Geographical Information Management and Applications

Master thesis

Weather conditions and route choice of cyclists:

Investigating the impact of weather conditions on the route choice behaviour of commuting cyclists in the Netherlands

GIMA Master Thesis

Date: 05-06-2019
Student: Jesse Gruijters – 3912337 - 6036929
Supervisor: dr. Kees Maat
Professor: prof.dr.ir. Peter van Oosterom
Subject: Impact weather conditions on route choice behaviour of cyclists

Acknowledgements

Growing up in the south of the Netherlands, I travelled multiple kilometres by bicycle every day. The routes ran through different landscapes, variating from warm urbanized city centres to cold and dark forests. Years later, I got the opportunity to investigate different factors that may influence cyclists route choice behaviour.

As part of the master Geographical Information Management and Applications, I proudly present to you my final thesis on the following topic: *Investigating the impact of weather conditions on the route choice behaviour of commuting cyclists in the Netherlands*.

This master thesis would not have been completed without the support of a number of people, whom I would like to thank.

First of all, I would like to thank my supervisor Kees Maat for the valuable conversations, feedback and answering my questions. Furthermore, I would like to thank Paul van de Coevering and Dirk Bussche for providing and explaining the cycling dataset. Also, I would like to thank the Fietsersbond for providing the crucial cycling network dataset.

Finally, I am very grateful to have friends and family who took time to motivate and support me during this process.

Summary

Until a decade ago, relatively little scientific attention has been paid to the impact of climate change and changing weather patterns on the transport sector (Koetse & Rietveld, 2009). However, Helbich et al. (2014) argue that active transport modes like cycling receive increasingly more scientific attention (e.g., Buehler & Pucher, 2010; Aldred, 2013; Heinen, Maat & Wee, 2013). Especially the effects of weather on these weather-exposed active transport modes are of interest, according to Helbich et al. (2014). Most studies, however, focus on weather extremes, while the effects of normal weather conditions on daily travel behaviours are often neglected (Sabir, 2011; Böcker, Dijst & Faber, 2014; Böcker et al., 2015; Liu et al., 2017). Also, variations in land use and transportation networks may also affect the impact of weather conditions on route choice behaviour of cyclists (e.g. Brandenburg et al., 2004; Phung and Rose, 2008; Thomas et al., 2013; Helbich et al., 2014).

This study is designed to investigate the effects of weather conditions on the route choice behaviour of commuting cyclists using GPS trajectories. The following factors are taken into account that are expected to influence route choice behaviour of cyclists: personal characteristics, weather conditions, the spatial environment and travel distance. Previous studies provided evidence of a relationship between the first three factors and cyclists' route choice behaviour, which directly influences travel distance. Still no evidence has been provided through GPS trajectories of actual differences in route choice and travel distance due to weather conditions.

This study concluded that not all weather conditions influence route choice behaviour of commuting cyclists that strong. The influence of wind, precipitation and temperature on age and gender are noticeable within this study, however, these effects remain minimal. The presence of daylight is no significant predictor of route choice behaviour. However, this study indicates positive effects of temperature and precipitation on absolute deviation, meaning higher temperatures and a higher amount of precipitation causes longer travel distances. Within this study, relative deviation cannot significantly be explained by weather conditions.

An interesting outcome of this study is that commuting cyclists tend to deviate from the shortest route when the temperature rises, up to a temperature of approximately 25 ˚Celsius, then the deviation becomes smaller. Another interesting outcome of the regression analyses is the negative effect of the combination of much rain and hard wind. This result is not remarkable, since much rain and hard wind are harsh circumstance for most cyclists. On the other hand, wind direction, expressed as head- or tailwind in this study, is no significant predictor of absolute or relative deviation within this study.

Besides, within this study, personal characteristics and weather conditions do not significantly influence commuters' choice to cycle through specific spatial environment categories. There are however some weather conditions that can explain differences in route choice between built and natural environments, however, these effects are fairly weak.

Further research could focus on differences between types of respondents, types of bicycle, seasons and infrastructure. It would also be interesting to use a mixed methods approach which could make it possible to gain more insight into the cyclists' preferences or motivations behind their route choices.

Contents

1. Introduction

1.1 Context and Problem Statement

Nowadays climate change adaptation and mitigation receive much attention, scientific as well as societal. Renewed climatological research has revealed evidence for global temperature rise, changes in precipitation patterns and increased frequencies of extreme weather conditions (IPCC, 2007; Böcker et al., 2015). Until a decade ago, relatively little scientific attention has been paid to the impact of climate change and changing weather patterns on the transport sector (Koetse & Rietveld, 2009). In view of the continuous exposure of transport activities to weather conditions, this is a remarkable fact, according to Böcker et al. (2015). The authors continue by stating that earlier studies on the relationship between climate, weather and transport mainly focused on network performance of transportation systems. For example, enough research has been performed on the impacts of extreme heat, frost, storm, fog, rain and snow on rail, air and road infrastructures. Unlike its effects on network performance, the effects of weather on individual travel behaviours on the micro-level (people's daily choices for outdoor and indoor activities, destinations and transport modes for instance) received much less attention. This applies especially to research into active open-air transport modes, such as walking and cycling, while these are directly exposed to weather (Sabir, 2011; Sabir et al., 2011; Böcker et al., 2013; Böcker et al., 2015).

However, Helbich et al. (2014) argue that active transport modes like cycling receive increasingly more scientific attention (e.g., Buehler & Pucher, 2010; Aldred, 2013; Heinen, Maat & Wee, 2013). Especially the effects of weather on these weather-exposed active transport modes are of interest, according to Helbich et al. (2014). The most investigated weather conditions are temperature, precipitation, wind speed and snowfall (in Nordic countries and Canada). The impacts of humidity, fog, sunshine and cloud cover are less often investigated (Liu et al., 2017). Most studies, however, focus on weather extremes, while the effects of normal weather conditions on daily travel behaviours are often neglected (Sabir, 2011; Böcker, Dijst & Faber, 2014; Böcker et al., 2015; Liu et al., 2017).

Recently, Zhao et al. (2018) discussed several studies in which the impacts of weather on cycling from both survey analysis and big data mining have been investigated. Main findings from survey data analysis indicate that cyclists are affected by adverse weather conditions more seriously compared to other travel modes (Winters et al., 2007; Müller et al., 2008; Sabir, 2011; Liu et al., 2015a, 2015b). Since cyclists are less protected against the bad weather compared to motorised travellers (Liu et al., 2015b), they are more likely to reduce cycling trips due to cold or hot temperature (Richardson, 2000), precipitation (Bergström and Magnusson, 2003; Winters et al., 2007), and strong wind (Flynn et al., 2012).

In their literature review, Böcker et al. (2013) discuss the knowledge on everyday weather and individual travel behaviours. The authors argue that no elaborative overview of impacts of weather conditions on individual daily travel behaviour existed up to that date, although some reviews (Heinen, Wee & Maat, 2010; Koetse & Rietveld, 2009) discussed the interactions between individual characteristics, weather and daily mode choices.

Some earlier studies indicated that cyclists' travel behaviour vary significantly depending on seasons (Richardson, 2000; Bergström and Magnusson, 2003). More recent studies (Böcker, et al., 2015; Liu et al., 2015a, 2015b, 2016) showed that the impact of weather on bicycle choice not only depends on seasons but also on regions and travel purposes. Also, the impacts of weather on non-commuters are much more significant than on commuters.

According to Zhao et al. (2018), these studies provide a solid foundation for analysis of associations of weather and the use of bicycles. Liu, Susilo and Karlström (2017) agree and argue that traditional survey data includes detailed personal characteristics. These types of data can help researchers to further understand how weather affects travel decision-making processes. Edmond et al. (2009) already stated that women, more than men, avoid risk and anticipate by either not traveling by bicycle, or by choosing a different route where or when the travel conditions are safer (Edmond et al., 2009). For instance, women cyclists care more about the presence of daylight than men (Bergström & Magnusson, 2003). Concerning the route choice of cyclists and the distribution between man and woman, it can also be suggested that women will mainly choose the shortest route, while men are more willing to take a detour (Heinen, Wee & Maat, 2010).

However, impacts of weather elements such as temperature and rainfall on cycling are difficult to be quantitatively estimated from studies based on survey data due to small sample size (Liu, Susilo & Karlström, 2017; Zhao et al., 2018). Besides, cyclists are often directly subject to weather and thus respond more complicatedly to the changes in weather conditions as compared to car users or transit passengers. In addition, cycling can be used for both utilitarian and leisure purposes. To analyse the weather-cycling relationship, it is therefore important to include the trip purpose in the analysis (Zhao et al., 2018). By using technology of smart counters or cards, bicycle usage can be recorded automatically, which makes it possible to examine weather-cycling relationship at a larger scale to complement the self-reported survey results (Zhao et al., 2015).

Compared to survey data analysis, Zhao et al. (2018) argue, the research on big data mining in terms of the weather influence on bicycle usage is relatively small. One frequent reason is the lack of data. To investigate the weather-cycling relationship, it is necessary to consider the impact of the weather on different days of week or times of day on bicycle usage, which results in large amounts of data within each period of interest. However, automatic data collection for cycling have become available only in recent years with the installation of smart cycling counters and the usage of smart cards (Zhao et al., 2018). Nowadays, passively generated data (such as GPS trajectories on travel routes) can record multi-day travels of one to many individuals (Liu, Susilo & Karlström, 2017). Collecting GPS trajectories saves both time and money and solves the problem of limited sample size (Li, 2017). This creates many scientific opportunities, since the variation of everyday weather influences individual travel patterns throughout the year (Van Leeuwen, Koetse, Koomen, & Rietveld, 2009). The seasonality of the weather conditions, reflecting seasonal habituation effects, plays a significant role in explaining variability in daily travel behaviour. This emphasizes the importance of incorporating seasonal effects in the analysis of meteorological impacts (Creemers et al., 2015).

Most studies have focused on a single travel behaviour dimension, such as mode choice or trip distance, while only a few have developed models involving several travel behaviour dimensions (e.g. Böcker et al., 2015; Liu et al., 2015a). Moreover, scientists have tended to focus on the push and pull factors at the origin and destination points, while the actual route between those two points remained rather unexplored (Duppen and Spierings, 2013).

Variations in land use and transportation networks may also affect the impact of weather route choice behaviour of cyclists. Helbich et al. (2014) also emphasize that the locational component and the spatial variations in weather effects on behavioural outcomes have been underexplored. Although some empirical results demonstrate different effects of weather conditions on cycling between different spatial settings (e.g. Brandenburg et al., 2004; Phung and Rose, 2008; Thomas et al., 2013). For instance, Phung and Rose (2008) demonstrated that cycling in suburban and weather-exposed areas is more sensitive to precipitation than cycling in inner-city and sheltered areas. Also, Díaz et al. (2002) mentioned a higher impact of heat on vulnerable population groups such as elderly. Although no direct link to travel behaviours was made, urban heat island could influence travel behaviour of relatively older cyclists. GPS trajectories can provide a more exact representation of the different spatial environments cyclists are exposed to. It would also open ways to directly assess the effects of land use patterns and lower-scale street designs on route choice behaviour. Such an approach could therefore provide a clearer understanding of localised weather conditions and interrelated route choices (Helbich et al, 2014).

As mentioned before, climate change adaptation and mitigation receive much attention nowadays, scientific as well as societal. The scientific relevance of this research has already been discussed in the previous paragraphs. Such scientific insights can be relevant for the achievement of policy goals concerning sustainable and healthy transportation via weather-exposed transport modes like cycling (Böcker, Dijst, Prillwitz, 2013; Böcker, Prillwitz, Dijst, 2013; Böcker et al., 2015). This certainly applies to the Netherlands, since the use of the bicycle in the Netherlands continues to grow (Kennisinstituut voor Mobiliteitsbeleid, 2017). Regional and local governments in the Netherlands are determined to improve the infrastructure for cyclists. For example, the province of Noord-Brabant and various municipalities are developing so-called 'cycle highways' (Provincie Noord-Brabant, 2016). It can be very useful to reveal the influence of weather and the spatial environment on the route choice behaviour of cyclists in order to determine the best trajectories for those cycle highways.

Despite the fact that more and more research has been performed into the effects of weather conditions on transport, cycling and route choice behaviour, there still exists a gap in scientific knowledge. No research has been performed yet into the effects of weather conditions on the route choice behaviour of cyclists in which the focus is on GPS trajectories and the differences in route choice throughout the seasons of the year. The aim of this research is to fill this gap and to provide scientific insights for the achievement of policy goals regarding cycling.

1.2 Research objectives

This study is designed to investigate the effects of weather conditions on the route choice behaviour of commuting cyclists in which the focus is on GPS trajectories and the actual differences in route choice throughout the seasons of the year. The following factors are taken into account that are expected to influence route choice behaviour of cyclists: personal characteristics, weather conditions, the spatial environment and travel distance. Previous studies provided evidence of a relationship between the first three factors and cyclists' route choice behaviour, which directly influences travel distance. Still no evidence has been provided through GPS trajectories of actual differences in route choice and travel distance due to weather conditions.

The factors 'personal characteristics', 'weather conditions', 'spatial environment' and 'travel distance' are eventually expressed in various variables in a statistical model. The first factor, 'personal characteristics', consists of only two variables: gender and age. Cycling motive would also belong to personal characteristics. This study is, however, limited to commuters only and therefore this variable has the same value for each respondent, which makes it useless to include it in the analysis. The second factor, 'weather conditions', consists of four variables: temperature, precipitation, wind speed/direction and daylight. The third factor, 'spatial environment', represents different types of 'land use'. Land use refers to the spatial properties of the direct surroundings of the roads such as the amount of greenery and buildings. The last factor, 'travel distance', is expressed in 'route length', 'deviation' and 'length per category'.

The impacts of weather conditions on commuting cyclists' route choice behaviour are examined within the province of Noord-Brabant in the Netherlands; the Metropolitan Region of Eindhoven (MRE) to be specific, a region committed to cycling (Provincie Noord-Brabant, 2016).

This study only focuses on commuting cyclists, due to the following arguments. Commuting is often undertaken on a daily basis and out of necessity (Gehl, 2011) and is usually undertaken in a repetitive mode, as efficient as possible (Duppen & Spierings, 2013). Yet there is more complexity to be found in daily commutes than just getting from origin to destination via the most efficient trajectories, according to Duppen and Spierings (2013). The authors give a detour through a quiet and green environment as an example, which could be chosen with the purpose of avoiding chaotic or dangerous situations and saving time. Since weather conditions directly affect active transport modes such as cycling, it can be expected that people would also take a detour due to precipitation or wind speed for instance. Additionally, commuting is a daily activity, which makes it possible to investigate the impact of weather conditions on travel patterns throughout the year. Hereby, this study can cover everyday weather as well as 'extreme' weather conditions.

1.3 Research questions

The research objectives lead to the following main research question:

To what extent do personal characteristics and weather conditions influence commuting cyclists' route choice behaviour?

In order to give this research structure, the main research question is divided into three sub-questions which will be explained in more detail.

1. *How are personal characteristics of commuting cyclists related to weather conditions and route choice behaviour?*

The first sub-question covers the factor 'personal characteristics'. Earlier studies have shown that personal characteristics as gender and age can influence route choice behaviour of cyclists. The answers to this question can complement that knowledge and therefore personal characteristics of commuting cyclists cannot be neglected.

2. *To what extent do commuting routes deviate under different weather conditions?*

The second sub-question covers the factors 'weather conditions' and 'travel distance'. The influence of temperature, precipitation, wind speed/direction and daylight on commuting cyclists' route choice behaviour has already been discussed. However, evidence for the actual deviation of commuting routes under different weather conditions is still lacking. The answers to this question can therefore complement and improve the existing knowledge.

3. *To what extent do commuters cycle through different spatial environments under different weather conditions?*

The third sub-question covers the factors 'spatial environment' and 'travel distance'. Earlier studies already provided insights into the influence of the spatial environment on cyclists' route choice behaviour. Still, there is a lack of knowledge on weather-oriented properties of the spatial environment. Also, factors such as protection from wind and rain could influence the route choices of commuting cyclists. GPS trajectories can provide new insights into the weather-oriented properties of the spatial environment. The answers to this question can thus complement and improve the existing knowledge.

2. Theoretical framework

This chapter serves as theoretical framework to identify the weather conditions and environmental factors that affect cyclists' route choice behaviour. First, the literature about the relationship between weather, transport and route choice behaviour in general will be discussed, after which the focus is on the relationship between weather and the active, open-air, transport mode cycling. In addition, the influence of personal characteristics and the spatial environment on route choice behaviour of cyclists is treated. Also, different ways of assessing the relationship between weather and cycling. This chapter concludes with a conceptual model, which gives an overview of weather, personal and environmental related factors that have shown to influence cyclists' route choice in earlier research.

2.1 Weather and travel behaviour

Studies focusing on weather impacts on travel behaviour differ from each other in a variety of perspectives (Liu et al., 2017). Böcker et al. (2015) mention that studies that have investigated the travel-behavioural effects of objectively measured weather conditions on transport mode choices generally conclude that cold, cloudy, wet and windy weather conditions stimulate motorised transport, while warm, sunny and dry weather conditions increase usage of active modes – with typically larger effects for leisure than for utilitarian trips (e.g. Sabir, 2011; Creemers et al., 2014). The impact of weather has been found to be stronger on active and open-air transport modes, particularly cycling, than in-vehicle modes (Sabir, 2011; Liu et al., 2017). Helbich et al. (2014) argue that the effects of weather conditions on active transport modes like cycling receive increasingly more scientific attention. The most investigated weather conditions are temperature, precipitation, wind and snowfall (in Nordic countries and Canada). The impacts of humidity, fog, sunshine and cloud cover are less often investigated (Liu et al., 2017). Most studies, however, focus on weather extremes, while the effects of normal weather conditions on daily travel behaviours are often neglected (Sabir, 2011; Böcker, Dijst & Faber, 2014; Böcker et al., 2015; Liu et al., 2017). Recently, Böcker et al. (2013), Liu et al. (2017) and Zhao et al. (2018) discussed the knowledge on everyday weather and individual travel behaviours. The authors argue that no elaborative overview of impacts of weather conditions on individual daily travel behaviour existed up to that date, although some studies discuss the interactions between personal characteristics, weather and daily transport mode choices.

In the following paragraphs, the current knowledge on weather and travel behaviour is discussed. First, the three most investigated weather conditions are treated; precipitation, temperature and wind respectively.

2.2 Precipitation

In their literature review, Böcker et al. (2013) discuss several studies on the effect of precipitation on trip generation. Scientists (i.e. Hanbali & Kuemmel, 1993; Hassan & Barker, 1999; Call, 2011) indicated car traffic reductions during rain- or snowfall. Others (Hofmann & O'Mahony, 2005; Tang & Thakuriah, 2012) analysed public-transport ridership statistics and detected decrease in public-transport use during rain- or snowfall. Saneinejad et al. (2012) showed an increasing probability of travellers choosing to walk on rainy days. These studies, which are mainly focused on North America, conclude negative precipitation effects on the use of motorized transportation. In contrast, a national travel survey study from the Netherlands indicated a positive relationship between precipitation and the use of car and public-transport, resulting from large-scale switching from active to motorized transport modes (Sabir, 2011).

Nevertheless, many studies were focused on active transportation, due to its direct exposure to weather (Böcker et al., 2013). As Sabir (2011) also stated, precipitation is mentioned as one of the most important reasons not to cycle (Bergström & Magnusson, 2003'; Winters et al., 2007). Snow is the major factor that negatively impacts on bicycle usage (Cools et al., 2010). Cross-sectional studies (Winters et al., 2007; Parkin et al., 2008) revealed lower levels of cycling for areas with higher annual precipitation, although they admit other physical factors such as hilliness have larger impacts. Others (i.e. Richardson, 2000; Phung & Rose, 2008; Goetzke & Rave, 2011; Miranda-Moreno & Nosal, 2011; Thomas et al., 2013) proved similar negative effects of precipitation on cycling.

While many studies indicate the relevance of precipitation for trip generation, transport mode and destination choices, few have analysed the effects of precipitation on a wider range of travel decisions (Böcker et al., 2013). Three studies (Cools et al., 2010; De Palma & Rochat, 1999; Khattak & De Palma, 1997) conclude negative weather effects on route choice and adjustment of departure time, indicating that people anticipate on differences in expected travel times. In addition, Aaheim and Hauge (2005) discovered reduction in travelled distance due to precipitation, indicating that people choose closer destinations or cancel trips to further destinations.

Precipitation does not affect travel choices in the same way in all situations (Böcker et al., 2013). Some studies (Ahmed et al., 2010; Brandenburg et al., 2004) revealed temporal differentiations in relative weather impacts for cycling counts. Generally, this seems to indicate a larger effect on leisure compared with utilitarian trips. This conclusion is supported by most studies measuring trip purpose directly (Sabir, 2011). For commute trips, respondents only adapt departure times, while for shopping and leisure also mode and destination changes or trip cancelling were determined (Cools et al., 2010). An exception is a German study by Goetzke and Rave (2011), demonstrating significant precipitation effects on cycling to work, but not on cycling for leisure. This study indicates that effects may also differ for different population categories and between different geographical contexts (Böcker et al., 2013).

2.3 Temperature

The effects of temperature on travel choices are generally lower than precipitation effects (Cools et al., 2010; Sabir, 2011). Nonetheless, these two studies conclude that temperature has significant positive effects on walking and especially cycling, and negative effects on car and public transport. Earlier studies confirm these positive effects of temperature on active open-air transportation and prove that warmer weather increases cyclist rate (Bergström & Magnusson, 2003; Brandenburg et al., 2004; Phung & Rose, 2008; Richardson, 2000). Studies in the Netherlands, Canada and Sweden have shown that bicycle usage positively correlates with temperature until the temperature reaches 25°C (Liu et al., 2015a; Sabir, 2011; Saneinejad et al., 2012). Additionally, studies from warmer climates, find that not only low temperatures, but also high temperatures, between 25 and 30 °C, are disadvantageous for cycling (Ahmed et al., 2010; Phung & Rose, 2008; Richardson, 2000). Also, Díaz et al. (2002) mentioned a higher impact of heat on vulnerable population groups such as elderly. Although no direct link to travel behaviours was made, urban heat island could influence travel behaviour of relatively older cyclists. Temperature effects on travel purposes such as leisure are stronger than for commuting

(Sabir, 2011; Thomas et al., 2012). Additionally, Aaheim and Hauge (2005) discovered that travelled distances for shopping are reduced with higher temperatures, whereas for recreational purposes trip distances increase. So, temperature effects on travel behaviour differ for trip-purposes, personal characteristics and geographical context (Böcker et al., 2013).

2.4 Wind

Compared to precipitation and temperature, wind it is often overlooked in studies on the effects of weather on travel choices, except when it comes to cycling (Böcker et al., 2013). Aaheim and Hauge (2005) detected wind as a deterrent for cycling, mentioned by most respondents in their study in Norway. In a cross-sectional comparison of Dutch municipalities Rietveld and Daniel (2004) found a weak negative correlation between average annual wind speed and cycling. Flynn et al. (2012) in the USA and Heinen, Maat, and Van Wee (2011) in the Netherlands discovered that wind negatively affects bicycle commuting. In another Dutch study, Thomas et al. (2012) found negative wind effects on cycle flows. Others pointed at the different effects of light and strong wind. Sabir (2011) discovered no changes in Dutch modal split shares for moderate wind speeds between 1 and 4 Beaufort. Only in the case of heavy wind (of approximately 5 Beaufort or higher), cyclist shares decreased from 30% to less than 25%, mostly increasing the share of walking. Also, an Australian study (Phung & Rose, 2008) mentioned cycling declines only for strong winds. As with precipitation and temperature, some studies indicated that wind effects differ for trip purposes and personal characteristics (Böcker et al., 2013). In Flanders, Cools et al. (2010) declared that half of their respondents mention storms as a reason to postpone or cancel shopping and leisure trips, whereas for commute trips, storms hardly lead to cancellations and only changed departure times. Despite the moderate effect of wind speed on cycling, the effect of wind direction has not been investigated yet.

2.5 Weather, seasons and climate

The impact of everyday weather on transport systems is not directly noticeable because the variation of everyday weather influences the individual's travel patterns throughout the year (Van Leeuwen, Koetse, Koomen, & Rietveld, 2009; Liu et al., 2017). Some earlier studies already indicated that cyclists' travel behaviour vary significantly depending on seasons (Richardson, 2000; Bergström and Magnusson, 2003). The seasonality of the weather conditions, reflecting seasonal habituation effects, plays a significant role in explaining variability in daily travel behaviour. This emphasizes the importance of incorporating seasonal effects in the analysis of meteorological impacts (Creemers et al., 2015).

Liu et al. (2017) emphasize that most studies did not separate the effects of weather and climate, while these cannot be separated. For instance, 10°C in summer in a country of cold climate may have a completely different effect from 10°C in winter in the same country. The former may be interpreted as 'cold in summer', while the latter may be interpreted as 'warm in winter'. In an attempt to solve this issue, Sabir (2011) used dummy variables to represent trips that took place in different seasons (seasonal dummies), together with the temperature variables in a travel behaviour model and found significant effects of the seasonal dummies. However, Liu et al. (2017) continue, this approach hinders the interpretation of the effect of temperature variables because their parameters should then be interpreted as the effects of changing temperature values in different temperature intervals after controlling for the season, which does not seem to be reasonable. Liu et al. (2015a) proposed an alternative approach; to allow the parameters of different temperature intervals to interact with the parameters of seasonal dummies. In other studies, Liu et al. (2014, 2015b) proposed another approach to separate the effect of climate and the effect of weather. They used the mean of historical meteorological variables or thermal comfort measures in a given month and given location of each trip as a variable to represent climate effect. They used the standardised deviation against this variable to represent the weather effect (Liu et al., 2017). Liu et al. (2017) conclude that there is still no consensus on how the effect of climate should be represented. Therefore, more relevant empirical evidence on the effect of weather on cycling travel behaviour is needed. Analysing GPS-based cycling routes throughout the year can certainly be a step in the right direction.

2.6 Personal characteristics and cycling

As mentioned before, weather effects on cycling behaviour may differ for trip purposes and personal characteristics (Böcker et al., 2013). According to Heinen, Wee and Maat (2010) the impact of gender on cycling appears to be country specific. In countries with low cycling rates, men tend to cycle more; while in countries with high cycling rates, such as the Netherlands and Belgium, cycling is also popular among women (Heinen, Wee & Maat, 2010). There are studies, however, that indicate differences in cycle behaviour between gender.

Edmond et al. (2009) state that women, more than men, avoid risk and anticipate by either not traveling by bicycle, or by choosing a different route where the travel conditions are safer (Edmond et al., 2009). Concerning the route choice of cyclists and the distribution between man and woman, it can be suggested that women will mainly choose the shortest route, while men are more willing to take a detour (Heinen, Wee & Maat, 2010).

In their study in Sweden, Bergström and Magnusson (2003) discovered that women mention precipitation more often as reason not to cycle than men. Keay (1992) detects considerable female cyclist reductions during light rain in Australia, while male cyclist reductions only occur during heavier rainfall. Also, women, recreationists and commuters have a greater aversion to rain (Bergström and Magnussen, 2003). Böcker et al. (2015) found evidence that women and older aged people have colder thermal experiences.

The impacts of weather on non-commuters are much more significant than on commuters. For instance, rising temperature in a 'colder than normal' day encourages non-commuters to ride in warm months, while that is not the case in cold winter (Böcker, Dijst, et al., 2013; Liu et al., 2015a, 2015b, 2016). Diáz et al. (2002) mentioned a higher impact of heat on vulnerable population groups such as elderly. Although no direct link to travel behaviours was made, urban heat island could influence travel behaviour of older cyclists. Precipitation and wind speed affect both commute and non-commute cycling in a negative way (Aaheim & Hauge, 2005; Flynn et al., 2012; Heinen et al., 2011; Liu et al., 2017).

Moreover, women cyclists care more about the presence of daylight than men (Bergström & Magnusson, 2003). However, this important aspect of cyclists' travel behaviour, sunlight, is still overlooked. Currently, no research exists that looks into changing routes between sunset and sunrise.

2.7 Spatial environment and cycling

Variations in land use and transportation networks may also affect the impact of weather route choice behaviour of cyclists. Most studies have focused on a single travel behaviour dimension, such as mode choice or trip distance, while only a few have developed models involving several travel behaviour dimensions (e.g. Böcker et al., 2015; Liu et al., 2015a). Moreover, scientists have tended to focus on the push and pull factors at the origin and destination points, while the actual route between those two points remained rather unexplored (Duppen and Spierings, 2013).

Helbich et al. (2014) also emphasize that the locational component and the spatial variations in weather effects on behavioural outcomes have been underexplored. Although some empirical results demonstrate different effects of weather conditions on cycling between different spatial settings (e.g. Brandenburg et al., 2004; Phung and Rose, 2008; Thomas et al., 2013). Aaheim and Hauge (2005) revealed larger precipitation effects on mode choice in the Bergen city centre compared with the outskirts, resulting from a more exclusive weather-independent car use in suburban areas. Phung and Rose (2008) demonstrated that cycling in suburban and weather-exposed areas is more sensitive to precipitation than cycling in inner-city and sheltered areas.

GPS trajectories can provide a more exact representation of the different environments cyclists are exposed to. It would also open ways to directly assess the effects of land use patterns and lower-scale street designs on route choice behaviour. Such an approach could therefore provide a clearer understanding of localised weather conditions and interrelated route choices (Helbich et al, 2014). Since the personal characteristics regarding cycling and the environment of cyclists in the Netherlands may differ from the rest of the world, exploring the relationship between weather and cycling on route choice behaviour in the MRE may bring new insights for the Dutch context.

2.8 Different ways of assessing the relationship between weather and cycling

According to Zhao et al. (2018), survey analysis and big data mining provide a solid foundation for analysis of associations of weather and the use of bicycles. Liu et al. (2017) agree and argue that traditional survey data includes detailed individual characteristics. These types of data can help researchers to further understand how weather affects travel decision-making processes. However, impacts of weather elements such as temperature and rainfall on cycling are difficult to be quantitatively estimated from studies based on survey data due to small sample size (Liu et al., 2017; Zhao et al., 2018). Besides, cyclists are often directly subject to weather and thus respond more complicatedly to the changes in weather conditions as compared to car users or transit passengers. In addition, cycling can be used for both utilitarian and leisure purposes. To analyse the weather-cycling relationship, it is therefore important to include the trip purpose in the analysis (Zhao et al., 2018). By using technology of smart counters or cards, bicycle usage can be recorded automatically, which makes it possible to examine weather-cycling relationship at a larger scale to complement the self-reported survey results (Zhao et al., 2015).

Compared to survey data analysis, Zhao et al. (2018) argue, the research on big data mining in terms of the weather influence on bicycle usage is relatively small. One frequent reason is the lack of big data. To investigate the weather-cycling relationship, it is necessary to consider the impact of the weather on different days of week or times of day on bicycle usage, which results in large amounts of data within each period of interest. However, automatic data collection for cycling have become available only in recent years with the installation of smart cycling counters and the usage of smart cards (Zhao et al., 2018). Nowadays, passively generated data (such as GPS trajectories on travel routes) can record multi-day travels of a given individual (Liu et al., 2017). Collecting GPS trajectories saves both time and money and solves the problem of limited sample size (Li, 2017). This creates many scientific opportunities, since the variation of everyday weather influences individual travel patterns throughout the seasons (Van Leeuwen, Koetse, Koomen, & Rietveld, 2009), which plays a significant role in explaining variability in daily travel behaviour (Creemers et al., 2015).

2.9 Conclusion

Studies regarding cyclists' route choice behaviour have shown that they do not always choose the shortest route to their destination. Various weather, personal and environmental related factors influence cyclists' route choice, which are shown in figure 2.1. Although previous already studied these factors, except for sunlight, GPS trajectories, collected throughout the year, can provide new interesting insights. To emphasize once again; these factors may vary between different geographical contexts.

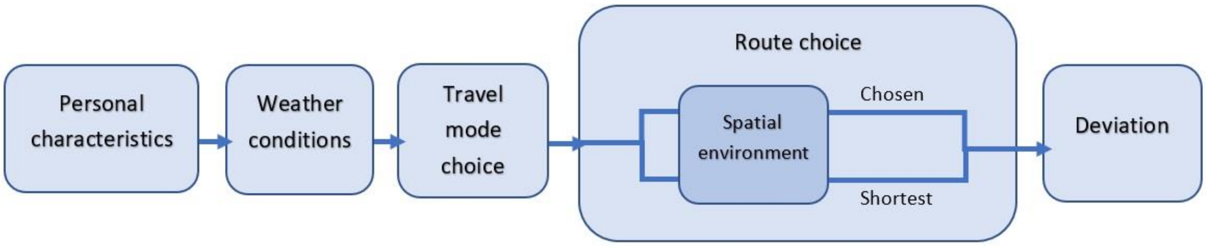

Figure 2.1: Conceptual model.

3. Methodology

This chapter describes how the research is carried out step by step and justifies the methods that are used for executing the study. First, the choice for the revealed preference method is explained by comparing stated and revealed preference studies. Thereafter, the study area is described. Followed by a detailed description of all the research methods within this study.

3.1.1 Stated preference studies

Route choice behaviour has been researched in many different ways. Until recent years, the literature on bike route choice was exclusively based on stated preference (SP) data (Zimmermann et al., 2017). For instance, respondent take part in a survey in which they were asked to evaluate routes based on their main characteristics (Winters et al., 2011; Zimmermann et al., 2017). Some SP based studies were limited to performing a descriptive analysis without estimating a formal model, while others used multinomial logit or regression analysis methods (Zimmermann et al., 2017). Although SP studies can be relatively inexpensively implemented and are able to evaluate alternatives that are not yet available, they also have a number of well-known shortcomings (Vedel et al., 2017; Zimmermann et al., 2017). The limitations of SP studies arise mostly from the difference between claimed and observed behaviour, as described by Sener et al. (2009). Therefore, SP studies does not always reflect the reality of the individual's route choice options and the preferences often do not manifest in reality (Broach et al., 2012; Tilahun et al., 2007; Winters et al., 2010). Zimmermann et al. (2017) confirm that it is indeed difficult for SP studies to put respondents in a setting where they can best reproduce the behaviour they exhibit in reality. Additionally, most stated preference studies are small-scale studies (Liu et al., 2017; Zhao et al., 2018). Therefore, they are often under sampled, biased and are not representative to generalize cycling behaviour because these studies deal with many limitations.

3.1.2 Revealed preference studies

Revealed preference (RP) studies were enabled by the emergence of geographic information systems (GIS) which gave access to new types of data. Data was then still collected through surveys, but instead of being put in hypothetical choice situations, participants had to recall their actual commuting routes, which were subsequently analysed with GIS (Zimmermann et al., 2017).

While providing useful insights, these first attempts to analyse bike route choice based on RP data never resulted in the estimation of a full route choice model, as observed by Broach et al. (2012). In particular, the models lack a comprehensive choice set of paths since the recalled route is compared mostly only to the shortest path (e.g. Harvey et al., 2008). The models focused on predicting specific aspects of route choice, such as the distance deviation from the shortest path or the presence of bike facilities but cannot be applied to predict path probabilities for a large set of routes. In other words, they are certainly useful for behavioural analysis, but not for trip distribution in a network (Zimmermann et al., 2017). The first RP study that overcame these limitations was the work of Menghini et al. (2010). Its main innovation was to exploit automatically processed GPS-based observations. Hence Menghini et al. (2010) were the first to obtain a large-scale GPS sample of cyclists' trajectories matched to a suitable network and to estimate a complete bike route choice model. Revealed preference studies are now considered as an established way to investigate route choice behaviour (Romanillos et al., 2015).

Some other studies followed the steps of Menghini et al. (2010), but overall the literature on bike route choice based on RP is still in its early stages compared to its car counterpart. Notably, Hood et al. (2011) extended the Zürich results of Menghini et al. (2010) to the US context, in a study based in San Francisco. Broach et al. (2012) contributed as well to the state of the art by estimating a model comprising a richer set of attributes.

Since GPS is getting more and more accurate and GPS is widely available large-scale studies can easily be set up at low costs (Hood et al., 2013). This provides opportunities to gain more insight into the previously under sampled cycling research. Also, possible gaps and ambiguity are avoided while identifying routes people use in reality (Zhu and Levinson, 2015).

Even though GPS studies are considered valuable, some challenges and limitations still exist. The quality of GPS data is influenced by external conditions, and therefore open space and clear skies are ideal for collecting accurate GPS data (Casello et al., 2011). This might lead to some inaccuracies in case a cyclist rides between or close to large structures, through tunnels or when it is clouded outside. Another challenge in revealed preference studies is the sampling of the cyclists, because it determines to a large extent for what purposes the data can be used and generalized. Many cycling studies deal with the sampling problem; the targeting of cyclists in studies from Sener et al. (2009) and Broach et al. (2012) lead to a sample of confident cyclists with a road warrior mentality for instance. Finally, Tilahun et al. (2007) state that revealed preference observes only the final consumer choice and does not take into account how the cyclists came to their final decision or how cyclists would act in case of future- or fictional situations.

3.1.3 Chosen method

In order to study the effects of weather on route choice behaviour, the choice has been made to use a GPS based revealed preference method. This method makes it possible to observe route choice behaviour by comparing the chosen routes of the cyclists to the shortest routes. The relationship between personal characteristics, weather, spatial environment and route choice behaviour is investigated by statistically comparing multiple attributes on the chosen and the shortest route. It is also examined whether the amount of deviation from the shortest route can be explained by the difference between the various attributes on both routes. The following paragraphs explain the used methods in more detail.

3.1.4 Research area

The impacts of weather conditions on commuting cyclists' route choice behaviour is examined within the Metropolitan Region of Eindhoven (MRE) to be specific. This area covers 21 municipalities (see figure 3.1), has over 750.000 inhabitants (CBS, 2018) and exists of different types of land use, such as agricultural land, forests, villages and a densely built up city centre. This region covers the same municipalities as the COROP region Noord-Brabant Zuidoost (CBS, 2017). COROP is a classification which is used for analytical purposes in the Netherlands.

The available GPS dataset called 'B-Riders' also covers this entire region. Finally, a weather station of the Royal Netherlands Meteorological Institute (KNMI) is located in Eindhoven, resulting in adequate and detailed information about weather conditions.

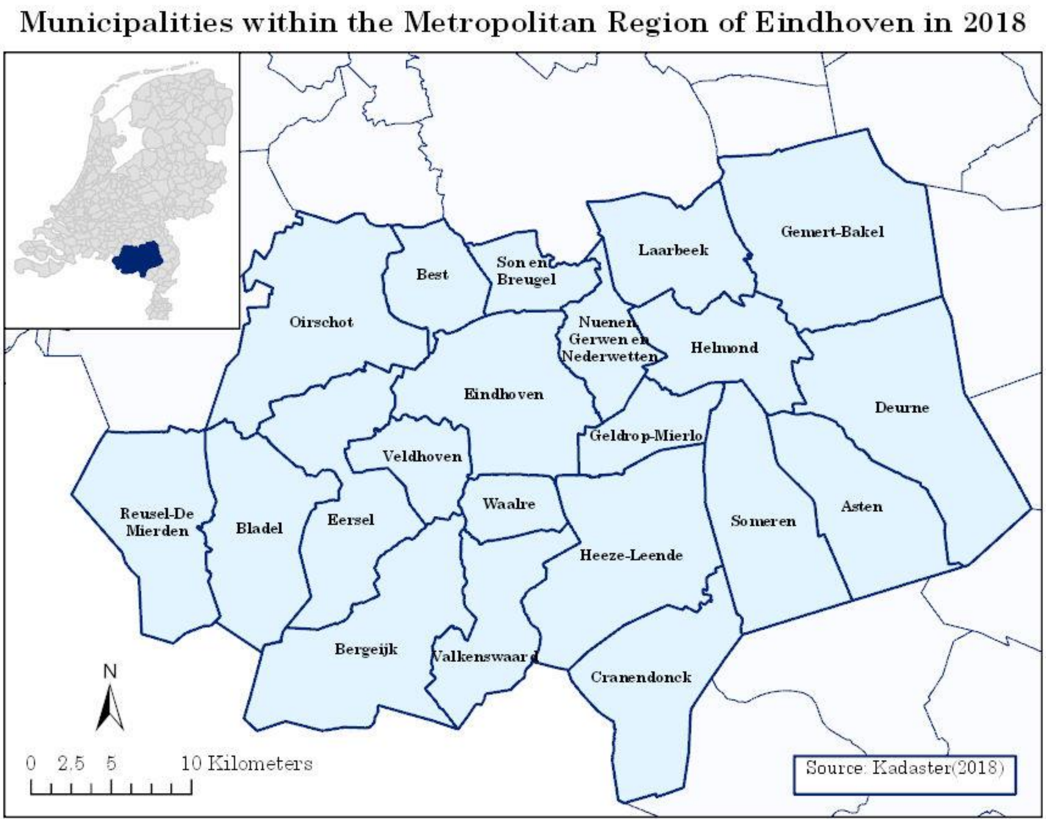

Figure 3.1: Municipalities within the Metropolitan Region of Eindhoven.

3.2 Research methods

This study can be divided into several activities. An overview of these activities, and thus the research process, are presented step-by-step in figure 3.2. The following paragraphs elaborate on the different steps of the research process, except for the literature research, since this has already been treated in section 2.

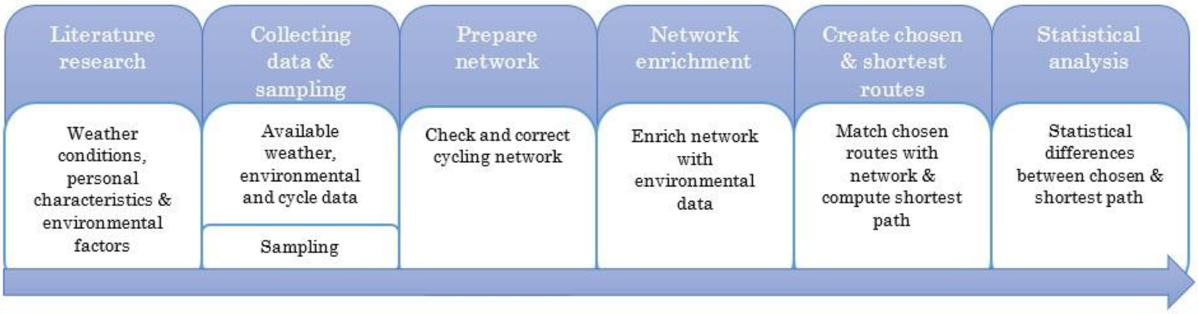

Figure 3.2: Research process.

3.3 Collecting data & sampling

For this study, several weather-, environmental- and cycle related datasets are needed. The next paragraphs describe the sources and formats of the used datasets.

3.3.1 B-Riders

- *Respondents, routes and cycle network*

The commuting cycle routes are derived from a GPS-based dataset on cycling behaviour, provided by B-Riders (2014). The B-Riders dataset emerged from an electric bicycle and speed pedelec stimulation program performed in the Dutch province of Noord-Brabant. The dataset covers the entire province of Noord-Brabant, which naturally covers the research area: The Metropolitan Area of Eindhoven. This program was originally initiated to generate interest for e-bikes, but also resulted in a lot of valuable GPS data. These GPS-tracks have been stored and numbered as 'links'. These links refer to unique line segments in a network dataset called 'Links', which is also obtained from B-riders. The quality of this network dataset will be checked using the cycle network dataset of the Fietsersbond (the Dutch Cycling Union), see section 3.3.2.

The B-Riders dataset is unique in its kind because it does not only provide GPS tracks, exact start and arrival times and cycling speed, but it also provides individual characteristics such as gender, age and trip purpose (e.g. commuting or recreational). Due to privacy concerns, the age of the respondents has been rounded to five years.

3.3.2 Fietsersbond and BBG

- *Cycle network and environmental data*

The second dataset that is used is the cyclist network dataset of the Fietsersbond, a national cycling association. This dataset consists of all digitized roads in the Netherlands that are accessible for cyclists and is composed by means of volunteered geographic information (VGI). According to Goodchild and Li (2012) VGI is a type of crowd-sourcing in which people create and contribute georeferenced facts about their spatial environment. The cycle network of the Fietsersbond, which they also use in their own route planner, can therefore be considered as a detailed and up-to-date source regarding cyclist networks in the Netherlands. This network is stored in vector format and contains additional attributes, such as road type, environment, degree of illumination and the presence of road salt during winter time. Because the network is built by volunteers, it is possible that values are subjective. That is why the environmental attributes are checked by comparing them to a reliable dataset that features information about the spatial environment: the BBG (the national land use dataset). After comparing the environmental attributes of both datasets in ArcGIS, the Fietsersbond datasets seemed to have many missing and misplaced values. Hence it has been decided to use the BBG as leading dataset for the spatial environment.

3.3.3 KNMI

- *Weather data*

The datasets regarding weather conditions are obtained from the Royal Dutch Meteorological Institute (KNMI). This Dutch national weather service maintains a database with detailed information on national, regional and local weather conditions of the present and previous years. One of the weather stations of the KNMI is located in Eindhoven. The KNMI therefore is the main source of hourly data on temperature, precipitation, wind speed/direction and daylight. These datasets are presented in csv or text format and have been converted to usable tables. Since the routes contained information about the data and the start and end time, it is possible to match the hourly weather conditions to the routes.

3.3.4 Data sample B-Riders dataset

As stated before, the B-Riders dataset contains information on age, gender, trip purpose and start and arrival time. In order to obtain the information needed for this study, one sample is initially drawn. Since the MRE is the research area, the dataset will be narrowed down to routes that have the origin and destination within this area. Each respondent has cycled at least one route within the MRE region. Subsequently, a random sample is drawn according to the following steps:

- The trip purpose is either from ‘home’ to ‘work’ or from ‘work’ to home.
- The dataset is firstly grouped by user id and then by route id.
- Per respondent, one random trip has been chosen.

This resulted in a sample containing one commuting trip per respondent that started and ended within the research area. Since this study focuses on commuters, it is important to compare its gender and age distribution to that of the employed labour force of the MRE region. As table 3.1 shows, it seems that the gender distribution is highly representative. Regarding age, the distribution is skewed: The respondents are substantially older than the average employed labour force of the MRE. This could be due to the fact that the B-Riders dataset consists of users of an electric bicycle or speed pedelec. For a detailed overview of the age distribution of the respondents, see Appendix I.

	Male	**Female**	**<25 years**	**25-44 years**	**>45 years**
MRE (N = 370.000)	55,4%	44,6%	14,8%	42,2%	43%
Sample (N = 734)	53,4%	46,6%	0,6%	26,7%	72,7%

Table 3.1: Age distribution MRE and sample.
Source: B-Riders (2014) & CBS (2019).

Initially, the sample exists of 734 routes. However, some routes will be excluded in order to achieve research objectives of this study as well as possible. All routes with a deviation of more than 10 kilometres will be excluded, since it is very unlikely that commuters would take a detour this large. This process is described in section 4.2.

3.4 Prepare and enrich cycling network

Before the chosen and shortest routes can be created it is necessary to prepare and enrich the cycling network. The B-Riders dataset is already provided with a routable cycle network called 'Links', existing of lines and vertices. These segments are all uniquely numbered and correspond with the variable 'linknumber' within the routes' csv files of the B-Riders dataset. This makes it possible to create the chosen routes. The methods used to compute the chosen and shortest routes are explained in more detail in section 3.5.

The quality of the 'Links' network is checked using the cycle network of the Fietsersbond. However, these two networks are geographically not perfectly matched. Which means that the computed routes would not match with the Fietsersbond network as well. In order to match the 'Links' network with the correct road segments of the Fietsersbond network, a buffer operation, with a distance of 5 meters, is performed. The line segments of the cycle network are selected when the centre of those segments lays within the buffer. The result is a geographically correct, routable cycle network. This network will subsequently be enriched with the spatial environment attributes of the BBG land use dataset (CBS, 2008), using geo-processing and spatial join methods. The original BBG land use dataset consists of multiple land use categories (see Appendix II), which are reclassified into six categories, as shown in table 3.2. Terrain that is impassable for cyclists will be excluded from the new attribute table. These six categories will be used both separately and combined. In order to investigate whether there are differences between urban and rural environment the six categories are also merged into just two categories: Built and Natural, shown in table 3.2 as well. These spatial environment categories are used to create the final dataset.

Reclassified category	**Merged category**
Residential, retail & public facilities	Built
Business & industrial	
Park & other public green	Natural
Open natural landscape	
Agricultural	
Closed natural landscape (mainly forest)	

Table 3.2: Reclassified and merged spatial environment categories.

3.5 Create chosen and shortest routes

After the routable cycling network is enriched with the spatial environment attributes, the chosen and shortest routes can be determined. As explained in section 3.3.1, the 'Links' network consist of uniquely numbered line segments, which correspond with the variable 'linknumber' in the routes' csv files of the B-Riders dataset. These csv files also contain the routeid for each unique route, a sequence code that shows the order in which the respondents have travelled specific road segments and the start and arrival time for each route. By using Modelbuilder in ArcGIS, a model (see Appendix III) is created which computes a shapefile for each chosen route within the sample. These shapefiles contain the length per line segment as well as per spatial environment category. Finally, all the chosen routes are merged into one shapefile that will be used to create the final dataset.

Before the shortest routes can be computed; the origin and destination points of each route have to be determined (see Appendix IV). Each chosen route exists of several links, stored in a specific order (sequence). This makes it possible to derive the beginning and end point of a route by selecting 'sequence 1' for origin point. The destination point was selected by a vice versa category; the 'sequence 1' in that category represented the end point of a route.

When the origin and destination points of the chosen routes have been determined, it is possible to calculate the shortest routes using the Network Analyst extension in ArcGIS, which can transform the cycle network into a route layer. During this process, the costs are assigned to the road's length in meters. Also, several restrictions such as travel direction and accessibility are set. Thereafter, a model (see Appendix V) is created using Modelbuilder in ArcGIS which computes the shortest route belonging to each chosen route. Just like the chosen routes, the chosen routes are also merged into one shapefile that will be used to create the final dataset.

Using another model (see Appendix VI), a straight line between the origin and destination points is also drawn, which made it possible to calculate the direction of travel for each route. This variable will be used to calculate to what extent the cyclists have head- or tailwind, the following section elaborates on this.

3.6 Creation of final dataset

To investigate whether the amount of deviation and the differences between environmental factors can be explained by personal characteristics and variation of weather conditions, the chosen and shortest routes have to be compared. Therefore, a final dataset will be created that can be used in the statistical analysis. The final dataset is a result of calculations with the output tables of the chosen and shortest routes, enriched with personal- and weather-related variables. To measure the deviation, both the absolute and relative deviation are calculated following these two formulas:

$$Absolute\ deviation\ (meters): \Delta length = \ length_{chosen} - length_{shortest}$$

$$Relative\ deviation\ (ratio): Ratio = \ length_{chosen}/length_{shortest}$$

For the relative deviation, a value of 1.00 indicates an identical length of the shortest compared to the chosen route, whereas values bigger than 1.00 indicate that the chosen routes are longer than the shortest routes. Due to the fact that the chosen routes are compared with the shortest possible routes, values below 1.00 cannot occur. The difference in length and coverage per spatial environment category is calculated as well. For all categories this is done in the following manner:

Difference in length (kilometres) per spatial environment category: $\Delta x = x_{chosen} - x_{shortest}$

Difference in coverage (%) per spatial environment category: $\Delta x = x_{chosen} - x_{shortest}$

The personal- and weather-related attributes do not differ between the chosen and the shortest routes, since they are both linked to the same respondent and timeframe. The final dataset (see section 4.3) will have the following personal- and weather-related attributes:

- Age in years (ratio)
- Gender (nominal)
- Head- or tailwind (ratio)
- Windspeed in km/h (ratio)
- Temperature in ˚Celsius (interval)
- Precipitation in mm (ratio)
- Dark/light (nominal)

For the nominal (dichotomous) variable attributes, dummy variables will be created to ensure that these variables can also be included in the statistical analysis. Before executing the statistical analysis, routes with an extreme deviation (>10km) will be eliminated from the final dataset. Table 3.3 is an example of the final attribute table of route 1575332.

Length	**Sample**	**Gender**	**Age**	**Temp.**	**Prec.**	**Windspeed**	**Headwind**	**Dark/ light**
9,37	Rain	F	24	12,8	12,1	14	96	Light
Abs. dev. Total	**Abs. Δ Built**	**Abs Δ Natural**	**Abs. Δ Cat. 1**	**Abs. Δ Cat. 2**	**Abs. Δ Cat. 3**	**Abs. Δ Cat. 4**	**Abs. Δ Cat. 5**	**Abs. Δ Cat. 6**
0,154	-0,146	0,3	-0,146	0,0	0,0	0,057	0,328	-0,085
Rel. dev. Total	**Rel. Δ Built**	**Rel. Δ Natural**	**Rel. Δ Cat. 1**	**Rel. Δ Cat. 2**	**Rel. Δ Cat. 3**	**Rel. Δ Cat. 4**	**Rel. Δ Cat. 5**	**Rel. Δ Cat. 6**
2,6%	-2,9%	2,9%	-2,9%	0,0%	0,0%	0,7%	4,3%	-2,1%

Table 3.3: Final attribute table for route 1575332.

3.7 Statistical analysis

Eventually, the influence of personal characteristics and weather conditions on route choice behaviour will be investigated using a multiple linear regression method. The main goal in this research for using regression analysis is to determine whether there is a significant relationship between the dependent and independent variables, and if so, how well this relationship can be explained in linear model(s). However, before such models can be built, several statistical tests are performed to determine whether there are significant differences between the chosen and shortest routes, between spatial environments and between males and females. After several statistical tests (such as T-test, ANOVA and Welch) are performed, an attempt can be made to create linear models.

3.7.1 Multiple linear regression

As a final product of this study, several multiple linear regression analyses will be executed. The models should explain average deviation or differences in length or coverage of the spatial environment categories by using a number of independent variables. Within these models, the focus lies on personal characteristics, weather conditions and the spatial environment. Although it seems logical that weather conditions and the spatial environment influence cycling route choice behaviour, these variables have not been studied before, using similar methods and amount of data. Therefore, this can be considered as an exploratory research. This makes it relevant to elaborate on the methodology of multiple regression models.

According to Braun and Oswald (2011) the combination of linear regression and exploratory research could cause some problems when analysing the impact of independent variables. The authors emphasize that one must be certain of the linearity and that one should be aware it is not confident whether a linear model fits the data the best. Besides Braun and Oswald (2011), Field (2013) also discusses multiple regression methodologies quite extensive. A common used way of interpreting and comparing variable effects within multiple linear regression is making use of the Beta (ß). In order to describe the studied phenomena correctly, it is important to gain insight into the relative importance of predictors within a regression model. However, determining the relative importance of predictors is always ambiguous when the predictors are correlating. Indices can therefore be calculated to reflect on the relative importance of predictors. Braun and Oswald (2011) mention three indices (*Incremental R^2*, *General dominance weights* and *Relative importance weights*), while Field (2013) only uses the incremental R^2 to validate predicator explanation of dependent variables. Since it is relatively easy to use the incremental R^2 in SPSS software, this method will also be used within the multiple regression analyses of this study.

Field (2013) describes several ways of entering predictors in a multiple regression model. The stepwise method in combination with automatic linear modelling may seem helpful when trying to find the best fitting model for exploratory predictors. Yet there are many disadvantages of this method. Model estimations are less reliable since this method usually results in missing significant predictors and in too high estimates of significance and R^2 values of the included predictors.

Besides, all weather conditions are interdependent in reality. In order to create a model that predicts reality as well as possible, it is important to include all weather-related variables simultaneously. Because of the exploratory elements of this study, the enter method is not ideal as well. Using the hierarchical method is a compromise between the two methods. Based upon the conceptual model (see section 2.9), the predictors are added to the model in the following sequence: (1) personal characteristics, (2) weather conditions.

3.7.2 Model variables

The dependent and independent variables in the linear model(s) are as follows:

Dependent (Y):	- The amount of deviation from the shortest route (in absolute or relative terms).
Independent (X):	- Personal characteristics. - The weather conditions at the time the route was driven.

OR

Dependent (Y):	- The difference in length or coverage of spatial environment category x from the shortest route (in absolute or relative terms).
Independent (X):	- Personal characteristics. - The weather conditions at the time the route was driven.

Resulting in models with the following structure: $Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + .. + b_kbX_k$

4. Data preparation

4.1 Prepare and enrich cycling network

Before the data can be analysed it is prepared to make sure there are no errors that may disrupt the research process. First of all, the cycling network is set up. This is an important step in order to create the chosen and shortest routes for all taken trips later on in the process. As stated in section 3.4, the 'Links' and Fietsersbond cycle networks were geographically not perfectly matched. The buffer operation resulted is a geographically correct, routable cycle network. Using geo-processing and spatial join methods, the network is enriched with the reclassified spatial environment attributes of the BBG land use dataset, as shown in figure 4.1.

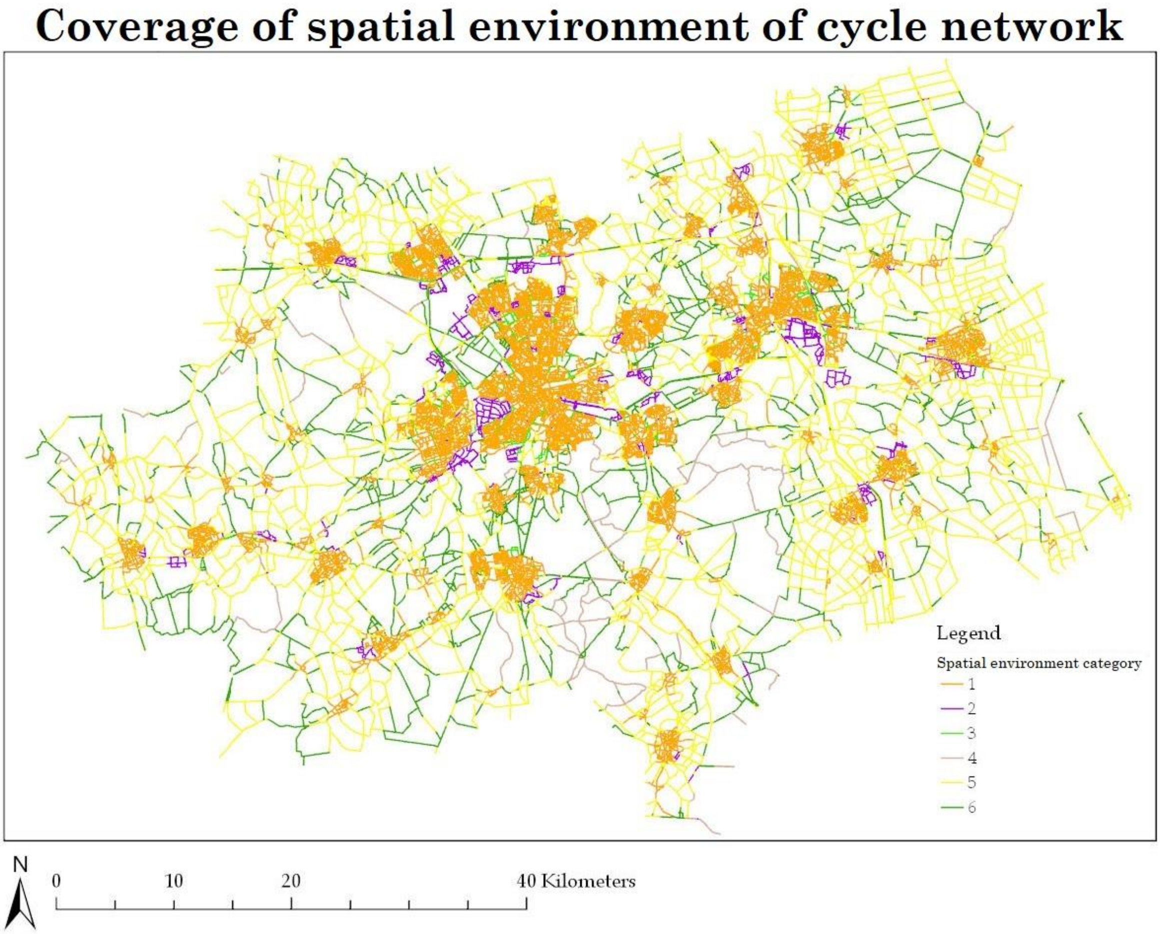

Figure 4.1: Coverage of spatial environment categories on cycle network.

A new attribute is added to the dataset; a column representing the newly calculated length per segment in meters. Using this variable, the travelled length per spatial environment category can be calculated. As figure 4.1 and table 4.1 show, a clear majority of the spatial environment of the cycle network consists of residential, retail and public buildings, followed by agriculture.

Spatial environment category	Description	Number of segments	Total length in km
1	Residential, retail & public facilities	36847	2591,273
2	Bussiness & industrial	3044	266,836
3	Park & other public green	1709	135,501
4	Open natural landscape	178	142,443
5	Agricultural	9833	1880,063
6	Closed natural landscape (mainly forest)	2604	881,679

Table 4.1: Distribution of spatial environment categories over cycle network.

4.2 Create shortest and chosen path

After the cycle network has been updated with the spatial attributes, the chosen routes can be determined. As already explained in section 3.4, the GPS-tracks have been numbered as 'links'. These numbers refer to unique line segments in the routable cycle network. By using ModelBuilder in ArcGIS, several geo-processing tools (see Appendix III) are deployed to compute the chosen routes. This resulted in 734 unique routes that are shown in figure 4.2.

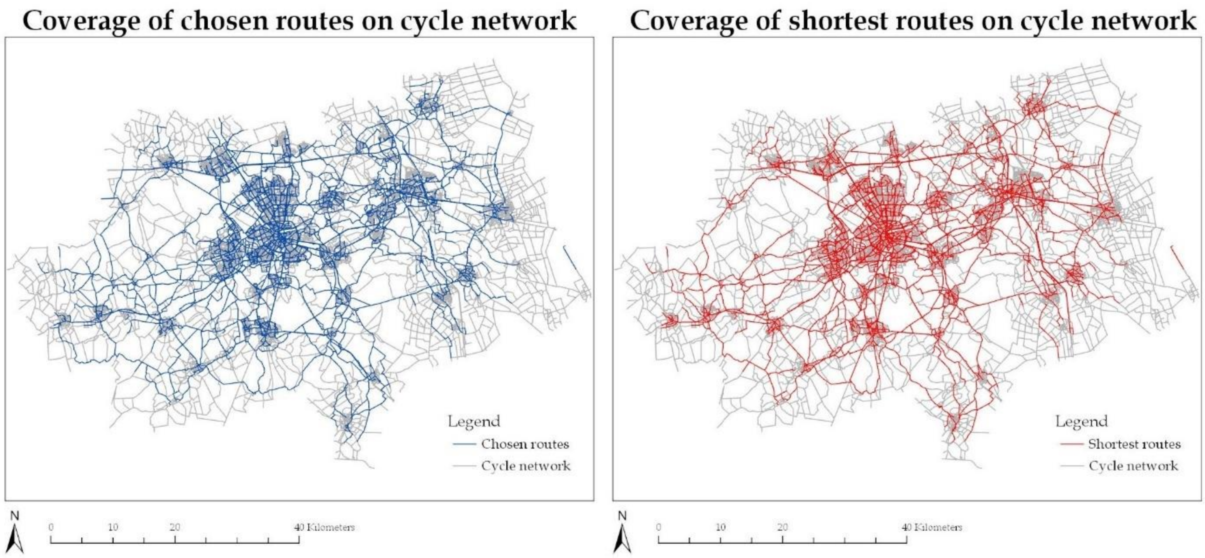

Figure 4.2: Coverage of chosen and shortest routes on cycle network, separate.

As explained in section 3.5, ArcGIS' Network Analyst extension and Modelbuilder are used to compute the shortest routes between the origin and destination point of each chosen route. Since the route layer does not feature environmental data, the shortest routes are spatially joined to the enriched cycle network in ArcGIS to add all attribute information. The shortest routes of the first sample are also shown in figure 4.2. At first glance, figure 4.3 shows that many chosen and shortest routes differ from each other. Figure 4.4 gives an example of the difference in spatial environment coverage between the chosen and shortest path of route 732592. Whether these differences are significant will be determined in section 5.

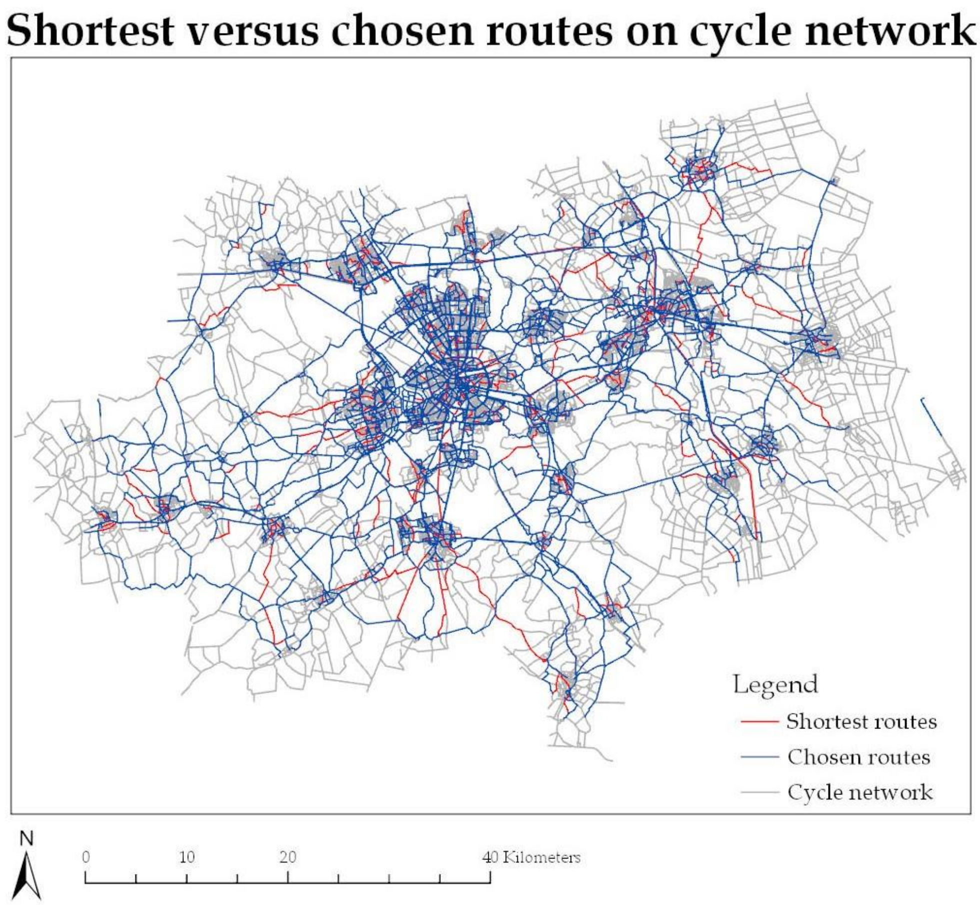

Figure 4.3: Coverage of chosen and shortest routes, combined.

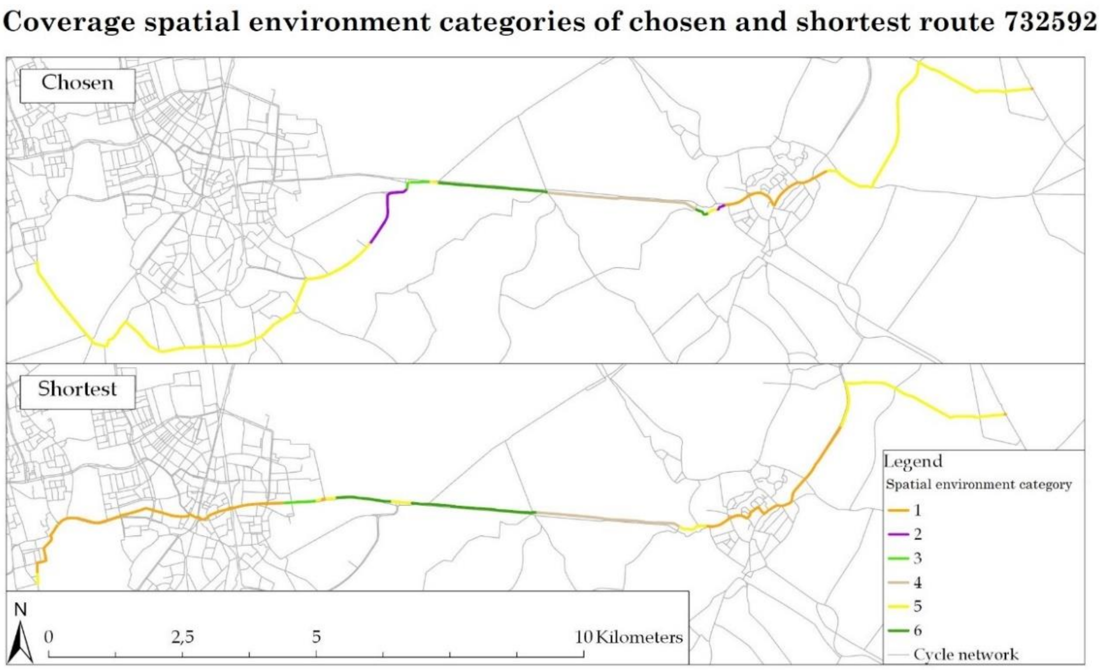

Figure 4.4: Example difference in spatial environment coverage between chosen and shortest route.

Before the data can be analysed, several outliers have been removed. Since this study is focussed on commuting cyclists, it can be assumed that routes which deviate more than 10 kilometres have more trip purposes than just commuting. Although a deviation of 10 kilometre is still very wide, no more routes have been removed in order to represent the reality as well as possible. Figure 4.5 gives an example of a route in which the cyclist has cycled more than twice as far; including this route, 27 cases have been removed, resulting in 707 routes.

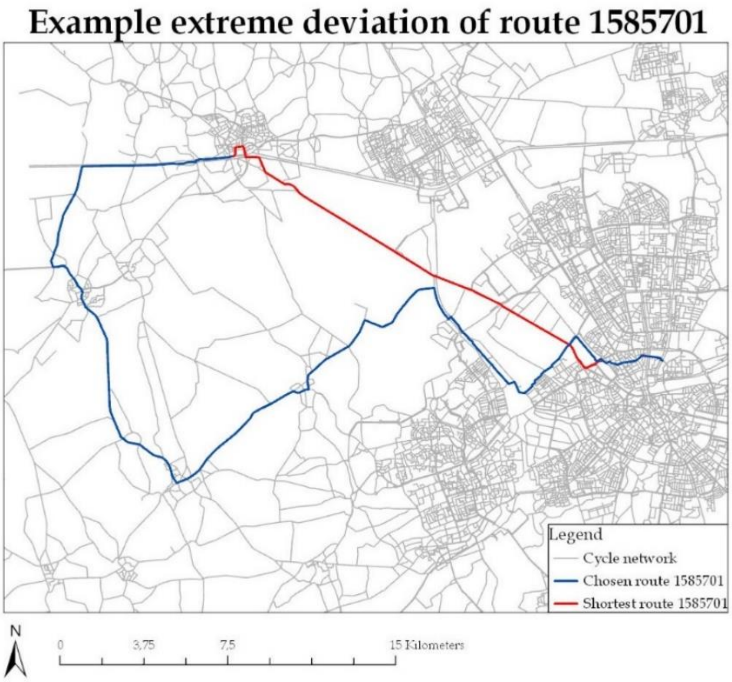

Figure 4.5: Example route with extreme deviation.

As already mentioned in section 3.5, the origin and destination points are not only used to compute the shortest routes. For each route, a straight line between the origin and destination points is drawn, which made it possible to calculate the direction of travel, as shown in figure 4.6. This variable is used to calculate to what extend the cyclists had head- or tailwind. The resulting variable is used in the final dataset, as will be explained in section 4.3.

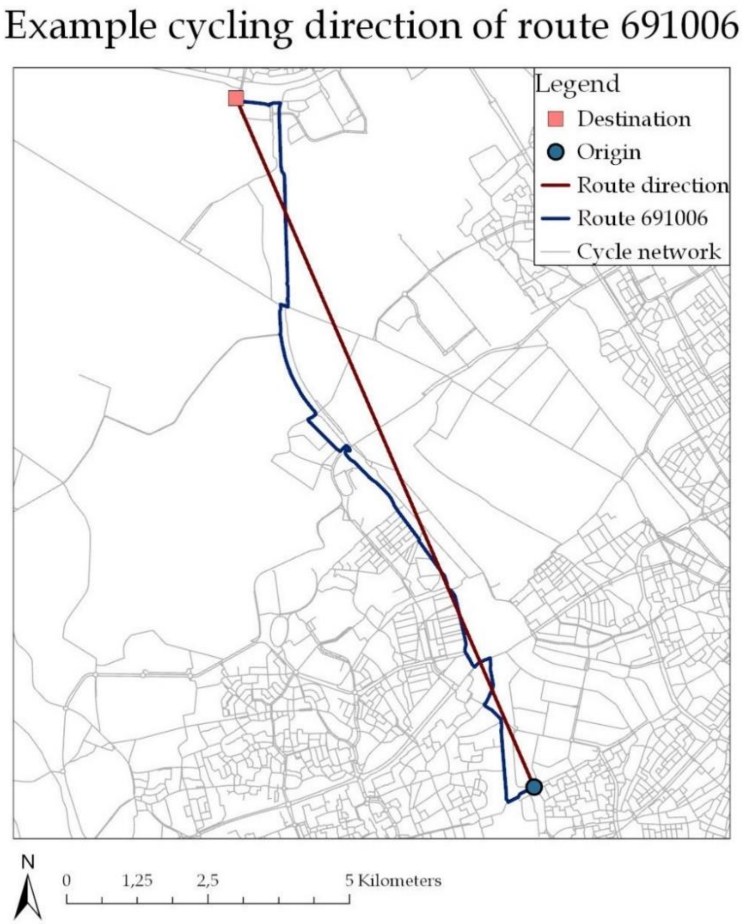

Figure 4.6: Example cycling direction of route 691006.

4.3 Final dataset

The chosen and shortest routes have already been linked to the enriched cycle network and the weather-related variables. Subsequently the differences (*Δ)* between the two routes, as shown in figure X, are calculated. The personal- and weather-related attributes do not differ between the chosen and the shortest routes, since they are both linked to the same respondent and timeframe. Besides, no calculations were needed for these variables, with the exception of 'Headwind'. This variable is calculated using the direction of travel and the wind direction.

There are differences between travel distance and length or coverage per spatial environment between the chosen and shortest routes. These differences are calculated according to the formulas discussed in section 3.6. Table 4.2 shows an example of the final attribute table of route 1575332. Initially, all length-related variables were calculated in meters. These variables have been converted from meters into kilometres since cyclists travel rather large distances and it can be assumed that they would not take distances that small into account when determining their route.

Length	**Sample**	**Gender**	**Age**	**Temp.**	**Prec.**	**Windspeed**	**Headwind**	**Dark/ light**
9,37	Rain	F	24	12,8	12,1	14	96	Light
Abs. dev. Total	**Abs. Δ Built**	**Abs Δ Natural**	**Abs. Δ Cat. 1**	**Abs. Δ Cat. 2**	**Abs. Δ Cat. 3**	**Abs. Δ Cat. 4**	**Abs. Δ Cat. 5**	**Abs. Δ Cat. 6**
0,154	-0,146	0,3	-0,146	0,0	0,0	0,057	0,328	-0,085
Rel. dev. Total	**Rel. Δ Built**	**Rel. Δ Natural**	**Rel. Δ Cat. 1**	**Rel. Δ Cat. 2**	**Rel. Δ Cat. 3**	**Rel. Δ Cat. 4**	**Rel. Δ Cat. 5**	**Rel. Δ Cat. 6**
2,6%	-2,9%	2,9%	-2,9%	0,0%	0,0%	0,7%	4,3%	-2,1%

Table 4.2: Final attribute table for route 1575332.

4.4 Sampling repeated

In order to increase the variance within the data, two sampling approaches are tested. As already explained in section 3.3.4, the first sample is drawn randomly. Subsequently, it is decided to repeat the same procedure for three new samples (see figure 4.7). However, these samples are not drawn randomly, but represent relatively extreme weather conditions. The three new samples meet the same requirements as the first sample, but additionally, the KNMI datasets are used to select the routes for these new samples. The following samples with one unique route per respondent, starting and ending within the study area, are drawn:

- Rain (399 routes with >1mm precipitation in the relevant hour)
- Warm (399 routes with >25˚C in the relevant hour)
- Wind (114 routes with >30km/h windspeed in the relevant hour)

The idea behind this is to investigate whether relatively extremer weather conditions can predict cyclists' route choice behaviour and to check if there are some mutual significant relations. Appendix VII compares the average weather conditions of the four samples with those of the research area. Including the first sample, 1619 unique routes from unique respondents will eventually be used in the statistical analysis. In the following sections, these samples will be referred to as follows: First, Rain, Warm and Wind.

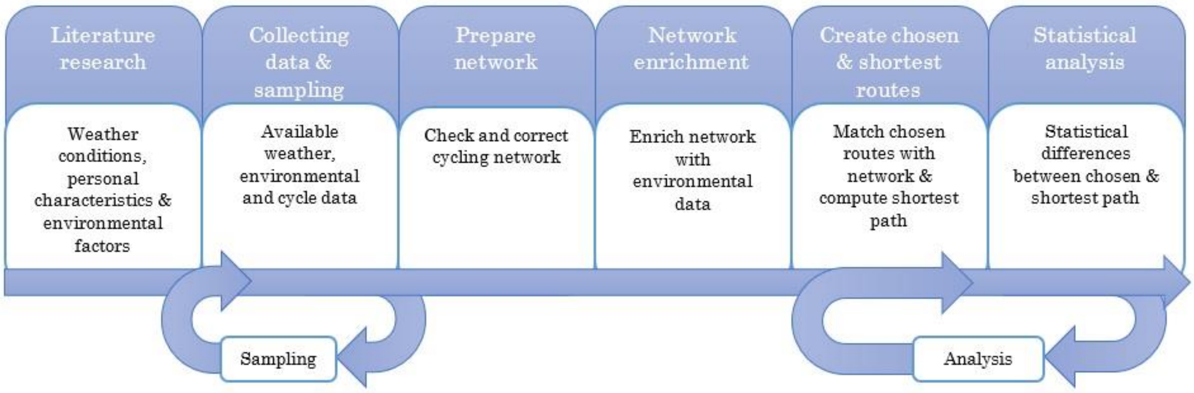

Figure 4.7: Partial repetition research process.

5. Data analysis

In this section, the results of the data analysis are presented. The first part will statistically investigate differences between the chosen and shortest routes, between spatial environments and between males and females. In the second part, the influence of personal characteristics and weather conditions on cyclists' route choice behaviour is examined using multiple linear regression models.

5.1 General route statistics

Before we can investigate what factors influence the deviation of the shortest routes, first the route lengths and deviations itself are investigated to see whether the respondents travelled further than the shortest alternatives at all. Also, the differences between spatial environments and personal characteristics are investigated.

5.1.1 Differences in route length

First of all, the routes' lengths are compared to investigate if the commuting cyclists actually travelled longer routes than the shortest possibilities. When comparing the descriptive statistics for both type of routes (see table 5.1), it shows at first glance that the shortest routes are shorter than the chosen routes in general. This difference, however, can be due to chance and therefore the difference needs to be proven statistically. To investigate whether there is a significant difference between the average deviation of the two types of routes a paired samples T-test is used (Field, 2009).

Therefore, the following null hypothesis is tested:

> *H0 = The average length of the shortest and chosen routes are not different from each other.*

The corresponding alternative hypothesis is:

> *HA = The average length of the shortest and chosen routes are different from each other.*

A paired samples T-test shows that on average, the length of chosen routes (M = 9,738) are significantly larger than the shortest routes (M = 8,922), $t(1619) = 28{,}8$, $p < 0.01$, r = 0. 0,582. The effect size can be classified as large.

Total sample	Length chosen (km)	Length shortest (km)	Abs. deviation (km) (chosen-shortest)	Rel. deviation (ratio) (chosen/shortest)
Count	1619	1619	1619	1619
Mean	9,738	8,921	0,817	1,144
St. deviation	5,673	5,352	1,143	0,375
Minimum	0,505	0,156	0	1,000
25%	5,463	4,869	0,194	1,029
50%	9,547	8,662	0,424	1,058
75%	13,000	12,158	0,942	1,114
Maximum	31,088	27,267	8,077	6,689

Table 5.1: Descriptive statistics trip length.

In general, the deviations from the shortest routes are not large. The first 50% of the chosen routes deviate less than 450 meters from the shortest route. Due to a number of outliers and variance in the data, the average deviation is 817 meters, or in relative terms;

the average deviation is 1.14 times as far as the shortest route. Even though many routes are concentrated around small deviations relative to the shortest route, the value of the standard deviation suggests that there is a large variance in the absolute deviation of all routes.

When comparing the descriptive statistics for the absolute deviation for the four different samples (see table 5.2), the average deviation differs between all four samples. The first sample seems to deviate less from the shortest route than the other three samples.

Absolute (km)	Total	1st sample	Rain	Warm	Wind
Count	1619	707	399	399	114
Mean	0,817	0,585	0,993	1,037	0,867
St. Deviation	1,143	0,637	1,401	1,461	1,052
Minimum	0	0	0,007	0,012	0,017
25%	0,194	0,169	0,227	0,211	0,202
50%	0,424	0,384	0,507	0,435	0,449
75%	0,942	0,790	1,117	1,172	1,090
Maximum	8,077	5,097	8,049	8,077	5,133

Table 5.2: Descriptive statistics absolute deviation.

To rule out if this is due to chance, the Welch test gives more insight into this matter. The Welch test is used as an alternative to the regular ANOVA test when the assumptions of the regular ANOVA are violated (Field, 2009). For the Welch test the *H0 = There is no difference in means between the four samples*; and the *HA = There is a difference in means between the four samples*. The Welch test shows that on average, the absolute deviation between the samples is significantly different: $F = 20.993$, $p < 0.01$. This allows to reject the null hypothesis with a 99% certainty. A Games Howell post hoc test is executed to find out which sample means significantly differ (Field, 2009). This test showed that the first sample significantly differed from the other three samples (see table 5.3). Negative mean differences suggest that on average the absolute deviation of the first sample is lower than the absolute deviation of the other three samples.

Games Howell		Mean difference	Sig.
First	Rain	-0,408*	0.000
	Warm	-0,453*	0.000
	Wind	-0,282*	0.031
Rain	First	0,408*	0.000
	Warm	-0,045	0.971
	Wind	0,126	0.724
Warm	First	0,453*	0.000
	Rain	0,045	0.971
	Wind	0,171	0.505
Wind	First	0,282*	0.031
	Rain	-0,126	0.724
	Warm	-0,171	0.505

Table 5.3: Results Games Howell test absolute deviation.

Relative (ratio)	Total	1st sample	Rain	Warm	Wind
Count	1619	707	399	399	114
Mean	1,144	1,090	1,179	1,200	1,135
St. Deviation	0,375	0,185	0,471	0,513	0,196
Minimum	1,000	1,000	1,001	1,003	1,004
25%	1,029	1,026	1,032	1,033	1,032
50%	1,058	1,052	1,059	1,066	1,070
75%	1,114	1,097	1,123	1,142	1,149
Maximum	6,689	4,022	5,266	6,689	2,119

Table 5.4: Descriptive statistics relative deviation.

When comparing the descriptive statistics for the four samples in relative terms (see table 5.4), the lowest average deviation is found again in the first sample. To find out if there is a significant difference between the average relative deviations (and between which samples), again the Welch Test and Games Howell post hoc tests are executed. For the Welch test the *H0 = There is no difference in means between the four samples*; and the *HA = There is a difference in means between the four samples*. The Welch test shows that on average, the means of the relative deviation between the samples are significantly different: $F = 9.223$, $p < 0.01$. The Games Howell test shows however that only the average deviation of the first sample is significantly different from the 2nd (Rain) and 3rd (Warm) sample (see table 5.5). This implies that on average the relative deviation of the first sample is smaller than in those two samples.

Games Howell		Mean difference	Sig.
First	Rain	-0,085*	0.003
	Warm	-0,107*	0.000
	Wind	-0,041	0.159
Rain	First	0,085*	0.003
	Warm	-0,022	0.926
	Wind	0,044	0.457
Warm	First	0,107*	0.000
	Rain	0,022	0.926
	Wind	0,066	0.162
Wind	First	0,041	0.159
	Rain	-0,044	0.457
	Warm	-0,066	0.162

Table 5.5: Results Games Howell test relative deviation.

5.1.2 Difference between spatial environments

Previous paragraph showed that there is a significant difference between the length of the chosen and shortest routes in general and that there are some significant differences in deviation between the four samples. However, possible variance between different types of spatial environment have not yet been discussed. Therefore, the spatial environment categories need investigation as well. Before the variables can be included in a regression model, this section examines whether the length and coverage per category is different on the chosen and shortest routes.

A method that is often used to do this is comparing means. To investigate whether the means are the same or significantly different for the chosen and shortest routes, a paired samples T-test is executed for all six categories that are investigated. Additionally, these categories are divided into two groups: built- and natural environment; a paired T-test is also executed for these two categories.

The following null hypothesis is tested first:

H0 = The average length per category of the shortest and chosen routes are not different from each other.

The corresponding alternative hypothesis is:

HA = The average length per category of the shortest and chosen routes are different from each other.

Spatial environment category	**Route**	**N**	**Mean (km)**	**Mean difference (chosen-shortest)**	**t**	**df**	**p value**	**r value (effect size)**
1	Chosen	1619	3,731	0,059	1,970	1618	0.049	0,049
	Shortest	1619	3,670					
2	Chosen	1619	0,809	0,120	6,978	1618	0.000	0,171
	Shortest	1619	0,689					
3	Chosen	1619	0,600	0,164	12,446	1618	0.000	0,296
	Shortest	1619	0,436					
4	Chosen	1619	0,181	0,035	2,571	1618	0.010	0,064
	Shortest	1619	0,146					
5	Chosen	1619	2,545	0,369	11,878	1618	0.000	0,283
	Shortest	1619	2,176					
6	Chosen	1619	1,874	0,070	2,638	1618	0.008	0,065
	Shortest	1619	1,804					
Built	Chosen	1619	4,539	0,180	6,090	1618	0.000	0,150
	Shortest	1619	4,359					
Natural	Chosen	1619	5,200	0,638	15,918	1618	0.000	0,368
	Shortest	1619	4,562					

Table 5.6: Mean differences of length per category (km).

As table 5.6 shows, the average length of all the categories of the chosen routes are significantly longer than those of the shortest routes. The null hypothesis of the T-test can be rejected with a 95% (or higher) certainty for all categories and the alternative hypothesis can be accepted; the average length per category of the shortest and chosen routes are different from each other. Although, they are considered to have a (extremely) small effect size. This does not guarantee that the coverage per category is also different between the chosen and shortest routes.

Therefore, the following null hypothesis is also tested:

H0 = The average coverage per category of the shortest and chosen routes are not different from each other.

The corresponding alternative hypothesis is:

HA = The average coverage per category of the shortest and chosen routes are different from each other.

Spatial environment factor	Route	N	Mean (%)	Mean difference (chosen-shortest)	t	df	p value	r value
1	Chosen	1619	42,9	-2,8	-8,427	1618	0.000	0,205
	Shortest	1619	45,7					
2	Chosen	1619	8,2	0,5	2,647	1618	0.008	0,066
	Shortest	1619	7,7					
3	Chosen	1619	6,4	1,2	8,003	1618	0.000	0,195
	Shortest	1619	5,2					
4	Chosen	1619	1,4	0,3	2,531	1618	0.011	0,063
	Shortest	1619	1,1					
5	Chosen	1619	24,4	1,5	5,619	1618	0.000	0,138
	Shortest	1619	22,9					
6	Chosen	1619	16,6	-0,7	-2,96	1618	0.003	0,073
	Shortest	1619	17,3					
Bebouwd	Chosen	1619	51,1	-2,3	-6,995	1618	0.000	0,171
	Shortest	1619	53,4					
Onbebouwd	Chosen	1619	48,9	2,3	6,995	1618	0.000	0,171
	Shortest	1619	46,6					

Table 5.7: Mean differences of coverage per category (%).

While the absolute differences are larger, the relative means of the categories do not differ that much. As table 5.7 shows, they differ between -2,8% and +2,3% and are considered to have a (extremely) small effect size. Still, the same applies to the difference in coverage per category as to the difference in length: the average coverage of all the categories of the chosen routes are significantly different from those of the shortest routes. The null hypothesis of the T-test can be rejected with a 98% (or higher) certainty for all categories and the alternative hypothesis can be accepted; the average coverage per category of the shortest and chosen routes are different from each other.

This section indicated significant differences between the length and coverage of the spatial environment categories in general. In addition, paragraph 5.1.1 already showed that there is a significant difference between the length of the chosen and shortest routes in general and that there are some significant differences in deviation between the four samples. The next step is to investigate whether there are significant differences in the length and coverage of the spatial environment categories between the four samples.

Absolute (km)	N	1	2	3	4	5	6	Built	Natural
Total	1619	0,059	0,120	0,164	0,035	0,369	0,069	0,179	0,637
First	707	0,025	0,076	0,146	0,035	0,321	-0,019	0,102	0,483
Rain	399	0,055	0,181	0,171	0,044	0,434	0,108	0,236	0,757
Warm	399	0,084	0,111	0,200	0,059	0,405	0,179	0,195	0,843
Wind	114	0,199	0,213	0,126	-0,086	0,314	0,102	0,411	0,455

Table 5.8: Mean differences in length per category.

Relative (%)	N	1	2	3	4	5	6	Built	Natural
Total	1619	-2,8	0,5	1,2	0,3	1,5	-0,7	-2,3	2,3
First	707	-1,8	0,2	1,0	0,3	1,6	-1,3	-1,6	1,6
Rain	399	-3,4	0,7	1,4	0,3	1,4	-0,3	-2,7	2,7
Warm	399	-3,8	0,4	1,4	0,4	1,7	-0,1	-3,3	3,3
Wind	114	-2,9	1,8	0,9	-0,5	1,3	-0,6	-1,2	1,2

Table 5.9: Mean differences in coverage per category.

When comparing the differences in length and coverage of the spatial environment categories of the four samples (see table 5.8 and 5.9) many (minor) differences can be observed. To find out whether these differences are significant, the Welch test and Games Howell post hoc tests are executed once more. The Welch test shows that the difference in length per category is significantly different between certain samples for category 2, 6, Built and Natural (see table X). Despite that, the Games Howell test indicates that the difference in length of the '2nd' category does not differ significantly between the samples. The Games Howell test still demonstrates the following significant differences:

- The difference in length of the '6th' category between the First and Warm samples (p = 0.022);
- The difference in length of the 'Built' category between the First and Wind samples (p = 0.020);
- The difference in length of the 'Natural' category between the First and Rain samples (p = 0.049); First and Warm samples (p = 0.003); Warm and Wind samples (p = 0.032).

As table 5.10 shows, the Welch test indicates that the average difference in coverage per category does not differ significantly between the samples. These results are confirmed by the Games Howell test as well.

Table 5.10: results Welch tests.

Difference in length per category		
Spatial environment category	**F**	**Sig.**
1	0.918	0.432
2	2.664	0.047
3	1.061	0.365
4	1.755	0.155
5	0.883	0.450
6	3.245	0.022
Built	3.231	0.022
Natural	5.628	0.001

Difference in coverage per category		
Spatial environment category	**F**	**Sig.**
1	2.265	0.080
2	1.783	0.150
3	0.529	0.663
4	2.534	0.056
5	0.067	0.977
6	1.507	0.212
Built	1.863	0.135
Natural	1.863	0.135

Until now the variables 'deviation' and 'spatial environment' have been analysed both separately and combined. We now know that the chosen routes are indeed significantly longer than the shortest routes and that there is a significant difference in deviation between the four samples. Also, there are significant differences in the length and coverage of the spatial environment categories between the chosen and shortest routes and between a part of the samples.

5.1.3 Differences in personal characteristics

As last step before all variables can be included in multiple regression models, personal characteristics need some investigation as well. Hence this section investigates whether there are differences in length and share per category between males and females. Also, the influence of age on differences in length and share per category is examined as well. At first glance, tables 5.11 and 5.12 show many differences between males and females. On average, men tend to deviate 153 meters further from the shortest route than women. To find out whether these differences are significant, the ANOVA and Welch test are executed once more.

Absolute (km)	N	Total	1	2	3	4	5	6	Built	Natural
Total	1619	0,817	0,059	0,120	0,164	0,035	0,369	0,069	0,179	0,637
Male	851	0,889	0,034	0,137	0,189	0,022	0,409	0,098	0,172	0,718
Female	768	0,736	0,087	0,101	0,136	0,050	0,325	0,038	0,188	0,548

Table 5.11: Mean differences in length per category.

Relative (%)	N	Total	1	2	3	4	5	6	Built	Natural
Total	1619	14,6	-2,8	0,5	1,2	0,3	1,5	-0,7	-2,3	2,3
Male	851	14,1	-3,3	0,5	1,4	0,1	2,0	-0,8	-2,8	2,8
Female	768	15,1	-2,3	0,5	0,9	0,4	1,0	-0,6	-1,8	1,8

Table 5.12: Mean differences in coverage per category.

Both tests show that on average, the means between males and females only significantly differ for the total absolute deviation, in route length in natural environment and in route length in category 3 (see table 5.13). A possible explanation for these small differences in deviation between males and females could be the use of e-bikes and speed pedelecs within the B-Riders program. When such bicycles are used, the travelled distance is less influenced by the amount of physical effort of the respondent.

Absolute				Relative		
Category	**Anova Sig.**	**Welch Sig.**		**Category**	**Anova Sig.**	**Welch Sig.**
Total	0.007	0.007		**Total**	0.653	0.654
Built	0.782	0.782		**Built**	0.133	0.134
Natural	0.035	0.034		**Natural**	0.133	0.134
1	0.385	0.386		**1**	0.129	0.13
2	0.31	0.311		**2**	0.965	0.966
3	0.043	0.042		**3**	0.115	0.115

4	0.304	0.299		**4**	0.129	0.136
5	0.177	0.172		**5**	0.075	0.074
6	0.255	0.252		**6**	0.772	0.771

Table 5.13: Results ANOVA and Welch tests.

5.1.4 Personal characteristics and route choice behaviour

To investigate whether there is a linear relationship between age or gender and deviation or the differences in length and share per spatial environment category, several simple linear regression analyses are executed within the total sample, the results are presented in Appendix VIII and IX. These analyses only proved a small significant relationship between age and the share in built or natural environment and between gender and absolute deviation, ΔBuilt and ΔCat. 3. The use of e-bikes and speed pedelecs could be a possible explanation for this as well. Although the influence of gender and age seems minimal, these variables will be included in the multiple regression models. On their own, gender or age do not explain any deviation or difference that strong. However, these variables will be included in the final model to see how they react to each other and to the weather-related variables.

5.1.5 Weather conditions and route choice behaviour

To examine whether there is a linear relationship between weather conditions independently and deviation or the differences in length and share per spatial environment category, several simple linear regression analyses are executed within the total sample, the results are presented in Appendix X. These analyses show that especially *dark/light* can significantly explain differences of several dependent variables on its own. Also, windspeed, temperature and precipitation seem to explain variance of some dependent variables significantly. However, the same applies to all significant models; with an R^2 lower than 0.01, the strength of these relations are very weak.

The next section will combine all variables in hierarchical multiple linear models to find out whether personal characteristics and weather conditions significantly explain absolute and relative deviation or differences between spatial environment categories on the chosen and shortest routes.

5.2 Regression analysis

This section will investigate whether people deviate further from the shortest route and if they deviate for a different spatial environment on the chosen route under certain weather conditions. The aim is to estimate hierarchical multiple regression models that can explain the deviation from the shortest route. Models are estimated for the four samples separately as well as combined.

The null and alternative hypotheses upon which the hierarchical regressions models are based are:

1. *H0 = There is no relationship between the amount of deviation from the shortest route and 'weather condition'.*

   HA = There is a significant relationship between the amount of deviation from the shortest route and 'weather condition'.

2. *H0 = There is no relationship between the difference in length of 'spatial environment category' and 'weather condition'.*

   HA = There is a significant relationship between the difference in length of 'spatial environment category' and 'weather condition'.

3. *H0 = There is no relationship between the difference in share of 'spatial environment category' and 'weather condition'.*

   HA = There is a significant relationship between the difference in share of 'spatial environment category' and 'weather condition'.

The null hypotheses will be rejected with 95% certainty when the model's probability of the corresponding F-statistic is $p \leq 0.05$. Table 5.14 shows the sequence in which the independent variables are included in the hierarchical multiple regression models.

Model	Independent variables	Scale of measurement
1	Age in years	Ratio
	Gender	Dichotomous (female = 1, male = 0)
2	Age in years	Ratio
	Gender	Dichotomous
	Head- or tailwind	Ratio
	Windspeed in km/h	Ratio
	Temperature in ˚Celsius	Interval
	Precipitation in mm	Ratio
	Dark/light	Dichotomous (dark = 1, light = 0)

Table 5.14: Independent variables of the hierarchical multiple regression models.

5.2.1 Correlation

The first step of building a regression model is to identify correlations of all possible predictors. The correlation of the predictors of absolute and relative deviation of the total sample is used as example. The correlation matrices for the predictors of absolute and relative deviation can be found in Appendix XI. These matrices indicate that *temperature*, *precipitation*, *dark/light* and *gender* show a significant correlation with *absolute deviation* and that only *temperature* show a correlation with *relative deviation*. In addition, these correlations are very low ($r < 0.1$). It is also important to look for high correlations amongst the predictors. It is interesting to mention that age only significantly correlates with gender, while gender also has a significant correlation with *absolute deviation*, *temperature* and *dark/light*. These correlations are also very low ($r < 0.1$). The threshold for excluding variables from a multiple regression is $r > 0.9$; none of the correlation values surpass this value. These correlation matrices are also created for the other eight dependent variables and for the four samples separately; less and lower correlations were found here.

5.2.2 Hierarchical multiple regression models

Despite the little and low correlations of the predictors, hierarchical multiple regression models are estimated for each dependent variable, which are presented in Appendix XII and XIV. As stated in section 3.7.1, the personal characteristic variables *age* and *gender* are entered first. Subsequently, the weather-related variables are entered into the model. No distinction is made between variables with or without correlation with the dependent variables. In total, 90 models are estimated; 18 per sample and 18 for the total sample.

The analysis resulted in 11 models that are proven to predict route choice significantly, of which only 9 include weather-related variables. Table 5.15 shows a summarized overview of the results per dependent variable. The green coloured cells represent significant models including the weather-related variables, the yellow coloured cells represent models that are only significant without the weather-related variables. The R^2 and the significance level of the models are presented as well.

Table 5.15: Summarized overview significant results simple linear regression analyses.

Absolute					
Dependent variable	Total	1	2	3	4
Total deviation	0,019**		0,040*	0,030*	
Δ Built	0,010*		0,055*		
Δ Natural					
Δ Category 1					
Δ Category 2					
Δ Category 3					
Δ Category 4					
Δ Category 5					
Δ Category 6					
Relative					
Dependent variable	Total	1	2	3	4
Total deviation					
Δ Built	0,008*		0,036*		0,128*
Δ Natural	0,008*		0,036*		0,128*
Δ Category 1					
Δ Category 2					
Δ Category 3					
Δ Category 4					
Δ Category 5					
Δ Category 6					

*** Significant at the 0,01 level*
** Significant at the 0,05 level*

Looking at table 5.15, it is interesting to see is that the three 'extremer' samples clearly contributed to the significance of the total models, since the first sample resulted in zero significant models. It is then clear to see that, within this study, personal characteristics and weather conditions do not significantly influence commuters' choice to cycle through specific spatial environment categories. There are however some significant models that can explain differences in route choice between built and natural environments. However, both *Built* and *Natural* are categories that have emerged from merging the six more specific spatial environment categories, which may explain why the relation with their predictors is stronger.

The following interesting models are interpreted and discussed:

- Model that predicts absolute deviation for the total sample;
- Model(s) that predict Δ Built rel. or Δ Natural rel. for the Wind sample;
- Model that predicts absolute deviation for the Rain sample;
- Model that predicts Δ Built abs. for the Rain sample.

Table 5.16: Model summary most significant model: absolute deviation total sample.

Dependent variable	Model	Independent variables	R^2	ΔR^2 - sign.	ANOVA
Absolute deviation	1	Age, gender	0,5%	0,5% *	0,017 *
	2	Age, gender Head/tailwind Windspeed Temperature Precipitation Dark/light	1,9%	1,4% **	0,000 **

*** Significant at the 0,01 level* *Sample: N = 1619*
** Significant at the 0,05 level*

Table 5.17: Linear models of predictors of absolute deviation of commuters in kilometres.

Absolute deviation	Predictors	B (confidence interval)	Standard error	Beta (β)	p-value
Model 1	Constant	1,063	0,190		0,000
	Age	-0,003	0,004	-0,023	0,351
	Gender	-0,159	0,057	-0,070	0,005
Model 2	Constant	0,792	0,215		0,000
	Age	-0,004	0,004	-0,029	0,246
	Gender	-0,176	0,057	-0,077	0,002
	Head/tailwind	0,000	0,001	-0,017	0,489
	Windspeed	0,006	0,004	0,045	0,077
	Temperature	0,013	0,004	0,092	0,001
	Precipitation	0,027	0,010	0,071	0,005
	Dark/light	-0,112	0,111	-0,027	0,313

Sample: N = 1619

Tables 5.16 and 5.17 present the most significant model in two hierarchical modelling steps: the model that explains the absolute deviation for the total sample. Since both models (1 and 2) are significant, they can be used separately to explain the absolute deviation of commuting cyclists. In column 'B' the coefficient of each predictor is shown, as well as the constant of the model. The rightest column gives the p-values for the B coefficients; a result of a t-test which tests whether the B coefficient significantly differs from 0. The lower the value, the stronger the variable affects the model. A non-significant p-value however, does not mean that there is no relationship between the predictor and the independent variable, absolute deviation in this case. Even though the correlation matrices (see Appendix XI) indicated a significant correlation of *temperature*, *precipitation, dark/light* and *gender* with absolute deviation, the model shows that *dark/light* is no significant predictor of absolute deviation. Age and wind direction, expressed as head- or tailwind in this model, are also no significant predictors of absolute deviation.

After adding the weather-related predictors, the negative influence of *age* and *gender* increased slightly, which also applies to almost all significant models (see Appendix XIV). In general, this model indicates that the addition of weather-conditions has a greater negative effect on older and / or female respondents, although this difference is minimal.

Although the B is not the best way to compare the predictors mutually, it is an easy way to have a glance at the effect size of each independent variable. Looking at the B values, it is clear to see that *gender* and *dark/light* have the greatest values. It is therefore worth mentioning that, after adding the weather-related variables, men tend to deviate 176 meters more than women. To compare: table 5.11 already showed that, without adding the weather-related variables, men deviate 153 further from the shortest route than women.

Another interesting result is that *temperature* has a positive significant influence on absolute deviation, meaning that commuting cyclists tend to deviate 13 meters when the temperature rises with one degree. For instance, whit a temperature of 20 ˚Celsius, commuters would cycle an average of 1/3 kilometre longer than with a temperature of -5 ˚Celsius. However, the significant model for absolute deviation of the Warm sample (see Appendix XIV), proves a negative significant effect of temperature on absolute deviation, meaning that commuting cyclists tend to cycle 72 meters shorter when the temperature rises with one degree. The routes of the Warm sample have been selected because the temperature at the moment the route was driven surpassed 25 ˚Celsius. It can be suggested that commuting cyclists tend to deviate from the shortest route when the temperature rises, up to a temperature of approximately 25 ˚Celsius.

In addition, windspeed (0,006) and precipitation (0,027) both have a positive effect on absolute deviation, meaning commuters tend to deviate from the shortest route when the windspeed or amount of precipitation increases. Whether they prefer built or natural environments under such weather conditions can partially be explained by the significant models for the Rain and Wind samples.

When interpreting a hierarchical multiple regression, in addition to the R^2, the ΔR^2 is also interesting. The ΔR^2 indicates whether or not adding the weather-related predictors led to an increased explanation of the dependent variable. When the number of predictors grows, the R^2 always grows. However, SPSS also presents the significance value for the ΔR^2. Despite the fact that with a ΔR^2 of 1,9% the R^2 almost quadrupled, the relation between the dependent variable and its predictors remains weak.

Table 5.18: Model summary of models with highest R^2.

Dependent variable	**Model**	**Independent variables**	**R^2**	**ΔR^2 - sign.**	**ANOVA**
Δ Built rel.	1	Age, gender	5,2%	5,2%	0,053
	2	Age, gender Head/tailwind Windspeed Temperature Precipitation Dark/light	12,8%	7,6%	0,038 *

Δ Natural rel.	1	Age, gender	5,2%	5,2%	0,053
	2	Age, gender Head/tailwind Windspeed Temperature Precipitation Dark/light	12,8%	7,6%	0,038 *

*** Significant at the 0,01 level* *Sample: N = 114*
** Significant at the 0,05 level*

Table 5.19: Linear models of models with highest R².

Δ Built rel.	**Predictors**	**B (confidence interval)**	**Standard error**	**Beta (β)**	**p-value**
Model 1	Constant	0,102	0,071		0,155
	Age	-0,002	0,001	-0,176	0,067
	Gender	0,024	0,021	0,108	0,261
Model 2	Constant	-0,025	0,142		0,863
	Age	-0,002	0,001	-0,132	0,180
	Gender	0,032	0,022	0,143	0,148
	Head/tailwind	0,000	0,000	0,033	0,726
	Windspeed	0,002	0,003	0,080	0,483
	Temperature	0,002	0,004	0,058	0,599
	Precipitation	-0,073	0,033	-0,245	0,029
	Dark/light	0,034	0,027	0,141	0,213
Δ Natural rel.	Predictors	B (confidence interval)	Standard error	Beta (β)	p-value
Model 1	Constant	-0,102	0,071		0,155
	Age	0,002	0,001	0,176	0,067
	Gender	-0,024	0,021	-0,108	0,261
Model 2	Constant	0,025	0,142		0,863
	Age	0,002	0,001	0,132	0,180
	Gender	-0,032	0,022	-0,143	0,148
	Head/tailwind	0,000	0,000	-0,033	0,726
	Windspeed	-0,002	0,003	-0,080	0,483
	Temperature	-0,002	0,004	-0,058	0,599
	Precipitation	0,073	0,033	0,245	0,029
	Dark/light	-0,034	0,027	-0,141	0,213

Sample: N = 114

Tables 5.18 and 5.19 show the two models with the highest R^2 (12,8%) and ΔR^2 (7,6%); the values of the other significant models can be found in Appendix XII. These two models can significantly explain the difference in share of the built or natural environment for the Wind sample. The models are almost the same; the only difference is that the coefficients and Beta's are the opposite of each other. This is explained by the fact that these categories complement each other: when the share of the natural environment increases, the share of the built-up environment decreases.

These models are an interesting example of the effect of using 'extremer' samples. The routes of the Wind sample have been selected because the average windspeed at the moment the route was driven surpassed 35 km/h. Therefore, it is not remarkable that the only significant predictor in these models is precipitation; after all, much precipitation and hard wind are harsh circumstance for most cyclists. Within this model there is a negative effect of precipitation (-7,3%) on the share of *Built* and a positive effect of precipitation (+7,3%) on the share of *Natural,* which implies that the respondents prefer to cycle outside the built-up area when there is strong wind and rain. As said before, these are fairly extreme samples.

Looking at the significant model that predicts absolute deviation for the Rain sample (see Appendix XIV), it shows a negative significant effect of windspeed (-0,047) on absolute deviation. Also, within the model that predicts the absolute difference in length of the built environment for the Rain sample, windspeed has a negative effect of -23 meters per mm precipitation increase. These models confirm the negative influence of precipitation and wind combined.

An interesting effect within the last model is the negative effect (-0,692) of dark/light on the absolute difference in length of the built environment, meaning that commuters tend to cycle 692 meters shorter through the built environment when it is dark.

There are three more significant models as a result of the total sample; one that can significantly explain the difference in travelled kilometres through built-up area and two models that can significantly explain the difference in share of the built or natural environment. However, these models are weak due to the fact they have an R^2 of 1% or lower and only one significant predictor (gender and age respectively). Moreover, the two models that explain the difference in share of the built or natural environment are only significant without including the weather-related variables. For a more detailed overview of these models, see Appendix XIII and XIV.

6. Discussion

This chapter will answer the research questions and discuss the limitations of this study and recommendations for future studies, to finish with the conclusion.

6.1 Answering research questions

This study was designed to investigate the effects of weather conditions on the route choice behaviour of commuting cyclists in which the focus is on GPS trajectories and the actual differences in route choice throughout the seasons of the year. Earlier studies suggested that personal characteristics, weather conditions and the spatial environment influence cyclists' route choice behaviour. Although travel distance is an important consideration when selection a route from A to B, no scientific evidence has been provided yet through GPS trajectories of actual differences in route choice and travel distance due to weather conditions. Before answering the main research question of this study, the answers to the sub questions are provided.

How are personal characteristics of commuting cyclists related to weather conditions and route choice behaviour?

On their own, age and gender barely explain route choice behaviour. In general, men tend to deviate further from the shortest route than women. After adding the weather-related predictors to the model for absolute deviation of the total sample, the negative influence of age and gender increased slightly, which also applies to almost all significant models (see Appendix XIV). This model indicates that the addition of weather-conditions has a greater negative effect on older and / or female respondents, although this difference is minimal. In the significant multiple regression models of the total sample, age is only significantly influencing the difference in share of the built or natural environment. Although the effect is (very) small, relatively older people tend to prefer cycling in a natural instead of a built environment. In the significant multiple regression models of the total sample, gender is only significantly influencing absolute deviation. This is in line with an earlier study of Heinen, Wee and Maat (2010) in which they suggested that women will mainly choose the shortest route, while men are more willing to take a detour.

Earlier studies (i.a. Aaheim & Hauge, 2005; Bergström & Magnusson, 2003; Böcker et al., 2015; Flynn et al., 2012; Heinen et al., 2011; Liu et al., 2017; Keay, 1992) mentioned rain and wind as primary negative influencers of route choice behaviour of female and / or older cyclists. Within this study, the effects of wind, precipitation and temperature on age and gender are noticeable, however, these effects remain minimal.

Since e-bikes and speed pedelecs are only recently emerging, earlier studies mainly focussed on regular bicycles. When electrical supported bicycles are used, the travelled distance is less influenced by the amount of physical effort of the cyclist, which could explain the minimal relations between personal characteristics and weather conditions or route choice behaviour.

To what extent do commuting routes deviate under different weather conditions?

Simple linear analyses showed that especially the presence of sunlight can significantly explain absolute deviation on its own. Also, windspeed, temperature and precipitation seem to explain absolute deviation significantly. These simple linear models were however very weak. While the correlation matrices indicated a significant correlation of temperature, precipitation and the presence of daylight with absolute deviation, the hierarchical multiple linear model for absolute deviation showed that the presence of daylight is no significant predictor. This model indicated positive effects of temperature and precipitation on absolute deviation, meaning higher temperatures and a higher amount of precipitation causes longer travel distances. Although the effect of these predictors remains low, these results fit an earlier study by Aaheim and Hague (2005) that presented reduction in travelled distance due to precipitation. Within this study, relative deviation cannot significantly be explained by weather conditions.

An interesting outcome of the regression analyses is the difference of the influence of temperature on absolute deviation between the total and Warm sample. For the total sample temperature has a positive significant influence on absolute deviation, meaning that commuting cyclists tend to deviate when the temperature rises. While the significant model for absolute deviation of the Warm sample, proves a negative significant effect of temperature on absolute deviation. Knowing these effects are weak, it can be suggested that commuting cyclists tend to deviate from the shortest route when the temperature rises, up to a temperature of approximately 25 ˚Celsius. These results complement earlier studies that have proven positive correlation between bicycle usage and temperature until the temperature reaches 25°C (Liu et al., 2015a; Sabir, 2011; Saneinejad et al., 2012). Additionally, the negative effect of temperature on absolute deviation of the Warm sample complement studies from warmer climates that already indicated that high temperatures, between 25 and 30 °C, are disadvantageous for cycling (Ahmed et al., 2010; Phung & Rose, 2008; Richardson, 2000).

Another interesting outcome of the regression analyses is the combination of much rain and hard wind. The significant model for absolute deviation of the Rain sample features a negative significant effect of windspeed. This result is not remarkable, since much rain and hard wind are harsh circumstance for most cyclists. The significant model that predicts absolute deviation for the Rain sample shows a negative significant effect of windspeed (-0,047) on absolute deviation. Also, within the model that predicts the absolute difference in length of the built environment for the Rain sample, windspeed has a negative effect of -23 meters per mm precipitation increase. These models confirm the negative influence of precipitation and wind combined. Previous studies (Heinen, Maat & Van Wee, 2011; Phung & Rose, 2008; Sabir, 2011; Thomas et al., 2012) already indicated negative effects of wind on cycling. The negative effect of the combination of wind and rain on absolute deviation is a worthy addition. On the other hand, wind direction, expressed as head- or tailwind in this study, is no significant predictor of absolute or relative deviation.

To what extent do commuters cycle through different spatial environments under different weather conditions?

Within this study, personal characteristics and weather conditions do not significantly influence commuters' choice to cycle through specific spatial environment categories. There are however some significant models that can explain differences in route choice between built and natural environments. Both Built and Natural are categories that have emerged from merging the six more specific spatial environment categories, which may explain why the relation with their predictors is stronger. Yet, the models that can significantly explain differences in share of the built or natural environment are very weak and have no significant weather-related predictors. Except for the models that explain the difference in share of the built or natural environment for the Wind sample, where precipitation is the only significant predictor. Within those models, there is a negative effect of 24,5% on the share of *Built* and a positive effect on the share of *Natural,* which implies that the respondents prefer to cycle outside the built-up area when there is strong wind and rain. As stated before, these are fairly extreme samples. This result is in contrast to earlier research of Phung and Rose (2008), in which the authors demonstrated that cycling in suburban and weather-exposed areas is more sensitive to precipitation than cycling in inner-city and sheltered areas.

To what extent do personal characteristics and weather conditions influence commuting cyclists' route choice behaviour?

Within this study, the effects of wind, precipitation and temperature on age and gender are noticeable, however, these effects remain minimal. The respondents used in this study were users of an e-bike or speed pedelec. When electrical supported bicycles are used, the travelled distance is less influenced by the amount of physical effort of the cyclist, which could explain the minimal relations between personal characteristics and weather conditions or route choice behaviour.

The hierarchical multiple linear model for absolute deviation showed that the presence of daylight is no significant predictor. However, this model indicated positive effects of temperature, precipitation and windspeed on absolute deviation, meaning higher temperatures and a higher amount of precipitation or stronger wind cause longer travel distances. Within this study, relative deviation cannot significantly be explained by weather conditions.

An interesting outcome of this study is, although the effects are weak, that commuting cyclists tend to deviate from the shortest route when the temperature rises, up to a temperature of approximately 25 ˚Celsius, then the deviation becomes smaller. Another interesting outcome of the regression analyses is the combination of much rain and hard wind. The significant model for absolute deviation of the Rain sample features a negative significant effect of windspeed. Also, within the model that predicts the absolute difference in length of the built environment for the Rain sample, windspeed has a negative effect. These models confirm the negative influence of precipitation and wind combined. This result is not remarkable, since much rain and hard wind are harsh circumstance for most cyclists. On the other hand, wind direction, expressed as head- or tailwind in this study, is no significant predictor of absolute or relative deviation within this study.

Within this study, personal characteristics and weather conditions do not significantly influence commuters' choice to cycle through specific spatial environment categories.

There are however some significant models that can explain differences in route choice between built and natural environments, however, these models are also fairly weak.

6.2 Limitations and recommendations

This study can be considered as a first attempt at data driven research to investigate the relation between cyclists' route choice behaviour and weather conditions using a GPS based revealed preference method. The lack of research within this subject was partly problematic when validating the results of this study. This method and the used data from the B-Riders program only made it possible to observe and investigate the final decision of the cyclist. Although the trip purpose is known, it would be interesting to use a mixed methods approach of stated- and revealed preference. This could make it possible to gain more insight into the cyclists' preferences or motivations behind their route choices. Despite this, the exploratory results of this study can be used as a first regarding route choice behaviour and GPS trajectories.

Using open- and freely available cycling data from the B-Riders program had many benefits, but also some limitations. The biggest advantage of using this data is that the data was already available and new data collection was therefore not necessary. Also the nature of the project made it possible to study cycling for everyday use, commuting in this case, which is often not possible with cycling data. However, using free and publicly available cycling data also had its limits due to privacy. First of all, the trips are anonymized by cutting off 200 meters from the start and end-location of a trip. This means that the derived origins and destinations are not the exact start and end locations of the routes. Besides, the sample was skewed concerning age, which makes it difficult to generalize the results of the data analysis. Also, the dataset is biased due to the fact that there was a motivational factor to participate in the B-Riders program. The research area is however representative for the whole of the Netherlands and the countries that have high mode shares of bicycle use and a comparable infrastructure and spatial environment. Different spatial environment and infrastructural characteristics (e.g. U.S.) make the results less interpretable.

Other limitations of this research are due to the availability and quality of spatial environmental data. Even though the Fietsersbond has the most accurate and rich cycling network that is created and kept up-to-date using volunteers (VGI), many roads still lacked usable data for spatial environment attributes. Especially detailed infrastructural or environmental information was lacking. Partly because the dataset is maintained by volunteers, which made many attributes fairly subjective. The used spatial environment dataset could have been operationalised better, to get more detailed and significant information on how weather conditions affect route choice within separate elements of the spatial environment.

The analysis showed however that a few models exist in which deviation or the difference in spatial environment categories could significantly be explained by personal characteristics and weather conditions. Previous research about cyclists' route choice behaviour corresponded with some of these outcomes. However, in general, the models did not perform that well which indicates that these relationships are not well captured in a hierarchical multiple regression model. This could be due to the used variables or the used sample, which consisted of electric supported bicycle users.

Even though some insight is acquired regarding the effect of personal characteristics and weather conditions on the amount of deviation from the shortest route, these models are not suitable for other purposes such as predicting route choice behaviour concerning the spatial environment. In order to gain more and better insight into the complex relationship between route choice and weather conditions, other (non-linear) methods should be explored and more detailed environmental factors should be taken into account. To model route choice behaviour, also other factors like infrastructure and safety should be considered as well.

More research on cyclists' route choice behaviour is needed in order to create accurate and flexible models that can provide more insight into the relationship between route choice behaviour and weather conditions by using multiple predictors. Further research could focus on differences between types of respondents, types of bicycle, seasons and infrastructure. Although the trip purpose and the GPS-trajectories are known, it would be interesting to use a mixed methods approach of stated- and revealed preference. This could make it possible to gain more insight into the cyclists' preferences or motivations behind their route choices.

6.3 Conclusion

This study is designed to investigate the effects of weather conditions on the route choice behaviour of commuting cyclists using GPS trajectories. The following factors were taken into account that were expected to influence route choice behaviour of cyclists: personal characteristics, weather conditions, the spatial environment and travel distance. Previous studies provided evidence of a relationship between the first three factors and cyclists' route choice behaviour, which directly influences travel distance. Still no evidence has been provided through GPS trajectories of actual differences in route choice and travel distance due to weather conditions.

The influence of wind, precipitation and temperature on age and gender are noticeable within this study, however, these effects remain minimal. The respondents used in this study were users of an e-bike or speed pedelec, which could explain the minimal relations between personal characteristics and weather conditions or route choice behaviour.

The hierarchical multiple linear model for absolute deviation showed that the presence of daylight is no significant predictor. However, this model indicated positive effects of temperature and precipitation on absolute deviation, meaning higher temperatures and a higher amount of precipitation causes longer travel distances. Within this study, relative deviation cannot significantly be explained by weather conditions.

An interesting outcome of this study is that commuting cyclists tend to deviate from the shortest route when the temperature rises, up to a temperature of approximately 25 ˚Celsius, then the deviation becomes smaller. Another interesting outcome of the regression analyses is the negative effect of the combination of much rain and hard wind. This result is not remarkable, since much rain and hard wind are harsh circumstance for most cyclists. On the other hand, wind direction, expressed as head- or tailwind in this study, is no significant predictor of absolute or relative deviation within this study.

Within this study, personal characteristics and weather conditions do not significantly influence commuters' choice to cycle through specific spatial environment categories. There are however some significant models that can explain differences in route choice between built and natural environments, however, these models are fairly weak.

7. References

Aaheim, H.A., & Hauge, K.E. (2005). Impacts of climate change on travel habits: A national assessment based on individual choices (CICERO Report 07). Oslo: Centre for International Climate and Environmental Research.

Ahmed, F., Rose, G. and Jacob, C. (2010) Impact of weather on commuter cyclist behaviour and implications for climate change adaptation. Proceedings of 2010, Department of Infrastructure and Transport (pp. 1–19), Canberra, Australia

Aldred, R. (2013). Cycling and society. Journal of Transport Geography, (30), 180-182.

Bergström, A., & Magnusson, R. (2003). Potential of transferring car trips to bicycle during winter. Transportation Research Part A: Policy and Practice, 37(8), 649-666.

Bierlaire, M., & Frejinger, E. (2008). Route choice modeling with network-free data. Transportation Research Part C: Emerging Technologies, 16(2), 187-198.

Böcker, L., Dijst, M., & Prillwitz, J. (2013). Impact of everyday weather on individual daily travel behaviours in perspective: a literature review. Transport reviews, 33(1), 71-91.

Böcker, L., Dijst, M., & Faber, J. (2014). Weather, transport mode choices and emotional travel experiences. In L. Böcker (Ed.), Climate, weather and daily mobility: Transport mode choices and travel experiences in the Randstad Holland (pp. 77–90). Dissertation. Utrecht University.

Böcker, L., Prillwitz, J., & Dijst, M. (2013). Climate change impacts on mode choices and travelled distances: a comparison of present with 2050 weather conditions for the Randstad Holland. Journal of Transport Geography, 28, 176-185.

Böcker, L., Dijst, M., Faber, J., & Helbich, M. (2015). En-route weather and place valuations for different transport mode users. Journal of Transport Geography, 47, 128–138.

Brandenburg, C., Matzarakis, A., & Arnberger, A. (2007). Weather and cycling—a first approach to the effects of weather conditions on cycling. Meteorological applications, 14(1), 61-67.

Braun, M. T., & Oswald, F. L. (2011). Exploratory regression analysis: A tool for selecting models and determining predictor importance. Behavior Research Methods, 43(2), 331–339.

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.

Buehler, R., & Pucher, J. (2010). Cycling to sustainability in Amsterdam. Sustain Fall/Winter, 21, 36-40.

Call, D.A. (2011). The effect of snow on traffic counts in western New York State. Weather Climate and Society, 3, 71–75.

Casello, J., Nour, A., Rewa, K. C., & Hill, J. (2011). An analysis of stated preference and GPS data for bicycle travel forecasting. In Proceedings of the Transportation Research Board 90th Annual Meeting, Washington, DC, USA (pp. 22-26).

CBS (Centraal Bureau voor Statistiek) (2017). Indeling van Nederland in 40 COROP-gebieden.

CBS (Centraal Bureau voor Statistiek) (2018). Retrieved from statline.cbs.nl

Cools, M., Moons, E., Creemers, L. and Wets, G. (2010). Changes in travel behavior in response to weather conditions: Do type of weather and trip purpose matter? Transportation Research Record: Journal of the Transportation Research Board, 2157: 22–28.

Creemers, L., Wets, G., & Cools, M. (2015). Meteorological variation in daily travel behaviour: Evidence from revealed preference data from The Netherlands. Theoretical and Applied Climatology, 120, 183–194.

De Palma, A. and Rochat, D. (1999). Understanding individual travel decisions: Results from a commuters survey in Geneva. Transportation, 26: 263–281.

Díaz, J., Garcı´a, R., Velázquez de Castro, F., Hernández, E., Lápez, C., & Otero, A. (2002). Effects of extremely hot days on people older than 65 years in Seville (Spain) from 1986 to 1997. International Journal of Biometeorology, 46, 145–149.

Duppen, van J. & Spierings, B. (2013). 'Retracing trajectories: the embodied experience of cycling, urban sensescapes and the commute between 'neighbourhood' and 'city' in Utrecht, NL', In: Journal of Transport Geography, 30, 234-243.

Edmond, C., Tang, W., & Handy, S. (2009) Explaining gender difference in bicycling behavior. Transportation Research Record: Journal of the Transportation Research Board, (2125), 16- 25.

Flynn, B. S., Dana, G. S., Sears, J., & Aultman-Hall, L. (2012). Weather factor impacts on commuting to work by bicycle. Preventive medicine, 54(2), 122-124.

Field, A. (2013). Discovering Statistics using IBM SPSS Statistics. Sage. Retrieved from https://books.google.nl/.

Gehl, J. (2011). Life between buildings: using public space. Island Press.

Goetzke, F. and Rave, T. (2011). Bicycle use in Germany: Explaining differences between municipalities through network effects. Urban Studies, 48(2): 427–437.

Hanbali, R. M., & Kuemmel, D. A. (1993). Traffic volume reductions due to winter storm conditions. Transportation Research Record, (1387).

Hassan, Y., & Barker, D. J. (1999). The impact of unseasonable or extreme weather on traffic activity within Lothian region, Scotland. Journal of Transport Geography, 7(3), 209-213.

Heinen, E., Wee, B. & Maat. K (2010) Commuting by Bicycle: An Overview of the Literature. Transport Reviews, 30:1, 59-96.

Heinen, E., Maat, K., & van Wee, B. (2013). The effect of work-related factors on the bicycle commute mode choice in the Netherlands. Transportation, 40(1), 23-43.

Helbich, M., Böcker, L., & Dijst, M. (2014). Geographic heterogeneity in cycling under various weather conditions: Evidence from Greater Rotterdam. Journal of Transport Geography, 38, 38-47.

Hofmann, M., & O'Mahony, M. (2005). The impact of adverse weather conditions on urban bus performance measures an analysis using ITS technology. IEEE Conference on Intelligent Transport Systems, Vienna.

Hood, J., Sall, E., & Charlton, B. (2013). A GPS-based bicycle route choice model for San Francisco, California. Transportation letters.

IPCC (2007). Report of the 19th session of the intergovernmental panel on climate change (IPCC). Geneva, 17-20.

Kennisinstituut voor Mobiliteitsbeleid (2017) Mobiliteitsbeeld 2017. Retrieved from https://www.kimnet.nl/mobiliteitsbeeld/mobiliteitsbeeld-2017#personenvervoer-article5

Khattak, A. J. and De Palma, A. (1997). The impact of adverse weather conditions on the propensity to change travel decisions: A survey of Brussels commuters. Transportation Research Part A, 31: 181–203.

Koetse, M. J., & Rietveld, P. (2009). The impact of climate change and weather on transport: An overview of empirical findings. Transportation Research Part D: Transport and Environment, 14(3), 205-221.

Liu, C., Susilo, Y. O., & Karlström, A. (2015a). The influence of weather characteristics variability on individual's travel mode choice in different seasons and regions in Sweden. Transport Policy, 41, 147-158.

Liu, C., Susilo, Y. O., & Karlström, A. (2015b). Investigating the impacts of weather variability on individual's daily activity–travel patterns: a comparison between commuters and non-commuters in Sweden. Transportation Research Part A: Policy and Practice, 82, 47-64.

Liu, C., Susilo, Y. O., & Karlström, A. (2016). Estimating changes in transport CO2 emissions due to changes in weather and climate in Sweden. Transportation Research Part D: Transport and Environment, 49, 172–187.

Liu, C., Susilo, Y. O., & Karlström, A. (2017). Weather variability and travel behaviour–what we know and what we do not know. Transport reviews, 37(6), 715-741.

Menghini, G., Carrasco, N., Schüssler, N., & Axhausen, K. W. (2010). Route choice of cyclists in Zurich. Transportation research part A: policy and practice, 44(9), 754-765.

Miranda-Moreno, L. and Nosal, T. (2011). Weather or not to cycle: Temporal trends and impact of weather on cycling in an urban environment. Transportation Research Record: Journal of the Transportation Research Board, 2247: 42–52.

Müller, S., Tscharaktschiew, S., & Haase, K. (2008). Travel-to-school mode choice modelling and patterns of school choice in urban areas. Journal of Transport Geography, 16(5), 342-357.

Parkin, J., Wardman, M. and Page, M. (2008). Estimation of the determinants of bicycle mode share for the journey to work using census data. Transportation, 35(1): 93–109.

Phung, J., & Rose, G. (2008). Temporal variations in usage of Melbourne's bike paths. In Proceedings of 30th Australasian Transport Research Forum, Melbourne.

Provincie Noord-Brabant (2016). Uitvoeringsprograme Fiets in de Vernelling 2016-2020.

Richardson, A. J. (2000). Seasonal and weather impacts on urban cycling trips. TUTI report 1-2000. the urban transport institute, Victoria. British Columbia, Canada.

Rietveld, P., & Daniel, V.E. (2004). Determinants of bicycle use: Do municipal policies matter? Transportation Research Part A, 38, 531–550.

Romanillos, G., Zaltz Austwick, M., Ettema, D., and De Kruijf, J. (2015). Big data and cycling. 36:1–20.

Sabir, M. (2011). Weather and travel behaviour.

Sabir, M., Van Ommeren, J., Koetse, M., & Rietveld, P. (2011). Adverse weather and commuting speed. Networks and Spatial Economics, 11(4), 701-712.

Saneinejad, S., Kennedy, C. and Roorda, M. (2012). Modelling the impact of weather on active transportation travel behavior. Transportation Research Part D, 17: 129–137.

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

Tang, L. and Thakuriah, P. V. (2012). Ridership effects of real-time bus information system: A case study in the City of Chicago. Transportation Research Part C: Emerging Technologies, 22: 146–161.

Thomas, T., Jaarsma, R., & Tutert, B. (2013). Exploring temporal fluctuations of daily cycling demand on Dutch cycle paths: the influence of weather on cycling. Transportation, 40(1), 1-22.

Tilahun, N. Y., Levinson, D. M., & Krizek, K. J. (2007). Trails, lanes, or traffic: Valuing bicycle facilities with an adaptive stated preference survey. Transportation Research Part A: Policy and Practice, 41(4), 287-301.

Van Leeuwen, E. S., Koetse, M. J., Koomen, E., & Rietveld, P. (2009). Spatial economic research on climate change and adaptation; Literature review for 'Knowledge for Climate'.

Vedel, S. E., Jacobsen, J. B., & Skov-Petersen, H. (2017). Bicyclists' preferences for route characteristics and crowding in Copenhagen–A choice experiment study of commuters. Transportation research part A: policy and practice, 100, 53-64.

Winters, M., Friesen, M. C., Koehoorn, M., & Teschke, K. (2007). Utilitarian bicycling: a multilevel analysis of climate and personal influences. American journal of preventive medicine, 32(1), 52-58.

Zhao, J., Wang, J., Xing, Z., Luan, X., & Jiang, Y. (2018). Weather and cycling: Mining big data to have an in-depth understanding of the association of weather variability with cycling on an off-road trail and an on-road bike lane. Transportation research part A: policy and practice, 111, 119-135.

Zhu, S., & Levinson, D. (2015). Do people use the shortest path? An empirical test of Wardrop's first principle. PloS one, 10(8).

Zimmermann, M., Mai, T., & Frejinger, E. (2017). Bike route choice modeling using GPS data without choice sets of paths. Transportation research part C: emerging technologies, 75, 183-196.