Improving the Assignment of Bicycle Traffic in a Countrywide Transportation Model by Considering Link Characteristics

Martin Pozdech

S2496712

Graduation Thesis
BSc of Civil Engineering
University of Twente

June 28, 2023

Supervised by:
Anna Grigolon
(University of Twente)
&
Sander van der Drift
(Dat.mobility)

Dat.mobility

UNIVERSITY OF TWENTE.

Abstract

With cycling being the second most-used mode of personal transport in the Netherlands (Statistics Netherlands, 2022), there is a clear need to be able to predict not only how many cyclists are on the roads right now, but also how their distribution across the network changes when the network changes. This is done today using traffic modeling, with one such model being the OmniTRANS Spectrum (*Mobiliteitsspectrum*). The project was commissioned by the model's operator Dat.mobility and they wanted to improve the accuracy of the bicycle traffic assignment in it. While this model can simulate traffic across the whole Netherlands, only the Arnhem-Nijmegen region was investigated in this work. Because the results from this area were to be applied to the whole model, restrictions were put in place such that any adjustments to improve accuracy must be based on attributes that are attached to the links and that the adjustments are done to the link speed values.

The investigation performed was done to answer research questions that asked a) if the observed counts of cyclists in this region could be used to determine the impacts of demand factors—characteristics of the link which influence its appeal—on traffic intensities, and b) if the traffic model in question could be made more accurate using the findings from the performed analysis.

To achieve this, three different approaches were tried. The first two were unsuccessful and involved regressions attempting to fit the demand factors onto some form of the observed traffic. The final approach—which succeeded in improving the accuracy—involved the use of a Calibration Coefficient (CC). It could identify the features which significantly impacted traffic, as well as scale the speeds to incorporate this impact in the model. This approach involved the use of a calibrated zone-to-zone model to provide observed intensities for near-all links in the node-to-node model being adjusted. Because of this differing construction, a novel metric was developed to validate the results in city centers where significant deviations from reality were observed. It compared the number of cyclists and the distances they cycled within these areas.

Ultimately, the adjustments found were not applied to the full model, due to limited time. While the identified CC values, and thus their resultant speeds, could improve the overall distribution of traffic in the full model, the values themselves are not representative of cyclists' behaviors throughout the whole Netherlands. The CC values are just a representation of the difference between the calibrated and default models within the study area per some characteristic—they give no insight into how appealing a characteristic is for a cyclist.

Preface

The research work described hereafter is the capstone to my studies as a Bachelor of Civil Engineering at the University of Twente. This project took place at Dat.mobility where I got an excellent look at how data analysis and transportation engineering can be combined in a commercial setting. To all those from Dat and Goudappel who helped me along this 10-week journey, I would like to express my overwhelming gratitude to you for helping me develop not only professionally, but personally as well—there are far too many of you to name individually here.

There is one name I cannot omit however: thank you to Sander van der Drift, my supervisor at Dat, who was always ready and able to provide a second point of view to any problems I was facing. Without his input, the results of the project would have ended up being far less satisfying.

Thank you to Anna Grigolon, my supervisor at the UT, who made sure I never lost sight of the academic aspect of this work. Without her, it may have ended up being a technical report rather than a thesis. I hope her introduction to supervising BSc students was as pleasant as it was for me to work with her.

Thank you to my parents Julius and Renata, who gave me the opportunity to study abroad at the UT. And finally, thank you to my all friends who supported me throughout the past three years.

If I was to sum up what my work at Dat entailed, it would be looking at maps, analysis, and programming—all things which I would do for fun in my free time, so this assignment was an excellent match for me. And I hope it will be as enjoyable for you to read as it was for me to execute.

Martin Pozdech
25th of June, 2023
Enschede, the Netherlands

Contents

1 Introduction

Transportation networks form the backbone of modern civilization, acting as a country's circulatory system by moving goods and people between different destinations. When dealing with such an important and expansive system, it is useful to be able to predict what the number of agents using said network is at any given point. Today, we accomplish this task by using traffic modeling; a powerful tool that allows actors to easily investigate various measures pertaining to the state of traffic on a network. Because of the predictive nature of traffic modeling, actors can also look at what effects changes to the network would have on traffic. This can be done quickly and without physical intervention, giving actors immediate insight into the interactions between the network and its—simulated—users.

These users being simulated is an important distinction to make, as every traffic model is only an abstraction of the real traffic situation. Thus it relies on many assumptions about the factors related to any person's travel behavior. Depending on the mode of travel, these factors can have varying impacts on the results of the model. For example, a car is a large vehicle whose users are protected from collisions by a metal frame. Compare this experience to that of riding a bicycle: you are far more exposed to your surroundings, meaning your perceived safety of a route will depend much more on the properties of the links used. Consider the benefits a solitary bike path has over an unprotected painted lane in this respect—one offers a physical separation between two modes of transport whose kinetic energies differ by orders of magnitude. This is just one example of how different modes will have different criteria with which their users will assess and choose their routes.

For cyclists, these criteria can range from physical factors such as the infrastructure present on a road to subjective factors such as the perceived beauty or comfort of a route. These factors will be further elaborated on in this work, but suffice it to say that there are many different types of features that influence a cyclist's route choice. While studies have been done into the effects of these features in the Netherlands (Meijning, 2019; Bernardi et al., 2018; Genugten and Overdijk, 2016), none have been able to define the impact they have on a cyclist's route choice in a form that can be easily translated into a traffic modeling framework—they lack concrete values on how specific features of a link affect the volume of traffic present on it. That is the knowledge gap that this research project will address.

The following chapter will go on to introduce the background of this research project as well clarifying its limitations, scope, and objectives. After that, the theoretical background will be discussed and the methodology for the project will be put forth. Lastly, the results are presented and discussed before finally concluding with remarks regarding possible future work in this area.

1.1 Project background

The project was commissioned by the firm Dat.mobility—a subsidiary of the transportation engineering consultancy group Goudappel—that specializes in mobility data analysis. They operate a traffic model called the OmniTRANS Spectrum (*Mobiliteitsspectrum*) that can provide traffic intensities as well as many other statistics about the whole Dutch transportation network, and it can do this for pedestrians, cyclists, cars, public transportation, and even freight traffic. In this model, Dat.mobility wants to improve the assignment of bicycle traffic such that is it more representative of the real situation. Specifically, they believe that the model is significantly overestimating the amount of cyclists in city centers. As such, their objectives for the project are to confirm if this is happening and, if it is, to implement a fix for this discrepancy into the model. This fix is to take the form of adjustments to the links' speed parameters.

1.2 Research scope

As the Spectrum model covers the whole of the Netherlands and Dat.mobility wishes to improve the accuracy for the whole model, at its broadest the scope covers this same area. However, as running and analyzing the model over such a large area would be computationally very intensive, a sub-section of the model was created. This version includes a detailed network—identical to that present in the full model—of the Nijmegen-Arnhem region while only having a simplified network outside of this study area. Dat.mobility hopes that the insights found using this region can be extrapolated to the whole Dutch network.

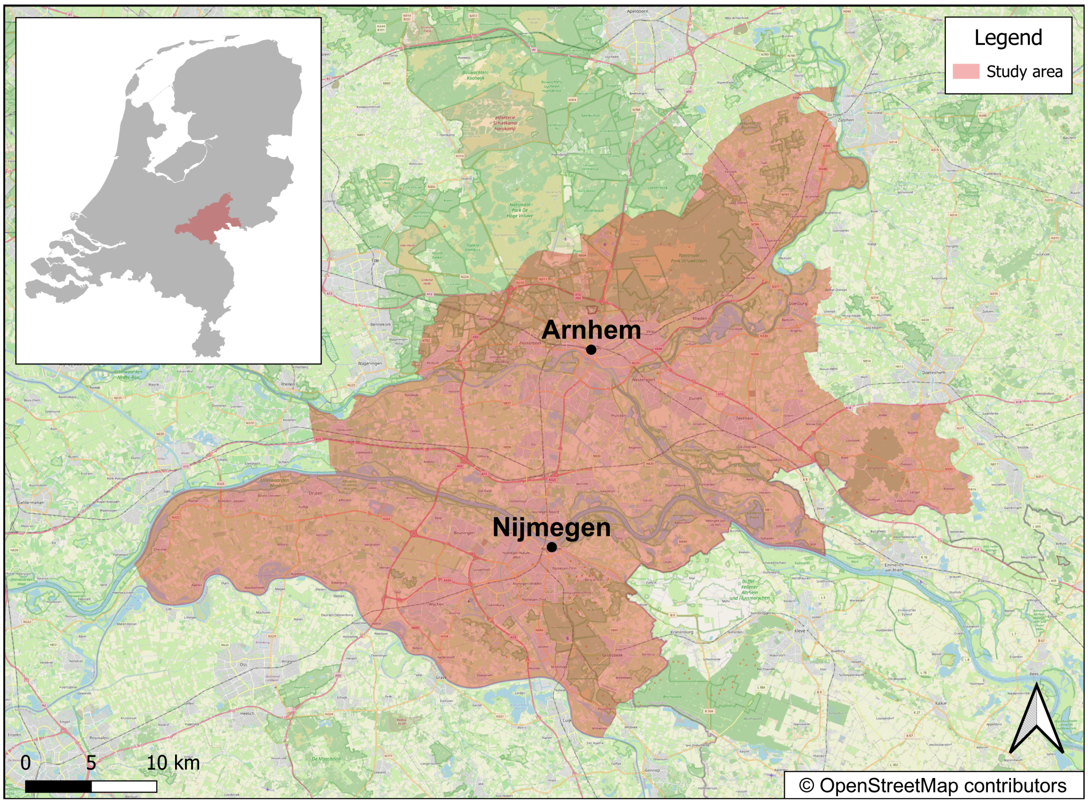

Figure 1: The Arnhem-Nijmegen area being studied

Because of this need to apply any changes to the countrywide network, any adjustments made to the study area network cannot be manual. That is to say, the adjustments must be dependent on some attribute either already attached to the link, or one that can be (spatially) attached, such that any link with said attribute would be affected by the change. This requires the use of existing databases that contain information about the Dutch cycling network. One such database is maintained by the Dutch Cyclists' Union (*Fietsersbond*), which contains detailed information about the infrastructure, quality, surroundings, and many other factors for all cycling links in the Netherlands. It will be the primary source of data regarding link characteristics referred to in this project.

1.3 Research objectives

The problem put forth by Dat.mobility is that they believe their model to not be as accurate as it could be, and thus they wish to improve it. The author's hypothesis is that this issue could be addressed by using the link characteristics to adjust their speeds. Therefore, the research objective is to improve the assignment of bicycle traffic in the OmniTRANS Spectrum model by adjusting the link speeds, based on the link's characteristics. To achieve this objective it is necessary to formulate research questions that the project will address, the first of which is about the identification and valuation of the factors that impact a cyclist's route choice.

- **Q1: Can a spatial analysis approach be used on count data to determine the impacts of demand factors on bicycle traffic intensities in the OmniTRANS model?**

In prior studies we saw the use of GPS route data to investigate the impacts of demand factors on route choice; however, could this same investigation be done with a spatial analysis on observed bicycle count data? This leads to following sub-questions:

- Q1a: What are the factors that impact bicycle traffic demand?
- Q1b: How would a spatial analysis of these demand factors look like?
- Q1c: What are the steps necessary to ensure the validity of the spatial analysis?

The second primary research question relates to the implementation of the factor impacts into the model.

- **Q2: Can the OmniTRANS model be made more accurate by considering demand factors related to a cyclist's route choice?**

While the literature review suggests that a bicycle traffic model can indeed be made more accurate by considering various factors, it is necessary to ask how the impacts of said factors can be connected with the model.

- Q2a: How can the influence of demand factors be implemented into the model?
- Q2b: Can the insights from a subsection of the entire model be extrapolated to the whole network?
- Q2c: Considering that all people have different (subconscious) valuations of the demand factors, is it appropriate to extrapolate a single weight for any one factor?

With the objectives for this research work laid out it is necessary to establish what the current state of knowledge in this area is. As such, let us now move on theoretical background of the project.

2 Theoretical background

2.1 Literature review

The idea of improving the route choice of cyclists in traffic models is not a novel one. There have been many studies conducted in various countries using varying methods to identify the features that influence a cyclist's route choice, often with the goal of improving traffic modeling for this mode. Table 1 contains an overview of the studies that were looked at in this literature review together with the characteristics they investigated and the method used in said investigation.

Table 1: List of select cyclist route-choice studies, their method of analysis, and their studied factors

Study	Method	Studied factors
Meijning, 2019	RP	Bicycle facilities, road quality, road surface, intersection type
Bernardi et al., 2018	RP	Bicycle facilities, road quality, surroundings, trip length, signalization
Jensen et al., 2018	RP	Bicycle facilities, surroundings, no. of car lanes, slope, turns, trip length
Broach et al., 2012	RP	Bicycle facilities, intersection type, slope, turns, trip length, car traffic volume
Hardinghaus & Nieland. 2021	RP	Bicycle facilities, road surface, intersection type, surroundings, lighting
Shin, 2016	SP	Bicycle facilities, surroundings, trip motive, trip length, cyclist comfort
Genugten & Overdijk, 2016	SP	Bicycle facilities, road quality, slope, travel time, intersection type, car speed

These investigations into what features of a link are desirable and undesirable for cyclists are conducted using two primary methods: stated-preference (SP) and revealed-preference (RP). Stated-preference studies use expert elicitation and surveys to directly ask cyclists to state the criteria they consider when selecting their routes. This approach is good at identifying the factors which cyclists themselves perceive as important for their route choice. Revealed-preference studies make use of collected data where cyclist behavior was observed and analyzed to reveal the characteristics that influence their routes. Such an approach allows the quantification of factors and their influence on cyclist behavior. In this section, what the factors that influence cycling behavior are, and the different approaches used to evaluate their impacts, will be presented through the existing literature on the subject. The result of this literature review will be a collection of factors that will be investigated to improve the assignment of bicycle traffic in the OmniTRANS Spectrum model.

2.1.1 Revealed-preference studies

A common trend for revealed-preference studies was to use GPS data of bicycle trips to investigate the characteristics of the routes cyclists take. How these studies differ in their approach is in regards to how they select the alternative routes against which the observed routes are compared. One option is to use the shortest route that can be taken between the observed trip origins and destinations. This approach was used by Meijing in their 2019 study assessing the impact of bicycle infrastructure in the Dutch province of Noord-Brabant. They used a linear regression analysis to test if there was a relationship between the amount of deviation from the shortest route and the differences between the network and infrastructural factors on the observed and shortest route. It resulted in a significant deviation only for signalized intersection, where cyclists would prolong their trip if it meant avoiding this feature. (Meijning, 2019)

One reason why Meijing's approach could have failed to produce significant results for more factors was the shortest alternate route selection—the shortest route is not always a realistic alternative for cyclists. In the 2012 report by Broach et al. they found that, when applied to their dense network, the shortest alternate paths generated would leave and return to the same corridor multiple times. Instead, they chose to generate alternate routes by maximizing individual criteria, subject to multiple distance constraint values. The process is described in Broach et al., 2010, where the generated alternatives were calibrated against the observed route deviations. To then assess the probability of one of these alternatives being chosen by a cyclist for a trip the Path-Size Logit (PSL) model is used. (Broach et al., 2012)

In the 2012 study by Broach et al. they found that dedicated bicycle facilities (protected paths and boulevards) had a significant positive increase on utility, meaning cyclists would prolong their trips to include these features. The presence of a bike lane was not found to be a significant driver; however, the authors stipulate that bike lanes appear to offset the negative effects of adjacent traffic but have no residual value of their own. They go on to state that, while bike lanes reserve a space for cyclists on the road, they are no more or less attractive than a basic low traffic volume street. For negative influences, they found slope, traffic control measures, and traffic intensity to be drivers of decreasing utility. Cyclists would be willing to significantly prolong their trips to avoid these features. (Broach et al., 2012)

One study that used a particularly interesting approach was the 2021 paper by Hardinghaus and Nieland. There they used user queries from a bike-routing engine for the city of Berlin (BBBike) to build their set of observed routes. These were grouped based on the settings the users would select in the engine, with options such as avoiding traffic lights, unlit streets, and bad surfaces. Users could also specify the type of road and if they preferred green pathways. To then evaluate which of these settings were significant drivers of cyclist routing, a hierarchical cluster analysis was performed. It could group the user generated routes based on their similarity which revealed five clusters of user preferences. They found that, for over more than half of their 450,000 analyzed data points, users selected to avoid bad surfaces. This was the most significant finding, but the authors stipulate that Berlin has much of its residential streets paved using cobblestone, which may be influencing the outcomes. Their other findings include the importance of side streets and dedicated facilities. (Hardinghaus and Nieland, 2021)

In the 2018 study by Jensen et al. they used a variation of the PSL model to better represent the behavior of cyclists in a traffic model of Copenhagen. Their results showed that the number of turns and increasing slope had a negative effect on perceived travel time. When investigating link surroundings, they found that there were no significant impacts on perceived travel time by the environment present on the left side of the link. For the right side of a link—when compared to reference surroundings consisting of parks, water features, cemeteries, and sports facilities—the perceived travel times increased in both high and low-density settlements, industrial areas, and forests. Regarding bicycle facilities, they found that protected bicycle lanes decreased the perceived travel time significantly, but that bicycle paths that are not adjacent to a road increase perceived travel time for slow cyclists. The authors try to explain this by suggesting that slow cyclists are especially sensitive to distance, therefore they are unwilling to deviate to include such a feature. This sentiment that cyclists are extremely sensitive to distance is echoed in other papers (Bernardi et al., 2018; Broach et al., 2012; Hardinghaus and Nieland, 2021).

Noticeable is how in the 2018 paper by Bernardi et al. they used observed alternate routes instead of generated (simulated) alternatives. They found that 41% of their recorded trips were repeated more than once, allowing them to build a choice set that, while smaller than in other studies, has all its alternatives be both realistic and considered as valid options by cyclists. As a reference route, they calculated the shortest possible route between each OD-pair and then proceeded to group the alternate routes based on their length. This approach resulted in traffic signalization to be a positive and significant factor, meaning cyclists would prolong their routes to include such features. Conversely, bicycle lanes and bike boulevards were found to be avoided by cyclists in relation to roadway links. This appears to coincide with the findings of Broach et al., 2012, where bicycle lanes were found to be no more attractive than a basic residential street. As for the influence of surroundings, from a baseline of "fair level of beauty", they found that cyclists who took the shortest route had a negative perception of beauty, meaning they would prefer shorter trips over those that had beautiful links. Again we can observe the sensitivity to distance. They also found that good-quality links had a strong positive influence on route choice, compared to a reference of fair-quality, which was to be expected. Finally, regarding trip motive, only leisure was found to be a significant positive driver of longer routes. (Bernardi et al., 2018)

2.1.2 Stated-preference studies

Another study that looked at trip motives and how they effected route choice was the 2016 study by Shin that grouped their observed routes into to-work and back-home trips. While this was a stated-preference study, they used GPS-mapped routes of selected cyclists that they would interview about the characteristics of said routes. They also recorded alternative routes that a cyclist would take, who were then questioned as to what compelled them to choose one alternative over another. Their findings showed that cyclists traveling to work would significantly prefer routes with safe, downhill sections that minimized the time spent travelling. For back-home routes, they found that cyclists significantly preferred routes that featured good scenery, which is in-line with expectations. When comparing primary routes with alternatives, they found minimizing distance and time was significantly more important for main routes while alternatives had significantly better scenery and—at a lower significance—better cycling facilities. Alternative routes were, on average, longer than main routes between 1 and 1.4 kilometers. (Shin, 2016)

Finally, we look at a stated-preference study by Genugten and Overdijk, 2016, where they asked 728 Dutch cyclists to evaluate a number of link features and rate their importance in route choice. These were, in order from most to least important: type of bicycle facility, road surface quality, slope, reduced travel time on short distances, non-priority intersections, speed of other car traffic, reduced travel time on long distances, signalized intersections, and priority intersections. Specific insights include that, as trip distance increases, so does the influence the road surface quality has on route choice. Older cyclists were also found to attach greater value to protected bicycle paths. We can also observe the importance of travel time, and that this importance is greater for short trips, which coincides with the findings of other papers discussed here (Shin, 2016; Bernardi et al., 2018). (Genugten and Overdijk, 2016)

Of note is that, when asked to state their preferences regarding intersection types, signalized intersections were found to be preferred at a 4% greater rate compared to uncontrolled intersections. This goes against the findings of Meijning, 2019, and Broach et al., 2012, but are conversely supported by the results of Bernardi et al., 2018. If we discard the findings of Meijing—due to a lack of realistic alternatives—this difference could be explained as a result of different cycling behaviors between the United States and the Netherlands. For example, even in US cities with better than average cycling facilities—such as Portland with its bicycle lanes and boulevards—intersections are not always optimized for use by cyclists. Turning left (across oncoming traffic) can be inconvenient and even unsafe if the intersection does not have a dedicated signal cycle that allows cyclists to make this turn. Even more so if the intersection has multiple car (turning) lanes. Such a problem is near-nonexistent in the Netherlands, where a great majority of intersections have good accommodations for cyclists. Therefore, when cyclists are given good facilities at intersections, we could infer that they seek out signalized intersections that guarantee them a safe and reliable crossing of busy car links.

2.1.3 Results of the review

One immediately common theme throughout the review was the dependence on route data to investigate the impacts of features on cyclist behavior. Having to collect data at such an individual level (per cyclist) goes to show just how individualistic cycling preferences are. How much a person values a certain feature will depend on their age, trip motive and distance, and even how experienced they are with cycling. Some of the papers discussed above aggregated the factor impacts based on these population groups, from which they were able to find if a certain population group was more likely to choose a path based on its characteristics. Take for example the 2018 study by Jensen et al., where they found slow cyclists to perceive solitary paths as 183% longer while fast cyclists perceived them as 203% shorter, compared to a baseline of no accommodations.

With the current state of knowledge regarding the factors that affect route choice for cyclists presented, it is now necessary to specify the factors that could influence bicycle traffic demand and what their possible values area. These are the factors that will be investigated by this research work and they are presented in Table 2.

Table 2: List of factors that influence bicycle traffic demand

Factor	Associated values
Bicycle facility	No accommodations, bicycle lane, bicycle path, bicycle street
Intersection control	Uncontrolled, signalized, roundabout
Car volume	Range from 0 to peak hourly intensity, per link, if applicable
Surroundings	Green (nature), blue (water features), urban, landmarks
Slope	Range for the average incline.
Lighting	Well lit, moderately lit, unlit
Road surface	Paved, unpaved
Surface quality	Good, middling, poor

Car volume is included as it can be a factor itself (Broach et al., 2012), but it can also be used as a proxy for car road classification—residential roads will have lower traffic intensities compared to primary links between towns. A bicycle lane is understood as an unprotected (painted) lane on a road accessible to cars, while the bicycle path is physically separated from car traffic.

2.2 Conceptual framework

To guide the project from a theoretical state to its conclusion, a conceptual framework was developed, shown in Figure 2. It illustrates how the knowledge taken from the literature review will be used to progress the project from conception to completion. These sources of knowledge are highlighted in red and they are connected to the stages of the project where they will be used. The framework terminates at the evaluation of the adjustments made to the model, which is the project result that Dat.mobility is interested in.

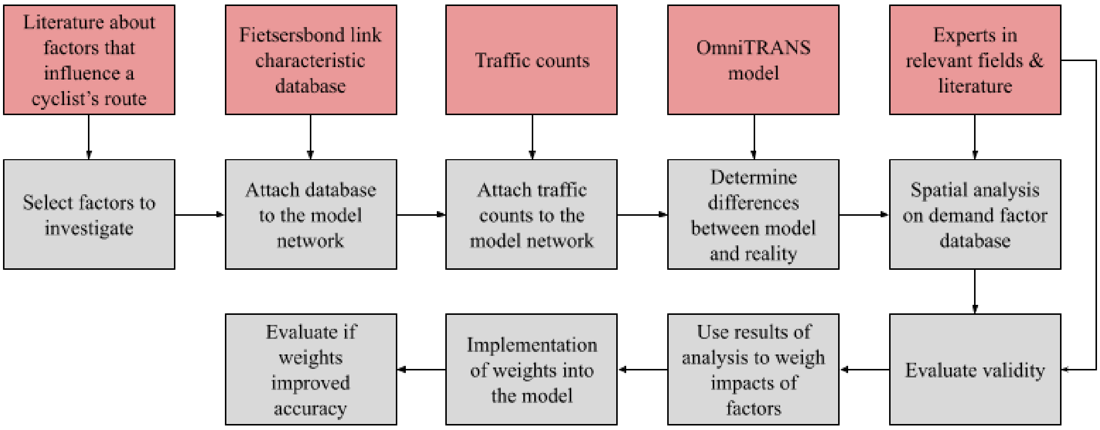

Figure 2: The conceptual framework for the project

2.3 Model construction

Throughout this work there are three different traffic models that will be referred to. This section will explain what the differences between these models are and how they are constructed, while also elaborating on some of the technical jargon used throughout the paper.

2.3.1 Node-to-node model

A node-to-node (n2n) model refers to the locations where agents—the cyclists being simulated—are spawned within the transportation network. These nodes are located at the connections between different links in the network. It is at these intersections where the agents start their trips and where traffic demand is first registered by the model, thus creating the traffic intensity present on a link. Figure 3 depicts an area from the model, where the total attraction and supply of agents from this zone is spread throughout the nodes in it. That is to say that any agent whose trip terminates in this zone can have their actual terminus be any one of the nodes.

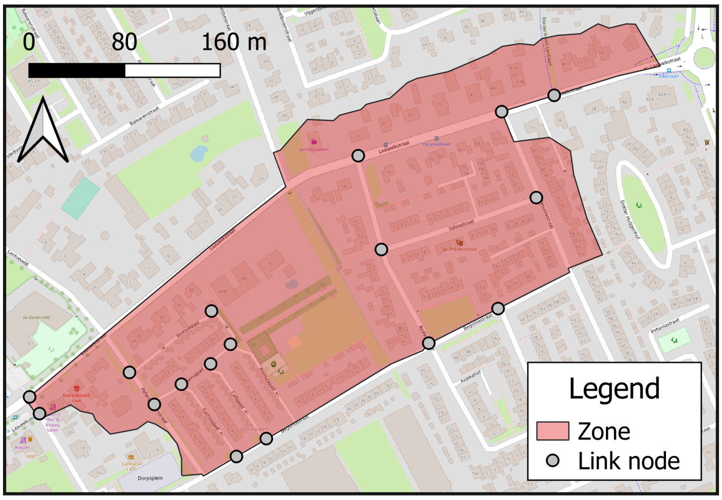

Figure 3: A sample zone and its intersection nodes

The amount of traffic going to or coming from each of these nodes is governed by a table where a percentage of the total traffic from the zone's Origin-Destination (OD) matrix is assigned to each node. That is how the OmniTRANS Spectrum model can depict traffic originating from and terminating at any address in the Netherlands, resulting a very granular representation of the traffic demand. Here, traffic demand and intensity both refer to the amount of traffic present on a link over some time period, often referred to as agents per hour. For this project, cyclists per day (c/day) will be used, as we are interested in the overall traffic present on the link throughout the whole day and not how it is distributed throughout the day.

Link speed theoretically refers to the speed at which an agent can move from the start of the link to its terminus, given in kilometers per hour. In practice, the speed is the cost that an agent must pay to move along a link. Which links the agent chooses to use depend on said cost, as the agent will attempt to minimize the total cost required to travel between his origin and destination. This is often referred to as all-or-nothing assignment and is how all of the models discussed here operate. By default, these speeds are assigned based on the facility, surface, and length of a link, with an additional modifier depending on if the link is inside or outside of a built-up area.

The version of the OmniTRANS Spectrum model provided to the author was written using batch script and SQL, with some dependencies on Java for calculations. The model is executed using a series of batch files from the command-line interpreter that is connected to a PostgreSQL database where the results are stored. It is this n2n model that should be optimized by adjusting the speeds.

2.3.2 Zone-to-zone model

What makes a zone-to-zone (z2z) model different from a n2n one is that the agents are not spawned at the nodes of the network, but on feeder links that connect the centroid of a zone to the road network. In the models used here, these zones take the shape of neighborhoods (*buurten*). Figure 4 depicts one such zone, with its associated feeder link shown using a black arrow. This feeder is connected to an intersection node on the network, where all of the traffic going to or from this zone will have their trips either terminate or originate.

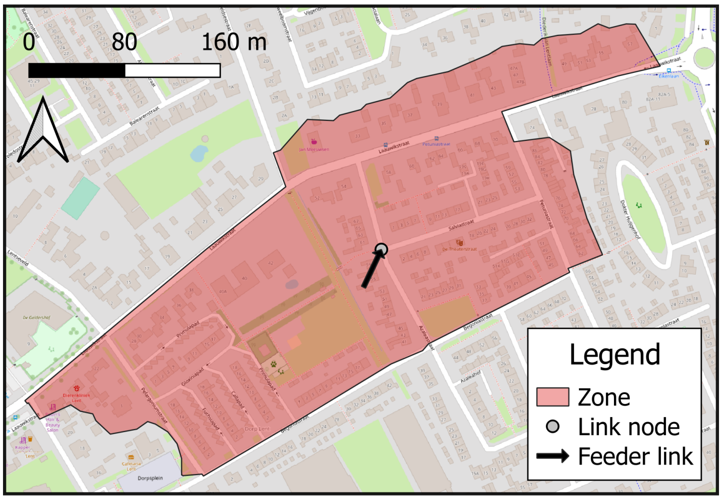

Figure 4: A sample zone and its feeder link

This is unlike a n2n model, where the attraction and supply is spread across all of the nodes in a zone, rather than being aggregated to a single point. The benefit of the z2z approach is that it is faster to run, but that comes at the cost of a reduced resolution in its outputs. These outputs are shown in Dat.mobility's proprietary OmniTRANS Desktop software which uses the Ruby programming language to calculate its results.

2.3.3 Calibrated commercial model

The final model discussed in this paper is a z2z model made by Goudappel for the transportation authorities of the Arnhem-Nijmegen region. Considering that this model was accepted by the client, it serves as a good benchmark for what an acceptable representation of the cycling situation in the study area is. However, as this model is older than those discussed previously, it has some differences in both the underlying network and the socio-economic data used. These differences can take the shape of different digitizations of the network, different values for link speed, or different OD matrices that govern the amount of traffic generated.

The model being calibrated means that its outputs were adjusted to fit a set of observed traffic counts for the region. Specifically, this calibration was done by adjusting the attraction levels for a zone until the traffic being spawned matched that which is observed (G. Wiersma, personal communication, May 25, 2023). What this project will attempt to do is a calibration of the n2n model, but instead of adjusting the OD matrix, it will attempt to influence the routing by changing the speeds. This is a novel method of doing a traffic model calibration, so let us move on to the methodology of how this will be accomplished.

3 Methodology

The following section will cover how this quantitative research project will achieve its results through the use of different approaches to determine what and how significant the relationships between bicycle traffic and link characteristics are. First, it is necessary to first present the state of both the available databases and the model, thus operationalizing what factors are possible to investigate and what the benchmarks for any improvements to the model are. Then, how each of the research questions will be addressed in the project workflow. Finally, the steps taken to pre-process the available data are discussed.

3.1 Operationalization

To establish what the factors that can be investigated using the Cyclists' Union database are, we can compare them with those stated in Table 2. These factors and their associated values within the database are shown in Table 3. It should be noted that the values stated here are translated from Dutch by the author and are filtered based on relevance to the project. An unedited breakdown, which includes the descriptive statistics for these characteristics, can be found in Appendix A - Fietsersbond database statistics.

Table 3: List of factors present for each link in the *Fietsersbond* database

Factor	Associated characteristics
Facility	Solitary path, protected path along a road, bicycle street, painted (unprotected) lane, normal road, service road, pedestrian area
Surface	Asphalt/concrete, tiles, bricks, other, semi-paved, unpaved
Surroundings	Rural village, fields, nature, forest, built-up (green), built-up (no greenery)
Water	Yes (present), no (not present)
Intersection*	Signalized, uncontrolled, roundabout, not applicable
Lighting	Well lit, somewhat lit (e.g. only at intersections), unlit
Hindrance	Very little, little, reasonable, a lot
Quality	Good, reasonable, bad
Beauty	Picturesque, nice, neutral, ugly, very ugly
Built-up area	Yes (inside), no (outside)
Max speed	15 km/h, 30 km/h, 50 km/h, 60 km/h, 70 km/h, 80 km/h,
Average slope	<1%, 1-2%, 2-4%, >4%
Max slope	< 1%, 1-2%, 2-4%, 4-6%, 8-10%, 10-20%, >20%
Crossing	Above grade, below grade, tunnel, ferry, none

* An intersection is understood as an at-grade crossing of two different modes.

When compared to Table 2, we see almost all of the factors represented in the database, with the only one missing being car traffic volume. This is something that could be found using a car traffic model from Dat.mobility, but the complexity of incorporating these results was considered to be too great for the insight it provided. Instead, the hindrance characteristic could be used, as it describes the nuisance other traffic poses on a link—examples of this hindrance could be cars parked on the bike lane or busy car roads with poor bicycle facilities.

It should be noted that this database is built using community-sourced information about the links, meaning it is often up to the contributor to decide how a characteristic is classified for a link. This subjective assessment can pose issues with consistency both within the database and between the database and reality. No attempts were made to try and correct any issues with this database, and it was left in the state that it was retrieved on the 28th of September, 2018.

3.2 Traffic counts

The best way to establish a model's accuracy is to compare it against observations of the modeled phenomena in reality. When talking about a traffic model's accuracy, these observations are traffic counts. That is what the second database of note for this project contains—the observed traffic counts of cyclists in the Arnhem-Nijmegen region. Where these counts took place can be seen in Figure 5.

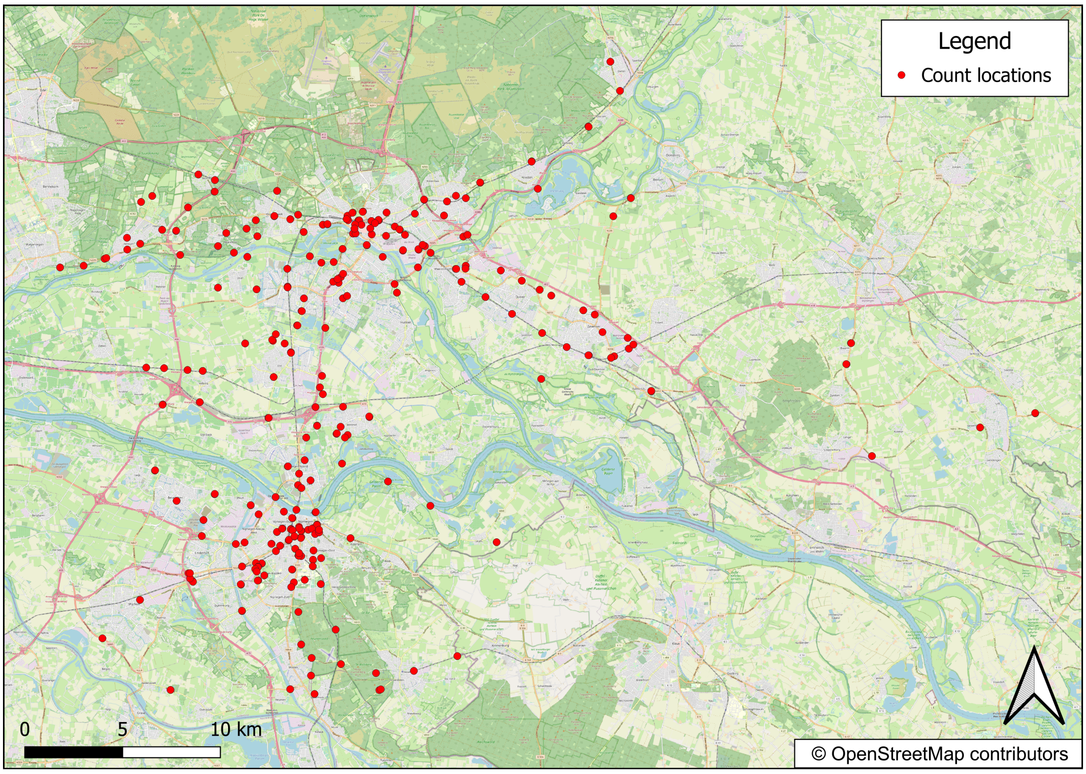

Figure 5: The locations of the cyclist counts

In total, there were 480 counts made within the study area, with one count for each direction of travel on a link. These were provided together with the calibrated z2z model and no information is available as to their source or date of collection. They contain the traffic intensity for the whole day in addition to intensities during the morning and evening peaks. Not all 480 counts could be used for the project however, as some were located outside of the study area or were on links whose shape was so different that they could not be easily connected to the assignment network. That left 466 count locations, which we will now use to evaluate the accuracy of the default n2n model assignment.

3.3 Default model accuracy

To evaluate the accuracy of the n2n model and any adjustments made to it, it necessary to first establish what these measures of accuracy are. One option is to plot the model intensities (x-axis) against the observed counts (y-axis) using a scatter plot, as seen in Figure 6. Here, each point represents one count location, with separate points for different directions on the same link. The red line is a fitted linear regression curve with an R^2 value of 0.6648, indicating that the curve explains 66.5% of the variation in the observed counts.

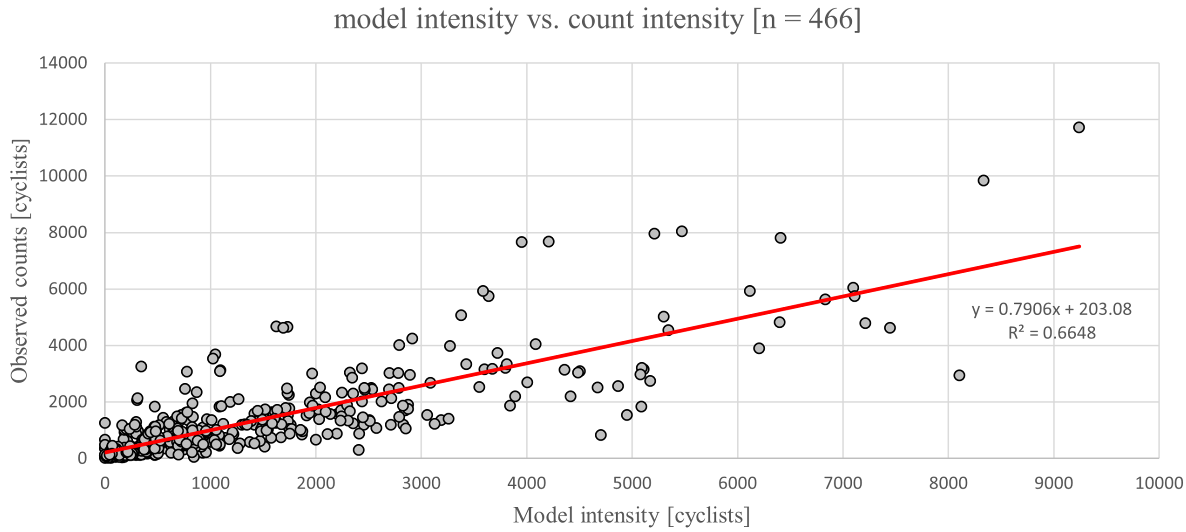

Figure 6: Scatter plot of the model data against the count data

Observe how the points are scattered closely around the regression line, indicating the model is already producing results of substantial accuracy. This is further supported by the high R^2 value. To check if the differences between the model and counts are randomly distributed, we can plot the standardized residual of traffic against its probability of occurrence, where traffic residual is the observed count subtracted from the model intensity. This is shown in Figure 7.

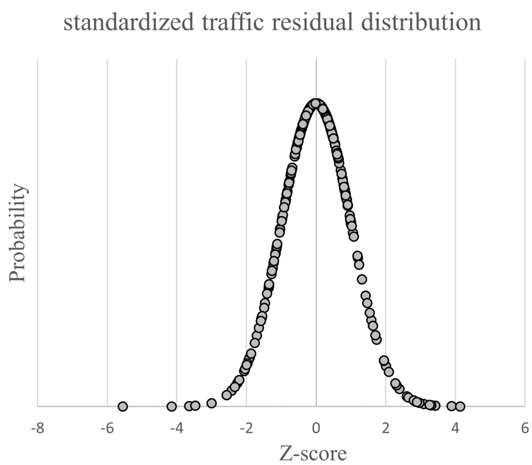

Figure 7: Standardized traffic residual probability plot

The distribution of residuals closely follows that of a bell curve, indicating that there is no pattern to their distribution. The slight left-tail skew indicates that the most extreme differences between the model and reality occur on links where traffic is being overestimated. This may explain why the average traffic residual is -78 cyclists for the count data set.

Earlier in this section the coefficient of determination R^2 and its use as a measure of model performance was introduced. As this value should be calculated for the model directly, and not in respect to some curve, it is necessary to go over how it is calculated.

3.3.1 Coefficient of determination - R^2

The coefficient of determination is calculated by subtracting the residual sum of squares (RSS) divided by the total sum of squares (TSS) from one. Eq. 3.1 shows the equation for the calculation of RSS, where y_i is the traffic count for location i (observed value) and x_i is the model intensity at location i (calculated value).

$$RSS = \sum_i (y_i - x_i)^2 \tag{Eq. 3.1}$$

Eq. 3.2 shows the equation for the calculation of TSS, where $\bar{y}$ is the mean traffic count over all traffic count locations i. The two are then combined in Eq. 3.3 to calculate R^2.

$$TSS = \sum_i (y_i - \bar{y})^2 \tag{Eq. 3.2}$$

$$R^2 = 1 - \frac{RSS}{TSS} \tag{Eq. 3.3}$$

What R^2 does is present the accuracy of the model as a percentage, rather than an absolute value. To get accuracy as an explicit amount of cyclists a different measure is computed, called root-mean-square error.

3.3.2 Root-mean-square error - *RMSE*

Root-mean-square error is a measure that aggregates the residuals at individual data points into a combined measure of precision, similar to that of standard deviation. Eq. 3.4 shows how RMSE is calculated, where n is the number of data points, or in this case, the number of count locations.

$$RMSE = \sqrt{\frac{RSS}{n}} \tag{Eq. 3.4}$$

Both RMSE and R^2 are measures common in data analysis, but this paper will also make use of some other measures that are specific to its subject matter.

3.3.3 Calibration coefficient - CC

One measure that is firmly rooted in traffic modeling is the calibration coefficient. It is a unitless quotient of the observed traffic over the modeled traffic, for a link, as seen in Eq. 3.5. The benefit of this measure is that it directly states by how the traffic on a link must be scaled. (Bostanci et al., 2018)

$$CC = \frac{y_i}{x_i} \qquad \text{(Eq. 3.5)}$$

Calculating this value for each link within the study area—with the calibrated model outputs acting as the observed counts y_i—their frequency of occurrence can be plotted, as seen in Figure 8. Note that these are filtered to exclude CCs equal to zero (n = 18009) and only consider intensities along the direction of digitization; however, both directions have very similar graphs. Additionally, the stated mean and median ignore CCs equal to or greater than 7, as these were considered to be extreme outliers.

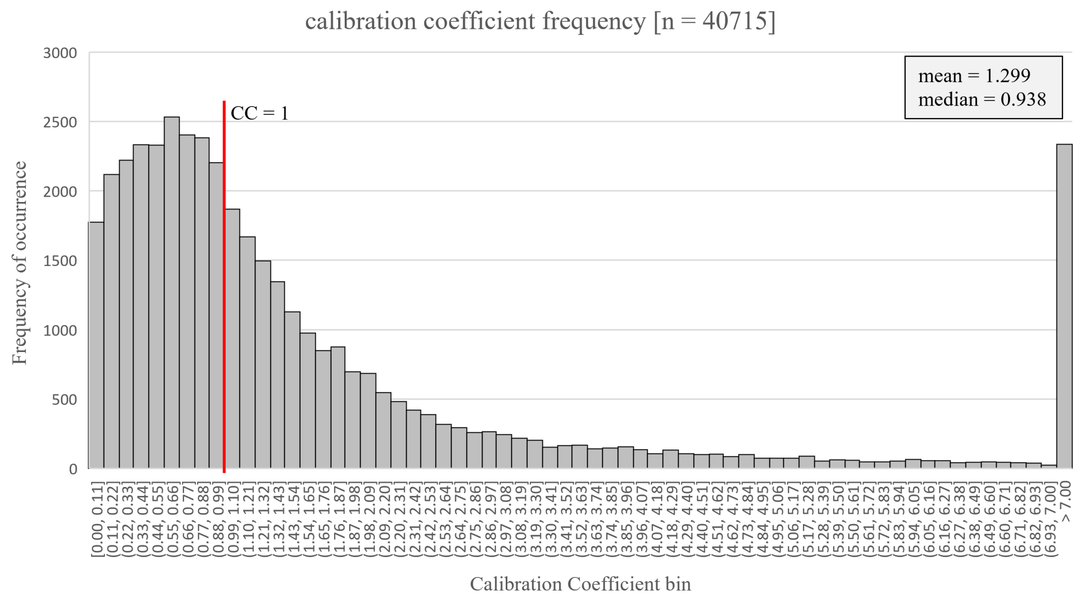

Figure 8: Histogram of the calibration coefficients compared against the calibrated model

To use the aforementioned measures to evaluate a model it is necessary that both the independent and dependent variable are able to be calculated. This is a non-issue when dealing with observed counts, but not so if you wanted to compare the default n2n model with the calibrated z2z model. Because of the different underlying model logic, some links would have no demand on them in the calibrated model, even though they would have traffic on them in reality. This would cause noise in the results, contaminating any analysis that would come after. To address this, a novel measure was suggested.

3.3.4 Cyclist-kilometers per day

This measure is computed by multiplying the total length of all links within an area with the total number of traffic generated on those links, as seen in Eq. 3.6. Here, l_i is the length of link i in kilometers and $x_{i,h}$ refers to the daily model intensity along the digitization direction, with $x_{i,t}$ being traffic against the direction of digitization. (T. Thomas, personal communication, June 1, 2023)

$$\text{cyclist-km per day} = \sum_i (l_i * (x_{i,h} + x_{i,t})) \qquad \text{(Eq. 3.6)}$$

The idea behind this measure is that, when computed for both the default and calibrated model, it compares the total distance traveled by all agents within a zone. Assume that both models have the same OD-matrices, but in one the cyclists travel along longer routes to leave the zone. This would be revealed by the higher c-km per day value, as the length of the links being cycled over by the same number of cyclists is higher. Eq. 3.7 shows how this can be computed as ratio, where values < 1 indicate underestimation and values > 1 indicate overestimation. The location and shape of the zones being investigated with this measure is shown in Figure 9, together with their values.

$$\text{c-km per day ratio} = \frac{\text{default c-km per day}}{\text{calibrated c-km per day}} \qquad \text{(Eq. 3.7)}$$

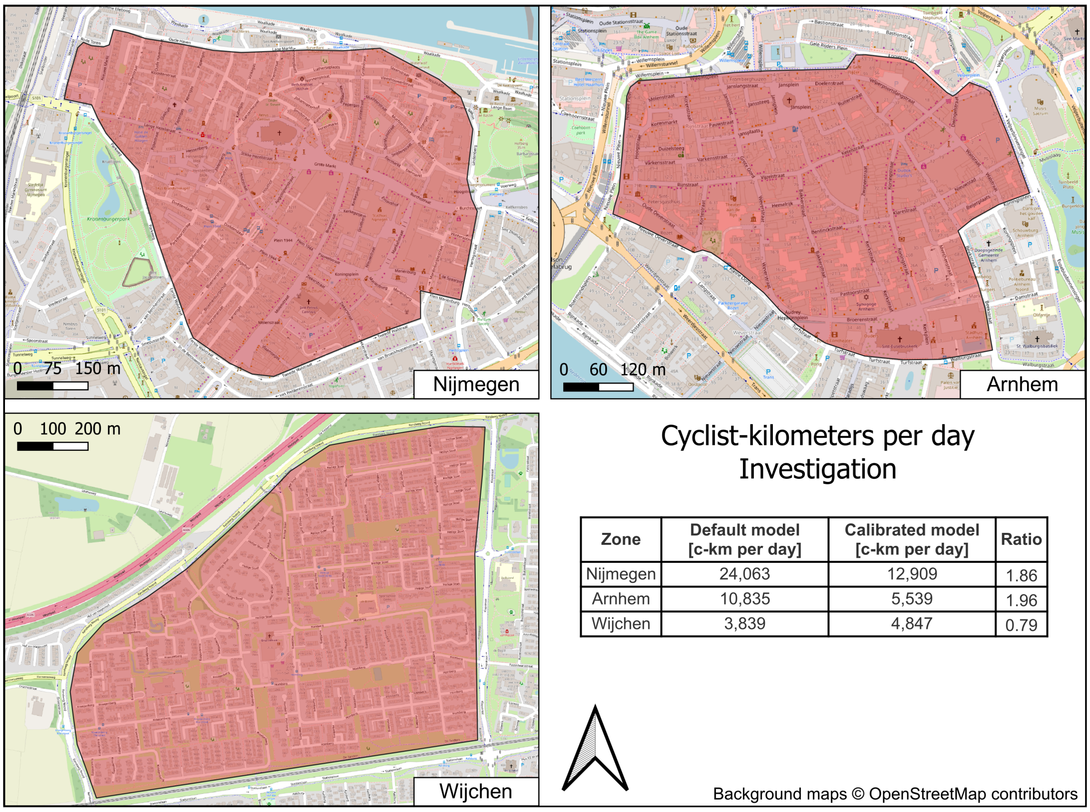

Figure 9: Areas selected for the comparison of the cyclist-km per day measure

These zones were shaped in such a way that a minimum amount of through-traffic would be present in reality, thus allowing the measure to investigate the effects of changes within the zone without external influence. We see significant overestimation happening in the city centers of Nijmegen and Arnhem, which is in line with expectations. As a control comparison, a neighborhood of the nearby town of Wijchen was looked at, where some slight underestimation of traffic was observed. This observation is confirmed using the CC measure with which an average value of 1.23 is found within the zone.

It should be noted that the OD-matrices between the two models (default and calibrated) differ because of the calibration process, causing the values to be skewed. However, as we are interested in the effects of adjustments to the model on this measure, we can ignore that issue and focus on the change in these ratios between the different model adjustments. With the measures introduced, the benchmarks for the model adjustments can be stated.

3.3.5 Adjustment benchmarks

To properly claim that any adjustments made to the model are improvements from the default version it is necessary to set some targets that the adjusted model must surpass. These are given in Table 4.

Table 4: List of measures and their benchmarks

Measure	Value	Ideal state	Performed on
R^2	0.615	1	count set
RMSE	917	0	count set
CC	0.919	1	study area
c-km per day	-	1	zones

3.4 Project workflow

Having introduced the measures used to evaluate the model accuracy we can now look at how these measures will be used in the analysis of the current traffic situation and the subsequent model adjustments. This section will cover how each of the research questions will be answered and what the limitations to each of the stated approaches are.

3.4.1 Question 1: Valuing the impacts of features on demand

The literature review revealed a number of factors that influence cyclist routing and these factors (or their proxies) were also found to be present in the available database. Given all that, it is necessary to establish what effect an individual value for a factor has on the speed of any link where that characteristic is present. One approach that achieves this is a regression analysis.

A multinomial regression model could give the contributions each of the factors have on the traffic, given that the independent variables were the factors on a link and the dependent variable was the amount of traffic generated on that link. This approach comes naturally with a number of limitations and assumptions. For one, if the dependent variable is traffic, then locations with high traffic intensities will have their features overvalued compared to locations with little traffic. To address this, a spatial regression model could be used. Such a model considers the values of the neighboring dependent variables to a point being regressed, thus decreasing the effects of large traffic intensities on the regression coefficients.

This could be done in the GeoDa software which can do Ordinary Least Squares (OLS) regressions as well as the mentioned spatial lag and spatial error models. These latter models require each data point to be spatially located on a x-y coordinate system, from which GeoDa constructs a set of polygons. The polygons dictate if two points are neighbors, and depending on the selection of the user, these neighbors can either be only along the edges of the polygon, or also across the vertices.

While using GeoDa would solve the issue with overvaluation, it would still have the problem of the coefficients being in units of cyclists, which would need to be processed somehow to get the equivalent effect on link speed. The alternative to this is setting the dependent variable to be speed but this comes with its own problems. One is that there is no way to convert traffic to speed, as traffic is consequential of the speed of a link, but there is no relation in the opposite direction. By assuming proportionality, a "theoretical speed" value could be calculated to produce such a relationship, as seen in Eq. 3.8.

$$\text{theoretical speed} = \frac{\text{count intensity}}{\frac{\text{model intensity}}{\text{default speed}}} \qquad \text{(Eq. 3.8)}$$

This theoretical speed could be regressed on directly (as the dependent variable) but it would also be possible to run the model first with these speeds assigned to their respective links, see if the accuracy was improved, and then perform the regression with the residuals now being calculated using the intensities the model just calculated. Doing this can verify if the theoretical speeds actually improve the model.

One limitation of this approach is that proportionality does not entirely apply here, and even if it did, the approach results in very high residuals returning extreme speeds (<4 and >200 km/hr). This could be accounted for by taking the average intensities per distinct default speed—of which there are only 28 in the study area—but such aggregation loses detail in any subsequent analysis.

Moving from the dependent to the independent variables, they too involve some assumptions, the most prominent of which involves the value of the variable itself. Take for example the facility factor; it has a range of values associated with it but there is no objective way to determine how much more appealing a solitary bicycle path is over a bicycle street. The immediate solution to is is using dummy variables which either take the value 1 when a characteristic is present or 0 when it is not. Issue then becomes the number of these variables, as most software is limited in how many independent variables it can consider—GeoDa has a maximum of 16 and Excel can compute up to 64. Considering that there are up to 69 unique characteristics over 13 factors that a link could have, Excel will have to be used when dealing with such variables.

One way to avoid all of these limitations with regression modelling would be to manually identify some relationship between the characteristics of the link, be it spatial or physical, and then adjust the speeds of links with this property based on some other measure. For example, Dat.mobility's theory is that traffic within city centers is being overestimated. With this hypothesis it would be possible to select some characteristic of this area (such as pedestrian spaces) and see if a change to the speed of all links with this property would improve the accuracy.

The immediate limitation with this approach is that it relies on some other calculation to find the effects of the characteristics on speed. One such option is to use the calibration coefficient and calculate its median value for each characteristic. Significance of this value could then be based its deviation from the mean value, but this may however cause the results to be over-fit to the data set. That is to say, the values found to improve the accuracy for this data set may not produce the same effects when applied to the whole Netherlands. It will also involve some trial-and-error to find which factors give the best results.

3.4.2 Question 2: Implementing the found impacts

The method used to find the effects of the features will partially dictate how they will be incorporated into the model, but there is some variety in the specifics. The most complex method involves the conversion of coefficients resulting from regressions involving traffic volumes. If the independent variable weights are in units of relative importance—i.e. 0 for the worst perceived characteristic and 1 for the best—then the coefficients will be the contribution this factor has towards the traffic volume in cyclists per day. To get a ratio in the form necessary to correctly adjust the speeds, the new expected traffic on a link must be divided by the current default traffic, similarly to the CC measure as seen in Eq. 3.5. By adding the resulting regression equation to the default speed, the new traffic volume can be calculated, which divided by the default speed returns a ratio which can be used to scale the speed by. Eq. 3.9 depicts this, where c_n is the coefficient for factor n, w_n is the weight for characteristic n, and x_i is the estimated traffic for link i.

$$\text{adjustment ratio} = \frac{x_i + c_1 * w_1 + c_2 * w_2 + ... + c_n * w_n + C}{x_i} \qquad \text{(Eq. 3.9)}$$

If the weights are set to the average residual for a characteristic, then the equation stays the same but c and w are flipped, with the regression fitting the importance instead of the traffic volume. When theoretical speed is the dependent variable, then this gets simplified further to just the fitted equation being how the speed is calculated—no ratio or other adjustment necessary.

The calibration coefficient was already introduced in prior sections as a measure of accuracy, but it can also serve as a weight with which to adjust the speeds by. This is because of its construction, which relates the ratio between the two intensities; default and calibrated. Note however that the CC can take values anywhere between 0.002 and 150,000, so some aggregation is required. Using the mean CC across all links with some characteristic would cause these extremes to significantly skew the adjustments. For example, the mean CC across all links is 24.092, but when values greater than 7 and those equal to 0 are removed, this drops down to a much more reasonable value of 1.299.

An average CC of 1.299 would indicate that the model is underestimating the amount of traffic, but based on the count data, it should be overestimating. This further confirms that using means would not be ideal. By comparison, the median for the filtered set is 0.938; a far more reasonable value which fits our expectations for the model. To check if a median CC for a feature deviates significantly from the overall median CC, we can compute the standard deviation across the median CCs for a factor. If one of these median CCs then falls outside of the median $\mp$ 1 standard deviation range, then the characteristic is said to have a significant impact for that factor. We can verify this by visually comparing the CC frequency plots for a characteristic against the frequency chart for the whole set (Figure 8).

3.5 Data pre-processing

Before any analysis can be done, the data must be processed and combined from their respective databases. This poses a challenge when the different databases use different methods to relate the information within them. The biggest problem here was connecting the count database with the model outputs, as the underlying networks of these two models was not identical, both in the digitization and the shapes of the links. To perform this connection, first the start and end nodes were compared to see if they lie within 0.1 meters of each other, and then azimuth of a link segment was calculated for both models. If they were identical, then the digitizations were also identical. If they were flipped by 180 degrees, then they were digitized in opposite directions. This process removed 875 links whose shapes were significantly different between the models.

One important step that is necessary to do before any regression analysis is checking the correlation between the values being investigated. If two independent variables are highly correlated with each other, then one of the fundamental assumptions of the regression is violated—no multicollinearity. The correlations between the characteristics in the count set was checked using Excel's CORREL function, and those with values $> \pm 0.8$ were excluded from subsequent investigations. The full correlation matrix can be seen in Appendix B - Factor correlation matrix, with the found highly correlated values being unknown surface; surroundings; hindrance; beauty; lighting, and when a link was marked as not being an intersection. This is quite reasonable considering that the unknown characteristics were almost always applied only to intersections.

4 Results

The results of the approaches outlined in the methodology chapter will be presented in this section. It is structured in a chronological manner, where each successive subsection is presented in the order its approach was used. Note however that many combinations of the methods were tried and the results presented here are only the most successful ones for that approach. For each of these subsections, the relevance of the approach in answering the research questions is given in addition to stating what caused the shift from one approach to another.

4.1 GeoDa regressions

Using GeoDa to do an spatially lagged regression gave the results seen in Table 5. The full report can be found in Appendix C - GeoDa regression report. Here, the dependent variable is the traffic residual and the independent variables use the average residual across all links with a specific characteristic as their attribute. This causes the coefficients to represent the scale of the impact the average residuals have for a characteristic within a factor. Negative coefficients imply opposite effects on traffic residuals than what the average residual (characteristic value) implies.

Table 5: Output of the best GeoDa regression using traffic residuals

Variable	**Coefficient**	**Probability**
Spatial variable	0.5157	0.00000
Constant	520.7160	0.36849
Length	-0.0834	0.49786
Speed	-12.8649	0.60348
Facility	0.5912	0.00087
Surface	0.5483	0.02150
Environment	0.0175	0.95584
Intersection type	-1.3223	0.56464
Max slope	0.7027	0.01220
Water	0.487409	0.77147
Nuisance	0.401715	0.07985
Lighting	0.4986	0.56028
Quality	2.4760	0.45904
Beauty	0.4586	0.19078
Road class	-1.9890	0.19944
Crossings	0.745415	0.01785
$R^2 = 0.3113$		

At a minimum 90% significance level, only the facility, surface, slope, nuisance, and crossings were found to be significant. This is reasonable, considering both facility and the surface are used by the default model to set the speeds. Crossings being significant could be explained as them being bottlenecks in the network which consistently cause large differences between the model and reality. These findings are also supported by the literature; the Broach et al., 2012, paper found both facility and slope to be significant drivers of increasing and decreasing utility, respectively. Genugten and Overdijk, 2016, found facility, surface, and slope to be the three most important factors for a cyclist's route choice.

Note that the R^2 value of 0.31 indicates a poor fit between the dependent and independent variables, not the accuracy of the adjusted model. When these coefficients were used to adjust the link speeds, both R^2 and RMSE were far worse than their benchmarks for the model, going as far as having a negative R^2.

Given all that, it would appear this method can indicate what factors are significant but the valuation of these factors is not appropriate for the purposes of this work. The biggest problem is believed to be the disconnect between the values being regressed and the target adjustment of speed; traffic is not 100% dependent on the link characteristics, but rather on the attraction and demand of areas close to it. To address this, it was decided to move from GeoDa to Excel and use different variables in future regressions.

4.2 Theoretical speeds in Excel

By moving to Excel, it was possible to use dummy variables for the characteristics instead of arbitrary weights. This has the regression coefficients directly relate how a specific characteristic influences the dependent variable. Regarding the concerns over the conversion of the coefficients into effects on speed, the previously introduced theoretical speed parameter is used as the dependent variable. The best results were found when this theoretical speed was the average per distinct default speed, rather than a individual theoretical speed per link. Unknown variables were ignored, due to their high correlation, and the constant in the exponential equation being fit was set to 1. The outputs significant at a 90% level are shown in Table 6, with the full version in Appendix D - Excel regression report.

Table 6: Partial output containing significant values from the LOGEST Excel regression

Variable	solitary moped path	protected moped path	protected bike path	bicycle street
Coefficient	0.9460	1.0923	1.0686	0.9217
Variable	painted gutter	normal road	service road	avg. slope > 4 %
Coefficient	0.6405	0.6287	0.7173	0.3998
Variable	max. slope 6-8%	max. slope 10-20%	below grade crossing	
Coefficient	1.3360	1.2134	0.7132	
	$R^2 = 0.9493$			

Notice how values from the factors that were found to be significant in Subsection 4.1 were again found to be significant here, with the only ones missing being nuisance and surface. For nuisance, that could be because the average residual for links with this characteristic was relatively high at -479, causing it to be more valued in the previous approach. Unlike in that method, the very high R^2 here shows that this method is very good at fitting the variables into an exponential function. Additionally, when the theoretical speeds—the dependent variable in the regression—are used in the model to check if they would improve the accuracy, significant improvements to both R^2 and RMSE were found with values of 0.69 and 818, respectively. This indicates that the theoretical speeds do improve the model performance, but with the caveat that this is true only when manually applied to select links. By using the exponential equation to calculate the new speeds for all links, the model returns an R^2 of 0.23 and an RMSE of 1295. These are both significantly worse than the benchmark and this was the best result for all regression attempts.

It is believed that a major contributing factor causing this to happen are the differences in the number of features between the count set and the study area set. Take for example the protected moped path characteristic: 8.6% of all links in the study area have this feature, while 27.5% have it in the count set. Because of the way traffic is distributed over the network—the agent optimizes their route to minimize cost—changes to link speeds with a particular characteristic will always have cascading effects on all other links on the network. The same amount of traffic is always generated going to the same destinations, so changing the speed only affects what path the traffic takes. And when the regression's 'mental model' differs from the situation its outputs are being applied to, then the resulting accuracy can never approach the theoretical one. The clear limitation here appears to be the reliance on count locations. So, the final approach used the calibrated model to investigate the differences across most of the links in the study area.

4.3 Calibration coefficient adjustments

Using the calibrated model, the CCs were computed for all links where the default model generated traffic in the study area. Table 7 shows the results of this approach for the facility factor in the direction of digitization. The full results are in Appendix E - CC per characteristic report.

Table 7: Calibration coefficient statistics of the facility factor

Facility	**Frequency**	**Median CC**	**Δ Median CC**	**Mean CC**
pedestrian crossing	123	0.562	-0.376	0.919
pedestrian area	137	0.614	-0.324	1.315
protected moped path	3882	0.685	-0.253	0.908
protected e-bike path	70	0.689	-0.249	0.677
service road	477	0.737	-0.201	0.983
ferry	22	0.849	-0.089	1.032
painted bike lane	2916	0.881	-0.057	1.107
on-ramp	1	0.939	+0.001	0.939
normal road	21595	0.946	+0.008	1.374
solitary bike path	2589	1.042	+0.104	1.418
protected bicycle lane	4872	1.088	+0.150	1.313
solitary moped path	692	1.132	+0.194	1.349
unknown	931	1.210	+0.272	1.616
bicycle street	104	1.240	+0.302	1.370
Value across all links	-	0.938	σ = 0.222	1.299

From those characteristics which had significant differences in their medians, two were found—through trial and error—to improve the accuracy when applied to the whole model. These were pedestrian areas and bicycle streets, and we can confirm their significance by looking at their frequency plots, seen in Figure 10 and Figure 11, respectively.

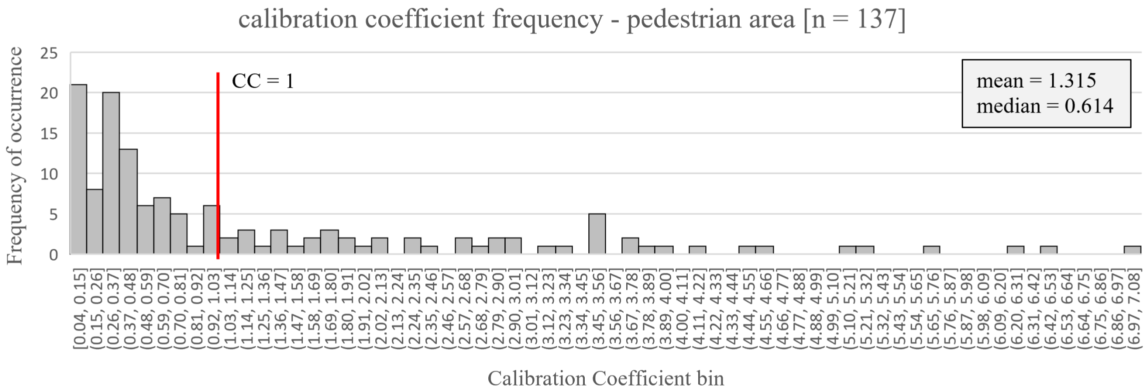

Figure 10: Histogram of the calibration coefficients for links with the pedestrian area feature

For pedestrian areas, there is a clear peak in the CCs below 1, with only 52 of the 137 links having a CC greater than 1. Bicycle streets have a less clean distribution of coefficients but they still show a majority of the CCs falling above 1.

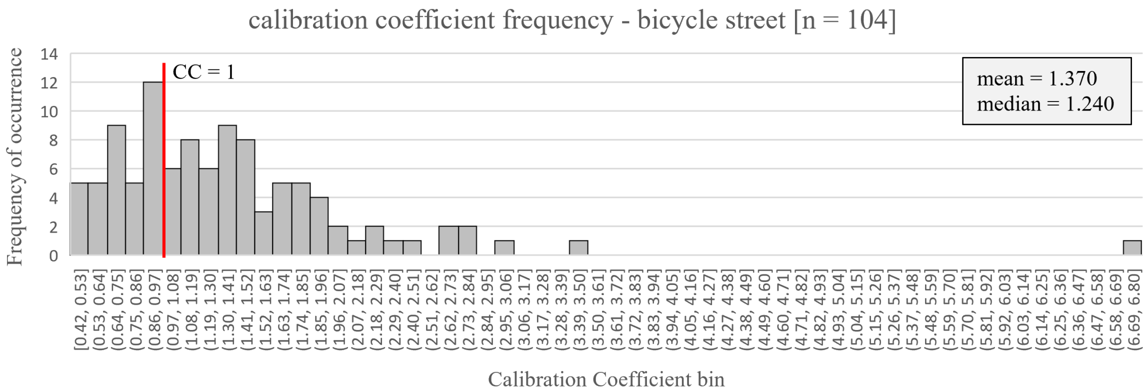

Figure 11: Histogram of the calibration coefficients for links with the bicycle street feature

Using a case expression, if a link had either the pedestrian area or bicycle street feature, then its default speed was multiplied by the relevant CC, thus satisfying the restriction on the method used to adjust the speeds. By doing this, the following effects on the absolute amounts of daily traffic were observed within the city centers of Arnhem (Figure 12) and Nijmegen (Figure 14).

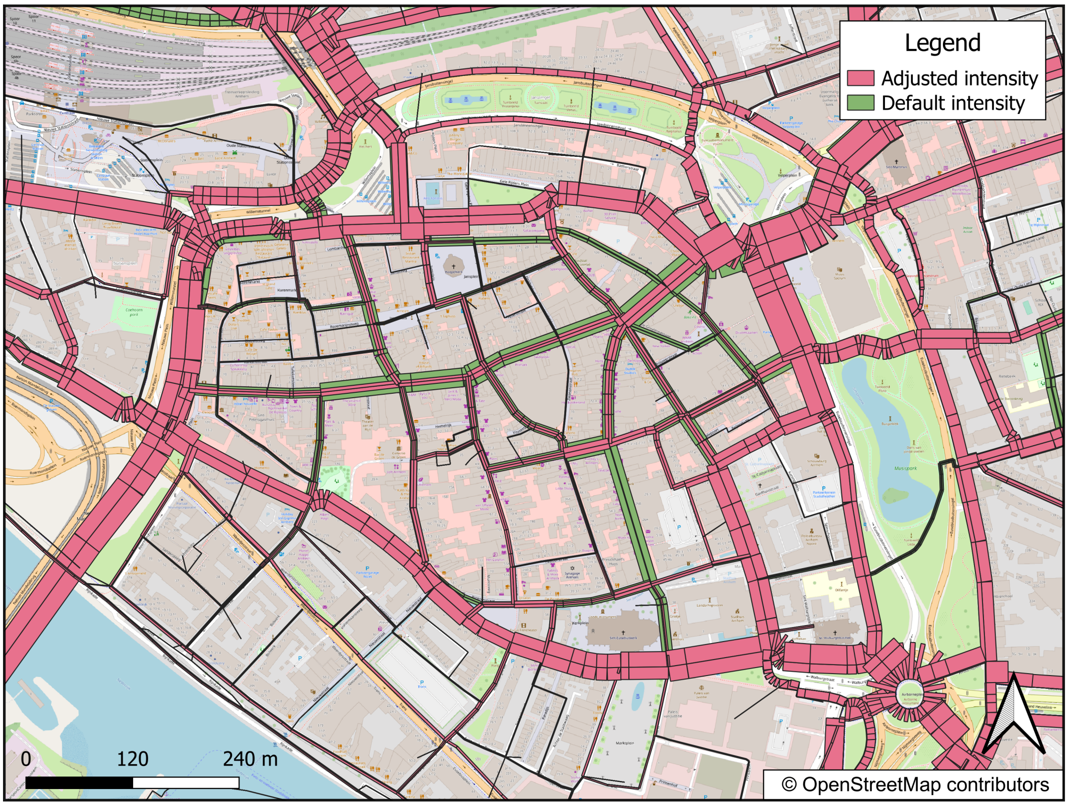

Figure 12: Total traffic within Arnhem's city center for the default and adjusted models

In Arnhem, the traffic decreases everywhere where less traffic is expected. All of the links in this area are appropriately marked as pedestrian areas, which is in line with what is observed in reality. Figure 13 shows a photo of a street in this area, where you can see how the lack of any bicycle facilities and narrow streets would make it unappealing for cyclists.

Figure 13: A street representative of the cycling situation in Arnhem's city center

However in Nijmegen, there is still a substantial amount of traffic going through the city center. Checking the facility property of these links reveals them to be marked as solitary bicycle paths, which is unexpected considering the area. Additionally, some of the links through which the traffic is being re-routed are sloped at a moderate average incline of 2 to 4 %. Such an incline is likely too unappealing to have the intensity projected by the adjusted model—recall how the Broach et al., 2012, paper found slope to significantly decrease the utility of a link.

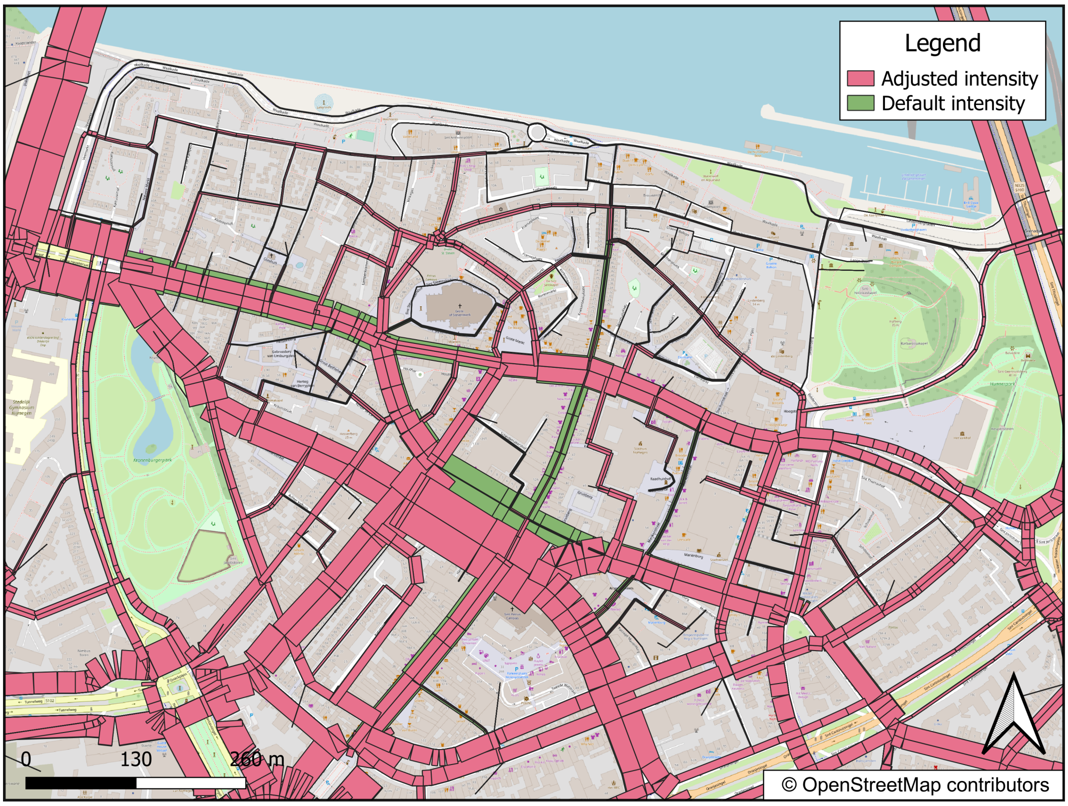

Figure 14: Total traffic within Nijmegens's city center for the default and adjusted models

To better highlight where the traffic is being routed, the difference in intensity between the two models was plotted in Figure 15 for Arnhem and Figure 17 for Nijmegen. In the first of these, a near-perfect adjustment is revealed; all of the pedestrian areas inside of the center show decreases, with the traffic being re-routed to the appropriate cycling links circling the city center where it is expected and observed in reality.

The changes outside of this area are routes which previously used some pedestrian link as a shortcut. How appropriate these changes depend on the location and the actual cyclist behavior there, but in general, having cyclists cycling on bicycle links is the preferred outcome of the model.

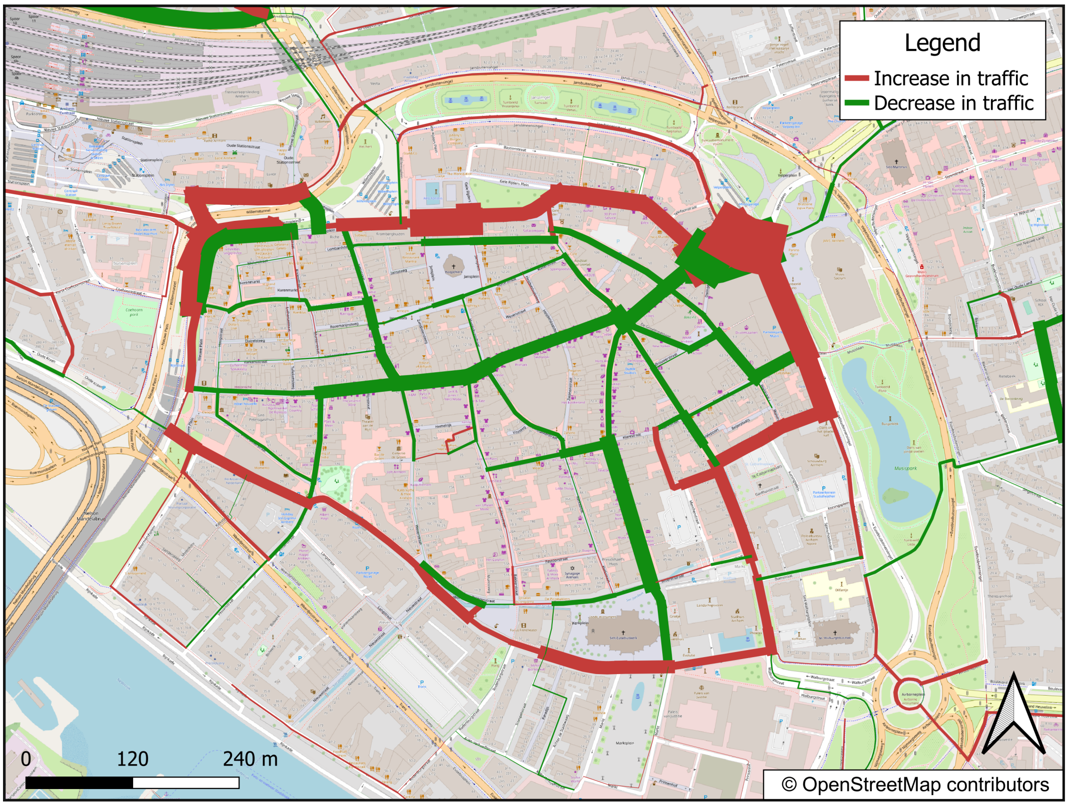

Figure 15: Difference in traffic within Arnhem's city center

In Nijmegen, the differences in traffic are far less neat. The biggest change shifts the traffic onto a street one block over, with the streets (called Plein 1944) being near identical in reality, but one is marked as a pedestrian area, and the other as a solitary bike path, as seen in Figure 16.

(a) A 'pedestrian area'

(b) A 'solitary bike path'

Figure 16: Highlighting the differences between reality and the database (Source: Google © 2023)

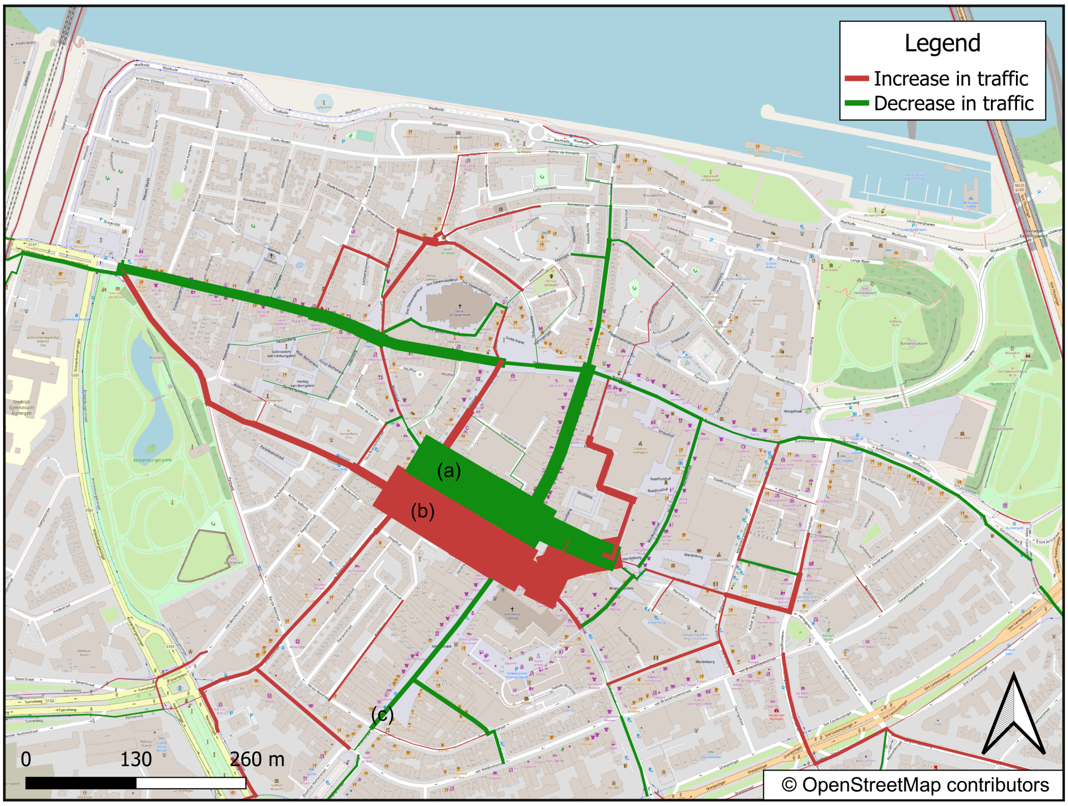

Figure 17: Difference in traffic within Nijmegen's city center

Finally, to check if the bicycle street adjustment had the desired effects, let us look at a long bicycle street running through the town of Duiven, near Arnhem. This is shown in Figure 18, and we can see increases along the bicycle street and decreases in the roads around it, which is in line with expectations.

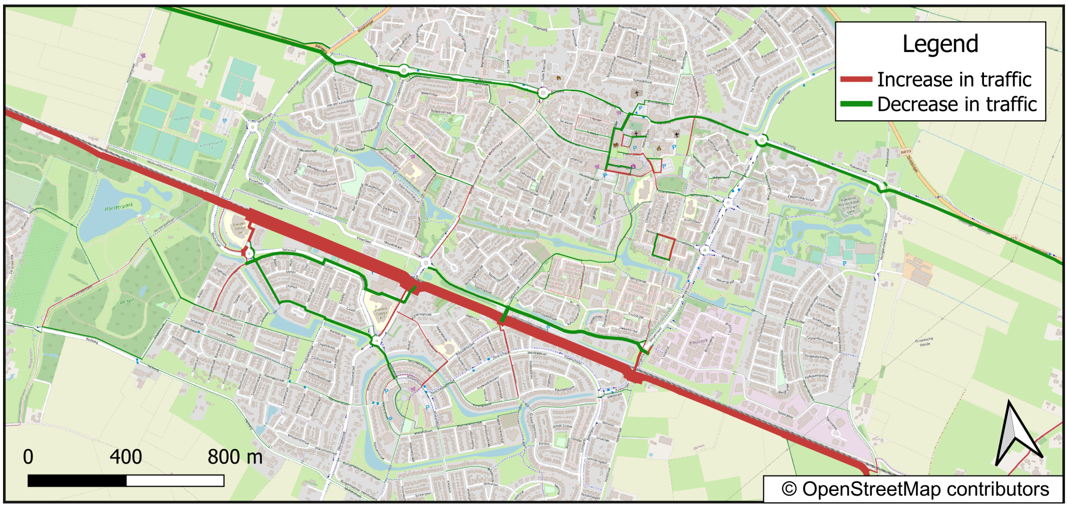

Figure 18: Difference in traffic for a bicycle street

4.4 Verification

To confirm that the unexpected traffic distribution in Nijmegen was due to the facility label (and its subsequently higher speed), the speeds for all pedestrian links was lowered to 5 km/hr from the adjusted value of circa 9 km/hr. The resulting traffic distribution is shown in Figure 19. There the traffic becomes concentrated on links labeled as solitary bicycle paths, with some links having daily intensities (circa 9000 c/day) greater than the two bridges that span the Waal river (circa 5200 c/day at most)—something that is clearly unreasonable. This distribution of traffic mirrors that shown in Figure 14, confirming it as a consequence of the different facility labels.

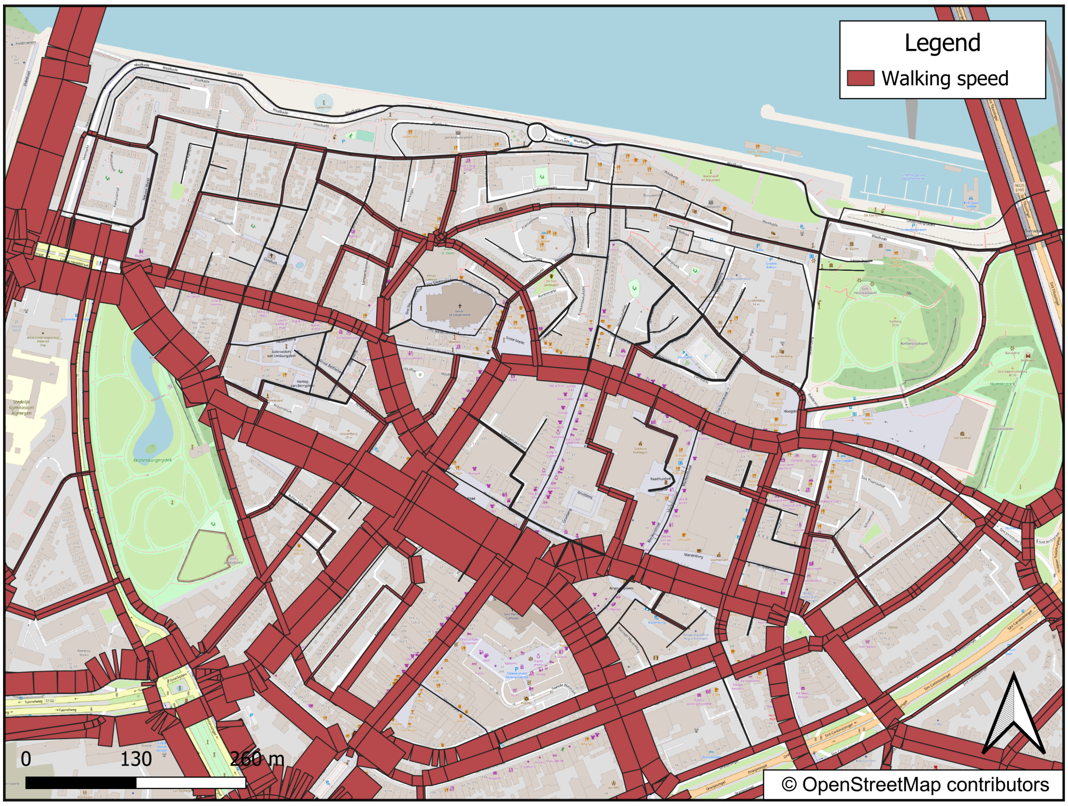

Figure 19: Total traffic within Nijmegen's city center with speed = 5 km/hr

While using such a low speed value would be representative of cyclists having to disembark and walk through the city center, that would not be accurate to the real situation, as cycling is still possible in this area. Indeed, some of the links marked as solitary bike paths in Nijmegen do have facilities for cyclists, such as lane markings on the ground to separate pedestrians from bicycles. In Figure 20 an example of these markings is shown. The location of this picture is labelled with (c) in Figure 17.

Figure 20: Cyclist facilities within Nijmegen's city center (Source: Google © 2023)

Here we can see two noteworthy traffic signs: the one in the top-right is the standard footpath sign (*voetpad*; G7), with the added information that cyclists are permitted as well, provided they stick to the right side of the path. On the left side of the figure, there is what appears to be a non-standard sign indicating that cyclists and pedestrians should "take each other into account" (*houd rekening met elkaar*). There are no clear rules in this situation as to who has the right of way, with the only legal requirement being that "It is prohibited for anyone to behave in such a way that [endangers others] or [traffic becomes hindered]." (Wegenverkeerswet 1994, § 1. Art. 5).

Clearly there are accommodations made for cyclists in Nijmegen's city center, however the default speeds assigned to the both the pedestrian areas and solitary bike lanes here range between 14 and 15 km/hr, which is far too high when you consider the hindrance posed by pedestrians and other obstacles. Thus a middle-ground value between the two—as was found with the CC median intervention—is ideal for representing what likely is the real cyclist speed in this area.

To statistically check if the adjustments improved the accuracy, let us look at the changes in the accuracy measures from Section 3.3. These are shown in Table 8. All of them (except for the Wijchen control) show improvements in their values between the default and adjusted model.

Table 8: Changes to the accuracy measures

Measure	**Initial value**	**New value**	**Change**
R^2	0.615	0.621	+0.006
RMSE	917	911	-7
CC	0.919	0.941	+0.022
Arnhem c-km/d	1.96	1.10	-0.86
Nijmegen c-km/d	1.86	1.81	-0.05
Wijchen c-km/d	0.79	0.79	-

The relatively minor changes to R^2 and RMSE stem from the fact that there are no count locations within the city center, so the changes here come from small differences in traffic at links around the city center, in addition to the changes to bicycle streets. In Arnhem's city center, a significant decrease is seen in the c-km/d measure, but much less so in Nijmegen. This is because of a poor shape selection when marking the borders of the city center, with it containing many normal roads. Additionally, because a substantial amount of the links in Nijmegen's city center are marked as solitary paths, no adjustments were applied to them.

One additional way to validate these results is by comparing the trends here with differences between different versions of the OmniTRANS Spectrum model. The version that was worked on during this project was made in 2018, with the newest version being from 2020. In the two cities looked at, there are large decreases in traffic in the same places as shown in this work. Unexpectedly however, the bicycle street shown in Figure 18 had a decrease in intensity between the model versions. This may be because of differing OD-matrices, which would be reasonable considering during 2020 travel demand was very abnormal.

5 Discussion

To best discuss the findings it would be good to first remind ourselves what this and any static traffic model does; they approximate the behavior of people through rules of supply and demand, with their pathing being decided using optimization algorithms that are far removed from the reality of human travel. The model has no concept of weather, time of day, or what season it is, all of which will have an impact on the behavior of cyclists. For example, when it is winter and it gets darker earlier, cyclists may tend to prefer roads with good lighting but the model cannot account for this, even through there is a factor related to street lighting.

Furthermore, the model is limited in the motives it can simulate. Specifically how it cannot model recreational travel. And with recreational cycling being a popular pastime in the Netherlands, this causes a significant disconnect between the model and reality. Especially as it relates to count data, because recreational cyclists will still be registered as demand on a counter even if the model cannot simulate that trip (motive). Another type of trip the model cannot simulate is last mile trip, where people traveling from a transportation hub to their eventual destination using a bicycle are not modeled as cyclist agents in the model. That is another significant disconnect between the model and reality, as it is very common for people who travel with the train to also use a bicycle at some point during their journey.

And there are disconnects not only in what trips the model can simulate, but also how travel behavior can vary within a motive. Consider the following situation: a parent with their child and a young adult both want to cycle to the train station. The parent will prefer to use safer, perhaps longer routes while the young adult will prefer the fastest route, because he is late for the train. To the model, these are the same agent, and both with have to optimize their route based on a single cost value. As such, attempting to adjust that cost by considering features which may be appealing to one of these people will cause the routing for the other to be distorted from reality. Recall the 2016 study by Shin where they found how routes to work differed to those going back home. In its current state, the model cannot account for this.

A difference in cycling behaviors may also be one of the reasons why it is not certain that the values found in this work will have the same effects when applied to the whole Netherlands. We saw in the literature review how cyclists perceived certain features differently between the US and the Netherlands, so it is not unreasonable to think there may be differences between the cyclists in Gelderland and Groningen, for instance.

Another issue is related to the classification of factors within the Cyclists' Union database. Because it relies on community contributors to assign values to the factors, there can and do arise differences between what is listed in the database and what is present in reality. We already saw an example of this in Nijmegen, but lets look at another one, shown in Figure 21. Here, the link is marked as a solitary bicycle path in the database, but, in the opinion of the author, the link is, at best, a protected bicycle path along a road.

However even then there is a problem: the protected lane is not wide enough for two cyclists to pass each other comfortably so even this label is dubious. Maybe this protected part is meant for pedestrians, in which case the facility characteristic would be a normal road with no accommodations for cyclists. But nothing stops cyclists from making use of the lane so that would not be representative either. What this should highlight is that the actual cycling situation on a road can vary greatly even within a single characteristic value.

Figure 21: A 'solitary bike path' near Hengelo

Regarding the failure of the approaches involving regressions, earlier it was introduced that one of the reasons for this failure was down to the disconnect between the regression's data set (n = 466) and the network that the outputs are applied to (n = 66783). This is partly because the distribution of characteristics between the two sets of data differ, but also because of a theoretical issue; people choose to cycle on a link not because it has a high speed, but because it will get them to their destination. Thus ascribing through the regression analysis that the resultant traffic on a link as being due to its features is false.

These approaches were however able to successfully identify features that coincided with the features found in the existing literature to significantly influence bicycle routing. The type of facility was found to be a significant driver of cyclist routing in all but one paper that was looked at, and this significance was confirmed using both GeoDa and Excel approaches. The type of surface was another factor that both the literature review and GeoDa revealed as being significant.

In that literature they used regression analyses extensively to attribute a characteristic's appeal for cyclists, often through a measure such as the distance traveled (by a cyclist) or the perceived travel time. The key part here is the use of routes, as it enables you to perform logistic regressions which return probabilities for an event either occurring or not. This then allows an analysis where the probability of a cyclist choosing a route can correctly be attributed to the characteristics along that route. Using solely count data to do this is not possible.

What the count data did allow for was a systematic adjustment of the default speeds that, for the pedestrian areas, landed perfectly between the too low speed of 5 km/hr and the too high speed of 15 km/hr at 9 km/hr. These findings were confirmed to be appropriate using both statistical measures and expert opinions.

6 Conclusion & recommendations

This paper opened by stating that the research goal was to concretely value the impacts some characteristics have on the amount of traffic on a link. Having presented the findings, it is clear that this was achieved: the calibration coefficient approach successfully identified features which had a significant impact on traffic, it provided the scale of these impacts, and it could be easily implemented into the model within the restrictions put forth by the client. These adjustments were found to improve the model using both traditional metrics as well as novel measures developed specifically to address the limitations related to the different model constructions. And while the typical metrics only showed minor changes, visual plots of the intensities showed significant improvement in the traffic situation wherever possible such that it was more representative of reality.

Ultimately, due to a lack of time, the adjustments could not be tested on the full model of the whole Netherlands. Doing so would confirm if the adjustments would improve the accuracy for the whole model, but based on the work done here, we can only be sure that the speeds for pedestrian areas would decrease and those for bicycle streets would increase. This is because the CC values are simply representations of the difference between the calibrated and default models for the study area per some characteristic—they do not give any insight into how appealing a characteristic is for a cyclist.

To continue investigating the relationship of bicycle traffic and link characteristics, this paper suggests a number areas where further research could be conducted. The first and most immediate suggestion is to continue testing what combinations of median CCs is the best, as this project just took the first combination that gave positive results. That testing could take the form of some optimization algorithm with the target measures being R^2 and RMSE, but trial and error together with some reasoned guesses would work as well.

Another suggestion that is relatively simple to do would be to improve the accuracy of the factor labels in the Cyclists' Union database. We saw how vastly different labels could be assigned to links with seemingly similar characteristics in Nijmegen, as well as how the cycling situation can vary within a label in Figure 21. To address the situation in Nijmegen, the characteristic value for these inner city links should be changed. Marking them as pedestrian areas would be the obvious solution, however that is not entirely representative; it is still possible to cycle in these areas and in some cases there even are facilities for cyclists (see Figure 20). But labeling them as solitary paths is wrong in the eyes of the author—the cyclists still need to interact and give way to other modes of traffic on these links. It may be best to use some new label like "pedestrianized street", where cyclists are permitted to cycle and/or there are some accommodations for cyclists in the area, but the obstacles present prevent cyclists from going as fast as they would on a typical bike path. It should be noted that, because this project used a version of this database from 2018, these issues may already be resolved in newer versions.

The Cyclists' Union also maintains a web-based route-planner application, in which users can specify some route preferences and the program returns a path for them to use. This is similar to the BBBike engine from the Hardinghaus and Nieland, 2021, paper. It may be of interest to perform a similar cluster analysis to identify what the important factors are for Dutch cyclists when selecting a route. Alternatively, this data could also be used to identify the groups of cyclists—and their proportion to the general population—who have distinctly different routing preferences.

Earlier it was mentioned how optimizing speeds for a feature important to one group of cyclists will decrease the accuracy for cyclists with different preferences when there is only one cost value per link. This could be addressed right now in the model, as it contains multiple OD-matrices for different groups of attraction and supply. For each of these, a different speed table could be used, thus optimizing the costs per population group. You could expand this even further and create multiple speed tables for different situations—such as inclement weather or if someone is travelling to or from work—to account for the differing cyclist behaviors in these settings. To do that would require extensive information about the distribution of cycling behaviors across the demographics modeled, which could be its own research project.

If that information was available however, then more options open up for trying to improve the model accuracy. One such idea would be to collect the average observed speeds of cyclists as they cycle, so that you could build a table with the real speeds of cyclists on links with certain features. This could then be used in a similar fashion to the CC, except now the rate that a link speed needs to be scaled by would be from observed speeds (over default speeds) instead of traffic. Alternatively, these observed speeds could be directly put as the speeds for any links with a given facility plus some modifiers—the same way the model has its default speeds assigned now.

Another limitation that could be addressed with this demographic behavior data is the issue of recreational cyclists polluting the count data. Its likely that any method which would record the trip speeds and locations would also record trip motives, so those which would be labeled as recreational could be discarded from subsequent analyses, as the model cannot simulate that trip motive. Should we choose to stick to the count data, then this could also be accounted for by investigating the distribution of traffic at a count location throughout the day. If we assume recreational trips avoid the typical morning and evening peaks, then if the traffic distribution at a count location does not follow the typical two-peak pattern it can be discarded.

Of course, the alternative to filtering out the recreational counts is to incorporate recreational trip motives into the model. That is easy to suggest, but if it was that easy to do, then it would have already been done. Keeping with the theme of difficult to implement ideas, another one would be to change the assignment algorithm so it can consider things like the characteristics of links between an agent's origin and destination. However this would again require data about the distribution of cyclist behavior to identify how many agents should get routed along paths with more greenery, less traffic nuisance, or some other factor. One paper that was not used in this work but that is relevant to the topic of large-scale bicycle traffic modeling was by Liu et al., 2020. There they used the results of the Broach et al., 2012, and Jensen et al., 2018, papers to create a very detailed traffic model where the preferences of cyclists was incorporated into the routing algorithm.

To summarize, almost all of the possibilities for further research related to this topic revolve around acquiring more data about the behavior of cyclists. And as we are interested in their routing specifically, collecting and analyzing route data should be at the forefront of any future research into the subject matter investigated by this paper. This lack of route data is why the work described here could not ascertain a relationship between the factors and cyclist preferences—it lacks the personalized data required to investigate the individualistic choices related to travel behavior.

References

Bernardi, S., Geurs, K., & Puello, L. L. P. (2018). Modelling route choice of dutch cyclists using smartphone data. *Journal of Transport and Land Use*, *11*(1), 883–900. https://doi.org/10.5198/jtlu.2018.1143

Bostanci, I., Gootzen, Y., & Lugtig, P. (2018). Data Linkage to Validate and Calibrate Traffic Estimations on a Nationwide Scale : A Framework for Official Statistics, 1–28. https://www.cbs.nl/en-gb/background/2023/11/validation-and-calibration-of-traffic-estimations

Broach, J., Dill, J., & Gliebe, J. (2012). Where do cyclists ride? A route choice model developed with revealed preference GPS data. *Transportation Research Part A: Policy and Practice*, *46*(10), 1730–1740. https://doi.org/10.1016/j.tra.2012.07.005

Broach, J., Gliebe, J., & Dill, J. (2010). Calibrated labeling method for generating bicyclist route choice sets incorporating unbiased attribute variation. *Transportation Research Record*, (2197), 89–97. https://doi.org/10.3141/2197-11

Genugten, W. V., & Overdijk, R. V. (2016). Het optimaliseren van fietsgedrag in verkeersmodellen. *PLATOS Colloquium*, 1–27.

Hardinghaus, M., & Nieland, S. (2021). Assessing cyclists' routing preferences by analyzing extensive user setting data from a bike-routing engine. *European Transport Research Review*, *13*. https://doi.org/10.1186/s12544-021-00499-x

Jensen, F., Fernández, A., Nielsen, O. A., & Rasmussen, T. (2018). Modellering af cykeltrafik. *Trafik & Veje*, *11*. http://asp.vejtid.dk/Artikler/2018/11/9142.pdf

Liu, C., Tapani, A., Kristoffersson, I., Rydergren, C., & Jonsson, D. (2020). Development of a large-scale transport model with focus on cycling. *Transportation Research Part A: Policy and Practice*, *134*(February), 164–183. https://doi.org/10.1016/j.tra.2020.02.010

Meijning, K. (2019). *Assessing the impact of bicycle infrastructure on cyclists ' route choice in the Dutch province of Noord-Brabant* (Master's thesis). University Utrecht. https://studenttheses.uu.nl/handle/20.500.12932/35223

Shin, D. K. (2016). *Explanation of factors influencing cyclists' route choice using actual route data from cyclists* (Doctoral dissertation). https://etheses.whiterose.ac.uk/13532/

Statistics Netherlands. (2022). Total transport performance in the Netherlands; modes of travel and regions [Accessed: 20-6-2023]. https://opendata.cbs.nl/statline/#/CBS/en/dataset/84687eng/table?ts=1687267295496

Wegenverkeerswet, Artikel 5, § 1 [Accessed: 27-6-2023]. (1994). https://wetten.overheid.nl/jci1.3:c:BWBR0006622&hoofdstuk=II&paragraaf=1&artikel=5&z=2023-06-01&g=2023-06-01

Appendix

Appendix A - Fietsersbond database statistics

Characteristic (group)	% in count set [n = 466]	% in study area [n = 66783]	Includes [FB characteristic]	% in count set [n = 466]	% in study area [n = 66783]
Solitary bike path	16.31	6.7	solitair fietspad	9.87	5.48
			solitair bromfietspad	6.44	1.22
Protected bike path	43.99	19.74	fietspad (langs weg)	16.52	10.99
			bromfietspad (langs weg)	27.47	8.6
			pedelec	0	0.15
Bicycle street	3.43	0.17	fietstraat	3.43	0.17
Painted gutter	14.38	5.02	weg met fiets(suggestie)strook	14.38	5.02
Normal street	17.81	61.21	normale weg	17.81	61.21
Service road	4.08	1.08	ventweg	4.08	1.08
Pedestrian area	0	3.15	voetgangersgebied	0	2.63
			voetgangersdoorsteekje	0	0.52
Excluded	0	2.93	rijbaanwissel	0	0.04
			veerpont	0	0.04
			onbekend	0	2.85
Paved	98.7	79.98	asfalt_beton	83.5	42.13
			klinkers	7.5	32.65
			tegels	7.7	5.2
Unpaved	0	2.16	halfverhard	0	0.6
			onverhard	0	1.21
			overig	0	0.35
Unknown	1.3	17.87	onbekend	1.3	17.87
Green surroundings	35.1	14.01	akkers/weilanden	12	4.33
			landelijke of dorps	12.4	6.52
			bos	6.4	2.37
			natuur (behalve bos)	4.3	0.79
Built up area - green	34.5	29.82	bebouwd (veel groen)	34.5	29.82
Built up area - normal	29	38.24	bebouwd (weinig of geen groen)	29	38.24
Unknown	1.3	17.92	onbekened	1.3	17.92
Water - yes	7.7	3.28	ja	7.7	3.28
Water - no	91.4	82.36	nee	91.4	82.36
Unknown	0.9	14.36	onbekend	0.9	14.36
Signalized intersection	0.4	2.43	kruispunt met VRI's	0.4	2.43
Uncontrolled intersection	0	8.24	kruispunt	0	8.24
Roundabout	0.4	3.13	rotonde	0.4	3.13
Not applicable	99.1	86.2	nvt	99.1	86.2

Characteristic (group)	% in count set [n = 466]	% in study area [n = 66783]	Includes [FB characteristic]	% in count set [n = 466]	% in study area [n = 66783]
<1% avg. incline	85	88.43	<1%	85	88.43
1-2% avg. incline	7.7	6.08	1-2%	7.7	6.08
2-4% avg. incline	5.2	3.69	2-4%	5.2	3.69
>4% avg. incline	1.3	1.59	>4%	1.3	1.59
Unknown	0.9	0.2	onbekend	0.9	0.2
<1% max. incline	61.4	76.6	<1%	61.4	76.6
1-2% max. incline	18.5	12.56	1-2%	18.5	12.56
2-4% max. incline	14.2	6.5	2-4%	14.2	6.5
4-6% max. incline	3.2	2.21	4-6%	3.2	2.21
6-8% max. incline	0.2	0.91	6-8%	0.2	0.91
8-10% max. incline	0.2	0.4	8-10%	0.2	0.4
10-20% max. incline	0.9	0.53	10-20%	0.9	0.53
>20% max. incline	0.6	0.08	>20%	0.6	0.08
Unknown	0.9	0.2	onbekend	0.9	0.2
Very little nuisance	12.7	14.14	zeer weinig	12.7	14.14
Little nuisance	62	54.47	weinig	62	54.47
Reasonable nuisance	19.5	12.5	redelijk	19.5	12.5
Much nuisance	4.5	0.9	veel	4.5	0.9
Very much nuisance	0	0.06	zeer veel	0	0.06
Unknown	1.3	17.93	onbekend	1.3	17.93
Good lighting	71.9	72.7	goed verlicht	71.9	72.7
Mediocre lighting	11.2	5	beperk verlicht	11.2	5
Not lit	16.1	7.81	niet verlicht	16.1	7.81
Unknown	0.9	14.48	onbekend	0.9	14.48
Good quality	82.8	63.75	goed	82.8	63.75
Reasonable quality	14.6	17.27	redelijk	14.6	17.27
Poor quality	1.3	0.99	slecht	1.3	0.99
Unknown	1.3	17.98	onbekend	1.3	17.98
Picturesque beauty	0.9	0.61	schilderachtig	0.9	0.61
Nice beauty	33.7	16.71	mooi	33.7	16.71
Neutral beauty	60.7	58.85	neutraal	60.7	58.85
Boring / ugly beauty	3.4	1.71	lelijk/saai	3.4	1.71
Very ugly beauty	0	0.05	zeer lelijk	0	0.05
Unknown	1.4	22.07	onbekend	0.4	3.08
			[null value]	1	18.99

Characteristic (group)	% in count set [n = 466]	% in study area [n = 66783]	Includes [FB characteristic]	% in count set [n = 466]	% in study area [n = 66783]
Other road [modifier]	52.4	72.11	overige weg	52.4	72.11
Important primary road	41	8.06	langs drukke weg	41	8.06
Along a busy road	5.8	2.76	belangrijke hoofdweg	5.8	2.76
Unknown	0.9	17.07	onbekend	0.9	17.07
Not (a crossing)	89.06	98.53	nee	89.06	98.53
Above-grade crossing	6.65	0.91	ergens overheen	6.65	0.91
Below-grade crossing	3.65	0.52	ergens onderdoor	3.65	0.52
Tunnel	0.64	0	tunnel	0.64	0
Ferry	0	0.04	veerpont	0	0.04

Appendix B - Factor correlation matrix

CORREL(X,Y)	X.length	speed	solitary_	solitary_	protecte	protecte	protecte	fietsstra	gutter	ordinary	service_	pedestri	facility_	asphalt	tiles	bricks	other_s	unpaved	unknow	fields
speed_Y	0.58	1.00																		
solitary_brom	-0.02	-0.02	1.00																	
solitary_fiets	0.11	-0.03	-0.09	1.00																
protected_brom	0.08	0.12	-0.16	-0.20	1.00															
protected_fiets	-0.19	-0.18	-0.12	-0.15	-0.27	1.00														
protected_pedelec	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!													
fietsstraat	0.01	-0.02	-0.05	-0.06	-0.12	-0.08	#DIV/0!	1.00												
gutter	-0.15	-0.20	-0.11	-0.14	-0.25	-0.18	#DIV/0!	-0.08	1.00											
ordinary_road	0.15	0.23	-0.12	-0.15	-0.28	-0.20	#DIV/0!	-0.09	-0.19	1.00										
service_road	0.00	0.07	-0.05	-0.07	-0.13	-0.09	#DIV/0!	-0.04	-0.08	-0.09	1.00									
pedestrian	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!								
facility_unknown	-0.05	0.01	-0.02	-0.02	-0.04	-0.03	#DIV/0!	-0.01	-0.03	-0.03	-0.01	#DIV/0!	1.00							
asphalt	0.26	0.39	0.02	0.07	0.04	-0.14	#DIV/0!	0.02	0.12	-0.06	-0.03	#DIV/0!	-0.15	1.00						
tiles	-0.14	-0.22	-0.08	-0.04	0.07	0.30	#DIV/0!	-0.05	-0.12	-0.13	-0.06	#DIV/0!	-0.02	-0.65	1.00					
bricks	-0.18	-0.29	0.06	-0.04	-0.14	-0.13	#DIV/0!	0.04	-0.02	0.23	0.11	#DIV/0!	-0.02	-0.64	-0.08	1.00				
other_surface	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!			
unpaved	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!		
unknown_surface	-0.09	-0.06	-0.03	-0.04	0.02	0.05	#DIV/0!	-0.02	-0.05	-0.05	-0.02	#DIV/0!	0.57	-0.26	-0.03	-0.03	#DIV/0!	#DIV/0!	1.00	
fields	0.25	0.46	0.06	-0.08	0.05	-0.13	#DIV/0!	0.00	-0.08	0.16	-0.01	#DIV/0!	-0.02	0.16	-0.11	-0.11	#DIV/0!	#DIV/0!	-0.04	1.00
rural	0.22	0.38	0.01	-0.08	0.10	-0.08	#DIV/0!	0.00	-0.12	0.10	0.05	#DIV/0!	-0.02	0.13	-0.06	-0.11	#DIV/0!	#DIV/0!	-0.04	-0.14
nature	0.25	0.15	0.03	0.14	-0.04	-0.09	#DIV/0!	-0.04	-0.03	0.07	-0.04	#DIV/0!	-0.01	0.09	-0.06	-0.06	#DIV/0!	#DIV/0!	-0.02	-0.08
forest	0.25	0.16	0.00	0.26	-0.08	-0.07	#DIV/0!	-0.05	-0.06	0.06	-0.05	#DIV/0!	-0.02	0.02	-0.08	0.06	#DIV/0!	#DIV/0!	-0.03	-0.10
green_city	-0.20	-0.19	-0.04	0.03	-0.01	0.14	#DIV/0!	0.06	0.02	-0.15	-0.01	#DIV/0!	-0.05	0.03	0.06	-0.07	#DIV/0!	#DIV/0!	-0.08	-0.27
city	-0.35	-0.54	-0.01	-0.12	-0.04	0.07	#DIV/0!	-0.02	0.17	-0.08	0.04	#DIV/0!	-0.04	-0.24	0.13	0.23	#DIV/0!	#DIV/0!	-0.07	-0.24
unknown_surroundings	-0.09	-0.06	-0.03	-0.04	0.02	0.05	#DIV/0!	-0.02	-0.05	-0.05	-0.02	#DIV/0!	0.57	-0.26	-0.03	-0.03	#DIV/0!	#DIV/0!	1.00	-0.04
signalized	-0.06	-0.13	-0.02	-0.02	-0.04	0.15	#DIV/0!	-0.01	-0.03	-0.03	-0.01	#DIV/0!	0.00	-0.15	-0.02	-0.02	#DIV/0!	#DIV/0!	0.57	-0.02
roundabout	-0.05	0.02	-0.02	-0.02	0.11	-0.03	#DIV/0!	-0.01	-0.03	-0.03	-0.01	#DIV/0!	0.00	-0.15	-0.02	-0.02	#DIV/0!	#DIV/0!	0.57	-0.02
uncontrolled	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
nvt	0.08	0.08	0.02	0.03	-0.05	-0.08	#DIV/0!	0.02	0.04	0.04	0.02	#DIV/0!	0.01	0.21	0.03	0.03	#DIV/0!	#DIV/0!	-0.81	0.03
avg_slope_none	0.14	0.24	0.00	0.09	-0.06	0.00	#DIV/0!	0.07	-0.13	0.06	0.05	#DIV/0!	0.03	0.12	-0.01	-0.08	#DIV/0!	#DIV/0!	-0.18	0.10
avg_slope_small	-0.05	-0.13	-0.07	-0.09	0.06	-0.07	#DIV/0!	-0.05	0.15	0.00	-0.01	#DIV/0!	-0.02	-0.05	-0.04	0.06	#DIV/0!	#DIV/0!	0.13	-0.10
avg_slope_mid	-0.11	-0.12	0.11	-0.07	0.02	0.04	#DIV/0!	-0.04	0.02	-0.05	-0.05	#DIV/0!	-0.01	-0.09	-0.03	0.09	#DIV/0!	#DIV/0!	0.15	-0.02
avg_slope_big	-0.07	-0.14	-0.03	0.09	0.02	0.05	#DIV/0!	-0.02	-0.05	-0.05	-0.02	#DIV/0!	-0.01	-0.10	0.18	-0.03	#DIV/0!	#DIV/0!	-0.01	-0.04
unknown_avg_slope	-0.03	-0.07	-0.02	-0.03	-0.01	0.02	#DIV/0!	-0.02	0.09	-0.04	-0.02	#DIV/0!	-0.01	0.04	-0.03	-0.03	#DIV/0!	#DIV/0!	-0.01	-0.03
max_slope_none	-0.21	-0.06	-0.15	-0.03	-0.09	0.05	#DIV/0!	0.06	-0.05	0.13	0.10	#DIV/0!	0.05	-0.10	0.07	0.10	#DIV/0!	#DIV/0!	-0.06	-0.09
max_slope_12	0.03	0.03	0.07	0.04	0.00	0.00	#DIV/0!	-0.02	0.11	-0.15	-0.04	#DIV/0!	-0.03	0.11	-0.05	-0.13	#DIV/0!	#DIV/0!	0.05	-0.06
max_slope_24	0.17	0.06	0.09	0.02	0.00	-0.02	#DIV/0!	-0.01	0.00	-0.02	-0.06	#DIV/0!	-0.03	0.01	-0.05	0.02	#DIV/0!	#DIV/0!	0.06	-0.06
max_slope_46	0.11	0.05	0.12	-0.07	0.12	-0.06	#DIV/0!	-0.04	-0.08	0.02	-0.04	#DIV/0!	-0.01	0.00	-0.02	0.02	#DIV/0!	#DIV/0!	-0.02	0.06
max_slope_68	0.26	0.12	-0.02	-0.03	0.01	-0.04	#DIV/0!	-0.02	-0.03	0.10	-0.02	#DIV/0!	-0.01	0.04	-0.02	-0.02	#DIV/0!	#DIV/0!	-0.01	-0.03
max_slope_810	0.04	-0.02	-0.01	-0.02	0.08	-0.02	#DIV/0!	-0.01	-0.02	-0.02	-0.01	#DIV/0!	0.00	0.02	-0.01	-0.01	#DIV/0!	#DIV/0!	-0.01	-0.02
max_slope_1020	-0.04	-0.07	-0.03	0.11	0.08	-0.05	#DIV/0!	-0.02	-0.04	-0.05	-0.02	#DIV/0!	-0.01	0.05	-0.03	-0.03	#DIV/0!	#DIV/0!	-0.01	-0.04
max_slope_big20	-0.05	-0.10	-0.02	-0.03	0.07	0.04	#DIV/0!	-0.02	-0.03	-0.04	-0.02	#DIV/0!	-0.01	-0.11	0.18	-0.02	#DIV/0!	#DIV/0!	-0.01	-0.03
unknown_slope_max	-0.03	-0.07	-0.02	-0.03	-0.01	0.02	#DIV/0!	-0.02	0.09	-0.04	-0.02	#DIV/0!	-0.01	0.04	-0.03	-0.03	#DIV/0!	#DIV/0!	-0.01	-0.03
water	0.24	0.25	0.00	-0.03	-0.08	-0.07	#DIV/0!	0.04	-0.10	0.29	-0.05	#DIV/0!	-0.02	0.19	-0.07	-0.07	#DIV/0!	#DIV/0!	-0.29	0.14
vlittle_hinder	0.22	0.18	0.16	0.31	-0.22	-0.03	#DIV/0!	0.07	-0.16	0.06	-0.01	#DIV/0!	-0.02	0.10	-0.11	-0.01	#DIV/0!	#DIV/0!	-0.04	0.06
little_hinder	-0.06	-0.01	0.03	-0.04	0.39	0.18	#DIV/0!	0.00	-0.50	-0.21	0.12	#DIV/0!	-0.08	-0.03	0.19	-0.10	#DIV/0!	#DIV/0!	-0.15	-0.04
reasonable_hinder	-0.03	-0.05	-0.13	-0.16	-0.23	-0.16	#DIV/0!	-0.03	0.54	0.22	-0.10	#DIV/0!	-0.03	0.00	-0.10	0.13	#DIV/0!	#DIV/0!	-0.06	0.05
much_hinder	-0.13	-0.13	-0.06	-0.07	-0.13	-0.10	#DIV/0!	-0.04	0.41	0.01	-0.04	#DIV/0!	-0.01	0.04	-0.06	0.02	#DIV/0!	#DIV/0!	-0.02	-0.08
unknown_hinder	-0.09	-0.06	-0.03	-0.04	0.02	0.05	#DIV/0!	-0.02	-0.05	-0.05	-0.02	#DIV/0!	0.57	-0.26	-0.03	-0.03	#DIV/0!	#DIV/0!	1.00	-0.04
well_lit	-0.39	-0.49	0.05	-0.18	0.07	0.18	#DIV/0!	0.12	0.17	-0.34	-0.02	#DIV/0!	-0.10	-0.15	0.18	0.11	#DIV/0!	#DIV/0!	-0.18	-0.24
somewhat_lit	0.19	0.33	-0.09	0.02	-0.02	-0.16	#DIV/0!	-0.07	-0.07	0.32	0.00	#DIV/0!	-0.02	0.16	-0.10	-0.10	#DIV/0!	#DIV/0!	-0.04	0.14
not_lit	0.34	0.34	0.03	0.21	-0.09	-0.10	#DIV/0!	-0.08	-0.15	0.15	0.03	#DIV/0!	0.15	0.10	-0.13	-0.04	#DIV/0!	#DIV/0!	0.05	0.18
unknown_lighting	-0.08	-0.08	-0.02	-0.03	0.05	0.08	#DIV/0!	-0.02	-0.04	-0.04	-0.02	#DIV/0!	-0.01	-0.21	-0.03	-0.03	#DIV/0!	#DIV/0!	0.81	-0.03
good_quality	0.06	0.07	-0.02	0.11	0.09	-0.07	#DIV/0!	0.02	0.14	-0.18	-0.08	#DIV/0!	-0.14	0.46	-0.23	-0.30	#DIV/0!	#DIV/0!	-0.25	0.10
reasonable_quality	-0.02	-0.06	0.04	-0.10	-0.12	0.03	#DIV/0!	-0.01	-0.12	0.23	0.10	#DIV/0!	-0.03	-0.32	0.13	0.34	#DIV/0!	#DIV/0!	-0.05	-0.08
bad_quality	-0.03	0.01	-0.03	-0.04	0.06	0.10	#DIV/0!	-0.02	-0.05	-0.05	-0.02	#DIV/0!	-0.01	-0.26	0.39	-0.03	#DIV/0!	#DIV/0!	-0.01	-0.04
unknown_quality	-0.09	-0.06	-0.03	-0.04	0.02	0.05	#DIV/0!	-0.02	-0.05	-0.05	-0.02	#DIV/0!	0.57	-0.26	-0.03	-0.03	#DIV/0!	#DIV/0!	1.00	-0.04
nice_beauty	0.31	0.30	0.00	0.19	-0.11	-0.05	#DIV/0!	-0.03	-0.12	0.25	-0.15	#DIV/0!	-0.05	0.13	-0.09	-0.07	#DIV/0!	#DIV/0!	-0.08	0.14
picturesque	0.00	-0.01	-0.02	0.13	-0.06	-0.04	#DIV/0!	-0.02	-0.04	0.08	-0.02	#DIV/0!	-0.01	0.04	-0.03	-0.03	#DIV/0!	#DIV/0!	-0.01	-0.03
neutral	-0.27	-0.27	0.00	-0.18	0.07	0.06	#DIV/0!	0.06	0.14	-0.23	0.17	#DIV/0!	-0.08	-0.07	0.10	0.06	#DIV/0!	#DIV/0!	-0.14	-0.18
boring	-0.01	0.00	0.05	-0.06	0.12	-0.05	#DIV/0!	-0.04	-0.01	-0.02	-0.04	#DIV/0!	-0.01	-0.01	-0.01	0.04	#DIV/0!	#DIV/0!	-0.02	0.15
unknown_beauty	-0.05	0.01	-0.02	-0.02	-0.04	-0.03	#DIV/0!	-0.01	-0.03	-0.03	-0.01	#DIV/0!	1.00	-0.15	-0.02	-0.02	#DIV/0!	#DIV/0!	0.57	-0.02
other_road	0.05	-0.01	0.18	0.29	-0.47	-0.28	#DIV/0!	0.18	0.06	0.41	-0.15	#DIV/0!	0.06	0.06	-0.24	0.17	#DIV/0!	#DIV/0!	-0.04	0.02
primary_road	-0.07	-0.06	-0.07	-0.08	-0.15	-0.11	#DIV/0!	-0.05	0.55	-0.07	-0.05	#DIV/0!	-0.02	0.11	-0.07	-0.07	#DIV/0!	#DIV/0!	-0.03	-0.04
along_busy_road	0.00	0.05	-0.15	-0.25	0.54	0.32	#DIV/0!	-0.16	-0.32	-0.38	0.18	#DIV/0!	-0.05	-0.08	0.28	-0.14	#DIV/0!	#DIV/0!	-0.10	0.00
unknown_class	-0.08	-0.08	-0.02	-0.03	0.05	0.08	#DIV/0!	-0.02	-0.04	-0.04	-0.02	#DIV/0!	-0.01	-0.21	-0.03	-0.03	#DIV/0!	#DIV/0!	0.81	-0.03
below_grade_crossing	-0.16	-0.18	0.14	0.09	-0.07	0.01	#DIV/0!	-0.04	-0.01	-0.09	0.08	#DIV/0!	-0.01	-0.10	0.03	0.12	#DIV/0!	#DIV/0!	-0.02	0.00
above_grade_crossing	-0.10	-0.12	0.00	-0.03	0.20	-0.07	#DIV/0!	0.04	-0.06	-0.12	0.03	#DIV/0!	-0.02	0.03	-0.01	-0.01	#DIV/0!	#DIV/0!	-0.03	-0.10
no_crossing	0.18	0.23	-0.08	-0.02	-0.15	0.06	#DIV/0!	-0.01	0.07	0.16	-0.07	#DIV/0!	0.02	0.08	-0.08	-0.06	#DIV/0!	#DIV/0!	0.04	0.09
tunnel_crossing	-0.02	-0.09	-0.02	-0.03	0.13	-0.04	#DIV/0!	-0.02	-0.03	-0.04	-0.02	#DIV/0!	-0.01	-0.18	0.28	-0.02	#DIV/0!	#DIV/0!	-0.01	-0.03

CORREL(X,Y)	fields	rural	nature	forest	green_c	city	unknow	signalize	roundab	uncontr	nvt	avg_slo	avg_slo	avg_slo	avg_slo	unknow	max_slo	max_slo
speed, Y																		
solitary_brom																		
solitary_fiets																		
protected_brom																		
protected_fiets																		
protected_pedelec																		
fietsstraat																		
gutter																		
ordinary_road																		
service_road																		
pedestrian																		
facility_unknown																		
asphalt																		
tiles																		
bricks																		
other_surface																		
unpaved																		
unknown_surface																		
fields	1.00																	
rural	-0.14	1.00																
nature	-0.08	-0.08	1.00															
forest	-0.10	-0.10	-0.06	1.00														
green_city	-0.27	-0.27	-0.15	-0.19	1.00													
city	-0.24	-0.24	-0.14	-0.17	-0.46	1.00												
unknown_surroundings	-0.04	-0.04	-0.02	-0.03	-0.08	-0.07	1.00											
signalized	-0.02	-0.02	-0.01	-0.02	-0.05	-0.04	0.57	1.00										
roundabout	-0.02	-0.02	-0.01	-0.02	-0.05	-0.04	0.57	0.00	1.00									
uncontrolled	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!								
nvt	0.03	0.04	0.02	0.02	0.07	0.06	-0.81	-0.71	-0.71	#DIV/0!	1.00							
avg_slope_none	0.10	0.11	0.08	-0.05	-0.03	-0.09	-0.18	-0.17	-0.17	#DIV/0!	0.24	1.00						
avg_slope_small	-0.10	-0.10	-0.05	0.15	-0.02	0.07	0.13	-0.02	0.25	#DIV/0!	-0.17	-0.66	1.00					
avg_slope_mid	-0.02	-0.02	-0.05	-0.06	0.03	0.01	0.15	0.29	-0.01	#DIV/0!	-0.20	-0.57	-0.06	1.00				
avg_slope_big	-0.04	-0.04	-0.02	-0.03	-0.04	0.14	-0.01	-0.01	-0.01	#DIV/0!	0.01	-0.29	-0.03	-0.03	1.00			
unknown_avg_slope	-0.03	-0.04	-0.02	-0.02	0.13	-0.06	-0.01	-0.01	-0.01	#DIV/0!	0.01	-0.24	-0.02	-0.02	-0.01	1.00		
max_slope_none	0.09	-0.08	-0.13	-0.08	0.00	0.11	-0.06	-0.08	-0.08	#DIV/0!	0.12	0.48	-0.32	-0.28	-0.14	-0.12	1.00	
max_slope_12	-0.06	0.16	-0.10	-0.07	0.03	-0.04	0.05	-0.03	0.15	#DIV/0!	-0.08	-0.03	0.17	-0.10	-0.05	-0.04	-0.56	1.00
max_slope_24	-0.06	-0.05	0.27	0.14	-0.01	-0.12	0.06	0.16	-0.03	#DIV/0!	-0.09	-0.34	0.12	0.45	-0.05	-0.04	-0.52	-0.19
max_slope_46	0.06	0.05	-0.04	0.08	-0.06	-0.04	-0.02	-0.01	-0.01	#DIV/0!	0.02	-0.21	0.17	0.11	0.07	-0.02	-0.26	-0.09
max_slope_68	-0.03	-0.03	0.25	0.09	-0.06	-0.05	-0.01	-0.01	-0.01	#DIV/0!	0.01	-0.05	0.09	-0.02	-0.01	-0.01	-0.10	-0.04
max_slope_810	-0.02	-0.02	-0.01	-0.01	0.06	-0.03	-0.01	0.00	0.00	#DIV/0!	0.00	0.02	-0.01	-0.01	-0.01	0.00	-0.06	-0.02
max_slope_1020	-0.04	-0.04	-0.02	-0.03	-0.03	0.12	-0.01	-0.01	-0.01	#DIV/0!	0.01	-0.21	0.15	-0.02	0.36	-0.01	-0.13	-0.05
max_slope_big20	-0.03	-0.03	-0.02	-0.02	-0.06	0.13	-0.01	-0.01	-0.01	#DIV/0!	0.01	-0.21	-0.02	-0.02	0.70	-0.01	-0.10	-0.04
unknown_slope_max	-0.03	-0.04	-0.02	-0.02	0.13	-0.06	-0.01	-0.01	-0.01	#DIV/0!	0.01	-0.24	-0.02	-0.02	-0.01	1.00	-0.12	-0.04
water	0.14	0.09	0.32	0.00	-0.08	-0.15	-0.29	-0.25	-0.25	#DIV/0!	0.35	0.16	-0.12	-0.12	-0.03	0.06	0.13	-0.11
vlittle_hinder	0.06	0.01	0.11	0.35	-0.06	-0.21	-0.04	-0.02	-0.02	#DIV/0!	0.04	0.05	-0.02	-0.02	-0.04	-0.04	-0.10	0.00
little_hinder	-0.04	0.05	-0.10	-0.14	0.11	0.02	-0.15	-0.08	-0.08	#DIV/0!	0.12	0.06	-0.09	-0.03	0.09	-0.02	0.00	0.01
reasonable_hinder	0.05	-0.05	0.00	-0.13	0.02	0.07	-0.06	-0.03	-0.03	#DIV/0!	0.05	-0.08	0.07	0.04	-0.06	0.07	0.09	-0.06
much_hinder	-0.08	-0.02	0.06	0.03	-0.16	0.20	-0.02	-0.01	-0.01	#DIV/0!	0.02	0.02	0.03	-0.05	-0.02	-0.02	0.05	0.07
unknown_hinder	-0.04	-0.04	-0.02	-0.03	-0.08	-0.07	1.00	0.57	0.57	#DIV/0!	-0.81	-0.18	0.13	0.15	-0.01	-0.01	-0.06	0.05
well_lit	-0.24	-0.15	-0.24	-0.38	0.27	0.36	-0.18	-0.10	-0.10	#DIV/0!	0.15	-0.07	0.00	0.05	0.07	0.06	0.03	0.00
somewhat_lit	0.14	0.28	0.13	-0.04	-0.10	-0.23	-0.04	-0.02	-0.02	#DIV/0!	0.03	0.06	-0.03	-0.01	-0.04	-0.03	-0.06	0.13
not_lit	0.18	-0.04	0.20	0.50	-0.23	-0.23	0.05	-0.03	-0.03	#DIV/0!	0.04	0.10	-0.02	-0.10	-0.05	-0.04	0.04	-0.14
unknown_lighting	-0.03	-0.04	-0.02	-0.02	-0.07	-0.06	0.81	0.71	0.71	#DIV/0!	-1.00	-0.24	0.17	0.20	-0.01	-0.01	-0.12	0.08
good_quality	0.10	0.02	-0.02	-0.07	0.07	-0.05	-0.25	-0.14	-0.14	#DIV/0!	0.20	0.03	-0.02	-0.03	0.00	0.04	-0.06	0.11
reasonable_quality	-0.08	0.01	0.03	0.09	-0.07	0.07	-0.05	-0.03	-0.03	#DIV/0!	0.04	0.02	-0.01	-0.01	0.01	-0.04	0.07	-0.14
bad_quality	-0.04	-0.04	-0.02	-0.03	0.08	0.01	-0.01	-0.01	-0.01	#DIV/0!	0.01	0.04	-0.03	-0.03	-0.01	-0.01	0.05	0.00
unknown_quality	-0.04	-0.04	-0.02	-0.03	-0.08	-0.07	1.00	0.57	0.57	#DIV/0!	-0.81	-0.18	0.13	0.15	-0.01	-0.01	-0.06	0.05
nice_beauty	0.14	0.09	0.25	0.33	-0.04	-0.40	-0.08	-0.05	-0.05	#DIV/0!	0.07	0.06	0.02	-0.09	-0.04	-0.02	-0.07	-0.04
picturesque	-0.03	-0.04	-0.02	0.17	0.03	-0.06	-0.01	-0.01	-0.01	#DIV/0!	0.01	0.04	-0.02	-0.02	-0.01	-0.01	-0.02	0.08
neutral	-0.18	-0.04	-0.22	-0.33	0.09	0.37	-0.14	-0.08	-0.08	#DIV/0!	0.12	0.00	-0.07	0.05	0.05	-0.02	0.08	0.02
boring	0.15	-0.07	-0.04	-0.05	-0.11	0.11	-0.02	-0.01	-0.01	#DIV/0!	0.02	-0.07	0.05	0.01	-0.02	0.11	0.01	-0.02
unknown_beauty	-0.02	-0.02	-0.01	-0.02	-0.05	-0.04	0.57	0.00	0.00	#DIV/0!	0.01	0.03	-0.02	-0.01	-0.01	-0.01	0.05	-0.03
other_road	0.02	-0.03	0.12	0.15	-0.14	0.03	-0.04	-0.07	-0.07	#DIV/0!	0.10	-0.01	0.03	0.01	-0.04	0.00	-0.01	-0.04
primary_road	-0.04	-0.04	-0.05	0.01	0.03	0.04	-0.03	-0.02	-0.02	#DIV/0!	0.02	-0.01	0.09	-0.06	-0.03	-0.02	0.05	0.03
along_busy_road	0.00	0.06	-0.09	-0.15	0.14	-0.04	-0.10	-0.05	-0.05	#DIV/0!	0.08	0.06	-0.11	-0.02	0.06	0.02	0.01	0.01
unknown_class	-0.03	-0.04	-0.02	-0.02	-0.07	-0.06	0.81	0.71	0.71	#DIV/0!	-1.00	-0.24	0.17	0.20	-0.01	-0.01	-0.12	0.08
below_grade_crossing	0.00	-0.07	-0.04	-0.05	0.03	0.08	-0.02	-0.01	-0.01	#DIV/0!	0.02	0.08	-0.05	-0.04	-0.02	-0.02	0.16	-0.09
above_grade_crossing	-0.10	0.00	-0.06	-0.07	0.06	0.08	-0.03	-0.02	-0.02	#DIV/0!	0.02	-0.10	-0.07	-0.06	0.28	0.35	-0.09	-0.07
no_crossing	0.09	0.05	0.07	0.09	-0.05	-0.14	0.04	0.02	0.02	#DIV/0!	-0.03	0.03	0.09	0.08	-0.20	-0.27	-0.04	0.12
tunnel_crossing	-0.03	-0.03	-0.02	-0.02	-0.06	0.13	-0.01	-0.01	-0.01	#DIV/0!	0.01	0.03	-0.02	-0.02	-0.01	-0.01	0.06	-0.04

CORREL(X,Y)	max_slo	max_slo	max_slo	max_slo	max_slo	max_slo	max_slo	unknow	water	vlittle_h	little_hi	reasona	much_h	unknow	well_lit	somewh	not_lit	unknow	good_q	reasona	bad_qua
speed, Y																					
solitary_brom																					
solitary_fiets																					
protected_brom																					
protected_fiets																					
protected_pedelec																					
fietsstraat																					
gutter																					
ordinary_road																					
service_road																					
pedestrian																					
facility_unknown																					
asphalt																					
tiles																					
bricks																					
other_surface																					
unpaved																					
unknown_surface																					
fields																					
rural																					
nature																					
forest																					
green_city																					
city																					
unknown_surroundings																					
signalized																					
roundabout																					
uncontrolled																					
nvt																					
avg_slope_none																					
avg_slope_small																					
avg_slope_mid																					
avg_slope_big																					
unknown_avg_slope																					
max_slope_none																					
max_slope_12	1.00																				
max_slope_24	-0.19	1.00																			
max_slope_46	-0.09	-0.09	1.00																		
max_slope_68	-0.04	-0.03	-0.02	1.00																	
max_slope_810	-0.02	-0.02	-0.01	0.00	1.00																
max_slope_1020	-0.05	-0.04	-0.02	-0.01	0.00	1.00															
max_slope_big20	-0.04	-0.03	-0.02	-0.01	0.00	-0.01	1.00														
unknown_slope_max	-0.04	-0.04	-0.02	-0.01	0.00	-0.01	-0.01	1.00													
water	-0.11	-0.10	-0.05	0.17	-0.01	0.05	-0.02	0.06	1.00												
vlittle_hinder	0.00	0.02	0.22	0.13	-0.02	-0.04	-0.03	-0.04	0.13	1.00											
little_hinder	0.01	-0.01	-0.04	-0.05	0.04	0.08	0.06	-0.02	-0.12	-0.49	1.00										
reasonable_hinder	-0.06	0.02	-0.10	-0.04	-0.02	-0.05	-0.04	0.07	0.11	-0.19	-0.63	1.00									
much_hinder	0.07	-0.09	-0.04	-0.02	-0.01	-0.02	-0.02	-0.02	0.02	-0.08	-0.28	-0.11	1.00								
unknown_hinder	0.05	0.06	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01	-0.29	-0.04	-0.15	-0.06	-0.02	1.00							
well_lit	0.00	-0.06	-0.02	-0.13	0.03	0.07	0.05	0.06	-0.12	-0.24	0.09	0.12	0.04	-0.18	1.00						
somewhat_lit	0.13	0.01	-0.04	-0.03	-0.02	-0.04	-0.03	-0.03	0.15	-0.01	0.00	0.03	-0.01	-0.04	-0.57	1.00					
not_lit	-0.14	0.05	0.06	0.18	-0.02	-0.05	-0.04	-0.04	0.10	0.31	-0.08	-0.16	-0.04	0.05	-0.70	-0.16	1.00				
unknown_lighting	0.08	0.09	-0.02	-0.01	0.00	-0.01	-0.01	-0.01	-0.35	-0.04	-0.12	-0.05	-0.02	0.81	-0.15	-0.03	-0.04	1.00			
good_quality	0.11	-0.05	0.01	-0.11	0.02	0.05	0.04	0.04	0.11	-0.07	0.01	0.08	0.07	-0.25	0.07	0.09	-0.11	-0.20	1.00		
reasonable_quality	-0.14	0.05	0.01	0.12	-0.02	-0.04	-0.03	-0.04	-0.01	0.10	0.01	-0.05	-0.06	-0.05	-0.04	-0.07	0.12	-0.04	-0.91	1.00	
bad_quality	0.00	-0.05	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01	-0.03	-0.04	0.09	-0.06	-0.02	-0.01	0.07	-0.04	-0.05	-0.01	-0.25	-0.05	1.00
unknown_quality	0.05	0.06	-0.02	-0.01	-0.01	-0.01	-0.01	-0.01	-0.29	-0.04	-0.15	-0.06	-0.02	1.00	-0.18	-0.04	0.05	0.81	-0.25	-0.05	-0.01
nice_beauty	-0.04	0.10	0.06	0.11	0.07	-0.03	-0.06	-0.02	0.29	0.26	-0.12	-0.04	-0.02	-0.08	-0.40	0.25	0.29	-0.07	0.06	-0.01	-0.08
picturesque	0.08	-0.04	-0.02	-0.01	0.00	-0.01	-0.01	-0.01	0.14	0.24	-0.12	-0.05	-0.02	-0.01	-0.15	-0.03	0.21	-0.01	-0.08	0.09	-0.01
neutral	0.02	-0.09	-0.08	-0.10	-0.06	0.04	0.06	-0.02	-0.24	-0.29	0.15	0.10	0.01	-0.14	0.47	-0.23	-0.35	-0.12	0.02	0.00	0.09
boring	-0.02	-0.05	0.08	-0.02	-0.01	-0.02	-0.02	0.11	0.00	0.00	0.05	-0.09	0.07	-0.02	-0.04	0.01	0.05	-0.02	-0.01	0.02	-0.02
unknown_beauty	-0.03	-0.03	-0.01	-0.01	0.00	-0.01	-0.01	-0.01	-0.02	-0.02	-0.08	-0.03	-0.01	0.57	-0.10	-0.02	0.15	-0.01	-0.14	-0.03	-0.01
other_road	-0.04	0.08	-0.04	0.02	-0.05	0.06	-0.08	0.00	0.14	0.25	-0.29	0.14	0.02	-0.04	-0.17	0.11	0.14	-0.10	-0.02	0.08	-0.12
primary_road	0.03	-0.05	-0.05	-0.02	-0.01	-0.03	-0.02	-0.02	0.00	-0.09	-0.32	0.29	0.34	-0.03	0.11	-0.03	-0.11	-0.02	0.09	-0.08	-0.03
along_busy_road	0.01	-0.08	0.07	-0.01	0.06	-0.04	0.10	0.02	-0.08	-0.20	0.46	-0.28	-0.18	-0.10	0.14	-0.09	-0.08	-0.08	0.02	-0.04	0.14
unknown_class	0.08	0.09	-0.02	-0.01	0.00	-0.01	-0.01	-0.01	-0.35	-0.04	-0.12	-0.05	-0.02	0.81	-0.15	-0.03	-0.04	1.00	-0.20	-0.04	-0.01
below_grade_crossing	-0.09	-0.08	-0.04	-0.02	-0.01	-0.02	-0.02	-0.02	-0.05	-0.01	0.06	-0.04	-0.04	-0.02	0.05	-0.07	0.01	-0.02	-0.03	0.05	-0.02
above_grade_crossing	-0.07	-0.09	0.12	-0.02	0.17	0.22	0.19	0.35	0.06	-0.10	0.14	-0.04	-0.06	-0.03	0.13	-0.04	-0.12	-0.02	0.05	-0.04	-0.03
no_crossing	0.12	0.13	-0.07	0.03	-0.13	-0.16	-0.14	-0.27	-0.01	0.09	-0.16	0.07	0.08	0.04	-0.14	0.08	0.10	0.03	-0.03	0.01	0.04
tunnel_crossing	-0.04	-0.03	-0.02	-0.01	0.00	-0.01	-0.01	-0.01	-0.02	-0.03	0.06	-0.04	-0.02	-0.01	0.05	-0.03	-0.04	-0.01	0.04	-0.03	-0.01

CORREL(X,Y)	reasona	bad_qua	unknow	nice_be	pictures	neutral	boring	unknow	other_r	primary_	along_b	unknow	below_g	above_	no_cros	tunnel_c
speed, Y																
solitary_brom																
solitary_fiets																
protected_brom																
protected_fiets																
protected_pedelec																
fietsstraat																
gutter																
ordinary_road																
service_road																
pedestrian																
facility_unknown																
asphalt																
tiles																
bricks																
other_surface																
unpaved																
unknown_surface																
fields																
rural																
nature																
forest																
green_city																
city																
unknown_surroundings																
signalized																
roundabout																
uncontrolled																
nvt																
avg_slope_none																
avg_slope_small																
avg_slope_mid																
avg_slope_big																
unknown_avg_slope																
max_slope_none																
max_slope_12																
max_slope_24																
max_slope_46																
max_slope_68																
max_slope_810																
max_slope_1020																
max_slope_big20																
unknown_slope_max																
water																
vlittle_hinder																
little_hinder																
reasonable_hinder																
much_hinder																
unknown_hinder																
well_lit																
somewhat_lit																
not_lit																
unknown_lighting																
good_quality																
reasonable_quality	1.00															
bad_quality	-0.05	1.00														
unknown_quality	-0.05	-0.01	1.00													
nice_beauty	-0.01	-0.08	-0.08	1.00												
picturesque	0.09	-0.01	-0.01	-0.07	1.00											
neutral	0.00	0.09	-0.14	-0.89	-0.12	1.00										
boring	0.02	-0.02	-0.02	-0.13	-0.02	-0.23	1.00									
unknown_beauty	-0.03	-0.01	0.57	-0.05	-0.01	-0.08	-0.01	1.00								
other_road	0.08	-0.12	-0.04	0.13	0.09	-0.11	-0.06	0.06	1.00							
primary_road	-0.08	-0.03	-0.03	0.00	-0.02	0.03	-0.05	-0.02	-0.26	1.00						
along_busy_road	-0.04	0.14	-0.10	-0.11	-0.08	0.12	0.08	-0.05	-0.87	-0.21	1.00					
unknown_class	-0.04	-0.01	0.81	-0.07	-0.01	-0.12	-0.02	-0.01	-0.10	-0.02	-0.08	1.00				
below_grade_crossing	0.05	-0.02	-0.02	-0.09	-0.02	0.06	0.09	-0.01	0.03	-0.05	0.00	-0.02	1.00			
above_grade_crossing	-0.04	-0.03	-0.03	-0.06	-0.02	0.07	0.00	-0.02	-0.14	-0.07	0.18	-0.02	-0.05	1.00		
no_crossing	0.01	0.04	0.04	0.12	0.03	-0.11	-0.05	0.02	0.12	0.09	-0.17	0.03	-0.56	-0.76	1.00	
tunnel_crossing	-0.03	-0.01	-0.01	-0.06	-0.01	0.06	-0.02	-0.01	-0.08	-0.02	0.10	-0.01	-0.02	-0.02	-0.23	1.00

Appendix C - GeoDa regression report

```
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set                : GeoDa_displaced_bigW
Spatial Weight          : GeoDa_displaced_v4_tagged_sym
Dependent Variable      :  relative_residual
Number of Observations:   466
Mean dependent var      :     -78.9614  Number of Variables    :   16
S.D. dependent var      :      913.707  Degrees of Freedom     :  450
Lag coeff.    (Rho)     :       0.5157

R-squared               :     0.311393  Log likelihood         :     -3763.97
Sq. Correlation         : -             Akaike info criterion :      7559.94
Sigma-square            :       574891  Schwarz criterion      :      7626.25
S.E of regression       :      758.215

-----------------------------------------------------------------------------
       Variable       Coefficient      Std.Error        z-value    Probability
-----------------------------------------------------------------------------
     W_relative_r        0.5157      0.0526228        9.79993      0.00000
         CONSTANT       520.716        579.022       0.899302      0.36849
           length    -0.0834345       0.123085      -0.677858      0.49786
            speed      -12.8649        24.7686      -0.519403      0.60348
         facility      0.691281       0.207664        3.32885      0.00087
          surface      0.548331       0.238498         2.2991      0.02150
      environment     0.0175929       0.317737      0.0553695      0.95584
intersection_type      -1.32232        2.29585      -0.575963      0.56464
        max_slope      0.702725       0.280395        2.50619      0.01220
            water      0.487409        1.67814       0.290446      0.77147
         nuisance      0.401715       0.229349        1.75154      0.07985
         lighting      0.498556       0.856004       0.582423      0.56028
          quality       2.47595        3.34395       0.740429      0.45904
           beauty      0.458605       0.350538        1.30829      0.19078
       road_class      -1.98904        1.55011       -1.28316      0.19944
        crossings      0.745415       0.314688        2.36875      0.01785
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                       DF      VALUE         PROB
Breusch-Pagan test                        14      280.2222      0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : GeoDa_displaced_v4_tagged_sym
TEST                                       DF      VALUE         PROB
Likelihood Ratio Test                      1       72.4318      0.00000
============================== END OF REPORT ================================
```

Appendix D - Excel regression report

tunnel_crossing	**no_crossing**	**above_grade_crossing**	**below_grade_crossing**	**along_busy_road**	**primary_road**	**other_road**
1	1.229168649	1.199185091	0.713248318	0.691490384	1	0.880320188
0	0.440676611	0.46234238	0.471180317	0.208651724	0	0.173380833
0.949355743	0.696871696	#N/A	#N/A	#N/A	#N/A	#N/A
171.1552541	420	#N/A	#N/A	#N/A	#N/A	#N/A
3823.435062	203.9646674	#N/A	#N/A	#N/A	#N/A	#N/A
boring	**neutral**	**picturesque**	**nice_beauty**	**unknown_quality**	**bad_quality**	**reasonable_qua**
0.924653083	0.995499143	1	1.317273224	5.498188912	1	0.784486685
0.432852901	0.396326172	0	0.38557549	0.596594803	0	0.329943352
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
good_quality	**not_lit**	**somewhat_lit**	**well_lit**	**much_hinder**	**reasonable_hinde**	**little_hinder**
0.821315144	0.677278737	1	0.750434251	1	1.031602711	1.107022868
0.327879674	0.147690961	0	0.133057712	0	0.183693122	0.216597137
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
vlittle_hinder	**water**	**max_slope_big20**	**max_slope_1020**	**max_slope_810**	**max_slope_68**	**max_slope_46**
1.211428262	1.289218871	3.034404871	1.213425084	1	1.335972802	1.575633176
0.241744195	0.15375339	0.959456529	0.807280961	0	0.844152908	0.734480229
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
max_slope_24	**max_slope_12**	**max_slope_none**	**avg_slope_big**	**avg_slope_mid**	**avg_slope_small**	**avg_slope_none**
1.75221474	1.407959328	1.309630281	0.267975559	0.670304798	0.800773709	0.724574497
0.726326925	0.724538175	0.719963429	0.931241376	0.810594914	0.805801609	0.789842754
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
nvt	**roundabout**	**signalized**	**city**	**green_city**	**forest**	**nature**
18.3791217	1.94039633	1	1.107552167	1.116420403	1.146638542	1
0.698332925	0.759546664	0	0.243053647	0.225721675	0.231315558	0
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
rural	**fields**	**unknown_surface**	**unpaved**	**other_surface**	**bricks**	**tiles**
0.94594329	0.984339787	1	1	1	0.791880687	1
0.224036008	0.223287418	0	0	0	0.203803355	0
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
asphalt	**facility_unknov**	**pedestrian**	**service_road**	**ordinary_road**	**gutter**	**fietsstraat**
0.871728382	1	1	0.876142395	0.767960995	0.782333387	1.125758315
0.162422621	0	0	0.614132937	0.577403762	0.573428671	0.606241207
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
protected_pedel	**protected_fiets**	**protected_brom**	**solitary_fiets**	**solitary_brom**	**speed**	**b**
1	1.30524249	1.052563554	1.334160686	1.155545881	1.0161287	1
0	0.57896158	0.580831033	0.580031419	0.590829655	0.031551075	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A

Appendix E - CC per characteristic report

Factor characteristic	*In direction of digitization*					*Against direction of digitization*				
facility	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**
voetgangersdoorsteekje	123	0.562	-0.376	0.919	TRUE	122	0.518	-0.42	0.878	TRUE
voetgangersgebied	137	0.614	-0.324	1.315	TRUE	135	0.574	-0.364	1.276	TRUE
bromfietspad (langs weg)	3882	0.685	-0.253	0.908	TRUE	2951	0.717	-0.221	0.962	FALSE
pedelec	70	0.689	-0.249	0.677	TRUE	26	0.785	-0.153	0.999	FALSE
ventweg	447	0.737	-0.201	0.983	FALSE	392	0.649	-0.289	0.942	TRUE
veerpont	22	0.849	-0.089	1.032	FALSE	22	0.91	-0.028	1.16	FALSE
weg met fiets(suggestie)strook	2916	0.881	-0.057	1.107	FALSE	2847	0.897	-0.041	1.13	FALSE
rijbaanwissel	1	0.939	0.001	0.939	FALSE	-	-	-	-	-
normale weg	21595	0.946	0.008	1.374	FALSE	21543	0.953	0.015	1.385	FALSE
solitair fietspad	2589	1.042	0.104	1.418	FALSE	2575	1.078	0.14	1.442	FALSE
fietspad (langs weg)	4872	1.088	0.15	1.313	FALSE	3277	1.056	0.118	1.298	FALSE
solitair bromfietspad	692	1.132	0.194	1.349	FALSE	699	1.053	0.115	1.371	FALSE
ONBEKEND	931	1.21	0.272	1.616	TRUE	919	1.199	0.261	1.633	TRUE
fietsstraat	104	1.24	0.302	1.37	TRUE	101	1.139	0.201	1.338	TRUE
surface	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**
onverhard	137	0.374	-0.545	0.754	TRUE	133	0.503	-0.416	0.824	TRUE
halfverhard	145	0.432	-0.487	0.664	TRUE	147	0.34	-0.579	0.679	TRUE
asfalt/beton	18626	0.897	-0.022	1.173	FALSE	17587	0.902	-0.017	1.185	FALSE
klinkers	8702	0.903	-0.016	1.194	FALSE	8626	0.92	0.001	1.209	FALSE
overig (hout/kinderkopjes e.d.)	120	0.918	-0.001	1.12	FALSE	117	0.936	0.017	1.246	FALSE
NULL	0	0.945	0.026	1.178	FALSE	0	0.937	0.018	1.193	FALSE
tegels	1644	1.052	0.133	1.276	FALSE	1371	1	0.081	1.259	FALSE
ONBEKEND	1021	1.154	0.235	1.388	FALSE	1012	1.154	0.235	1.408	FALSE
surroundings	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**
akkers/weilanden	1840	0.367	-0.552	0.622	TRUE	1710	0.346	-0.573	0.592	TRUE
landelijke of dorps	2804	0.548	-0.371	0.818	TRUE	2630	0.543	-0.376	0.814	TRUE
bos	797	0.55	-0.369	1.071	TRUE	738	0.609	-0.31	1.103	FALSE
natuur (behalve bos)	304	0.785	-0.134	0.967	FALSE	288	0.852	-0.067	1.097	FALSE
ONBEKEND	6709	0.98	0.061	1.212	FALSE	5343	0.983	0.064	1.237	FALSE
bebouwd (veel groen)	11325	0.991	0.072	1.247	FALSE	10633	0.996	0.077	1.264	FALSE
bebouwd (weinig of geen groen)	12301	1.013	0.094	1.296	FALSE	11980	1.014	0.095	1.302	FALSE
water	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**
ja	1455	0.659	-0.26	0.95	TRUE	1429	0.676	-0.243	0.963	TRUE
nee	28939	0.926	0.007	1.199	FALSE	27564	0.928	0.009	1.211	FALSE
ONBEKEND	5686	0.945	0.026	1.178	FALSE	4329	0.937	0.018	1.193	FALSE
lighting	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**
beperkt verlicht (bijvoorbeeld allee	2285	0.441	-0.478	0.751	TRUE	2157	0.433	-0.486	0.754	TRUE
niet verlicht	2544	0.485	-0.434	0.85	TRUE	2379	0.501	-0.418	0.834	TRUE
ONBEKEND	5712	0.945	0.026	1.18	FALSE	4357	0.937	0.018	1.194	FALSE
goed verlicht	25539	0.988	0.069	1.26	FALSE	24429	0.992	0.073	1.274	FALSE

quality	frequency	median_cc	Δ median cc	avg_cc	one_sd_delta	frequency	median_cc	Δ median cc	avg_cc	one_sd_delta
slecht	236	0.631	-0.288	0.925	TRUE	205	0.654	-0.265	0.937	TRUE
redelijk	5420	0.891	-0.028	1.177	FALSE	5222	0.879	-0.04	1.19	FALSE
goed	23702	0.912	-0.007	1.183	FALSE	22538	0.919	0	1.194	FALSE
ONBEKEND	6722	0.98	0.061	1.211	FALSE	5357	0.983	0.064	1.236	FALSE
hinderance	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**
zeer veel	22	0.348	-0.571	0.502	TRUE	23	0.241	-0.678	0.53	TRUE
zeer weinig	2769	0.892	-0.027	1.233	FALSE	2738	0.913	-0.006	1.245	FALSE
redelijk	6403	0.903	-0.016	1.152	FALSE	6359	0.905	-0.014	1.162	FALSE
weinig	19711	0.908	-0.011	1.183	FALSE	18413	0.91	-0.009	1.194	FALSE
veel	468	0.974	0.055	1.172	FALSE	447	1.028	0.109	1.204	FALSE
ONBEKEND	6707	0.98	0.061	1.212	FALSE	5342	0.982	0.063	1.236	FALSE
beauty	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**
schilderachtig	210	0.428	-0.491	0.809	TRUE	220	0.455	-0.464	0.916	TRUE
mooi	5508	0.621	-0.298	0.952	TRUE	5246	0.619	-0.3	0.958	TRUE
lelijk/saai	449	0.923	0.004	1.384	FALSE	432	0.965	0.046	1.319	FALSE
NULL	0	0.945	0.026	1.178	FALSE	0	0.937	0.018	1.193	FALSE
neutraal	23198	0.968	0.049	1.234	FALSE	22076	0.971	0.052	1.247	FALSE
ONBEKEND	1030	1.155	0.236	1.397	TRUE	1019	1.166	0.247	1.423	TRUE
within_built_up	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**
nee	6504	0.475	-0.444	0.779	TRUE	5907	0.482	-0.437	0.799	TRUE
ja	29576	1.014	0.095	1.275	FALSE	27415	1.009	0.09	1.284	FALSE
intersection_type	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**
nvt	30525	0.912	-0.007	1.186	FALSE	29122	0.917	-0.002	1.198	FALSE
kruispunt	3050	0.933	0.014	1.191	FALSE	805	0.931	0.012	1.196	FALSE
rotonde	1397	0.938	0.019	1.141	FALSE	2596	0.937	0.018	1.202	FALSE
kruispunt met VRI's	1108	1.013	0.094	1.227	FALSE	799	1.01	0.091	1.209	FALSE
avg_slope_fwd & bwd	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**
>4%	545	0.852	-0.067	1.138	FALSE	696	0.928	0.009	1.206	FALSE
ONBEKEND	81	0.86	-0.059	1.03	FALSE	69	0.834	-0.085	1.139	FALSE
1-2%	2227	0.864	-0.055	1.137	FALSE	2110	0.863	-0.056	1.16	FALSE
2-4%	1363	0.882	-0.037	1.16	FALSE	1496	0.924	0.005	1.226	FALSE
<1%	31864	0.927	0.008	1.192	FALSE	28951	0.924	0.005	1.2	FALSE

max_slope_fwd & bwd	frequency	median_cc	Δ median cc	avg_cc	one_sd_delta	frequency	median_cc	Δ median cc	avg_cc	one_sd_delta
8 - 10%	118	0.581	-0.338	0.908	TRUE	142	0.919	0	1.102	FALSE
>= 20%	30	0.601	-0.318	0.886	TRUE	23	0.606	-0.313	0.936	TRUE
4 - 6%	749	0.807	-0.112	1.106	FALSE	896	0.846	-0.073	1.185	FALSE
2 - 4%	2291	0.812	-0.107	1.105	FALSE	2322	0.863	-0.056	1.168	FALSE
10 - 20%	160	0.831	-0.088	1.168	FALSE	151	0.73	-0.189	1.024	TRUE
6 - 8%	293	0.838	-0.081	1.142	FALSE	355	0.884	-0.035	1.149	FALSE
ONBEKEND	82	0.857	-0.062	1.022	FALSE	70	0.824	-0.095	1.129	FALSE
1-2%	4519	0.858	-0.061	1.127	FALSE	4064	0.859	-0.06	1.167	FALSE
<1%	27838	0.946	0.027	1.207	FALSE	25299	0.94	0.021	1.209	FALSE
long_distance_route	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**
LF12b	3	0.091	-0.828	0.359	TRUE	1	0.943	0.024	0.943	FALSE
LF3-2b	12	0.221	-0.698	0.423	TRUE	6	1.337	0.418	1.292	FALSE
LF12a	3	0.263	-0.656	0.466	TRUE	1	0.901	-0.018	0.901	FALSE
LF3-2a, LF3-2b	13	0.567	-0.352	0.527	TRUE	13	0.748	-0.171	0.663	TRUE
LF4b, LF4a	39	0.591	-0.328	0.975	TRUE	196	0.804	-0.115	1.035	FALSE
LF3-3b	27	0.624	-0.295	1.001	TRUE	33	0.612	-0.307	0.817	TRUE
LF3-2b, LF3-2a	127	0.667	-0.252	0.868	TRUE	122	0.704	-0.215	0.891	TRUE
LF3-2a, LF3-3a, LF3-2b, LF3-3b	25	0.737	-0.182	0.736	TRUE	25	0.852	-0.067	0.823	FALSE
LF4a, LF4b	195	0.805	-0.114	1.028	FALSE	37	0.604	-0.315	1.04	TRUE
LF3-3b, LF3-3a	101	0.818	-0.101	1.091	FALSE	102	1.2	0.281	1.688	TRUE
LF12b, LF12a	18	0.832	-0.087	0.863	FALSE	18	0.724	-0.195	0.637	TRUE
LF4b	33	0.852	-0.067	1.123	FALSE	40	1.164	0.245	1.267	FALSE
GEEN	35254	0.922	0.003	1.188	FALSE	32548	0.922	0.003	1.2	FALSE
LF3-2a, LF4a, LF3-2b, LF4b	6	0.983	0.064	0.837	FALSE	6	0.947	0.028	0.801	FALSE
LF3-3a, LF3-3b	81	1.001	0.082	1.359	FALSE	83	1.018	0.099	1.202	FALSE
LF3-2b, LF3-3b, LF3-2a, LF3-3a	6	1.023	0.104	1.229	FALSE	6	0.877	-0.042	1.049	FALSE
LF12a, LF12b	24	1.043	0.124	1.339	FALSE	23	0.977	0.058	1.074	FALSE
LF3-3a	32	1.067	0.148	1.402	TRUE	20	1.192	0.273	1.229	FALSE
LF4a	63	1.266	0.347	1.529	TRUE	19	0.894	-0.025	1.071	FALSE
LF3-3a, LF3-2a, LF3-3b, LF3-2b	17	1.363	0.444	1.546	TRUE	17	1.433	0.514	1.517	TRUE
LF3-3b, LF4a, LF3-3a, LF4b	1	2.886	1.967	2.886	TRUE	1	3.426	2.507	3.426	TRUE
LF3-2a	-	-	-	-	-	5	0.246	-0.673	0.265	TRUE
primary_route	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**
beide	5977	0.838	-0.081	1.099	FALSE	5937	0.834	-0.085	1.107	FALSE
terug	118	0.869	-0.05	1.17	FALSE	2088	0.916	-0.003	1.078	FALSE
heen	3445	0.904	-0.015	1.092	FALSE	177	1.022	0.103	1.172	FALSE
geen	26540	0.943	0.024	1.218	FALSE	25120	0.942	0.023	1.23	FALSE

node_route	frequency	median_cc	Δ median cc	avg_cc	one_sd_delta	frequency	median_cc	Δ median cc	avg_cc	one_sd_delta
terug	157	0.624	-0.295	1.141	FALSE	1513	0.868	-0.051	1	FALSE
beide	5630	0.7	-0.219	0.965	TRUE	5620	0.708	-0.211	0.969	TRUE
heen	2468	0.85	-0.069	1.018	FALSE	257	0.848	-0.071	1.127	FALSE
geen	27825	0.974	0.055	1.246	FALSE	25932	0.97	0.051	1.26	FALSE
max_car_speed	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**Δ median cc**	**avg_cc**	**one_sd_delta**
70	2	0.315	-0.604	0.315	TRUE	-	-	-	-	-
60	2319	0.377	-0.542	0.7	TRUE	2316	0.365	-0.554	0.698	TRUE
80	685	0.424	-0.495	0.705	TRUE	690	0.47	-0.449	0.769	TRUE
	0	0.948	0.029	1.2	FALSE	0	0.942	0.023	1.208	FALSE
30	12932	0.95	0.031	1.232	FALSE	12834	0.962	0.043	1.254	FALSE
50	3405	0.957	0.038	1.222	FALSE	3293	0.974	0.055	1.238	FALSE
stapvoets (15)	455	1.037	0.118	1.357	FALSE	460	0.986	0.067	1.31	FALSE
ONBEKEND	4301	1.064	0.145	1.3	FALSE	4323	1.038	0.119	1.308	FALSE
crossings	**frequency**	**median_cc**	**median_diffe**	**avg_cc**	**one_sd_delta**	**frequency**	**median_cc**	**median_diffe**	**avg_cc**	**one_sd_delta**
ergens overheen	466	0.762	-0.157	0.962	TRUE	405	0.744	-0.175	0.985	TRUE
veerpont	21	0.82	-0.099	0.927	TRUE	21	0.782	-0.137	0.984	TRUE
tunnel	1	0.834	-0.085	0.834	TRUE	1	0.607	-0.312	0.607	TRUE
ergens onderdoor	271	0.88	-0.039	1.121	FALSE	249	0.817	-0.102	0.988	FALSE
nee	35321	0.922	0.003	1.189	FALSE	32646	0.923	0.004	1.203	FALSE