, the results for the influence of the variables describing monotony do agree with findings from previous studies (Thiffault & Bergeron, 2003; Zhao & Rong, 2013): cyclists prefer more variation in the surrounding built environment. The second regression model takes as input the differences between the values of the observed routes and their corresponding shortest paths for every variable included in the model. Although the predictive coefficients take slightly different proportions, similar conclusions can be drawn on the effect of openness and monotony of the built environment.

In general, the established regression models have a relatively small explanatory strength for the divergence from the shortest path. This can be explained by the fact that only a small amount of predicting variables has been included in the model, while previous studies have identified many potential influencing factors. The proportion of explanatory strength covered by the variables describing openness and monotony of the built environment is small, since the greater part is explained by the travel time. Also, for both regression models, the number of distinct land uses has been found to have a statistically insignificant influence on the divergence from the shortest path.

The results of the linear regression analyses show that openness of the built environment has a negative influence on the amount of distance people are willing to diverge from the shortest path. Therefore, it can be concluded that cyclists in the sample area prefer 'less open' routes, but the proportional influence of this factor can be considered small. On the contrary, cyclists in the sample area prefer more variation in the built environment (expressed in the number of land use changes and the number of distinct building functions). However, as is the case for the openness of the built environment, the proportional influence can be considered as low.

7

Limitations & Recommendations

This section covers the limitations and recommendations that arise from the implementation process. The main steps of the process include a review of the existing literature on the concepts of openness and monotony, and based on that, the creation of the base route model, the openness model and the monotony model. The resulting values from the models were considered per each route and were compared with other controlled variables with linear regression. The process of implementation and the results highlighted some limitations around the used approach and methods. Based on the limitations some recommendations are proposed.

Base route model
Limitation: All bicycle lanes are considered bidirectional
Recommendation: Consideration of the bicycle lane direction

In the base route model, the shortest path between all distinct combinations of a starting point and an ending point is calculated. For the shortest path, all the segments of the bicycle lane network were considered without taking into account the direction of the bicycle lane. This was based on the idea that most of the main roads support the travelling in both directions. Nevertheless, taking into consideration the actual direction of a bicycle lane could indicate a different shortest path between two points. However, even considering the direction of the bicycle lanes, the statistical model (average openness per links and later per routes) and the openness model map would not be affected.

Openness model
Limitation: The openness model is applied only on the buildings
Recommendation: Include more elements of the built environment

The calculation of openness is implemented by considering only the surrounding buildings, excluding the vegetation and other possible artefacts. The assumption behind this approach is based on the intention to consider elements that have the same kind of impact. That means that vegetation could affect openness in a big extent, but it could have an opposite impact on a cyclist perception than buildings.

Limitation: The direction of route of a cyclist is not considered.
Recommendation: Consideration of the real visible angle

The current approach for openness considers all the buildings around a specific point within an angle of 360°. With this method also the buildings that are behind the cyclist contribute to the final openness of each sampling point. Ideally, the direction of the cyclist should be considered and only the buildings interfering with his visual field should affect the result of openness of the sampling point. This recommendation presupposes that also the direction of the bicycle lane is introduced in the model.

Monotony model

Limitation: Monotony is a factor of land use and building functions
Recommendation: Include more aspects of the built environment

Currently, the monotony of the built environment is a parameter of the variation of the land use and building function. Apart from these parameters, for the calculation of monotony more aspects could be considered, like the variation of the facades or the type of the buildings.

Statistical analysis

Limitation: Aggregation of values for openness and monotony per each route
Recommendation: Consider the detail in the measured values

The values of openness and monotony were calculated in detail on the road segments. More specifically, the openness was calculated on sample points along the bike lane network in an interval of 20 meters. The monotony value was extracted by the variation of the land use and land covers per meter of line segment. On the other hand, the statistical analysis was implemented on each route that consists of many road segments, requiring the aggregation of the calculated values to an average value. As a result the different values of openness and monotony along a route are not considered and the final outcome is generalised.

Limitation: The purpose of travel is not considered for the analysis
Recommendation: Distinction between mandatory and non mandatory routes

In the current approach, the purpose of travel for each route is not taken into account. Though, purpose of travel is affecting the travel behaviour and a distinction between mandatory and non-mandatory travels could result in different correlation values. This approach will lead to more representative results as travels for commuting are usually not affected by factors like openness or monotony but are based on shortest distance and the minimum travel time.

Limitation: Mostly control variables for the built environment were used
Recommendation: Include extra control variables

The current control variables are mostly parameters of the built environment infrastructure. Complementary, extra value could be added by using real data concerning the weather conditions, traffic volumes or facilities on the roads. Due to time and data limitations this fell out of the scope of this research.

Limitation: The ideal extent of openness is not identified
Recommendation: Further research on the minimum/maximum accepted openness value.

The statistical analysis provides information on whether openness influences the route choice or not. A next step could be to find the the minimum and maximum accepted openness values for a cyclist.

Limitation: Differences of rural and urban areas are not considered
Recommendation: Distinction of rural and urban areas for the analysis

The basis of the statistical analysis is the difference of the shortest path and the chosen route from a starting to an ending point. In the rural areas the alternative routes are restricted as the road network is less dense and with smaller number of road intersections than in the urban areas. The restricted number of alternatives in the rural areas result in overlapping of shortest route and chosen route, reducing the differences in the measured values. These results may in extent counterbalance the measured differences of the measured the values in the urban areas. Thus, a distinction in the analysis could present a different image for the influence of openness and monotony on the route choice.

Limitation: Results are not applied on other areas
Recommendation: Result validation

An application of the models and the statistical analysis in another sample area could lead to different result. The validation implies the application of the model on different sample data, which are not already used.

References

AHN. (2018). Actueel Hoogtebestand Nederland. Retrieved from: http://www.ahn.nl/index.html

Akar, G., Clifton, K.J. (2009). Influence of individual perceptions and bicycle infrastructure on decision to bicycle. Transportation Research Record 2140, 165–171.

Balci, P. (2017). Route Choice Preference of Cyclists:an agent-based simulation model for the city of Utrecht. Master's Thesis Degree Programme in Geographical Information Management & Applications. Utrecht University, Delft University of Technology, Wageningen University and University of Twente.

Batty, M. (2001). Exploring isovist fields: space and shape in architectural and urban morphology. Environment and planning B: Planning and Design, 28(1), 123-150.

Benedikt, M. L. (1979). To take hold of space: isovists and isovist fields. *Environment and Planning B: Planning and design*, *6*(1), 47-65.

Broach, J., Dill, J., Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. Elsevier: Transportation Research Part A, vol. 46, pp. 1730-1740.

Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: density, diversity, and design. Transportation Research Part D: Transport and Environment, 2(3), 199-219

Cervero, R., Sarmiento, O. L., Jacoby, E., Gomez, L. F., & Neiman, A. (2009). Influences of built environments on walking and cycling: lessons from Bogotá. International Journal of Sustainable Transportation, 3(4), 203-226.

Cichański, A., & Wirwicki, M. (2010). Ergonomics analysis of anthropo-technical system In the environment of CATIA program.". Journal of Polish Cimac, 5(3), 19-25.

Dill, J., & Gliebe, J. (2008). Understanding and measuring bicycling behavior: A focus on travel time and route choice.

Ehrgott, M., Wang, J. Y. T., Raith, A., & Van Houtte, C. (2012). A bi-objective cyclist route choice model. Transportation Research Part A: Policy and Practice, 46(4), 652–663. Retrieved from: https://doi.org/10.1016/j.tra.2011.11.015.

Fisher-Gewirtzman, D., & Wagner, I. A. (2003). Spatial openness as a practical metric for evaluating built-up environments. *Environment and Planning B: Planning and Design*, *30*(1), 37-49.

Handy, S. L., & Xing, Y. (2011). Factors correlated with bicycle commuting: A study in six small US cities. International Journal of Sustainable Transportation, vol 5(2), pp. 91-110.

Heath, T., Smith, S. G., & Lim, B. (2000). Tall buildings and the urban skyline: The effect of visual complexity on preferences. Environment and Behavior, 32(4), 541-556.

Heinen, E., Maat, K., & Van Wee, B. (2011). The role of attitudes toward characteristics of bicycle commuting on the choice to cycle to work over various distances. Transportation research part D: transport and environment, 16(2), 102-109.

Hopkinson, P., Wardman, M. (1996) Evaluating the demand for cycle facilities. Transport Policy vol. 3, pp. 241–249.

Hur, M., Nasar, J. L., & Chun, B. (2010). Neighborhood satisfaction, physical and perceived naturalness and openness. *Journal of Environmental Psychology*, vol. *30*(1), pp. 52-59.

Lawrence, D. L., & Low, S. M. (1990). The built environment and spatial form. *Annual review of anthropology*, *19*(1), 453-505.

Ministry of Transport, Public Works and Water Management. (2009). Cycling in the Netherlands. Den Haag: Ministry of Transport, Public Works and Water Management.

Moudon, A. V., Lee, C., Cheadle, A. D., Collier, C. W., Johnson, D., Schmid, T. L., & Weather, R. D. (2005). Cycling and the built environment, a US perspective. Transportation Research Part D: Transport and Environment, 10(3), 245-261.

MTR Corporation Limited (2017). Environmental Impact assessment report: Proposed Comprehensive Residential and Commercial Development atop Siu Ho Wan Depot. Retrieved from: https://www.epd.gov.hk/eia/register/report/eiareport/eia_2522017/EIA/pdf/Text/Ch%2011%20-%20Landscape%20and%20Visual.pdf

MTR Corporation Limited (2017). Landscape and visual impact assessment. Retrieved from: https://www.epd.gov.hk/eia/register/report/eiareport/eia_1252006/html/eiareport/Part3/Section11/sec3_11.htm

NCase. (2018). Github. Sight and light. Retrieved from: https://github.com/ncase/sight-and-light

Noord-Brabant. (2017). Fietspaden. Retrieved from: https://www.brabant.nl/dossiers/dossiers-op-thema/verkeer-en-vervoer/fiets/fietspaden

Oke, T. R. (1981). Canyon geometry and the nocturnal urban heat island: comparison of scale model and field observations. *International Journal of Climatology*, *1*(3), 237-254.

Oke, T. R. (1988). Street design and urban canopy layer climate. *Energy and buildings*, *11*(1-3), 103-113.

Papinski, D & Scott, D. M. (2011). A GIS-based toolkit for route choice analysis. Journal of Transport Geography, vol. 19, pp. 434-442.

Pradhan, B. (2010). Remote sensing and GIS-based landslide hazard analysis and cross-validation using multivariate logistic regression model on three test areas in Malaysia. Advances in Space Research, vol. 45 (2010), pp. 1244–1256.

Province of Noord-Brabant (2009). *Fiets in de versnelling.* Provincie Noord-Brabant.

Qing, S., Chen, P., Schmiedeskamp, P., Bassok, A., Childress, S. (2014). Bicycle Route Choice: GPS Data Collection and Travel Model Development. Seattle: University of Washington.

Strauss, J., Miranda-Moreno, L.F., Morency, P. (2015). Mapping cyclist activity and injury risk in a network combining smartphone GPS data and bicycle counts, in: Accident Analysis and Prevention, vol. 83 (2015), pp. 132–142.

Thiffault, P. & Bergeron, J. (2003). Monotony of road environment and driver fatigue: a simulation study. Accident Analysis and Prevention, vol. 35, pp. 381-391.

Unwin, N. C. (1995). Promoting the public health benefits of cycling. Public Health, 109(1), 41-46.

Watson, I. D., & Johnson, G. T. (1987). Graphical estimation of sky view-factors in urban environments. *International Journal of Climatology*, *7*(2), 193-197.

Winters, M., Brauer, M., Setton, E.M., Teschke, K. (2010). Built Environment Influences on Healthy Transportation Choices: Bicycling versus Driving. Journal of Urban Health: Bulletin of the New York Academy of Medicine, Vol. 87, No. 6, pp. 969-993

Winters, M., Davidson, G., Kao, D., Teschke, K. (2011). *Motivators and deterrents of bicycling: comparing influences on decisions to ride*. Transportation, vol. 38, pp. 153–168.

Winters, M., Teschke, K., Grant, M., Setton, E.M., Brauer, M. (2010). How Far Out of the Way Will We Travel?: Built Environment Influences on Route Selection for Bicycle and Car Travel. Transportation Research Record: Journal of the Transportation Research Board, No. 2190, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 1–10.

Wikipedia. (2018). Explained variation. Retrieved from: https://en.wikipedia.org/wiki/Explained_variation

Wikipedia. (2018). Parametric equation. Retrieved from: https://en.wikipedia.org/wiki/Parametric_equation

Zhao, P. (2014). The impact of the built environment on bicycle commuting: Evidence from Beijing. Urban Studies, 51(5), 1019-1037.

Zhao, X. & Rong, J. (2013). The relationship between Driver Fatigue and Monotonous Road Environment, in: W. Wang and G. Wets, Computational Intelligence for Traffic and Mobility, Atlantis Computational Intelligence Systems, vol. 8, pp. 19-36.

Appendix A - Tables base route model

path_id integer	routeid numeric	source bigint	target bigint
0	295527	895	9968
1	295698	6545	2769
2	295734	1302	7849
2	295766	1302	7849
4	295778	7885	4147
5	295780	8560	5674
5	306623	8560	5674
5	351380	8560	5674
5	363347	8560	5674
5	399588	8560	5674
5	422660	8560	5674
5	442308	8560	5674
5	497554	8560	5674
5	527253	8560	5674
5	539776	8560	5674
5	636481	8560	5674
5	686409	8560	5674
5	696494	8560	5674
5	721997	8560	5674
5	809063	8560	5674
5	845485	8560	5674
5	874861	8560	5674
5	962102	8560	5674
5	1026522	8560	5674
5	1049526	8560	5674
5	1110781	8560	5674
5	1144238	8560	5674
5	1371130	8560	5674
5	1445516	8560	5674
5	1452387	8560	5674

Table A1. *Source - target per path_ids,routeids*

path_id integer	linknummer numeric	source double precision	target double precision
34179	1238556	12600	4759
34179	1279421	12600	4759
34179	1156890	12600	4759
34179	36636	12600	4759
34179	907173	12600	4759
34179	467994	12600	4759
34179	895969	12600	4759
34179	703894	12600	4759
34179	71844	12600	4759
34179	1238594	12600	4759
34179	953785	12600	4759
34179	729864	12600	4759
34179	443815	12600	4759
34179	1070438	12600	4759
34179	1050230	12600	4759
34179	1253616	12600	4759
34179	1035858	12600	4759
34179	764948	12600	4759
34179	98233	12600	4759
34179	119036	12600	4759
34179	587051	12600	4759
34179	485081	12600	4759
34179	989302	12600	4759
34179	933337	12600	4759
34179	933338	12600	4759
34179	671262	12600	4759
34179	362586	12600	4759
34179	301910	12600	4759
34179	851132	12600	4759
34179	831060	12600	4759

Table A2. *Example linknummer per shortest path_ids*

linknummer numeric	len_m integer	speed_kmh numeric	seconds double precision
989434	5	19.0886	0.942971197468646
601580	132	19.7136	24.1051862673484
622136	47	17.5237	9.65549512945326
683033	118	21.5648	19.6987683632586
703713	56	17.4311	11.5655351641606
703838	59	13.5898	15.6293690856378
709218	71	21.0704	12.1307616371782
709261	93	23.8202	14.0552976045541
709398	9	17.7215	1.82828767316536
710528	7	23.9648	1.05154226198424
724764	159	12.6124	45.3839079001617
805408	80	16.2319	17.7428397168539
785278	35	23.5443	5.35161376638932
805710	209	15.2164	49.4466496674641
825796	60	19.7101	10.95884850914
846302	196	20.7572	33.9930241072977
871389	36	18.5148	6.99980556095664
989397	69	20.4627	12.1391605213388
887025	53	17.1759	11.1085881962517
907332	70	18.2235	13.8282986254013
927876	81	19.4384	15.0012346695201
975383	10	12.8652	2.7982464322358
975522	132	22.2755	21.3328544813809
989267	29	21.6843	4.81454324096236
1009615	122	22.0666	19.9033833939075
1029643	69	20.4945	12.1203249652346
1050153	160	12.3408	46.6744457409568
1095687	100	14.0257	25.667168127081
1114780	1221	19.9024	220.857785995659
1116174	9	23.9824	1.35099072653279

Table A3. *Travel times per linknummer.*

Appendix B - Vertical field of view

For the vertical view, the natural line of sight is normally a 10° cone, below the the horizontal line (Figure A1). The impact of a structure of specific height reduces as the distance between the person and the structure increases (MTR Corporation Limited, 2017). This distance is also affected by the angle of view of a person. To choose the most suitable angle of vertical view, we used data from ergonomics on anthropometric data for the angle of view of a car driver, considering that a cyclist and a car driver will have the same visual field while driving. According to Cichański, Artur & Wirwicki, Mateusz (2018) this angle is 15° (Figure A2).

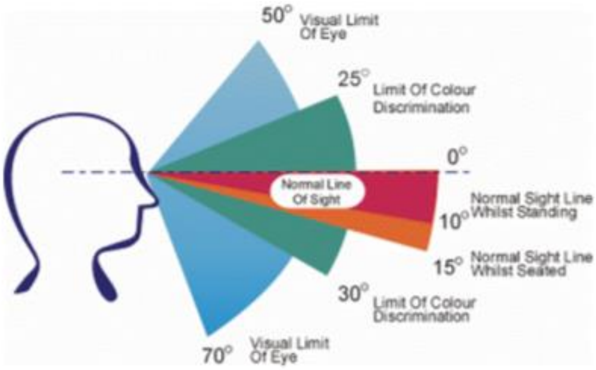


***Figure B1**. Vertical Field of view (*MTR Corporation Limited,2017., pg1 Appendix)

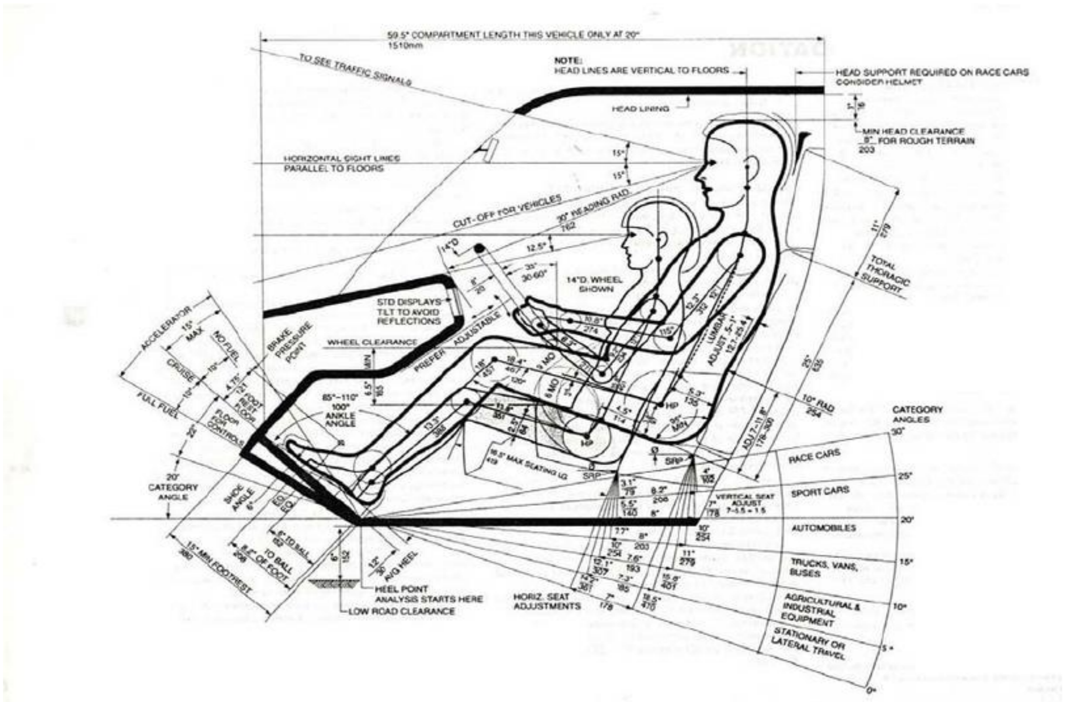
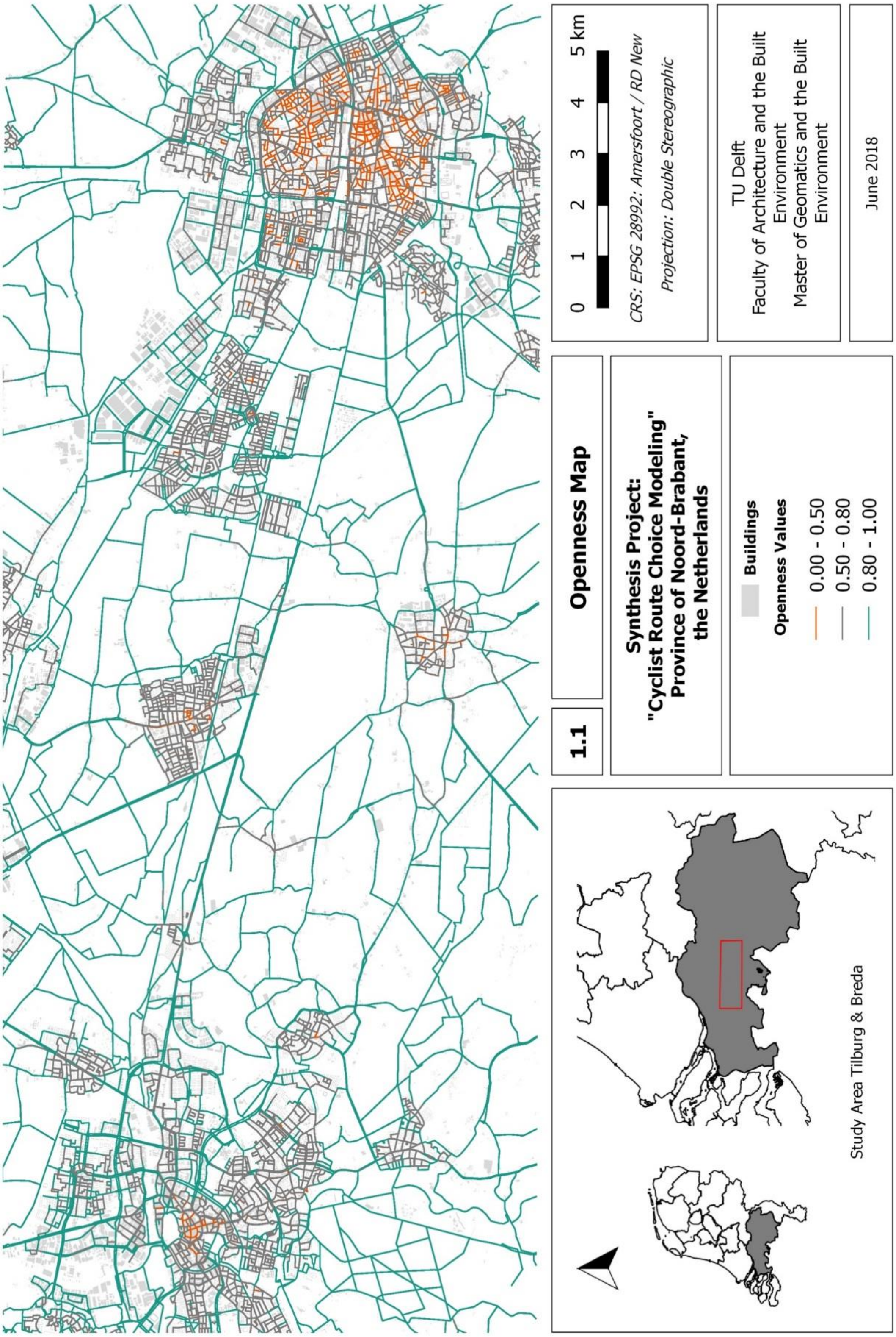


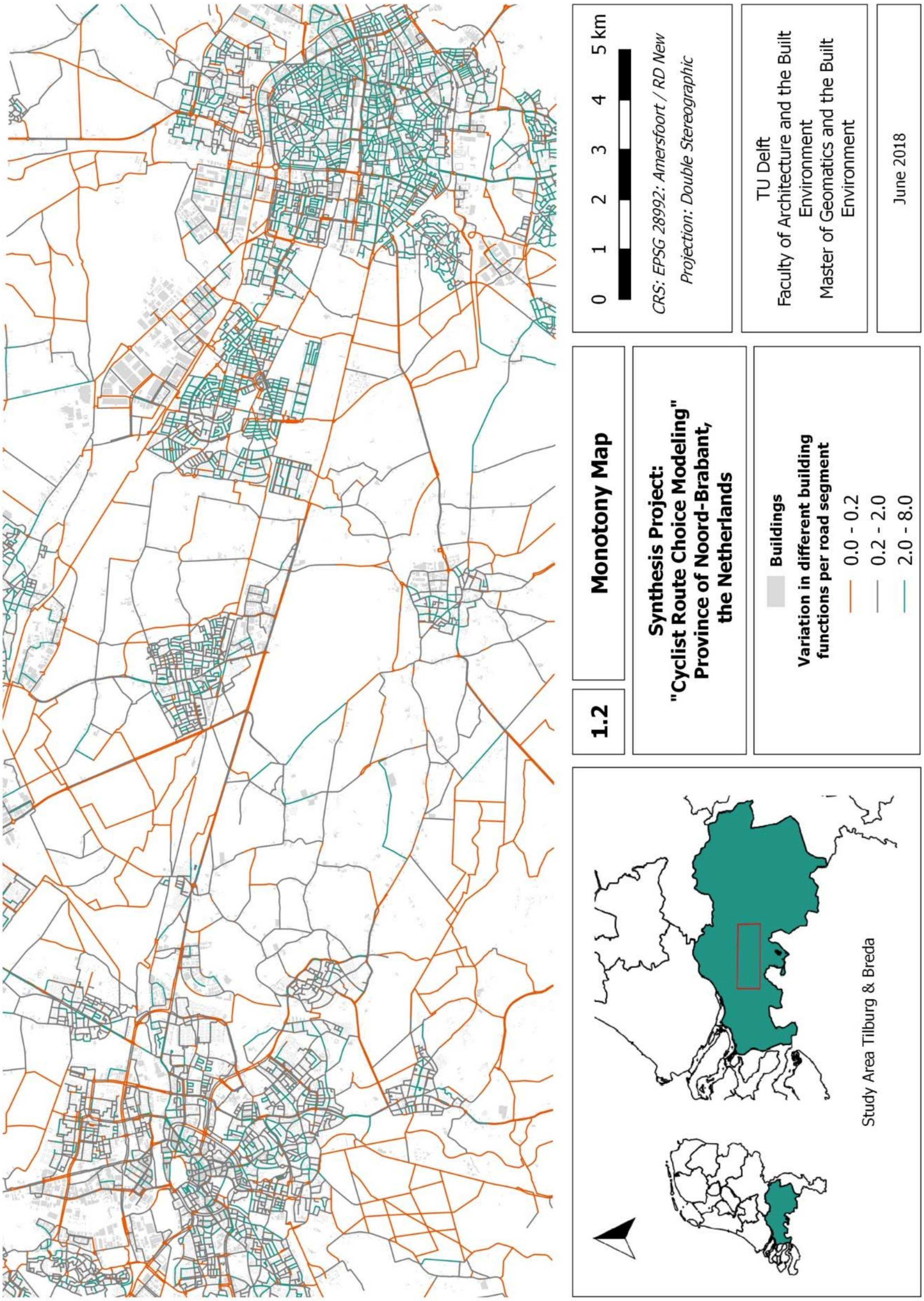
***Figure B2**. Driver ergonomics (*Cichański, Artur & Wirwicki, Mateusz. (2018), pg 2)

Appendix C - Openness map



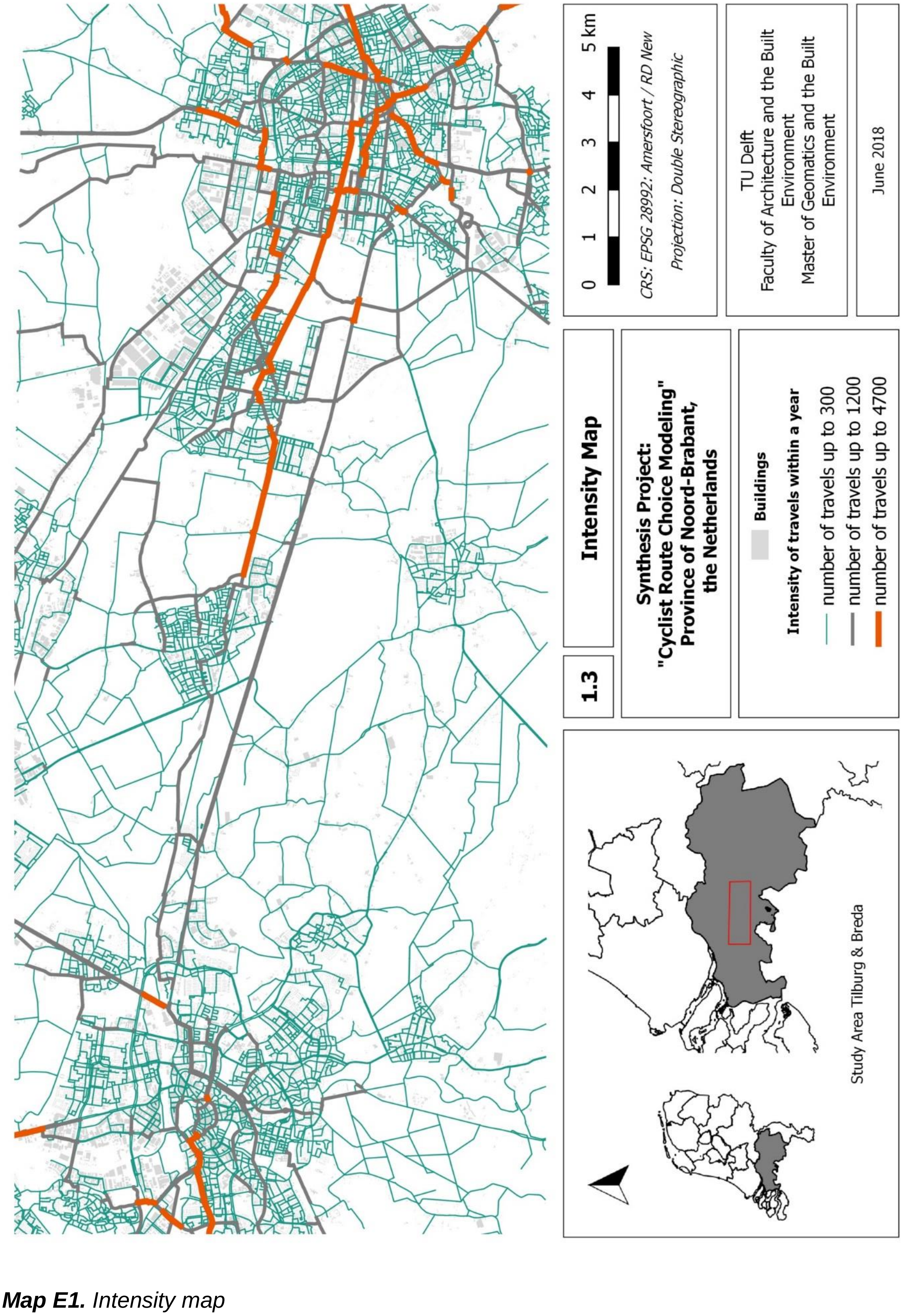
Map C1. *Openness map*

Appendix D - Monotony map



Map D1. *Monotony map*

Appendix E - Intensity map



Map E1. *Intensity map*

Appendix F - Algorithm to compute shortest path

```
import psycopg2

#connection with the database
hostname = 'localhost'
username = 'postgres'
password = '1234'
database = 'bikes'

   #CALCULATING THE SHORTEST PATHS

def doQuery( conn ) :
        cur = conn.cursor()
        cur2 = conn.cursor()


        #downloading source and target  for every route
        cur.execute(
        "SELECT distinct routeid,linkstart,
edge_start,source_start,target_start,linkend,edge_end,source_end,target_end FROM end_start_route limit
54183")
        nodes_list = []

        for element in cur.fetchall():
                        nodes_list.append(((float(element[3]), float(element[7])), float(element[0])))

                file_shortest = open("shortest_final", "w")
                file_shortest.write("{0},{1},{2},{3},{4}\n".format("path_id", "edge", "routeid", "source",
"target"))


        #shortest path functions from pgRouting
        def ashortest2(n1, n2):

                cur.execute(
        "SELECT * FROM pgr_bdAstar('SELECT id, source, target, cost, reverse_cost, x1,y1,x2,y2 FROM
road_topology',%s, %s); ",(int(n1), int(n2)))

                shortest = []

                for element in cur.fetchall():
                file_shortest.writelines(
                str(i) + ',' + str(element[3]) + ',' + str(routeid) + ',' + source + ',' + target + "\n")
        lista1 = []
        lista2 = []

        #eliminating duplicates (same source-target)
        for i in range(54183):
        routeid = (nodes_list[i][1])
        source = str(nodes_list[i][0][0])
        target = str(nodes_list[i][0][1])
        duplicates1 = (nodes_list[i][0][0], nodes_list[i][0][1])
        duplicates2 = (nodes_list[i][0][1], nodes_list[i][0][0])

        if duplicates1 not in lista1:
               if duplicates2 not in lista2:
               ashortest2(nodes_list[i][0][0], nodes_list[i][0][1])
               lista2.append(duplicates2)
               lista1.append(duplicates1)



   #COMPARING NUMBER OF LINKS IN SHORTEST PATHS VS NUMBER OF LINKS IN EVERY ROUTE
        def comparisons_route_shortest():

        #loading links per shortest path
        cur.execute("SELECT path_id,edge,linknummer, source,target FROM short_link_edge")

         map_path = {}
        for element in cur.fetchall():
            if map_path.has_key(float(element[0])):

map_path[float(element[0])].append((float(element[2]),float(element[3]),float(element[4])))
   else:
          links_shortest = []
                links_shortest.append( (float(element[2]),float(element[3]),float(element[4])))
                map_path[float(element[0])]=links_shortest

        # loading links per routes
                cur2.execute("SELECT path_id,routeid,linknummer,source_start,source_end FROM
paths_routes_linknummer")  # 54183

                map_route= {}
```

```
        for element in cur2.fetchall():
        if map_route.has_key(float(element[0])):

        map_route[float(element[0])].append((float(element[1]),float(element[2]),float(element[3]),float(el
ement[4])))
                 else:
                links_routes = []

links_routes.append((float(element[1]),float(element[2]),float(element[3]),float(element[4])))
                map_route[float(element[0])]=links_routes

        coincidences={}
        file_shortest = open("frequency_paths", "w")
        file_shortest.write("{0},{1},{2},{3},{4},{5}\n".format("path_id", "coincidences", "total_links",
"percentage","source","target"))

        for key, value in map_route.iteritems():
                lst_route = value
                lst_path=[]
                if map_path.has_key(key):
                lst_path = map_path[key]
         count=0
                for item in lst_route:
           for ln in lst_path:
         if item[1] == ln[0]:
             count=count+1

coincidences[key]=(count,len(map_route[key]),round(float(count)/len(map_route[key])*100,2),ln[1],ln[2])

 if map_path.has_key(key) and coincidences.has_key(key):


file_shortest.write(str(int(key))+','+str(coincidences[key][0])+','+str(coincidences[key][1])+','+str(coinc
idences[key][2])+','+str(coincidences[key][3])+','+str(coincidences[key][4])+"\n")

else:
                        file_shortest.write(str(int(key)) + ',' + str(0) + ',' + str(0) + ',' +
str(0)+"\n")


myConnection = psycopg2.connect( host=hostname, user=username, password=password, dbname=database )
doQuery( myConnection )
myConnection.close()
```

Appendix G - Algorithm to find neighbour points

```
import psycopg2
import math
from shapely import wkb
#reading road points from the database
def points (min,max):
        hostname = 'localhost'
        username = 'postgres'
        password = '1234'
        database = 'geo1101'
        conn = psycopg2.connect(host=hostname, user=username, password=password, dbname=database)
        cur = conn.cursor()
        cur.execute("select unique_id,geom,linknummer from links_fietser_points20 where unique_id >= %s and 
unique_id < %s;",(min, max))

        lista_points= []
        i = 0
        l2 = []
        for element in cur.fetchall():
                element = (element[0],wkb.loads(element[1], hex=True),element[2])
                lista_points.append(element)
        return lista_points
#reading building centroids from the database
def read_centroids ():
        hostname = 'localhost'
        username = 'postgres'
        password = '1234'
        database = 'geo1101'
        conn = psycopg2.connect(host=hostname, user=username, password=password, dbname=database)
        cur = conn.cursor()
        cur.execute("select gml_id,geom,mediaan_ho from centroids_buildings;")

        lista_centroids= []
        i = 0
        l2 = []
        for element in cur.fetchall():
               element = (element[0],wkb.loads(element[1], hex=True),element[2])
                lista_centroids.append(element)
        return lista_centroids
#FDN ALGORITHM
def FDN_algorithm(point,centroids):

   neighbours={}
   h_threshold= 55.2
   for pt in point:
       lst=[]
       file_nm.write("%s," % pt[0])
       for cnt in centroids:
           height_building= cnt[2]
           distance = math.sqrt((pt[1].x - cnt[1].x) ** 2 + (pt[1].y - cnt[1].y) ** 2)
           if distance <210 and height_building<h_threshold :
               lst.append(cnt[0])
               file_nm.write("%s," % cnt[0])
           if distance < 600 and height_building > h_threshold:
               file_nm.write("%s," % cnt[0])
       file_nm.write("\n")

file_nm= open("point_neighbours.csv","w")
def _test ():
   min=0
   max=113688
   road_points_list = points(min,max)
   centroids= read_centroids()
   FDN_algorithm(road_points_list,centroids)

if __name__ == "__main__":
   _test ()
```

Appendix H - Algorithm to find intersections with buildings (openness model)

```
import shapefile as shp
import psycopg2
import math
import copy
from shapely import wkb
from shapely.geometry import Point, LineString
import shapely

#POINTS INTERSECTIONS CALCULATIONS FOR THE OPENNESS MODEL

#1 creating the classes:
class Point:
        def __init__(self,x,y):
        self.x=x
        self.y=y
        self.near_buildings_id=[]
        self.near_segments=[]
        self.id=[]
        self.link_nummer=-99
        def __str__(self):
        return "x:"+str(self.x)+" y:"+str(self.y)
        def __repr__(self):
        return "x:" + str(self.x) + " y:" + str(self.y)

class Poin_param:
        def __init__(self,x,y,t,id,h):
        self.x=x
        self.y=y
        self.t=t
        self.building_id = id
        self.building_h = h

class Ray:
        def __init__(self,a,b):
        self.a=a
        self.b=b
        def __str__(self):
        return str(self.a)+" "+str(self.b)
        def __repr__(self):
        return "a: "+str(self.a)+",b: "+str(self.b)

class Segment:
        def __init__(self,a,b):
        self.a=a
        self.b=b
        self.building_id=""
        self.h =""

class Building:
        def __init__(self,id,h):
        self.id=id
          self.height=h
        self.segments=[]

#2 creating building segments lists
def buildings_by_id(building_list):
        buildings_by_id = {}
        for building in building_list:
                building_id = building[-1]
                building_h  = building[-2]
                del building[-1]
                del building[-1]

                for item in building:
                        if building_id in buildings_by_id.keys():
                                b_obj = buildings_by_id[building_id]
                                b_obj.segments.append((item[0],item[1]))
                        else:
```

```
                            segment_list = []
                            segment_list.append((item[0],item[1]))
                            building_obj = Building(building_id,building_h)
                            building_obj.segments = segment_list
                            buildings_by_id[building_id] = building_obj
        return buildings_by_id

#3 Intersection function
def get_intersection(ray,segment):

        #RAY in parametric: Point + Delta*T1
        r_px = ray.a.x
        r_py = ray.a.y
        r_dx = ray.b.x-ray.a.x
        r_dy = ray.b.y-ray.a.y

        #SEGMENT in parametric: Point + Delta*T2
        s_px = segment.a.x
        s_py = segment.a.y
        s_dx = segment.b.x-segment.a.x
        s_dy = segment.b.y-segment.a.y

        #Checking if Ray and segments are parallel
        r_mag = math.sqrt(r_dx*r_dx+r_dy*r_dy)
        s_mag = math.sqrt(s_dx*s_dx+s_dy*s_dy)

        tolerance =0.00001
        if r_dx/r_mag==s_dx/(s_mag+ tolerance) and r_dy/r_mag==s_dy/(s_mag+ tolerance) :
        # Unit vectors are the same.
        return None

        # Solving the systems of equations ans calcutating parameters T1 and T2
        T2 = (r_dx*(s_py-r_py) + r_dy*(r_px-s_px))/((s_dx*r_dy - s_dy*r_dx)+ tolerance)
        T1 = (s_px+s_dx*T2-r_px)/(r_dx+ tolerance)

        # checking if there are point of intersections, by evaluating the parametric values.
        if T1<0 :
                return None
        if (T2<0 or T2>1):
                return None

        # Return the point of intersection
        return Poin_param(r_px+r_dx*T1, r_py+r_dy*T1,T1,segment.building_id,segment.h)

#4 Reding the buildings polygons
def read_building_sample(number):
        data= shp.Reader("Building_in_sample.shp") #use the name of your shapefile
        feat2=data.iterShapeRecords()
        pol=[]
        for i in range(number): #142314 range is the number of features you parse
                rec=next(feat2)
                geom=rec.shape.points
                geom.insert(len(geom), rec.record[12])  # re.record[12] has the heights of the buildings
                geom.insert(len(geom), rec.record[0]) #re.record[0] has the unique id of the building
                # print pol
                pol.append(geom)
        return pol

#5 reading sample points from the database
def points (min,max):
        hostname = 'localhost'
        username = 'postgres'
        password = '1234'
        database = 'geo1101'
        conn = psycopg2.connect(host=hostname, user=username, password=password, dbname=database)
        cur = conn.cursor()

        cur.execute("select unique_id,geom,linknummer from links_fietser_points20 where unique_id >= %s and
unique_id < %s;",(min, max))

        lista_points= []
        i = 0
        l2 = []
        for element in cur.fetchall():
                element = (element[0],wkb.loads(element[1], hex=True),element[2])
                lista_points.append(element)
```

```
        return lista_points

#6 transform building polygons to segments
def transform_to_segments(building,building_id):
        raw_point_list = copy.copy(building.segments)
        segment_list=[]
        i=0
        while i < len(raw_point_list):
                 point_a = Point(raw_point_list[i][0], raw_point_list[i][1])
                point_b = Point(raw_point_list[i+1][0], raw_point_list[i+1][1])
                segment = Segment(point_a, point_b)
                segment.building_id = building_id
                segment_list.append(segment)
                segment.h = building.height
                i = i + 1
                if i == len(raw_point_list)-1:
                         point_a = Point(raw_point_list[i][0], raw_point_list[i][1])
                         point_b = Point(raw_point_list[0][0], raw_point_list[0][1])
                         segment = Segment(point_a, point_b)
                        segment.building_id=building_id
                         segment.h = building.height
                        segment_list.append(segment)
                         break

        return segment_list

def frange(start, end, step):
        tmp = start
        while(tmp < end):
        yield tmp
        tmp += step

#remove end of lines
def chomp(x):
        if x.endswith("\r\n"): return x[:-2]
        if x.endswith("\n") or x.endswith("\r"): return x[:-1]
        return x

#7 reading neighbours from buffering file
def builgins_in_points_buffer():
        f = open('areas_smart.txt','r')
        relation = {}
        for line in f:
                line = chomp(line)
                line_array = line.split(",")
                relation[line_array[0]]=line_array[1:len(line_array)]
        return relation

#8 creating buildings segments
def create_segments_inside_buffer(road_points_list,buildings_by_id,builgins_in_points_buffer):
        point_list = []
        for point in road_points_list:
                p_id = point[0]
                p = Point(point[1].x,point[1].y)
                p.id=p_id
                p.link_nummer = point[2]
                if builgins_in_points_buffer.has_key(str(p_id)):
                        p.near_buildings_id=builgins_in_points_buffer[str(p_id)]
                        point_list.append(p)

                        for building_id in builgins_in_points_buffer[str(p_id)]:
                              if buildings_by_id.has_key(str(building_id)):

                                   p.near_segments.extend(transform_to_segments(buildings_by_id[building_
                                   id],building_id))
        return point_list

#9 opening  the files
file_lines = open("testing_report", "w")
file_openness = open("testing_report.csv","w")
file_openness.write("{0},{1},{2},{3},{4},{5},{6},{7},{8}\n"

        .format("point_id","linknummer","point_x","point_y","building_id","intersection_x","intersection_y"
,"height","distance"))

10 #calling the functions read_building_sample/ buildings_by_id
```

```
buildings = read_building_sample(2000) #49789
buildings_by_id = buildings_by_id(buildings)

#Adding conditions when querying to the database
min=0
max=500
ii=1
count = 0
while min <= 113688:#number of road points
        road_points_list = points(min,max)
        min= min+500
        max=max+500

        #Intersection final algorithm
        builgins_inside_buffer = builgins_in_points_buffer()
        road_points_list_with_segments=                                                         =
create_segments_inside_buffer(road_points_list,buildings_by_id,builgins_inside_buffer)

        #Add buildings segments to each point
        for point in road_points_list_with_segments:
                for angle in frange(0, math.pi * 2, (math.pi * 2) / 36):
                        # Creating the Rays
                        dx = math.cos(angle)
                        dy = math.sin(angle)
                        point_a = Point(point.x, point.y)
                        point_b = Point(point.x + dx, point.y + dy)
                        ray = Ray(point_a, point_b)
                        closestIntersect = None

                        if len(point.near_segments) == 0:
                                line = str(point.id) + "," + str(point.link_nummer) + "," + str(point.x)
        + "," + str(point.y)+ ','+','+','+','+','+','+"\n"
                                file_openness.write(line)
                        for segment in point.near_segments:
                                intersect = get_intersection(ray, segment)
                                if intersect is None:
                                        continue
                                if not closestIntersect or intersect.t < closestIntersect.t:
                                        closestIntersect = intersect

                        if closestIntersect is not None:
                                distance  = math.sqrt((point.x  -  closestIntersect.x)  **  2  +  (point.y  -
        closestIntersect.y) ** 2)

                        #Last buffering within resulting zones (buffering)
                                if distance is not None:
                                    if distance > 0.9:
                                       h = closestIntersect.building_h
                                       h1 = 12.3
                                       h2 = 18.4
                                       h3 = 55.2
                                       h4 = 152.5
                                       b1 = 45.53
                                       b2 = 68.28
                                       b3 = 205.6
                                       b4 = 568.66
                                       write_flag = False
                                       if h < h1 and distance < b1: #range [-0.7  - 12.3)
                                               write_flag=True
                                       if h > h1 and h <= h2 and distance < b2: #range (12.3  - 18.4]
                                               write_flag=True
                                       if h > h2 and h <= h3 and distance < b3: #range (18.4 - 55.2]
                                               write_flag=True
                                       if h > h3 and distance < b4: #range [55.2 - 568.66)
                                               write_flag=True

                                        #writing the files
                                       if write_flag:
                                        # convert points to wkt
                                        intersect_point_geom    =    shapely.geometry.Point(closestIntersect.x,
        closestIntersect.y)

                                        point_geom= shapely.geometry.Point(point.x, point.y)
                                        count=count+1

                                        #create linestring
```

```
                            lines_intersections_draw          =          LineString([point_geom,
intersect_point_geom]).wkt
                                    file_lines.writelines(lines_intersections_draw + "\n")


                              line  =  str(point.id)  +  ","  +  str(point.link_nummer)  +  ","  +
        str(point.x)  +  ","  +  str(point.y)  +  ","  +  str(closestIntersect.building_id)  +  ","  +
        str(closestIntersect.x) + "," + str(closestIntersect.y) + ","+str(closestIntersect.building_h)+","+
        str(distance) + "\n"

                                    file_openness.write(line)

                            if closestIntersect is None:
                                    line  =  str(point.id)  +  ","  +  str(point.link_nummer)  +  ","  +
                str(point.x) + "," + str(point.y) + ',' + ',' + ',' + ',' + ',' + ',' + "\n"
                                    file_openness.write(line)

        file_openness.close()
```

Appendix I - Algorithm to compute openness values

```
import pandas as pd

df=pd.read_csv('point_intersections_final.csv',error_bad_lines=False)   #read csv file as a dataframe (df)
df['formula']= df.distance/(df.distance+df.height)   #add a column with the formula
df['formula'].fillna(1, inplace=True)   #replace with 1 all the empty columns
openness=df.groupby('point_id')['formula'].mean()  #average openness group by pointid
df=df.set_index(['point_id'])
df['openness']=openness
df=df.reset_index()
##print df
df=df.drop_duplicates('point_id')  #distinct point id


df.to_csv('openness_pandas.csv', sep='\t')
```

Appendix J - Results linear regression analysis: shortest path as input

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95,0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	118,836	10,654		11,154	,000	97,954	139,717					
	cross_roads	-,763	,055	-,157	-13,986	,000	-,870	-,656	-,195	-,066	-,064	,166	6,020
	av_distbuilpm	210,907	91,462	,021	2,306	,021	31,639	390,174	,061	,011	,011	,264	3,792
	av_luchangepm	466,591	110,131	,094	4,237	,000	250,732	682,450	,013	,020	,019	,043	23,500
	av_dlandusepm	-102,843	118,145	-,019	-,870	,384	-334,409	128,723	,002	-,004	-,004	,042	23,780
	openness	-,670	,119	-,050	-5,616	,000	-,903	-,436	-,064	-,026	-,026	,260	3,843
	perc_separated	-,427	,031	-,073	-13,838	,000	-,487	-,366	-,082	-,065	-,063	,755	1,325
	travel_time	-,016	,002	-,074	-6,862	,000	-,020	-,011	-,213	-,032	-,031	,180	5,542

a. Dependent Variable: difference_perc

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,241[a]	,058	,058	136,376840%	,058	396,108	7	45011	,000

a. Predictors: (Constant), travel_time, av_distbuilpm, av_dlandusepm, perc_separated, openness, cross_roads, av_luchangepm

Correlations

		difference_perc	cross_roads	av_distbuilpm	av_luchangepm	av_dlandusepm	openness	perc_separated	travel_time
Pearson Correlation	difference_perc	1,000	-,195	,061	,013	,002	-,064	-,082	-,213
	cross_roads	-,195	1,000	,226	,212	,226	-,214	,090	,851
	av_distbuilpm	,061	,226	1,000	,190	,136	-,839	-,195	-,059
	av_luchangepm	,013	,212	,190	1,000	,975	-,079	,385	,118
	av_dlandusepm	,002	,226	,136	,975	1,000	-,045	,415	,124
	openness	-,064	-,214	-,839	-,079	-,045	1,000	,226	,099
	perc_separated	-,082	,090	-,195	,385	,415	,226	1,000	,105
	travel_time	-,213	,851	-,059	,118	,124	,099	,105	1,000
Sig. (1-tailed)	difference_perc	.	,000	,000	,003	,303	,000	,000	,000
	cross_roads	,000	.	,000	,000	,000	,000	,000	,000
	av_distbuilpm	,000	,000	.	,000	,000	,000	,000	,000
	av_luchangepm	,003	,000	,000	.	,000	,000	,000	,000
	av_dlandusepm	,303	,000	,000	,000	.	,000	,000	,000
	openness	,000	,000	,000	,000	,000	.	,000	,000
	perc_separated	,000	,000	,000	,000	,000	,000	.	,000
	travel_time	,000	,000	,000	,000	,000	,000	,000	.
N	difference_perc	45019	45019	45019	45019	45019	45019	45019	45019
	cross_roads	45019	45019	45019	45019	45019	45019	45019	45019
	av_distbuilpm	45019	45019	45019	45019	45019	45019	45019	45019
	av_luchangepm	45019	45019	45019	45019	45019	45019	45019	45019
	av_dlandusepm	45019	45019	45019	45019	45019	45019	45019	45019
	openness	45019	45019	45019	45019	45019	45019	45019	45019
	perc_separated	45019	45019	45019	45019	45019	45019	45019	45019
	travel_time	45019	45019	45019	45019	45019	45019	45019	45019

Table J1-J3. *Output regression analysis shortest path, including the explained variance and correlation matrix*

Appendix K - Descriptive statistics predictors: shortest path as input

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
cross_roads	45019	192	0	192	42,66	,136	28,890	834,632
av_distbuilpm	45019	.2515	.0000	.2515	.019797	.0000645	.0136852	,000
av_luchangepm	45019	.6816	.0020	.6836	.081571	.0001333	.0282922	,001
av_dlandusepm	45019	.6808	.0020	.6828	.075270	.0001250	.0265303	,001
openness	45019	71,247%	28,753%	100,000%	79,09249%	0,049804%	10,567150%	111,665
perc_separated	45019	100,000%	0,000%	100,000%	42,62113%	0,113037%	23,983799%	575,223
travel_time	45019	6402,405	1,088	6403,492	1032,35134	3,125885	663,240267	439887,652
Valid N (listwise)	45019							

Table K1. *Descriptive statistics of predictors with shortest path values as input*

Appendix L - Results linear regression analysis: difference observed route - shortest path as input

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	23,648	,983		24,062	,000	21,721	25,574					
	cross_roads_diff	,149	,076	,012	1,957	,050	,000	,297	,088	,009	,009	,596	1,678
	av_distbuilpm_diff	-939,023	95,762	-,056	-9,806	,000	-1126,719	-751,328	-,008	-,046	-,045	,664	1,506
	av_luchangepm_diff	-597,444	151,762	-,099	-3,937	,000	-894,900	-299,988	-,063	-,019	-,018	,034	29,349
	av_dlandusepm_diff	194,140	163,738	,030	1,186	,236	-126,790	515,070	-,060	,006	,005	,034	29,544
	openness_diff	-3,102	,165	-,114	-18,774	,000	-3,426	-2,778	-,082	-,088	-,087	,576	1,736
	percentage_sep_diff	,438	,045	,049	9,790	,000	,350	,526	,019	,046	,045	,856	1,169
	avg_travel_time_diff	,156	,006	,138	27,156	,000	,144	,167	,142	,127	,126	,827	1,210

a. Dependent Variable: difference_perc

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	,190[a]	,036	,036	137,965307%	,036	239,824	7	45011	,000

a. Predictors: (Constant), avg_travel_time_diff, av_distbuilpm_diff, av_dlandusepm_diff, percentage_sep_diff, cross_roads_diff, openness_diff, av_luchangepm_diff

Correlations

		difference_perc	cross_roads_diff	av_distbuilpm_diff	av_luchangepm_diff	av_dlandusepm_diff	openness_diff	percentage_sep_diff	avg_travel_time_diff
Pearson Correlation	difference_perc	1,000	,088	-,008	-,063	-,060	-,082	,019	,142
	cross_roads_diff	,088	1,000	,346	,123	,129	-,498	-,019	,347
	av_distbuilpm_diff	-,008	,346	1,000	,069	,045	-,532	-,237	,000
	av_luchangepm_diff	-,063	,123	,069	1,000	,982	,002	,257	-,027
	av_dlandusepm_diff	-,060	,129	,045	,982	1,000	,015	,270	-,027
	openness_diff	-,082	-,498	-,532	,002	,015	1,000	,173	-,003
	percentage_sep_diff	,019	-,019	-,237	,257	,270	,173	1,000	-,044
	avg_travel_time_diff	,142	,347	,000	-,027	-,027	-,003	-,044	1,000
Sig. (1-tailed)	difference_perc	.	,000	,047	,000	,000	,000	,000	,000
	cross_roads_diff	,000	.	,000	,000	,000	,000	,000	,000
	av_distbuilpm_diff	,047	,000	.	,000	,000	,000	,000	,464
	av_luchangepm_diff	,000	,000	,000	.	,000	,330	,000	,000
	av_dlandusepm_diff	,000	,000	,000	,000	.	,001	,000	,000
	openness_diff	,000	,000	,000	,330	,001	.	,000	,262
	percentage_sep_diff	,000	,000	,000	,000	,000	,000	.	,000
	avg_travel_time_diff	,000	,000	,464	,000	,000	,262	,000	.
N	difference_perc	45019	45019	45019	45019	45019	45019	45019	45019
	cross_roads_diff	45019	45019	45019	45019	45019	45019	45019	45019
	av_distbuilpm_diff	45019	45019	45019	45019	45019	45019	45019	45019
	av_luchangepm_diff	45019	45019	45019	45019	45019	45019	45019	45019
	av_dlandusepm_diff	45019	45019	45019	45019	45019	45019	45019	45019
	openness_diff	45019	45019	45019	45019	45019	45019	45019	45019
	percentage_sep_diff	45019	45019	45019	45019	45019	45019	45019	45019
	avg_travel_time_diff	45019	45019	45019	45019	45019	45019	45019	45019

Table K1-K3. *Output regression analysis difference, including the explained variance and correlation matrix*

Appendix M - Descriptive statistics predictors: difference observed route - shortest path as input

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean		Std. Deviation	Variance
	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Statistic
cross_roads_diff	45019	179	-66	113	3,62	,052	11,100	123,220
av_distbuilpm_diff	45019	,4157	-,1774	,2383	-,000992	,0000393	,0083321	,000
av_luchangepm_diff	45019	,7598	-,3983	,3615	,001710	,0001094	,0232118	,001
av_dlandusepm_diff	45019	,7472	-,3855	,3617	,001903	,0001017	,0215856	,000
openness_diff	45019	60,298%	-25,040%	35,258%	2,83929%	0,024432%	5,183990%	26,874
percentage_sep_diff	45019	200,000%	-100,000%	100,000%	0,39072%	0,074068%	15,715458%	246,976
avg_travel_time_diff	45019	3671,828	-588,418	3083,410	113,10438	,587785	124,714312	15553,660
Valid N (listwise)	45019							

Table M1. *Descriptive statistics of predictors with the differences between the observed route and shortest path as input*

Appendix N - Results of the paired-samples t-tests

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	cross_roads - cross_roads_s	3,618	11,100	,052	3,515	3,720	69,146	45018	,000
Pair 2	av_distbuilpm - av_distbuilpm_s	-9,921696E-4	8,332083E-3	3,926953E-5	-1,069138E-3	-9,152007E-4	-25,266	45018	,000
Pair 3	av_luchangepm - av_luchangepm_s	,0017097717	,0232117944	,0001093984	,0014953492	,0019241943	15,629	45018	,000
Pair 4	av_dlandusepm - av_dlandusepm_s	,0019026008	,0215856045	,0001017340	,0017032005	,0021020012	18,702	45018	,000
Pair 5	avg_openness - avg_openness_s	2,839291%	5,183990%	0,024432%	2,791404%	2,887179%	116,210	45018	,000
Pair 6	percentage_separated - percentage_separated_s	0,390722%	15,715458%	0,074068%	0,245548%	0,535896%	5,275	45018	,000
Pair 7	travel_time - travel_time_s	113,1043845	124,7143116	,5877848280	111,9523168	114,2564522	192,425	45018	,000

Table L1. *Output of the paired-samples t-tests for every predictor*

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	cross_roads	46,28	45019	28,339	,134
	cross_roads_s	42,66	45019	28,890	,136
Pair 2	av_distbuilpm	1,880518E-2	45019	1,265011E-2	5,962063E-5
	av_distbuilpm_s	1,979735E-2	45019	1,368523E-2	6,449916E-5
Pair 3	av_luchangepm	,0832807638	45019	,0261352293	,0001231767
	av_luchangepm_s	,0815709920	45019	,0282922486	,0001333428
Pair 4	av_dlandusepm	,0771727464	45019	,0242031650	,0001140707
	av_dlandusepm_s	,0752701456	45019	,0265302821	,0001250386
Pair 5	avg_openness	81,93178%	45019	10,647421%	0,050182%
	avg_openness_s	79,09249%	45019	10,567150%	0,049804%
Pair 6	percentage_separated	43,01185%	45019	22,728538%	0,107121%
	percentage_separated_s	42,62113%	45019	23,983799%	0,113037%
Pair 7	travel_time	1145,455724	45019	685,2440825	3,229589853
	travel_time_s	1032,351340	45019	663,2402673	3,125884765

Table L2. *Paired-samples statistics for every pair of predictors*