The influence of weather on travel behaviour - a multi-method analysis

R.M. Faber August, 2019

TU Delft

Kennisinstituut voor Mobiliteitsbeleid | KiM

THE INFLUENCE OF WEATHER ON TRAVEL BEHAVIOUR – A MULTI-METHOD ANALYSIS

A thesis submitted to the Delft University of Technology in partial fulfillment of the requirements for the degree of

Master of Science in Engineering and Policy Analysis

by

Roel Faber
Student number: 4307631

To be defended in public on August 26th, 2019

Supervisor: dr. Eric Molin, TU Delft
Committee: dr. ir. Maarten Kroesen, TU Delft
dr. ir. Hadi Asghari, TU Delft
dr. Olaf Jonkeren, Kennisinstituut voor Mobiliteitsbeleid
ir. Mathijs de Haas, Kennisinstituut voor Mobiliteitsbeleid

ACKNOWLEDGEMENTS

In many respects this document marks the conclusion of my educational career. Six years of primary education, six years of high school and six years of higher education at Delft University of Technology have taught me everything I needed to know to be able to write the document that is in your hands or on your screen right now. This is thus the official end of the beginning, marking the start of 'real life'. What follows is a short and incomplete list of the people that have been involved and to whom I owe my gratitude.

First and foremost I'd like to thank my committee members, who have been involved extensively during the research and writing processes that in the end culminated in this thesis. Maarten, I distinctly remember our conversation in May of last year when I first approached you about possible thesis research projects. After a short quip that I was very early with this request indeed you immediately pointed me in the direction of the Kennisinstituut voor Mobiliteitsbeleid (KiM). This was the first piece of great advice in what has turned out to be a very long list of very helpful tips, pointers, discussions, and prompts that have been indispensable during my thesis. Hadi, I left every meeting we've had with a sense of both complete doubt and exhilaration, no matter how short our talks or how much of my material you had read beforehand. Somehow you were always able to make me aware of the problems I was facing, nearly causing something like a mental breakdown on multiple occasions. Luckily you were also somehow able to make me feel that I could overcome these problems at the end of our conversation. Thank you! Finally Eric, I think I understood you correctly that I'm the first student on whose committee you only served as the chair. This meant only a limited amount of your time was available, but I'm very grateful for every single time you have looked at my materials as you always found both the key weaknesses in my argumentation and provided me with the information necessary to address them.

Of course, my committee also includes two KiMmers, who were my daily supervisors throughout the project: Olaf and Mathijs. Both of you have been absolutely vital to the whole research process. You were always available for my questions and helped me deal with many obstacles I encountered in the last six months. Thanks so much to both of you for your time, your advice, our discussions and our conversations. I'd like to thank Danique Ton for her essential help in developing my choice sets and Frits Bijleveld for the pleasant conversations and his advice at the very start and the very end of these six months. Writing a thesis however is a culmination of a process that is far longer than the six months you spent writing the final document. During my time at the university I have had the opportunity to meet and learn from many great people. In the first place I'd like to thank Mark de Reuver, for the role you've played in my development as a student and a researcher. I'm not sure if you remember, but I think the first time I was in your office you had called me in to discuss whether or not I had contributed enough to a group project to deserve the same grade as the other members. The last time I was in your office we discussed our conference paper and how to proceed to perhaps turn it into a journal paper. I think that progression says it all. Other researchers and staff at the university I'd like to thank are Ricardo Matheus, Marijn Janssen, Stella Boess, Amir Ebrahimi Fard and Haye Hazenberg. Through our collaborations you've allowed me to learn many vital skills and to develop the confidence in myself I needed to start and finish this thesis.

Finally for my friends and family: you know who you are. I couldn't have done this without your unwavering support.

Let me conclude by thanking all KiMmers for providing an environment in which I always felt both welcome and productive and for the many times one of you was available to answer my questions and provide me with helpful advice. I'm very much looking forward to joining you this year.

But first a long overdue holiday!

Roel Faber
Delft, August 2019

EXECUTIVE SUMMARY

Societal and political attention to the effects of climate change and possible mitigation and adaptation policies has increased sharply in the last decades, resulting partly from increasing awareness about the role of humanity and partly from the ever more noticeable changes in our world caused by climate change. This societal interest has highlighted a lack of knowledge about the effects that a changing climate will have on many aspects of our lifes, one of which is the transport section. To understand the impact climate change will have on our transport system we need to know how travel behaviours are affected by weather circumstances, which is the main topic of our research.

We focus on four aspects of the relationship between weather and travel behaviour: (1) how weather is taken into account in the decision-making process (2) if the influence of singular weather variables (such as temperature) depend on the value of other parameters (3) if the influence of weather is different for urban and rural areas and (4) whether there are groups of people whose response to travel behaviour are distinctly different from one-another. This knowledge can be used for climate change adaptation measures, such as ensuring that our supply of travel infrastructure will be able to cope with changes in travel demand resulting from a changed climate, and mitigation measures, such as increasing the number of people that use more sustainable travel options like the bicycle.

For our analyses we use travel data provided by the KiM Netherlands Institute for Transport Policy Analysis, which is the result from a travel diary survey held in autumn. We use weather data as measured by weather stations, provided by the Royal Netherlands Meteorological Institute (KNMI). This data is used to estimate the influence of weather on travel demand and mode choice in the Netherlands, using regression and choice models respectively. Within our analyses we try to find factors that moderate the relationship between weather and travel behaviour, such as urban density and socio-demographics.

With respect to the four aspects identified above, we report the following findings:

(1) that people use a general perception of the weather during the whole day for their mode choice travel decisions, which contrasts with the most common practice of using the weather at the trips' departure time.

(2) by accounting for the fact that meteorological variables always co-occur in our models we are able to more accurately capture its effect on travel behaviour. The difference with the current practice of estimating separate effects for each weather variable is particularly stark for days at the extreme end of the observed range of weather variables.

(3) The influence of weather on travel behaviour differs more qualitatively between rural and urban areas: the total effect size of the weather similar, but they are brought upon by different weather variables. The difference is also very specific to travel modes. For bicyclists the effects of wind speed seem to be more sizeable in urban environments, whilst temperature, rain, and sunshine have smaller effects in urban environments.

(4) We find multiple groups of travellers whose responses to weather variations are different from one another. These differences seem to be caused by the set of travel modes that are used during average weather conditions. People that only use the car during average conditions are not very affected, with only enjoyable weather conditions prompting increased bicycle use. If the car and the bicycle are used often people swap between the modes, although use of the bicycle during inclement conditions is relatively much higher than for the other two groups. The last group has a more multi-modal travel pattern, which results in the largest variations caused by weather. Inclement conditions favour both public transport and the car, with car use increasing quite sharply during wet weather with high wind speeds.

Additionally we find that weather variations account for differences in travel behaviour across both the spatial and temporal dimensions. A particularly surprising finding is that the smaller number of bike trips in the western provinces of the Netherlands can be fully explained by the fact that there are

higher average wind speeds and lower temperatures in this part of the country.

Our results have several implications for the research community and policy makers. We advice researchers to account for the fact that the weather is perceived as a whole and thus that the effect of one single variable (such as temperature) will depend on the values of other variables. We also found interesting subgroups with different reactions with regards to weather. We advice researchers to more closely investigate the effects of weather for the separate subgroups. Finally we find sizeable differences in the effect of weather between different regions, even within the relatively small country of the Netherlands. Researchers studying a relatively large study area would do well to estimate separate effects for regions within their study area, for example based on population density and geographical location.

For policy makers our findings imply that there is a sizeable effect of weather that could be used to improve the forecasts of future travel demand, both in the short- and long terms. Whilst policy makers obviously can't control the weather, we have found that changing travel patterns or attitudes to travel modes will have repercussions for the effect weather has on travel behaviour. We think that policies aimed at allowing commuters to gain experience with using the bicycle for their daily commute during summer, coupled with temporary financial incentives when weather conditions become less favourable, could be one way of achieving more cyclists during inclement conditions. Policy makers could even target younger professionals specifically, as they are much more likely to have already developed such habits during their education.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

With the effects of climate change becoming ever more noticeable in daily life WMO [2019]; IPCC [2019] the awareness of climate change has become much greater in the last few years. This has prompted national governments to address the danger presented by climate change in two ways [Wiebes, 2019; Rijksoverheid, 2018b]: by trying to mitigate its effects by reducing greenhouse gas emissions and by preparing the adaptation of society to a changed climate. These ambitions also have repercussions for the domain of transport, as we strive for sustainable mobility in the near- and distant futures [Rijksoverheid, 2018a].

Despite these ambitions the car is by far the most used travel mode in the Netherlands, as it is used for nearly 60% of all travelled kilometers [Kennisinstituut voor Mobiliteitsbeleid, 2017]. This usage has even grown during the last decade, with an all-time high of 147.6 billion kilometres travelled with the car in 2017 [Centraal Bureau voor de Statistiek, 2018]. This has caused considerable increases of congestion by up to 20% in the last few years [ANWB, 2018], increases that are only projected to continue into 2019 [Kennisinstituut voor Mobiliteitsbeleid, 2018]. This congestion is the cause of considerable societal economic losses, ranging from 2.8 to 3.7 billion Euro in 2016. [Kennisinstituut voor Mobiliteitsbeleid, 2017] and even further increases in greenhouse gas emissions [Ligterink and Smokers, 2016].

Decreasing the dependence on- and use of the car, for example by increasing the use and availability of other travel modes, is necessary to achieve more sustainable transport and would cause many other positive effects, such as reduced congestion and increased health due to a reducton of particular matter emissions [de Nazelle et al., 2010]. Further health benefits might arise from corresponding increases in the use of active modes such as cycling [Grabow et al., 2011].

For these reasons the Dutch government aims to realize a modal shift away from the car, especially for commute traffic. The concrete ambition is to move 200,000 people out of the car and onto the bicycle by 2022, which is to be realized by an investment plan of € 100 million [van Veldhoven-van der Meer, 2018].

These goals can be accomplished by increasing the attractiveness of other travel modes and/or by increasing the costs of car usage [Gärling and Schuitema, 2007; Meyer, 1999]. One factor that needs to be taken into consideration when designing such demand policies is the influence of weather. Research shows that weather (forecasts) systematically alter both the amount of travel demand as a whole and the preferred travel mode [Böcker et al., 2013a; Liu et al., 2017]. The more exposed active modes (bicycling and walking) are especially affected by weather circumstances : this exposure to inclement weather is one of the main barriers preventing further bicycle adoption [Zhao et al., 2018; Rérat, 2018] and use [Heinen et al., 2010]. Besides these short-term influences of weather a local change in climate is likely to cause travel patterns to shift in the absence of any policies. To enable adaptation policies we need to identify the strength and direction of such a shift, enabling a design of the infrastructural network that is able to handle increasing changes in mobility patterns as a result of climate change.

The influence of the weather on travel behaviour is thus the topic of our study. We are interested in how variations in weather circumstances, higher or lower temperatures, dry or wet weather, and gentle breezes or windstorms, cause changes in travel behaviour of people. Whilst intuition provides some expectations, quantitative estimates are needed to aid mitigation and adaptation policies. We will explain the relevance of this topic for policy makers in Section 1.1, whilst providing the scientific relevance in Section 1.2. Based on this information we've set up a research design, which is described in Section 1.3. The approach used to carry out the research can be found in Section 1.4 and finally the structure of this thesis can be found in Section 1.5.

1.1 RELEVANCE FOR POLICY

The weather of course is a phenomenon that policy makers can not change, prompting the question why research into its effects on travel behaviour should be carried out. In this section we give an answer to this question, describing how such research can be used to attain goals of policy makers and/or other actors in the field of transport to ensure that it is relevant for society and not just as a scientific endeavour.

There are two principally different avenues through which knowledge about the effect of the weather on travel behaviour could be useful to accomplish either policy goals, such as accommodating increased mobility [Rutte et al., 2017; van Nieuwenhuizen Wijbenga and van Veldhoven-van der Meer, 2019] and realizing modal shifts to active modes [van Veldhoven-van der Meer, 2018], or goals of other actors within the transport field, such as providing congestion forecasts [ANWB, 2019b] and adapting supply to expected travel demand.

The first avenue directly uses information about the effect of weather on travel behaviour, for example in travel forecasting or determining the effect of infrastructural developments. The second uses information on how to change the strength or direction of the effect of weather on travel behaviour, for example by making changes in the built environment to shield travellers from inclement weather conditions. Below we will explain these avenues in more detail.

There are two distinct possible avenues for policy implications that result from knowledge about the effect of weather on travel behaviour:

1. Using information about the effect of weather on travel behaviour itself, for example to more accurately predict future travel behaviour and evaluate or explain past travel behaviour.
2. Knowledge about how to change the effects of the weather on travel behaviour. For example, to increase the number of cyclists during high wind speeds, one could place wind-barriers next to bike paths.

The first avenue is the ability to use insights into the effect of weather on travel behaviour directly, which can be useful when forecasting future travel demand (both in the short- and long-term) and whilst evaluating the effect of other variables, such as new infrastructural projects.

More accurate forecasts in the short-term might be used to adaptively increase or decrease travel supply depending on weather conditions, especially by local public transport authorities. For example, if the research finds that there is a modal shift from bicycles to local transit options during inclement weather conditions, it might be possible to increase the frequency of buses when these conditions are forecasted and vice-versa. Slightly different is the use of information with respect to the use of the car to forecast future congestion levels. These forecasts are provided by both the Royal Dutch Touring Club (ANWB) and Rijkswaterstaat, part of the Ministry of Infrastructure and Water Management. Currently forecasts are used to give some indication of expected congestion and to advice motorists to adapt their behaviour according to weather events, but a better understanding of how and where weather impacts travel behaviour of motorists could lead to improved congestion forecasts.

Information about the weather can also be used to more accurately evaluate and explain variations in past travel behaviour. Weather is structurally different between parts of the Netherlands: the northern- and western parts are closer to the sea than the eastern- and southern parts of the country, resulting in colder temperatures, higher wind speeds, and more rain in the north and west compared to the South and East. These structural changes in weather might partly explain some of the differences in travel behaviour between these regions. This enables the identification of the true effect of other factors on these differences, which can be used to give more accurate information about the effect that might be achieved through policies.

The weather can also be used to explain variations in travel behaviour across time. Policies are often based on future extrapolations of current mobility trends. To ensure these extrapolations are accurate, we need to separate the underlying trend from noise in the observations. If a year contains many days of snow for example the observed number of bicycle trips during the year will probably decline quite considerably. This decline is not structural however: by accounting for the negative influence of the number of snow days the underlying trend can be estimated. Dutch Statistics and the KiM produce a long-term mobility trend [Boonstra et al., 2019], based on annual travel surveys. Currently a few annual weather statistics are used when estimating this mobility trend, but there is some potential for improvement for which we need to know more about how the weather affects travel behaviour. If we are able to more accurately assess the influence of weather anamolies we are able to give a more accurate trend, which leads to better predictions about future travel behaviour.

In the long-term we can use our knowledge about the effect of weather on travel behaviour to see how projected climate change scenarios will impact travel behaviour. This information can be used to create adaptive plans to ensure that travel supply will match travel-demand when the climate in the Netherlands has changed. If research shows that the effects of climate change might result in an increase of bicycle travel demand, policies can be designed that ensure the addition of bicycle capacity during the next decades. Furthermore these expected effects of climate change can be used to more accurately assess the impact of infrastructure projects, which is particularly important considering the often long-term nature of most of these projects.

Within the first avenue three possible areas where more knowledge about the effect of weather on travel behaviour can be useful are identified.

The first is the ability to predict short-term travel demand based on weather forecasts, information that can be used to adaptively change the supply of travel options (such as public transport frequency) based on weather or to provide more accurate congestion forecasts.

The second is the ability to explain variations in recorded travel behaviour across space (for example between provinces in the Netherlands) and time (for example across years), which can be used to improve long-term trend estimations.

The third is the ability to estimate the effects of climate change on travel behaviour, which is useful when planning infrastructural projects that will change the shape of our infrastructure for the decades to come.

The second avenue relates to the ability of policy makers to change the strength of the effects of weather. We might for example consider to provide a cover for cycling paths in the Netherlands, thus reducing the exposure of cyclists to the weather. Doing so would change the effect of weather on the number of cyclists. In this context of bicycling, Eva Heinen already made the observation that "future research should focus not only on climate and weather conditions, which cannot be changed, but also on measures and facilities that might lessen the weather's negative effects" [Heinen et al., 2010, p. 69]. We will thus include these measures and facilities in our analyses, providing information about which (type of) measures can be taken to change the effect of the weather. Below we'll describe two kinds of measures and facilities that can be used by policy makers to influence the effect of weather on travel behaviour.

The first kind concerns the people whose travel behaviour is affected by the weather and more specifically whether certain groups of people react differently to the same set of weather circumstances than other people. As an example, Motoaki and Daziano [2015] found that more experienced cyclists are more likely to keep using the bike during inclement weather conditions than less experienced cyclists. If this is the case for the Netherlands as well, policy makers could try to improve cycling experience or habits during nice weather circumstances knowing that this experience will also mean that these people are more likely to keep cycling during inclement conditions. Other research has found that people with positive attitudes towards the bicycle are more likely to use the bicycle during inclement conditions as well [Nordbakke and Olsen, 2019]. Policy makers could try to change the attitudes towards the bicycle, for example by getting people to try the bicycle at least temporarily or by ensuring that bicycle facilities at work are improved. Other options are to provide temporary monetary rewards to people for bicycling during peak-hours, as is experimented with by the Dutch government during so-called bicycle stimulation projects [Van Baaren and Dijksterhuis, 2015; Goudappel Coffeng, 2017]. During these projects people sign up to use an application which measures their bicycle transportation habits and rewards them for biking longer distances to and from work. This application can be designed to give additional prompts or reasons to cycle during inclement weather conditions [Dijksterhuis and Van Baaren, 2017], which could be especially useful during the transition from dry and warm summer months to the colder and wetter autumn months. Such policies are designed to allow people to get familiar with using the bicycle for their commute, with the idea that people are able to build positive experiences with the bicycle and thus will keep using the bicycle after the project has ended.

The second kind concerns the built environment within which travel behaviour decisions are made. This built environment might change the relationship in two different ways. The first of these is the ability of the built environment to act as a shelter against weather circumstances, as was exemplified before by the extreme idea of installing a roof on bicycle paths. Whilst of course this is an extreme and impractical example, concrete policies have been made to protect cyclists at exposed parts of their route, often in the form of wind-screens at bridges [Rijkswaterstaat, 1995; StudioSK, 2017]. This use of the built environment as cover extends beyond the use of cycling however, for example by providing more shelter at public transport access/egress points. Furthermore it's important to realize that cover can also be found in ordinary buildings: the buildings within a city trap heat and provide a cover for wind speeds creating a micro-climate that might be more suited to the active transport modes of cycling and walking. The second way in which the built environment might change the effect of weather is by providing a more closely connected network of shops, houses, and offices. The resulting reduced distances or changed travel patterns might mean that the effect of weather is less pronounced in denser areas than in less dense areas. Finally the infrastructure might be changed to decrease the travel times of cyclists, for example by reducing wait times at traffic lights during rain. This is now being implemented in multiple cities in the Netherlands [Fietsersbond, 2019], but its effects are not very clear.

> Policy makers might be able to change the strength of the effect of weather through two different means.
> The first uses information regarding how different groups in the population respond to weather. If more experienced cyclists show less sensitivity to weather policies aimed at increasing cycling experience might reduce the strength of the effect of weather on bicycle use.
> The second is based on changes that can be made in the built environment. Providing travellers with cover from weather might reduce the strength of the effect of the weather. Another possibility is that a more closely connected built environment causes travellers to be less sensitive to weather, for example by reducing the distance that they need to travel.

1.2 KNOWLEDGE GAPS

In this section we will introduce the main knowledge gaps emerging from the literature. To do so we will discuss review papers and summarise the state of the field. This section thus does not contain a comprehensive literature review, although such a review is made as part of this thesis and can be found in Chapter 2.

Three literature reviews have been published on the topic of weather and travel behaviour: Koetse and Rietveld [2009], Böcker et al. [2013a], and Liu et al. [2017]. These reviews enable the identification of remaining knowledge gaps and contradictory findings in the global literature. Since the influence of weather on travel behaviour is region-specific (due to local climates and customs) it is useful to also take into account which knowledge gaps exist specifically within the Dutch context.

Previously two PhD tracts have been carried out with the goal of estimating the influence of weather and travel behaviour, resulting in doctoral dissertations and published papers by Sabir [2011] and Boöcker [2015]. Other researchers have mostly published single papers on the topic, using a multitude of approaches. Bocker has followed up on the recommendations in his own review article by looking into perceptions of weather [Böcker et al., 2015], which is very rich but limited to the general-Rotterdam area. Sabir's research is spatially broader (using data from the entirety of the Netherlands), but has been carried out a decade ago and doesn't address the more recent knowledge gaps. The issue of weather expectations and forecasts is addressed in a study by Cools and Creemers [2013], which is situated in Flanders, the Dutch-speaking part of Belgium. Some work on the moderating impact of individual characteristics is done by Heinen [2011], but their research is limited to the bicycle and only compares occasional and frequent cyclists. They also address perceptions and attitudes with regards to weather and its influence, but again the focus is solely on the bicycle.

Within the context of the Netherlands, we have identified four knowledge gaps given here and described in more detail below.

1. There is not enough evidence on how weather is incorporated into the decision making process
2. Studies have not looked at the co-occurring effects of weather
3. The moderating effect of spatial variation on the response of travel behaviour to weather is not taken into account.
4. We lack information regarding the moderating effects of individual characteristics and resulting heterogeneity with regards to the response of travel behaviour to weather

The first knowledge gap concerns the incorporation of weather into the decision-making process, more specifically looking at *which* weather influences decisions. The question thus is how the very complex phenomenon of weather with its significant variation over both the spatial and temporal dimensions is perceived and used by the traveller to inform his or her travel behaviour decisions. We need to know the answer to this question to decide which values for weather parameters (such as temperatures, rain, wind speeds, etc.) to assign to the trips or days for which travel behaviour decisions have been made and the answer is thus relevant for all research into the effects of weather on travel behaviour. The conclusions can also be used in our own further analyses, providing evidence for two of the more contentious steps of the modelling process (conceptualization and operationalization). This issue is described in detail in a section of the literature review by [Liu et al., 2017] on the matching of weather to travel survey data. (p. 8). The most common method is to use the weather at the time and location of trip origin, but the decision to use this method seems to be based mostly on intuition. Evidence from revealed preference studies is needed [Liu et al., 2017, p. 8], as "there is still no consensus on how weather is incorporated into travel decision processes" [Liu et al., 2017, p. 22].

The second knowledge gap relates to the fact that whilst most studies estimate a singular influence of a weather variable, say wind, the actual effect of an in- or decrease of this variable on weather depends on the values of other weather variables. Current practices thus do not reflect the decision-making process of travellers, who are swayed by an overall perception of the weather based on the combined values for all co-occurring weather parameters. For example we expect that increases in wind speeds will have a stronger negative effect on the use of active modes during a cold rain shower than during warm and dry conditions. The point is made most clearly in the literature review by [Böcker et al., 2013a], who state that "future analyses should not single out weather parameters, but also incorporate combined weather effects" (p. 87). By accounting for the combined and holistic nature of weather parameters we can thus more closely resemble the actual decision-making process, increasing the validity of our results.

Whereas the first and second knowledge gaps focus on the estimation of the effect of weather on travel behaviour, the third and fourth knowledge gaps relate to knowledge about factors that might moderate the relationship between weather and travel behaviour.

The third knowledge gap relates to the variation in travel behaviour responses that are expected between different regions, and particularly between urban/very dense areas on the one hand and rural/less dense areas on the other. Comparing the effect of weather on travel behaviour between urban and rural areas enables better forecasting and allows us to gain further understanding on the way in which weather impacts travel behaviour. The knowledge gap is identified by both Böcker et al. [2013a], who state that "Comparisons could also be made between the impacts of weather on travel behaviour in rural, suburban and urban areas (...), which may hold important implications for policies"(p.86) and Liu et al. [2017], who identify improved understanding of this spatial variation as one of their research directions (p. 22). If we find differences between regions we might be able to find possible reasons for these differences. Some of these differences might be under the control of policy makers, giving them a way to control the effect of the weather on travel behaviour.

The fourth knowledge gap is similar in some respects to the third, as we are again looking at moderating factors. Rather than looking at spatial variation however, this knowledge gaps considers the variation in responses to weather circumstances as a result of varying individual characteristics. We expect that people will respond differently to the same weather circumstances based on numerous characteristics, including their lifestyles, genders, health conditions, etc. [Böcker et al., 2013a]. This heterogeneity within the population due to numerous factors has not been researched often, and the final recommendation by [Böcker et al., 2013a] describes this knowledge gap perfectly: "More attention should be paid to the mediating roles of personal backgrounds and socio-demographic characteristics" (p. 87). Again more information on how personal characteristics influence the relation between weather and travel behaviour gives possible options for policy makers to control the reaction of travellers to weather circumstances, but it might also lead to a greater understanding of how the reactions might change in the future as a result of bigger demographic or attitudinal changes in the population.

1.3 RESEARCH DESIGN

Using the information from both Section 1.1 and Section 1.2 the research will be designed so that it both addresses the four remaining knowledge gaps ensuring that it is scientifically relevant and is able to provide relevant information to policy makers ensuring that it is societally relevant.

It's important to point out that whilst scientific and societal relevance are separate objectives, they can be achieved through the same means. The first avenue of policy implications is using our best understanding of the relation between weather and travel behaviour to forecast future behaviour or evaluate and explain past travel behaviour. It stands to reason that any new knowledge could be used to improve these forecasts so that addressing any scientific knowledge gaps automatically enhances our ability to forecast future behaviour. Of course, some knowledge is more valuable for predictive and explanatory purposes and in particular addressing the remaining knowledge gaps pertaining to how weather is incorporated into the travel decision making process (knowledge gap 1) and the combination or interaction of weather variables (gap 2) will aid forecasting abilities by ensuring that our analyses are in line with the decision-making processes of travellers.

The second policy avenue relates to how policy makers might change the effect of the weather on travel behaviour. Information about how the spatial and personal characteristics moderate the relationship between weather and travel behaviour (knowledge gaps 3 and 4 respectively) is absolutely necessary to give policy recommendations related to this avenue.

In summary then, addressing the knowledge gaps should (results permitting) also lead to a research that has practical implications for policy makers in the mobility field. Whilst the general objective of this thesis is thus - as any thesis - two-fold, namely being both scientifically and societally relevant, we can specify this objective in the context of our study with the following sentence:

Research Objective: *Determine the moderating effects of individual, trip, and spatial characteristics on the effect of the weather on travel behaviour in the Netherlands.*

To achieve this research objective we need to provide the answer to a couple of questions. First and foremost we need to be able to estimate the effect of weather on travel behaviour, which comes with two main conceptual difficulties.

> To determine moderating effects we first need to be able to estimate the effect of weather on travel behaviour in general.

The first is to gain knowledge about the different effects that might occur. In other words we need to establish how the causal relationship from weather to travel behaviour is constituted: which factors are passed along the way? Are there multiple causal paths? To make this more concrete, we might envision that rain both dissuades one from cycling because you are not protected and thus will get wet, whilst it might also dissuade from using the car because roads get slippery and less safe. It's clear that these are two very different causal paths between weather and behaviour. Before we perform quantitative analysis we thus need to get a clear conceptual map of the relations between weather and travel behaviour that we are investigating.

With the first sub-question we try to hurdle this conceptual difficulty above by finding the causal paths that exist between weather and travel behaviour. Answers to this question can be found in the findings and conceptual discussions of existing literature on the relationship between weather and travel behaviour. In doing so we be able to conduct analyses that build upon the knowledge gained by previous researchers, properly embedding our research within the broader literature and ensuring that our work can meaningfully contribute to the current state of the field, furthering our understanding of the relationship between weather and travel behaviour. This leads to the following question:

1. *In which ways can weather affect travel behaviour and what is known about the existence and strength of these effects?*

The second conceptual difficulty relates to how objective weather affects the mental state and the decision making process of individuals across spatial and time dimensions. Weather is a very complex phenomenon, which changes - sometimes very dramatically - across these two dimensions: it might rain at my current location, but be dry 5 km away and the rain might go away entirely in 20 minutes. The question thus revolves around how humans use this very rich information to inform their decisions revolving around travel behaviour. Decisions could be based around current weather and/or future weather and/or daily weather and/or past weather at the location of trip origin, my current location, the location of my destination, etc. For our estimations we need to chose one approach, preferably the one that is closest to the actual way in which weather influences our mental states and around which we base our decisions. We have formulated the following question:

2. *Which conceptual connection between weather and travel behaviour is best able to capture the influence of weather on travel behaviour?*

> Before we can determine the effect of weather on travel behaviour we need to overcome two conceptual difficulties. The first of these difficulties is how we conceptualise the causal relation from weather to travel behaviour. The second difficulty is how weather is incorporated into the decision-making process of a single decision.

When these two conceptual difficulties are dealt with we can turn our attention to the estimation of the moderating effects. Travel behaviour is also a very broad concept, consisting of many different types of behaviours such as decisions whether or not to travel, the speed and route which are used for travelling, and much more. Here we choose to estimate the effect of weather on two mainstays of travel behaviour theory: trip generation and mode choice. The first concept relates to the aforementioned decision whether or not to travel at all. Here we are thus first trying to estimate the effect of weather on this decision, and then second to estimate the moderating effects of individual and spatial characteristics. The second travel behaviour concept we study is the mode choice, which is the decision of how to travel: with the car, public transport,bicycle, or another mode. Again we first need to estimate the effect of the weather and then find possible moderators.

Whilst we are studying the effect of weather on two dimensions of travel behaviour (trip generation and mode choice), we have three remaining knowledge gaps. Addressing all three knowledge gaps for both dimensions would lead to very complex models with a multitude of parameters, which makes both the estimation and interpretation very complex if not impossible. We have thus decided to split the three remaining knowledge gaps over the two dimensions of travel behaviour.

For travel demand we have decided to focus on addressing two knowledge gaps: the combination of weather parameters (knowledge gap 2) and the spatial variation (knowledge gap 3). This leads to the following question:

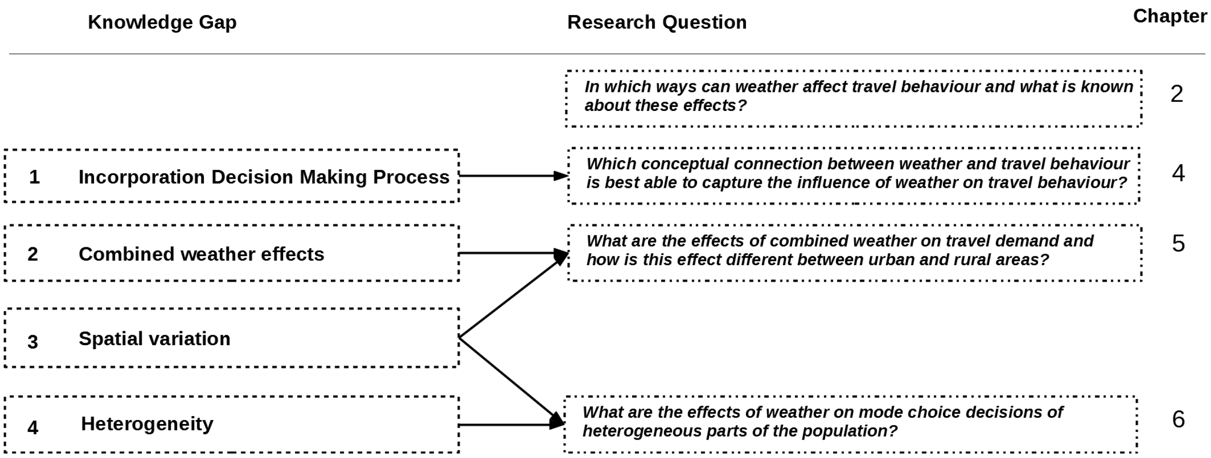
3. *What are the effects of combined weather on travel demand and how is this effect different between urban and rural areas?*

Only one knowledge gap is remaining, which is the fourth knowledge gap investigating the heterogeneity within the population with respect to the reaction in travel behaviour as a result of weather variation. As part of the possible individual variables explaining this heterogeneity we include information pertaining to the density level of the residential location of people, thereby addressing knowledge gap 3 again to some extent.

4. *What are the effects of weather on mode choice decisions of heterogeneous parts of the population?*

A visualisation of the relation between the knowledge gaps and the research questions is given in Figure 1.1. The visualisation shows how the knowledge gaps are addressed by the research questions and how these questions are then answered per chapter in this thesis. A flow diagram of the thesis is given further below in Figure 1.2.

Figure 1.1: The connection between knowledge gaps and research questions



1.4 RESEARCH APPROACH

The main objective of the proposed research is to estimate moderating effects of possible moderating factors on the effect of weather on travel behaviour. To provide these estimates a quantitative approach is needed. Quantitative analysis is particularly well fit to transport analysis research [Clifton and Handy, 2003], with many different possible statistical and modelling approaches [Golob, 2003]. The best specific approach, as always, depends on the objective.

To enable this quantitative approach, we first need high-quality data sources for weather, travel behaviour, and possible moderating factors. For the weather data we use data from weather stations in the Netherlands, measured and distributed by the Royal Netherlands Meteorological Institute (KNMI), which is the Dutch national weather service. They supply many different data sets, but we'll be using six different data sets (each with an overarching theme, such as temperature or wind) where values are recorded in 10 minute intervals. The data sets are open, and can be found and accessed through the KNMI open data portal. Data is available from 2003 onwards, with data being added every month.

For our travel behaviour data, we'll use use a relatively new longitudinal panel data set, the Mobility Panel Netherlands [MPN; Hoogendoorn-Lanser et al., 2015]. The MPN presents a few important improvements over traditional travel survey data, the first being the longitudinal nature where people both fill in multiple days per year and are followed for multiple years. This allows us to more accurately establish a causal relationship between weather and travel behaviour, as respondents' decisions are known across time. The second benefit is the increased depth of information in the MPN, as more background variables and statistics are known for the respondents. Using these variables we can look into many variables (socio-demographics, attitudes, residential location, etc.) that might moderate the influence of weather on travel behaviour.

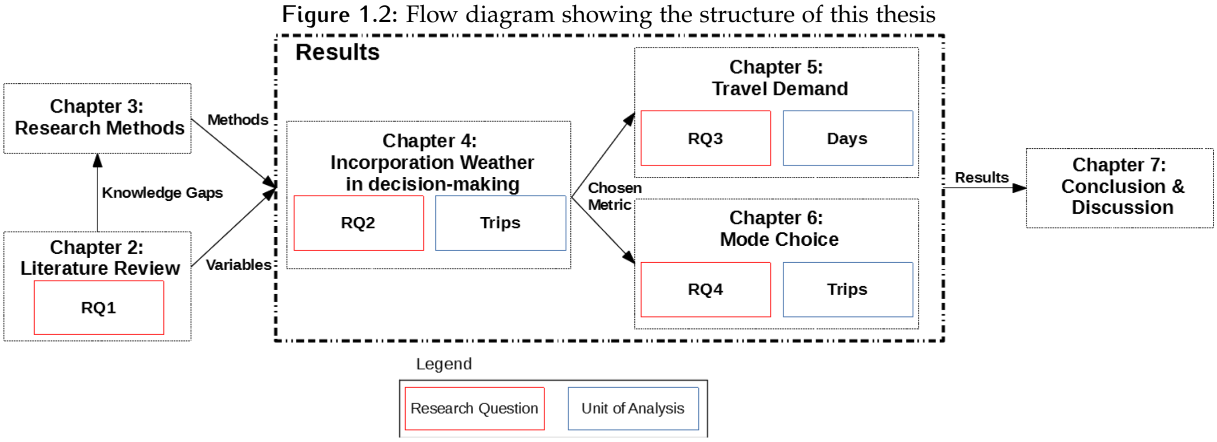
The combination of these data sets enables rich and varied research into the effect of weather on travel behaviour, allowing us to address all four knowledge gaps identified above to some extent. In particular, two dimensions of travel behaviour are of interest: mode choice and travel demand. Mode choice will be estimated using a discrete choice model, which has two important strengths: (1) the approach enables direct comparison of the strength of multiple indicators [Ben-Akiva and Bierlaire, 1999] and (2) the approach ties in with econometric valuations such as the expected Consumer Surplus of policies, projected market shares and point elasticities. [McConnell, 1995]. The choice model is a typical example of a deductive approach: one chooses variables that are likely to influence choices based on theory and tests the theorised effects against empiric data.

Travel demand will be estimated using regression models. Since the main dependent variable will be the number of trips made, which is a count variable, either Poisson or Negative Binomial regressions will be the best fit [Gardner et al., 1995]. Poisson regressions assume that the variance of the dependent variable is equal to the mean, whereas Negative Binomial regressions are able to use a separate variance [Lawless, 2006]. If the variance thus does not equal the mean, Negative Binomial regressions are more appropriate [Ver Hoef and Boveng, 2007].

1.5 THESIS STRUCTURE

This thesis will be structured slightly differently than most, since three quantitative analyses are performed that are in some sense independent of one-another, although later chapters are able to use information gained from earlier chapters during the modelling procedure. For this reason there are three results chapters, namely Chapter 4, Chapter 5 and Chapter 6. These chapters are all written as more or less stand-alone mini-papers, containing a short description of the conceptual relations, data, method, model, results, and conclusions. Of course, these mini-papers are still embedded within the overall work of this thesis, just like a normal results chapter would be. This means that the thesis contains a single introduction, literature review, research methods, and conclusion and discussion. These chapters contain and set the overarching context in which the modelling has taken place and in which the model results can be used for policy analysis and recommendations. The structure is visualised in Figure 1.2 below.

Figure 1.2: Flow diagram showing the structure of this thesis



Chapter 2 contains an extensive review of the available literature. An overview of previous review and research papers is made to see which topics have or have not been studied in-depth. Then the results from the research papers are compared to find which results are consistent and thus where contradictory findings exist. Finally the literature is scanned for their conceptual ideas of how weather might influence travel behaviour. These ideas are discussed and lead to a conceptual model. The research methods are given in Chapter 3, together with an introduction and discussion of the data that is used to estimate the models and a description of the processing of this data. Then in Chapter 4 different conceptualisations of the link between weather and travel behaviour, resulting in different operationalizations, are compared to find which conceptualisation is best able to capture the effect of weather on travel behaviour. The conclusions from this chapter are used to prepare the data that is used to estimate both the the effect of weather on travel demand in Chapter 5 and on mode choice in Chapter 6. Using these results the answers to the research questions are given in Chapter 7, followed by a discussion of the research, its implications for policy makers and recommendations for future research.

2 LITERATURE REVIEW

In this chapter our goal is to find possible conceptual connections between weather and travel behaviour, as well as the strength of the effects as found in previous literature. To gather this data we need to perform a comprehensive review of the literature. Following the guidelines for a structured literature approach by Webster and Watson [2004] and Van Wee and Banister [2016] we start the chapter by explicitly stating how we found the papers we analyse in the literature review in Section 2.1. We follow that up by an overview of the theories and concepts that are central to the topics of travel behaviour and weather in Section 2.2. Then we discuss the papers, starting with the review papers in Section 2.3 and then the research papers in Section 2.4. We conclude the chapter with a conceptual view of the relation between weather and travel behaviour and a scoping and conceptual model for this thesis in Section 2.5.

2.1 LITERATURE SELECTION

To select papers for this literature review the top journals for the transportation field were identified using opinions from scholars in the field and journal statistics as given by publishers. These top ranking journals were the following:

- Transportation Research Parts A, D, and F,
- Journal of Transport Geography
- Transport Policy
- Transport Reviews

Google Scholar and Science Direct were used to search these journals using a simple query: 'weather', which ensured that papers were returned if title, abstract, or key-words included this word. This procedure enables the identification of the most influential papers researching the influence of weather on transport. The papers resulting from this search were scanned for their focus on travel behaviour, as opposed to for example travel infrastructure or road safety. This resulted in a count of 28 research papers and 3 review articles. These papers were backward and forward snowballed through citations (both older papers cited by- and newer papers citing the paper in question). The final review consists of 55 papers, including three review papers.

2.2 CORE THEORIES AND CONCEPTS

Before we discuss the articles we will first examine some of the core theories and concepts for both travel behaviour and weather. This discussion clarifies the definition of some concepts whilst enabling us to embed the research findings into the proper context. We'll first address travel behaviour theories and concepts, then moving on to weather theories and concepts.

2.2.1 Travel Behaviour Theories and concepts

Travel behaviour research can roughly be divided into two streams: demand and supply. Demand is measured in the number of trips made using a certain mode during a certain time period, whilst supply is the number of trips that can be accommodated by this mode during a certain time period. For the car, supply would thus be the number of cars that can be accommodated by a certain road.

Within demand modelling, the so-called four-step model is the dominant modelling approach McNally [2000]. This four-step model would allow one to forecast travel demand at specific locations. The four steps are 1) Trip Generation, which is often measured by a frequency of trips, 2) Trip Distribution, where origin and destination are matched, 3) Mode Choice, where trips are factored to mode-specific trips, and 4) Route choice, where the exact route from origin to destination is modelled.

Of these four steps we are interested in modelling the effect of weather on trip generation and mode choice. This decision is mostly informed by the strenghts and limitations of our travel behaviour data, which is not large enough to ensure a high-resolution trip distribution model could be built and doesn't contain information about the routes travellers used to get from their origin to their destination. Furthermore the effects of weather on mode choice and trip generation are more easily comparable with other countries, which enables a more scientific contribution to the general literature can be made.

Another point of attention is the concept of the trip, which we have mentioned above. A trip is the complete journey from origin to destination. These trips can consist of trip-legs, which is the travel during the trip made with one specific mode. Trips with public transport for example often consist of three trip-legs: one to the public transport access point, the public transport journey, and then the trip from public transport egress point to the final destination. Finally a tour is a special kind of trip, where the trips' origin and destination are the same.

The concepts discussed above are summarized in Table 2.1.

Table 2.1: Travel Behaviour concepts and their definitions

Concept	Definition
Travel Mode	Fundamentally different means of transportation. Examples are the car and bicycle
Trip Generation	The number of trips made at a certain moment and time
Mode Choice	The decision to use a certain travel mode for a trip
Leg	Travel using one transport mode or means of transport
Trip	Complete travel from point of origin to point of destination
Tour	Trip that departs from- and arrives to the same location

2.2.2 Weather Theories and concepts

Weather is a very complex concept, consisting of many different phenomena like temperature, wind, and precipitation that vary considerably across spatial and temporal dimensions. We have made two decompositions of weather: the first is based on the time-scale, the second is based on the different phenomena that together form our concept of the weather

When we de-compose based on time, we identify three different forms of weather: short-term weather, seasonality, and climate. Short-term weather are variations in terms of minutes, hours, or days. Think of a passing rain shower, morning fog, or the temperature of a day. Seasonality refers to yearly changes to the weather patterns, due to the varying angles of the earth in relation to the sun during its yearly orbit. These angles result in a change in the amount and duration of solar radiation falling on specific locations on earth, triggering the seasons of summer, autumn, winter, and spring. Even longer term are shifts in local and/or global climate, causing long term weather patterns to change [Böcker et al., 2013b]. Local climates vary due to different latitudes and the spatial relation to the sea [Barry and Chorley, 2009].

This information is summarized in Table 2.2.

Table 2.2: Weather concepts and their definitions

Concept	Definition
Short-term weather	Weather patterns that vary across days or weeks
Seasonality	Variability in weather patterns throughout the year
Climate	Weather patterns over a long period of time

The other decomposition of weather is based on the type of weather phenomenon. Main examples of such types are temperature, precipitation, and wind speed. These main weather types and some others are given in 2.3, together with their unit of measurement.

Table 2.3: Types of weather and their unit of measurement

Weather Type	Measurement Unit
Temperature	Degrees Celsius
Rain	millimeter per hour
Wind Speed	meter per second
Humidity	% of maximum humidity
Solar Radiation	Watt per square meter
Cloud Cover	Okta
Optical Range	100m

Instead of the more general precipitation we have included a sub-form of precipitation in the form of rain. Other forms of precipitation, such as snow or hail, were not recorded separately and were also very unlikely to fall during the weeks the MPN was collected (late September through early November). During these weeks we have recorded no days with an average temperature below zero degrees.

Whilst we assume that the reader will be familiar with temperature, rain and precipitation and how they are measured we will give a quick explanation of the other variables. Humidity is the presence of water vapour within the air, often measured as the relative amount of water vapour compared to the maximum amount of water vapour that could be present. This maximum amount increases when the temperature of the air increases. Humidity is important for humans, as we rely on evaporation of our sweat to cool us down. This evaporation becomes more difficult when humidity levels get closer to 100%. Areas with high humidity thus feel less comfortable when temperatures increase.

Solar radiation or sunshine is a combination of the duration and strength of the sun's radiation, which is measured in the amount of energy radiated by the sun per square meter. Cloud Cover is a measure of how much of the skies are covered by clouds. There are eight different measurements (which explains the unit name), ranging from a 1 for fully visible skies to an 8 for fully clouded skies. The meteorological optical range meanwhile is a measure for the visibility, which could be reduced by dark conditions, rainfall, smog, or fog. The unit here is the distance it takes for light from a lightsource to be reduced to 5% of its original intensity.

2.3 REVIEW PAPERS

The review papers give broad research gaps and corresponding directions for future research. The three review articles are introduced in Table 2.4, which also includes their key findings about the state of the scientific field.

It is interesting to acknowledge the increasing amount of evidence through time: Koetse and Rietveld [2009] had to conclude that empirical evidence on the influence of weather on travel behaviour is still very scarce, whilst both later reviews had much more evidence to consider. Some of the conclusions from Koetse and Rietveld [2009] can now thus be considered out-dated, as research into the effect of weather on travel behaviour has become more commonplace throughout the last decade. The findings that destination choice, route choice, and departure time are under-researched do however still hold true to this day, probably partially due to the difficulty of estimating the effect of weather on these variables and the fact that this influence is extremely specific to one location's infrastructure. The scarce research that has been performed with this focus has found that there is an impact of weather on these behavioural parameters, although due to the limited amount of research no conclusive statements can be made about this effect.

The other two reviews are more recent and they find similar conclusions about the state of the field. They have both indicated multiple knowledge gaps that need to be addressed by researchers, both to gain a more complete understanding of the relation between weather and travel behaviour and to provide research that is more relevant to the policy-maker.

Table 2.4: Review papers and their key findings

Review Paper	Findings
Koetse and Rietveld [2009]	Destination choice, route choice, and departure time are under-researched Empirical evidence is scarce in general Most focus is placed on short-term weather
Böcker et al. [2013a]	No consideration of weather perceptions Only a few weather parameters (rain, temp, wind speed) researched extensively Weather parameters are often not combined Moderating contexts (such as socio-demographics) rarely considered Prevalent assumption of linear-in-parameters relationships
Liu et al. [2017]	No consideration of weather perceptions Most studies only research a single travel behaviour variable Prevalent assumption of linear-in-parameters relationships Weather parameters are often not combined Unclear how weather is incorporated in travel behaviour decisions

The first knowledge gap discussed by both papers relates to the incorporation of weather into the travel behaviour decision making process. Weather as a phenomenon differs a lot across spatial and temporal dimensions. Researchers that use weather data observed at weather stations need to combine this data with behavioural data. This should be done in such a way that the weather data is as accurate for the behaviour in question as possible, whilst reflecting the decision-making process of the traveller. Most research uses the time of origin, effectively arguing that people base their decision based on the weather at the moment of the decision [Liu et al., 2017, p. 8], whilst past or future weather might also impact the decision [Böcker et al., 2013a, p. 80]. For a commute to work a traveller might for example also consider the expected or forecasted weather during his return journey, rather than just the weather he/she can observe during the trip to work. If we want to provide valid results we thus need to know more about how people use information about the weather to inform their travel decisions.

The second knowledge gap identified by both review papers is the fact that most research estimates the separate effect of singular weather types (wind, rain, temperature, etc.), whilst weather is a co-occurring phenomenon of these singular types. The effect of a change in one type (an increase in wind speed) probably depends on the value for other types (high or low temperatures). An increase in wind speed in hot temperatures might be perceived as a gentle and comfortable breeze, whilst higher wind speeds in cold temperatures can decrease perceived temperatures even further [Böcker et al., 2013a, p. 81]. Researchers thus need to find the effects of combined or holistic weather, instead of reducing complexity by estimating separate effects for each weather type. Similarly research often specifies linear-in-parameters effects of weather types [Liu et al., 2017, p. 9]. In effect this means that research commonly pre-suppose that the effect of an in- or decrease in a weather type is constant throughout the observed range. As an example, this would mean that the effect of a 2 degree rise from -1 to 1 degree Celsius is the same as the same 2 degree increase from 15 to 17 degrees Celsius. In reality decision-makers are probably more affected by the first increase from freezing to non-freezing weather than the second increase, which is barely noticeable. Researchers are advised by the review papers to include non-linear estimates, as they are likely to more closely reflect actual decision-making behaviour.

The third knowledge gap identified by the review papers concerns the fact that weather will probably have a different impact in varying geographical locations, which is not considered or explained often enough [Böcker et al., 2013a, p. 86]. Of course these differences can be expected when considering countries that are far apart, due to the fact that local climates, customs, and travel patterns will be different between two such locations. In a desert climate a decrease in temperature might lead to more comfortable outdoor weather, whilst in a temperate climate an increase might be more beneficial instead. Similarly some countries' travel patterns are more or less reliant on the car, whilst in other countries alternative modes (such as public transport or cycling) might be easier to use instead. If the bicycle is an option, then increases in temperature might lead to substitutions to the bicycle which are of course impossible or more difficult if local customs and infrastructure are not suited to this mode.

Importantly differences are also expected within a country, and especially between rural and urban areas and coastal and in-land areas [Liu et al., 2017, p. 22]. The former difference can be explained both by different travel patterns and the fact that urban areas provide a very local micro-climate by trapping heat and shielding travellers from wind and precipitation. Differences between coastal and in-land areas are expected due to the different local climates, as coastal areas are often colder and suffer from higher wind speeds and more precipitation.

Three knowledge gaps are identified by both of the recent review papers on the topic of weather and travel behaviour.

1. How weather is incorporated into travel decision-making processes is not understood well. Weather changes constantly, both across time and space. We need to know which weather affects travel behaviour to be able to understand its effect.

2. Researchers often estimate linear effects of singular weather variables (wind, rain, etc.). In reality decisions are probably based on the combination of multiple variables providing a single view of the weather as good or bad.

3. We don't know enough about the differences in response of travel behaviour to weather across geographic regions, such as coastal and in-land areas or rural and urban areas.

2.4 RESEARCH PAPERS

The review articles conclude on a high conceptual level about the state of the field and future directions for research. It is useful to complement the analysis of review articles by an analysis of research articles for two reasons: 1) to see the developments in the field that took place after the last review article was published and 2) to get an insight on a lower conceptual level. The most interesting insights relate to findings about the influence of different weather elements and patterns and the techniques and data used to find these insights. Combining the two analyses will then provide clear and convincing knowledge gaps that need to be addressed. The analysis of the research papers will be described per travel mode.

2.4.1 Research Overview

Here an overview of the literature will be made and discussed. The overview gives a comprehensive view of the research design of previous studies. We are thus not yet looking at the research findings, as those will be discussed later on in Section 2.4.2. Two different overviews are given: the first is the overview of recent literature, which extends an overview table from the literature review by Böcker et al. [2013a] to the present using new research articles. The second is an overview of all previous research done in the context of the Netherlands. These views are presented in Section 2.4.1 and Table 2.4.1 respectively. Combining both views will enable the identification of knowledge gaps in the global scientific community that are also relevant in the context of the Netherlands.

We are specifically looking at research in the context of the Netherlands because the influence of weather on travel behaviour is very region-specific for three reasons. Firstly due to the fact that the local climate might influence how people respond to changes in weather: for example, higher temperatures often lead to increases in use of non-sheltered modes, but this finding might not hold for already very hot desert climates. Secondly the local travel customs might have an influence on the reactions to weather: the viability of the bicycle might impact whether inclement weather leads to an increase- or decrease of public transport use. If the bicycle is viable, then this relatively exposed mode is one of the competitors of public transport and inclement weather might lead to increased use of public transport. If it is not viable then the main competitor of public transport is the more sheltered car, and inclement weather might lead to a decrease of public transport usage. Thirdly moderating variables, such as the built-environment, attitudes, or other customs, also vary per region and thus can change the weather - travel behaviour relationship.

Overview of recent research

For the overview of recent research we have made use of a framework that has been used previously by Böcker et al. [2013a] to create an overview table. This table gives an overview of research by visualizing which variables and modelling themes have been addressed in previous research. The columns are research articles and the rows are methodological themes with regards to the weather, travel behaviour, or the general modelling procedure. Cells are left white if a study did not investigate a theme, whilst they are filled if the study did research that particular theme. We have used the papers that are more recent than the review article by Böcker et al. [2013a] in our table, which thus effectively extends that table to the present day. We have also made slight adjustments to the themes we analysed to reflect both our research interests and recent developments within the field. The overview is visually presented in Table 2.5. This table will be interpreted below.

We will discuss this overview in three sections, starting with the meteorological attributes and moving on to the mobility attributes before shortly discussing some modelling themes.

In terms of meteorological attributes we can see that most studies have used objective weather, with relatively fewer studies (also) using perceived weather. Of course data sources for perceived weather are completely different than sources for objective weather. Gaining reliable data about perceived weather in relation to mobility is a very specific task, which costs significantly more time for the researcher than using objective weather that often is already gathered by a national meteorological institute. Such qualitative data could also be gathered in stated preference studies, but we can see that such studies are also relatively scarce. Not surprisingly then not many researchers have paid attention to the effects of weather forecasts on travel behaviour. This question is relatively closely linked to the question of perceived weather, as perceptions of weather are likely to be influenced by the expectations people form about the weather. These expectations must be formed based on some form of forecast, ranging from rudimentary expectations based on their own observations (”it is cloudy, so it’s likely to rain”) or expectations gathered from meteorologists through television, newspapers or online. We can also see that most studies use absolute weather values, instead of relative weather values. Relative weather values are values against some kind of baseline, for example the average weather for a certain season. Use of relative weather thus assumes that people make decisions based on changes in weather, rather than absolute levels.

Looking further to the considered weather parameters we see that the effects of precipitation, temperature, and wind have been reported extensively, possibly because these variables have the biggest effect on travel behaviour. Humidity has also been included often and whilst some studies use other weather variables these are often very interchangeable: no other variables have been used even nearly as often as the previously mentioned ones. We see that very few studies have combined weather phenomena or used some classification of good/bad overall weather: most studies thus research the effect of singular variables as if they are separate entities. As discussed before we expect that the combination of weather parameters would allow researchers to better replicate the fact that weather values always co-occur in their models, enabling a more accurate appraisal of the effect of weather on travel behaviour.

Research overwhelmingly explains mode choice or trip generation, with less attention paid to destination/route choices or departure times. We see that some papers have tried to find an effect of weather on travel distance, whilst others use travel distance as an explanatory variable. In the former case the theory is that people travel smaller distances in inclement weather, in the second that people who travel shorter distances are less affected by weather circumstances. Perhaps both theories are true at the same time, but no papers have investigated both effects simultaneously. The relative lack of attention paid to destination or route choices can be explained in multiple ways. First is that estimating such impacts necessitates information about existing routes and destinations, which is relatively scarce compared to mode choice and trip generation data. Secondly knowing the effect of weather on mode choice and trip generation might lead to more concrete policy recommendations and finally these effects are easier to generalize to other regions and countries.

Other explanatory variables that are only used sparingly are people’s attitudes, health, urbanization, and peak-hour effects. For the former two this can be explained by the fact that most research has been

Table 2.5: Overview of recent studies, lay-out adapted from [Böcker et al., 2013a, P. 73]

			Bocker et al. 2013	Cools & Creemers 2013	Hjorthol 2013	Sabir et al. 2013	Thomas et al. 2013	Arana et al. 2014	Gebhart & Noland 2014	Helbich et al. 2014	Nosal & Miranda-Moreno 2014	Singhal et al. 2014	Wadud 2014	Bocker et al. 2015	Creemers et al. 2015	Liu et al. 2015	Liu et al. 2015	Mathisen et al. 2015	Motoaki & Daziano 2015	Stralen et al. 2015	Zanni & Ryley 2015	Bocker et al. 2016	Kashfi et al. 2016	Meng et al. 2016	Schmiedeskamp & Zhao 2016	Shaaban et al. 2016	Tao et al. 2016	Li et al. 2017	Gallop et al. 2018	Hyland et al. 2018	Ma et al. 2018	Rerat 2018	Tao et al. 2018	Ton et al. 2018	Wei et al. 2018	Zhao & Wang 2018	Miao et al. 2019	Nordbakke & Olsen 2019
Meteorological Attributes	Measurement Themes	Objective Weather	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1				1	1	1	
		Perceived Weather			1									1							1	1		1								1		1				
		Absolute Weather	1	1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1					1	1		1	1		1				1			
		Relative Weather					1					1				1	1						1				1									1		
		Recorded Weather	1			1	1	1	1	1	1	1	1	1	1	1	1	1					1		1	1	1	1	1		1				1	1	1	
		Stated Weather		1														1	1										1								1
		Rolling Values									1													1					1							1		
		Weather Forecasts		1																1				1														
		Sub-Hourly Data																																				
		Hourly Data				1			1		1	1	1	1	1								1				1	1		1		1		1			1		
		Daily Data					1	1		1	1	1			1	1	1	1					1		1		1	1			1			1	1	1	1	
		Regional stations	1			1	1	1	1	1	1	1	1	1	1	1	1	1					1		1	1	1	1	1		1				1	1	1	
	Weather Parameters	Seasonality	1		1		1				1		1	1			1	1					1		1	1	1	1	1	1			1	1		1		
		Precipitation		1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1		1	1	1				1		1	1	1	1
		Temperature		1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1		1		1		1	1	1	1
		Wind		1		1	1	1	1	1		1	1	1	1	1		1				1	1	1		1	1		1		1		1		1	1		
		Humidity				1	1	1	1		1		1		1	1							1	1		1	1		1		1		1		1	1		
		Others	1	1		1	1		1			1	1	1	1	1	1		1	1	1	1				1			1		1						
		Weather Impression																			1			1						1							
		Combined parameters	1											1	1								1				1						1		1			
Mobility Attributes	Explained Variables	Trip Generation		1	1		1	1	1	1	1	1	1			1		1	1	1			1		1	1	1	1	1			1	1	1	1	1	1	1
		* Car				1								1	1		1			1										1	1							1
		* Public Transport						1				1		1	1		1			1			1				1	1		1	1		1		1		1	
		* Bicycle					1		1	1	1		1	1		1	1	1	1	1				1	1				1		1	1		1		1		
		* Pedestrian												1		1	1									1					1			1				1
		* Other													1		1																				
		* Multi-Modal																																				
		Mode Choice	1	1		1								1	1	1	1			1		1								1	1							
		Destinations		1																							1						1					
		Routes		1														1	1							1												
		Departure Time		1															1	1																		
		Distances/Travel Times	1		1				1						1										1								1			
	Explanatory Variables	Socio-Demographics	1	1	1	1				1			1	1		1	1		1		1	1									1	1			1				1
		Travel Cost/Travel Time				1								1			1		1										1							
		Mode availability	1	1	1					1			1	1		1	1					1													1				1
		Attitudes												1							1	1																1
		Health			1					1				1					1			1															
		Urbanization	1	1										1		1	1																				
		Weekday/Holiday	1			1	1		1		1	1		1			1	1			1	1		1	1	1	1	1				1		1			
		Day/Night/Peakhour	1						1					1								1				1	1		1				1			1		
		Trip Purpose	1	1	1	1	1	1				1		1		1				1		1								1			1	1				
		Built Environment	1						1			1	1	1					1								1						1	1			1	
	Measurement Instrument	Travel Diary	1		1	1				1				1	1							1										1			1			
		GPS																																				
		Road Counters					1				1		1												1				1								1		
		Smart-Card' data						1	1			1												1				1	1					1		1		1	
		Survey	1	1	1								1	1	1	1		1	1	1			1							1	1	1		1				1
		Others																								1											
Modelling	Modelling Themes	Panel								1				1						1										1				1			1	
		Cross-Sectional	1														1								1	1										1			
		Time-Series									1	1						1										1	1								
		Individual	1	1	1	1				1				1			1		1	1	1	1		1						1	1	1		1				1
		Aggregate					1	1	1		1	1	1										1		1	1	1	1		1				1		1		1	
		Heterogeneity						1		1				1		1			1	1										1							

done using aggregate travel behaviour data where nothing is known about the choices of individual travellers. To estimate the effect of attitudes or health we do need this individual data, for example collected using travel diaries. We need the same information to estimate if there are different groups of travellers who respond differently to weather circumstances, which explains why only few studies have addressed this question as well. In general questions pertaining to the heterogeneity of the reaction of travel behaviour to weather, whether that be heterogeneity between travellers or between regions, are not researched often. Most research thus estimates a singular average effect that is assumed to be true across the entire population. We expect that there is a sizeable amount of heterogeneity. This heterogeneity might be on the level of the traveller, caused by varying travel patterns, attitudes, or the availability of alternative modes of travel. It might also be investigated between various regions, such as urbanized vs. rural or coastal vs. in-land regions.

In conclusion we can state that our overview of recent research leads us to believe that a couple of points are under-researched in the global research field. We seem to lack information about perceived weather, both how it is affected by (forecasts of) objective weather and how it affects travel behaviour. Furthermore we need to look beyond precipitation, wind, and temperature, to ensure that all possibly relevant variables are considered. Finally research has only sparingly investigated the effect of heterogeneity with respect to the reactions to weather in the population.

Overview of Dutch research

In the context of the Netherlands, three groups of researchers have published papers on the influence of weather on travel behaviour in the last decade. The first group is based primarily on research done by Muhammad Sabir, who produced a doctoral dissertation on weather and travel behaviour in 2011 [Sabir, 2011]. The dissertation contains three parts: weather and individual travel choices; weather and travel time; and weather and road safety. Sabir et al. use metereological information retrieved from KNMI weather stations throughout the Netherlands and match this to individual trips/days, with data ranging from 1996 through 2005. The travel behaviour data is the MON (Mobiliteitsonderzoek Nederland) / OViN (Onderzoek Verplaatsingen in Nederland), which is a data-set that is widely used to study travel behaviour in the Netherlands. The data-set consists of around 50.000 respondents who all fill out a travel diary for one day of the year and answer a basic questionnaire, providing information about socio-demographics, availability of modes, etc.

The second group is the result of research of Lars Böcker, who similarly wrote a doctoral dissertation on climate, weather, and daily mobility which was succesfully defended in 2014 [Böcker et al., 2013a]. In contrast to the country-wide studies by Sabir, Böcker focuses his research on the very urbanized Randstad region in the Netherlands. Böcker also makes use of MON/OViN for one of his chapters investigating the possible effects of climate change on travel demand [Böcker et al., 2013b], although this study makes use of a sub-sample that is limited to the Randstad region. Due to limitations of MON/OViN, including it being too broad to effectively study weather and the lacking account of subjective experiences, Böcker et al. also gathered their own data within the greater Rotterdam area. This data set consists of a questionnaire and three repeated waves of travel diaries distributed in consecutive seasons. This data set allows Bocker to tailor their research towards the more subjective experiences and perceptions of weather, linking them to place valuations and the change brought upon these by variations in weather circumstances.

The third group is spearheaded by Lieve Creemers, a Flemish researcher who first wrote a Master's thesis on the effect of weather conditions and forecasts on travel behaviour [Creemers, 2010] and then followed a PhD track that seems to have a broader focus on travel demand forecasting. The research of Creemers is partly based around Flemish data and partly uses Dutch data. The Flemish-based papers mostly use a stated-preference approach to investigate the impacts of forecasts [Cools and Creemers, 2013], trip purpose, and type of weather [Cools et al., 2010]. The Dutch-based papers take a different approach, as they use revealed preference data from MON and weather stations from the KNMI [Creemers et al., 2015] in an approach that closely resembles that of Sabir. The main innovation of this paper is the use of multiple thermal indices, rather than directly observable weather variables.

Besides these PhD tracts where the relationship between weather and travel behaviour was at the forefront of the research, weather has sometimes been a relatively small part of other research. Examples are the findings by Eva Heinen, whose PhD research revolved around use of the bicycle [Heinen, 2011]. As part of this research, she studied the influence of weather on bicycle using two different research designs. In the first an analysis is made of commute to work choices in the Dutch cities of Delft and Zwolle [Heinen et al., 2011]. Travel behaviour data based on an internet survey is coupled to the weather stations that are closest to these two cities. Interestingly separate models are estimated for occassional and frequent cyclists, enabling comparison of the effect of weather between these two groups. The other study uses interviews to qualitatively assess the influence of attitudes and norms on bicycle commuting [Heinen and Handy, 2012]: these attitudes and norms also encompass the weather and its perception in relation to bicycle commuting.

There are two further individual papers that study the effect of the weather on travel behaviour in the Netherlands. The first researches the effects of weather conditions on cycling demand and couples weather station data to nearby bicycle road counters collected at 16 bicycle paths near the Dutch cities of Gouda and Ede [Thomas et al., 2009]. The second paper researches the effect of weather on the probability of congestion, doing so by estimating both the effects of adverse weather conditions on both road capacity - travel supply - and on the number of trips made by car in these conditions and thus travel demand [van Stralen et al., 2015]. Combining these analyses returns the probability that demand exceeds supply and thus that congestion forms. The travel demand portion is estimated using a choice model, the data for which is gathered using a stated preference approach.

We see that in the context of the Netherlands a relatively large amount of time is spent on the effects of forecasts and weather perceptions on travel behaviour, making this a less pressing knowledge gap. In contrast no information is known about heterogeneities within the population, both between different regions of the Netherlands and between groups of travellers.

Knowledge Gaps

Based on our overviews of recent and Dutch research we have found a few additional knowledge gaps. Below we will describe these knowledge gaps. Unfortunately we're not able to address all of them within our research, so we'll also motivate shortly why we decided not to pursue some of them.

The first additional knowledge gap relates to the perception of weather and how these perceptions influence travel behaviour. We know that for the weather to have any influence on behaviour it needs to be perceived by the decision-makers. We expect that these perceptions are thus more direct drivers for behaviour. It would be very interesting to know how objective weather is linked to perceived weather and how these perceptions of weather in turn influence travel behaviour. Unfortunately we are unable to pursue this research during this thesis, as the necessary data is not collected as part of the Mobility Panel Netherlands. This MPN does address weather with one question, but it's only asked if people have not travelled during the day at all. People only answered positively a few times, making effective research impossible. To address this question we would have needed to collect our own data and link it to the travel behaviour data within the MPN, which is not feasible in the time available for this thesis. Closely related is knowledge pertaining to how forecasts of weather influence travel behaviour. We think such research could be combined very well in a more qualitative setting where people are interviewed about how they use forecasts when planning a trip. The nature of such research is thus drastically different than the quantitative research performed in this thesis, making it very difficult if not impossible to address this knowledge gap.

> Research into the effects of perceived weather is still scarce. Unfortunately such research requires data that is not available to us so we are unable to address this knowledge gap within this thesis.

The heterogeneity of travellers' reactions to weather has also been researched only very sparingly, both in recent and in Dutch research. With heterogeneity we mean that people will respond differently to the same weather circumstances. We might for example envision that people with more experience with the bicycle are more comfortable to use the bicycle during slippery conditions than less experienced cyclists. But it might also be the case that people who use either the bicycle or public transport start using public transport more during rain, whilst people who use either public transport or the car might use public transport less. This information can be used by policy makers to change the effect of weather on travel behaviour, for example by ensuring people get experienced with the bicycle.

Taking this heterogeneity into account we thus estimate a more complex model, with the assumption that this complexity is a better representation of reality. The data we use is perfect for these kind of models: the travel behaviour data from the MPN gives us information about multiple consecutive choices by the same traveller and includes many background variables for each traveller. We are thus able to see the decisions people make under various weather circumstances and we know who makes these decisions. This allows us to uncover patterns related to the background variables. We thus do pursue this knowledge gap, as well as the three knowledge gaps found in the review papers described in Section 2.3.

> The heterogeneity of travellers' reaction to weather has not been researched extensively. Such knowledge could be a valuable addition to the scientific field and it could be used by policy makers to change the effect of weather on travel behaviour. Furthermore our data is perfect for the models necessary to uncover the heterogeneities. We thus focus on this knowledge gap in addition to the three knowledge gaps identified by the review papers.

2.4.2 Research Findings

Here the actual research results will be discussed, to see which findings are replicated often and which findings are seemingly contradictory. Contradictory findings might provide a further research area where this thesis might be able to provide needed clarity. The first four sections cover one travel mode and discuss the research into the effect of weather variables on this specific mode. The last section is more general and focusses on modelling trends, mediators, and non-weather variables that play a role in the conceptual link between weather and travel behaviour.

Car

The first travel mode that will be discussed is the most ubiquitous: the private car. The car is a sheltered mode that often has the ability to either cool or heat travellers to shield them from the weather and create a pleasant micro-climate. In theory, the car should thus be the mode that is least sensitive to weather influence. This theoretical notion is backed up by Böcker et al. [2016], who find that the travel experience and place valuation of trips using sheltered modes is less sensitive to inclement weather compared to non-sheltered modes. This doesn't mean of course that car use and travel behaviour is completely insensitive to weather changes: below the main findings of the literature will be presented.

The first weather variable that might influence car use is the temperature. In general, research finds that cold weather leads to reduced car usage [Liu et al., 2017]. Al Hassan and Barker [1999] finds that traffic volume decreases as a result of severely cold weather, perhaps explained by the barriers for travel behaviour and mobility that can be caused by winter weather Hjorthol [2013], such as frozen or snowy roads causing slippery conditions which dissuade people from travelling by car [Kilpeläinen and Summala, 2007] (in a climate where temperatures in winter drop below 0 degrees Celsius). Even in a climate where freezing temperatures are an anomaly there seems to be a reduction of traffic volume caused by cold and wet seasons, as in the Melbourne-based study by Keay and Simmonds [2005]. In terms of mode-share, the car seems to lose out to the bicycle when temperatures increase [Khattak and De Palma, 1997; Liu et al., 2015b]. In a projection of the impacts of climate change Böcker et al. [2013a] finds that the softer winters will correspond to a decreased use of the car, whilst hotter and wetter summers might prompt people to use the car more.

Rain and general precipitation also seems to drive people to using the car more frequently, as a substitute for less-sheltered travel modes. Hyland et al. [2018] finds that commuters shift towards using the car during inclement weather conditions. Similarly precipitation seems to increase car usage in the study by Liu et al. [2015a]. By making a distinction between light rainfall and heavy rainfall, van Stralen et al. [2015] find a non-linear relationship between precipitation and travel demand in their stated preference study. Whilst light rainfall conditions increase car usage, heavy rainfall conditions seem to decrease the number of trips with the car.

Car usage seems to not be affected much by any other conditions, although the modal share of the car might increase due to a decrease in the use of other modes. In general, one of the main overall conclusions is that car behaviour is in fact relatively insensitive to the direct influence of weather variations. Weather variations that impact the infrastructure however do indirectly impact car use, as can be seen in the example of winter weather causing slippery conditions, which in turn prompt people to reduce their car usage. A variation on this theme can be found in the research by Sabir et al. [2013], who look into travel behaviour of trips made to the beach. They find that hot conditions increase the likelihood of congestion, prompting some people to use public transport or bicycle instead.

> The car offers a relatively sheltered means of transport. In inclement weather conditions (cold and wet weather) people switch from more exposed modes, such as the bicycle, to the car. Car usage is thus increased in these conditions.

Public Transport

These papers have studied the effects of weather on the amount of passengers of public transport modes: bus, tram, metro, or train. Public Transport itself is sheltered, but the travel to and from entry- and exitpoints is often more exposed. The general conclusion from these papers is thus that there will be less total trips made with public transport during inclement weather conditions [Tao et al., 2016; Wei et al., 2018], such as cold, rainy, and/or windy days.

Many papers on the use of the bus find that there is a positive relationship between temperature and transit usage [Stover and McCormack, 2012; Arana et al., 2014; Li et al., 2017]. Some papers however present conflicting evidence: Singhal et al. [2014] for example finds that hot days have a negative effect on the number of transit trips, and Miao et al. [2019] finds that both very cold and very hot days result in less trips. Furthermore Kashfi et al. [2016] and Tao et al. [2016] don't find a significant relation between temperature and the amount of trips using transit modes at all. These differences might be explained by local differences; such as the climate of the location or the transport mode that are used as substitutes. In the context of the Netherlands Creemers et al. [2015] finds that increases in temperatures (more specifically, the physiologically equivalent temperature [PET]) have a negative effect on the mode share of public transport trips. Sabir [2011] finds a similar result using a regression model: higher temperatures decrease transit usage for both the train and bus, tram and metro (btm) in the Dutch context. This might be the result of the unique position of the bicycle in the Dutch context, where higher temperatures might lead to the substitution of public transport for the bicycle.

Some studies find that increasing wind speeds also negatively affect transit ridership [Guo et al., 2008]. This effect seems to hold up for more than one form of public transport, as both studies researching bus [Arana et al., 2014] and underground [Singhal et al., 2014] use report this finding. Some studies however find no significant effect [Kashfi et al., 2016], whilst another study only finds a negative effect in winter, spring and autumn but not during summer months [Stover and McCormack, 2012].

> There is relatively more uncertainty about the effect of weather on public transport use, resulting from contradictory findings. Within the Netherlands previous research has found that increasing temperatures reduce public transport usage.

Cycling

Cycling is the single mode that is researched most often, which makes intuitive sense: cyclists are often more vulnerable to inclement weather and the expected effects of weather on cycling behaviour is thus probably stronger than that on more sheltered modes. Most papers do indeed seem to agree that cycling counts are affected by changes in the weather, especially by changes in temperature, precipitation, and wind speed [Zhao et al., 2018; Rérat, 2018; Gallop et al., 2018].

This consensus extends for the most part to the direction of the effects. The presence of rain decreases the bicycle count, as do increasing wind speeds. Most papers use data collected in locations with relatively mild climates and find that increasing temperatures have a positive effect on the number of trips using the bicycle. An interesting side-note comes from the findings of two papers researching the effects of weather in the hot climate locations of Dar-Es-Salaam and Singapore : they find that increasing temperatures reduce cycling counts [Nkurunziza et al., 2012; Meng et al., 2016]. This is evidence to suggest there is an ideal temperature for cycling and that both colder and hotter temperatures deter the use of the bicycle. More evidence is found in papers that find that the effects of temperature on bicycle counts changes as the temperature increases past a certain point. Wadud [2014] finds that the positive effects level off after 25 °C, whilst the results from Miranda-Moreno and Nosal [2011] suggest that the effect becomes negative after 28 °C. Khattak and De Palma [1997] reports on the same effect, but finds that 24 °C is the point where bicycle counts peak. So whilst there doesn't appear to be an exact value for the ideal cycling temperature, the general idea that there is such an ideal temperature for cycling seems to be replicated in the literature. Most of the cycling papers however do not report this ideal temperature, mostly resulting from the fact that the model that is used doesn't include non-linear effects. It is unclear whether these papers have estimated a quadratic/polynomial effect but found it to be insignificant or whether they simply assumed an effect that is linear in parameters.

There is less consensus about the absolute or relative strengths of the effect of the weather parameters. Nankervis [1999] for example indicates that the effects of weather are not as strong as commonly thought, whilst Winters et al. [2011] and Rérat [2018] indicate that weather is one of the most important factors. Some papers meanwhile find that temperature is the strongest predictor [Thomas et al., 2009].

> The bicycle is often found to be the most affected by weather circumstances, probably due to its relatively high exposure to weather. Increases in rain and wind lead to reduced bicycle usage, whilst higher temperatures induce more bicycling usage up to a certain ideal temperature.

Pedestrian

Pedestrians are, similar to cyclists, relatively exposed to the weather when compared to motorised modes. Whilst bicycling use has been the topic of many papers, relatively fewer papers have looked into the effects of weather on pedestrian behaviour.

Shaaban and Muley [2016] finds a log-linear relationships of temperature and pedestrian volume. They find that a higher temperature reduces pedestrian volume, which can be explained by the fact that their research is set in a desert climate. Saneinejad et al. [2012] find in their Toronto-based study that colder temperatures reduce the amount of walking, whilst precipitation seems to cause an increase, with a possible explanation being that people prefer walking over cycling in rainy conditions.

Seasonal effects are also found, for example by Ton et al. [2018] who found that the relatively warmer month of september prompted more people to walk than the colder months of october or november. Opposite findings are presented in the study by Liu et al. [2015b], who indicate that winter increases the mode share of walking.

Böcker et al. [2013b] shows that projected climate change scenarios will increase walking in the winter and spring seasons (probably due to increased temperatures), but cause a decrease in summer and autumn seasons (due to increased precipitation).

> The effect of weather on the number of pedestrian trips is also relatively sizeable, although pedestrians seem to be affected less than cyclists. There is also less research focused on pedestrian behaviour compared to cyclists, which means that there is less evidence for the effects. Findings seem to indicate that there is an ideal temperature for walking, with temperatures possibly being either too hot or too cold. Precipitation also seems to have a negative effect on the number of pedestrian trips.

Moderators, explanatory variables and modelling themes

There have been some investigations into the differences in travel behaviour responses to weather variation across the spatial dimension, for example differences between urban and rural areas. Within the Brisbane area, Tao et al. [2018] conclude that the effects of weather are more pronounced for peripheral regions compared to the inner-city. Similarly in research conducted in the Rotterdam area, Helbich et al. [2014] find considerable spatial differences between the inner-city and periphery. Some research has investigated the role of additional transit stations and found that people in areas that are close to stations are less sensitive to weather changes [Miao et al., 2019]. Other studies find that areas with underground (and thus sheltered) stations are less susceptible than areas with elevated (less-sheltered) stations [Singhal et al., 2014].

Another moderator that is often included is the type of day, where typically a difference is made between weekday and weekend trips under the assumption that weekend trips are on average made for leisure purposes more often than weekday trips. These studies all report that weekend use is more susceptible to weather influences than weekday use [Arana et al., 2014; Tao et al., 2018; Wei et al., 2018; Guo et al., 2008, and more.]. Other approaches that have investigated the moderating effect of trip purpose directly also report that leisure trips are more sensitive than utilitarian trips [Helbich et al., 2014; Kilpeläinen and Summala, 2007; Cools and Creemers, 2013, and more.].

Two papers have already incorporated heterogeneity between travellers. Motoaki and Daziano [2015] uses a latent class approach and finds that there are two distinct groups whose cycling behaviour responds differently to weather. One group of more experienced riders is more resilient to weather changes compared to the less-experienced group. This inclusion of heterogeneity is however rare: since most research uses anonymous travel behaviour data in the form of automatic counters they are unable to estimate separate effects for subgroups in the population. Nordbakke and Olsen [2019] uses a different approach to estimate heterogeneity: their dependent variable is 'weather tolerance', which they estimate using personal characteristics, attitudes, and habits. They find that environmental attitudes and transport habits are related to weather tolerance and conclude that policy measures aimed at these variables might be able to increase weather tolerance.

2.5 CONCEPTUAL CONCLUSIONS

The above sections give a good idea of the strength of the effect of various weather variables on travel behaviour, in particular the effect on trip generation and mode choice. We have also found some knowledge gaps that are still remaining in the literature and some contradictory findings, especially with respect to the influence of weather on public transport.

What is not so clear yet is how this influence works and which factors are part of the causal path between the exogenous weather and the endogenous travel behaviour. In the sections below we will first discuss the paths between weather and travel behaviour. We will work towards a conceptual model of the relations that are studied within this research, which can be seen in Figure 2.1. We then discuss how the different paths within this conceptual model relate to the knowledge gaps and where the paths within the model are addressed within our research in Section 2.5.2.

2.5.1 Conceptual Model

The question at hand is thus how weather affects travel behaviour, where we are interested in the qualitative pathways between these two concepts rather than the quantitative estimates of such a relationship. We're thus describing the causal paths between weather and travel behaviour and the factors that can be found in between the exogenous concept of weather and the endogenous concept of travel behaviour. These causal paths can become very complex, containing many factors, paths, and feedback loops. A comprehensive study of the entire range of possible influences is thus impossible. Below we highlight the most important paths and give our reasons and arguments for including or excluding these paths within our analysis. The final conceptual model is thus a simplified version of reality, which serves as a useful framework throughout the thesis ensuring that all of our analyses fit within an overarching story.

Weather can affect travel behaviour in many different ways. Below we will describe the most important of these pathways, ultimately leading to the formulation of a conceptual model that is used within this thesis.

The first question we need to answer is how humans make (travel-behaviour) decisions, which of course is the subject of entire research fields. It is clear that our decisions are the result of mental processes, but there is no consensus on what these processes entail (or even how much control we have over them). Following the econometric approach we pre-suppose that this mental process is a somewhat rational comparison of different alternatives where we estimate the utility of each alternative. The utility can be seen as the usefulness or gain that the decision-maker would attain when he chooses an alternative. The decision-maker will then choose the alternative with the highest utility. These alternatives are the different travel modes in the mode-choice models and going on a trip or staying at the current location for the travel demand model.

Given these assumptions we now need to know how the weather affects the utility of the different alternatives. We can identify two distinctly different paths, the first of which is the effect the weather has on the infrastructure (which in turn affects the utilities) and the second is the effect of weather on the utility of the modes directly. The effect of weather on the performance of travel infrastructure can be profound, as for example is the case when public transport companies change their schedule due to extreme weather events. Whilst we should be aware that travel behaviour is affected by the state of the infrastructure, we have not included these effects in our thesis. We have made this decision based on the fact that such research would require a different set of data, methods, and background knowledge, making it difficult to do well and to include it within the overall context of the other analyses.

The effect of the weather can be classified into two separate paths. The first goes directly from weather to the utility of an alternative. When it is raining for example using the exposed bicycle will probably become less attractive. The second path is the effect of weather on infrastructure, which then leads to behavioural change. If it's freezing trains can run less often, which causes travel times to increase and thus make the train less attractive. Within this research we have focussed on the former pathway.

The second path mentioned above is the effect of weather on the utilities of the modes and trips. As said before, for the weather to affect utility it needs to first be translated into a mental state by people, which could be conceptualised as their perceptions. Objective weather conditions thus influence perceptions of weather in the first place. This relationship is moderated by individual and local characteristics, as well as custom: for example a temperature of 15 °Celsius might be perceived as cold by people living in a desert climate, but as hot by people living in an Arctic climate. These perceptions then in turn have an influence on the utility associated with trips and modes. Perceptions of cold weather might lead to reductions in utility associated with the more exposed active modes for example. Note that the influence of weather on utility is thus always linked through a mental state: this is the translation of the objective, measurable weather to the subjective experience that is the result of this weather.

This causal path from weather to travel behaviour thus runs through two factors: the perceptions of weather and the resulting assigned utilities to each of the alternatives. Unfortunately we do not have the data necessary to include the perceptions of weather in the analysis, which means that we'll have to forego this link in the chain and determine the direct effect of weather circumstances on the utility of alternatives.

We hypothesize that the strengths of these paths are moderated by many variables related to the individual making the decisions and the trip in question. In terms of individual characteristics we investigate socio-demographics, attitudes, mode availabilities, and spatial variables. We'll give a theoretical example of expected moderation effects for all of these variables. For socio-demographics we suppose that older people's bicycle use will be more affected by increasing wind speeds, as their lower average physical fitness means that cycling against the wind would be more of a hindrance in comparison to younger people. We expect that people with positive attitudes to the bicycle are more likely to keep using the bicycle, either because they are more experienced and thus more capable of dealing with wind speeds or because their perceptions of negative influences on bicycle use will be reduced.

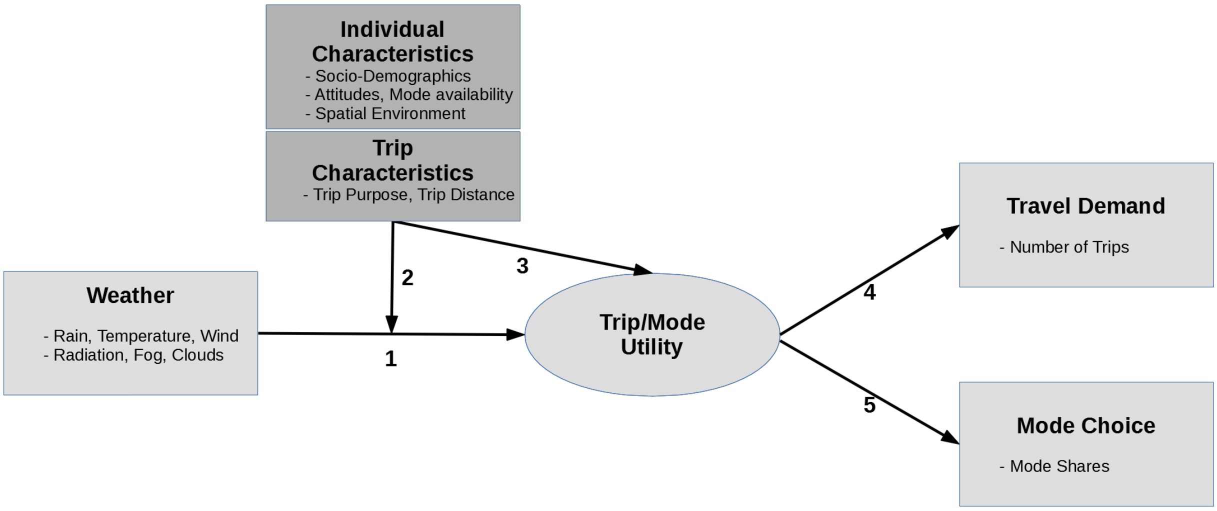
People who own an electronic bicycle are expected to be less susceptible to increasing wind speeds, as less effort is required on the part of the bicyclist. Finally we think that there might be a difference between areas with a high population density and more rural areas, for example because the buildings in denser areas might act as a shelter, creating more pleasant cycling conditions.

Finally we also think that trip characteristics might mediate the path between weather and utility. Finally the purpose of the trip is expected to matter as well, with leisure trips being affected to a greater extent than utilitarian trips.

> We expect that the effect of weather on travel behaviour is not the same for all travellers. Factors related to the individual and the trip are expected to moderate the relationship between weather and travel behaviour. As examples, we expect people with more bicycling experience to be less sensitive to inclement weather and we think that trips with a leisure purpose are more affected by the weather than utilitarian trips.

The above discussions lead to the conceptual model that is visualised in Figure 2.1. Note that this is thus a conceptual model that is used within this thesis and is not ment as a complete conceptual overview of all possible relations between weather and travel behaviour.

Figure 2.1: Conceptual Model of the relationships that will be studied in this thesis



2.5.2 Thesis sections

All of the links presented within Figure 2.1 will be studied within this thesis. We use different models to estimate various parts of the conceptual model, meaning that each results chapter focuses only on a subsection of the complete model.

First we will try to find how we can conceptualize and operationalize the link between weather and mode utility in Chapter 4. To do so we need to figure out how weather is incorporated into the decision-making process of travellers, and more specifically which weather is used to make the decisions: weather at the time of departure, on-route, daily weather, or other options are all possible. This chapter does not try to estimate accurate effects on mode choice, nor does it include moderating factors.

The effects of weather on travel demand are estimated in Chapter 5, whilst the effects on mode choice are estimated in Chapter 6. Both chapters actively control for moderating and mediating effects of other variables and include variables related to socio-demographics and the spatial environment. They focus on different parts of the conceptual model however, differing to some extent into how and which variables are used. Which link is studied where is informed by two differences between these chapters: the first is the unit of analysis (the day for Chapter 5 and the trip for Chapter 6 and the second concerns the strengths and weaknesses of the different modelling techniques.

It's easier to estimate and interpret direct relations between socio-demographics, including spatial environments with regression models. Within the latent class models including such effects would make the utility function much more complex, making it much more difficult to interpret the findings of the model. Conversely one of the main strengths of the latent class approach is its ability to allow for many moderators of a relation, where the discrete class structure ensures relatively straightforward interpretation. There are still some moderating effects investigated within the regression model (such as the moderating effect of the spatial environment) and some direct influence of trip characteristics on utility within the choice model, but they are limited.

Table 2.6 contains an overview which relationships will be studied in which chapter. The main chapter is mentioned in bold, other chapters that study the relationship in less detail are included in parentheses.

Table 2.6: Chapters where theoretical relationships are discussed

Number of link	Thesis Chapter
1	**4**
2	**6** (5)
3	**5** (6)
4	**5**
5	**6**

3 RESEARCH METHODS AND DATA COLLECTION

This chapter will describe the research methods, tools, and data necessary to provide answers to the research questions described in Chapter 1. First the data sources and preparation will be described in Section 3.1. Then the methods used to research the effects of weather on trip generation and mode choice will be described in Section 3.2.1 and Section 3.2.2 respectively.

3.1 DATA SOURCES AND PREPARATION

Of equal importance to the model specification is the data that is used as input. To estimate the effects of weather on travel behaviour one would need (at least) two data-sets: one pertaining to weather and one pertaining to travel behaviour. The data sources for both of these types of data can be found in section 3.1.1. The data preparation procedures are described in Section 3.1.2. A more in-depth explanation of the data prepation process can be found in Appendix A.

3.1.1 Data Sources

High quality data is an absolute necessity to provide quality scientific work. For this thesis we use objectively measured weather data provided by the Royal Netherlands Meteorological Institute (KNMI), which is the leading Dutch weather-based institution. Travel behaviour data is obtained from the Mobility Panel Netherlands (MPN), which consists of multiple parts including a travel diary and more general surveys. More information on both data sources is given below.

Weather Data

For my objective weather data this research uses open data collected and distributed by the Royal Netherlands Meteorological Institute (KNMI) at the KNMI open data portal. The data-set contains information on 45 different weather attributes, ranging from rainfall to solar radiation, as collected by 50 different weather stations. Note that not all of the weather stations collect all types of weather: some might measure only temperature, whilst others might measure only wind speed. The KNMI thus provides different data sets for the overarching weather parameters (such as temperature). A total of six different data sets are used in this research, with the following themes: precitipation, humidity & temperature, atmospheric pressure, wind, clouds, and sunshine. The geographic spread of the weather stations is visualized in Figure 3.1, which also shows the number of themes captured by each of the stations. The data-sets contain a very high temporal resolution, providing values for every 10 minutes. This temporal resolution allows for more precise matching of these objective weather data to travel behaviour data.

Travel Behaviour Data

The data pertaining to travel behaviour is also already collected, in the form of the Mobility Panel Netherlands (MPN). The MPN is a longitudinal panel data-set, where respondents are tracked across multiple years [Hoogendoorn-Lanser et al., 2015]. Collection began in 2013, so that as of January 2019 six yearly waves have been collected, of which five have been fully processed and are ready for analysis whilst the sixth can serve as an external test data dest. The MPN consists of two parts: the first consists of multiple questionnaires, which contain questions about amongst others demographics, attitudes, locations, and opinions of the respondents. This information is provided based on the unit of analysis: there is one data set for days, one for people, and one for households. Every two years a special

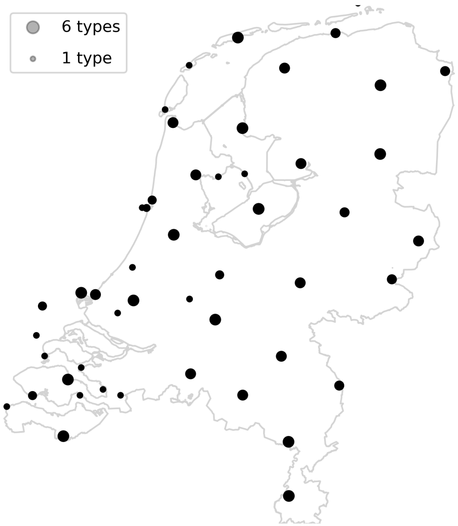


Figure 3.1: Location of weatherstations in the Netherlands

subject is investigated in more detail. For waves 2 and 4 this special subject is the opinion of people with regards to travel modes, where people's attitudes are collected. The second part is a travel diary, where each respondent collects information about their travel behaviour during three consecutive days each year. The data from the travel diaries is collected in the 2 weeks before and after the end of the Dutch school holiday that usually takes place in the middle of October ('autumn-holiday'). No data is collected during the holiday itself, as doing so would distort the results. This means that we have travel behaviour data for the Autumn months only, which is one of the main drawbacks of this data source.

3.1.2 Data Preparation

All of the necessary data for estimating the models is thus collected. A remaining challenge that should not be underestimated is to combine weather data with travel behaviour data. Weather data needs to be matched to individual trips for Chapter 4 and Chapter 6, whilst they need to be matched based on date for Chapter 5. Besides this challenge, all the usual data preparation processes needed for working with such data sets are still also required: data cleaning, creating a comprehensive travel behaviour data set from the various data sets of the MPN within one year and then combine this data set across multiple years, creating a single weather data set, etc.

Below we will give a short explanation of the algorithms used to connect weather data to travel behaviour data in . A more comprehensive overview can be found in Appendix A. Critical tools for the data preparation process are the programming languages Python and R [R Core Team, 2017], respectively mostly using the library pandas [McKinney, 2010] and the tidyverse package [Wickham, 2017].

Matching weather data to travel behaviour data

Weather needs to be matched to travel behaviour based on spatial and temporal dimensions: we want to assign weather to travel behaviour based on the weather that impacted that specific piece of travel behaviour. To make this more concrete, in this thesis weather needs to be matched to two separate travel behaviour units: trips and days. Trips have the distinct advantage that we know where they originate from and what their destination is, whilst for days we only know the residential or work location of respondents. We will quickly explain the algorithm used for both units.

For trips we can use the locations of the trip. Which location needs to be used exactly is the subject of Chapter 4 and we will not discuss this here, so for simplicities sake we will assume we use the location where the trip originates. Due to privacy, we don't know the specific address, but rather the postcode-4 (pc4) area of the origin. These zipcode based areas can be very small, especially in dense urban areas. In rural areas however they can be quite spacious (for a visualisation, see Figure A.3). The spatial challenge is thus to connect trips originating from these pc4 areas to the closest weather station, based on the station's point locations. A complication here is the fact that the pc4 areas are subject to changes across years and although changes are often not very large they can be impactful.

The first step in solving the problem is calculating a point location for each pc4 area. To do so we downloaded the Basisregistratie Adressen en Gebouwen (BAG) dataset from the Dutch Kadaster, which is the land registry, cadastre, and mapping agency of the Netherlands. This data set contains the locations of all addresses and their zip codes. We downloaded the BAG for each year separately, using the November or October data set based on availability. Then for each year separately we calculate the average location of the addresses within one pc4 area and use this location as the pc4 point location going forward.

The second step is to calculate distance matrices from each pc4 location to each weather station. The coordinates are used to calculate Euclidean distances (straight-line distances on a round surface) between each pc4 area and the weather stations. Keep in mind that this process needs to be repeated six times, one time each for each kind of weather station (not all weather stations collect temperature for example, so to assign temperature to a trip we need to find the closest weather station that does collect temperature). These distance matrices can then be used to provide a rank-order of weather stations for each pc4 based on the distance between the two.

With these rank-orders in place, we can start to think about which values we want to read from this weather station's data. Specifically for which time (or time period) we want to connect data to the trip. Again, many possibilities here that are discussed in Chapter 4, so for simplicities sake we'll use the time of departure. We then try to read the weather value of interest from the closest weather station at the time of departure. If this value is recorded we collect it and assign it to the trip. If the value is not recorded, we use the next closest weather station. This process goes on until we have either recorded a value or a pre-specified maximum distance is reached. In this thesis the maximum distance is set to 30 km. This estimate is based on the fact that 30 km is far away enough so that most of the Netherlands is covered by at least one weather station for the most important weather variables, whilst being close enough that the measurements for temperature, wind speed, and sunshine are still valid.

> Data about the weather is connected to travel behaviour data For each trip or day we use a postcode-4 area and select the data from the weather station that is the closest to this area. Different times and locations relevant to the trip or day can be used to select the final data set. We compare these methods in Chapter 4.

The average distance from a trips' location to the weather station that is used to collect the data varies per type of weather, as not all weather stations collect all weather variables. For wind-related variables the distance is the smallest at 12.8 km, whilst the distance is largest for cloud-related variables at 14.1 km. The distribution of the distance is plotted for all six types in Figure 3.2. Please note that these are the distributions for the trips when a connection was made, which means that the maximum distance is 30 km. Of course some trips were located further away from weather stations, leading to unsuccessful connection attempts. Again the percentage of trips that failed to connect to a weather station is different for each type of weather. This failure percentage is lowest for wind data, being as low as 3%. For temperature, rain, and sun the failure rate is close to 5%, drastically increasing to 30% for cloud data and 40% for air data.

Data Cleaning

To get valid results from our statistical models, we need high-quality data as input. We thus need to recognise possible mistakes in the data and take actions to remedy them. A difficulty here is that the pre-existing data sources have already been handled to varying extents. A further complication is that the data cleaning and preparation processes are distinctly different for each of the models run and thus for each of the results chapters. Here we will give the data cleaning processes that are shared amongst chapters.

With respect to the weather data we have performed no significant cleaning efforts under the assumption that recorded data would be accurate at most times. Missing data fields were dealt with by trying to find the data in the next weather station. If there are no other weather stations within the

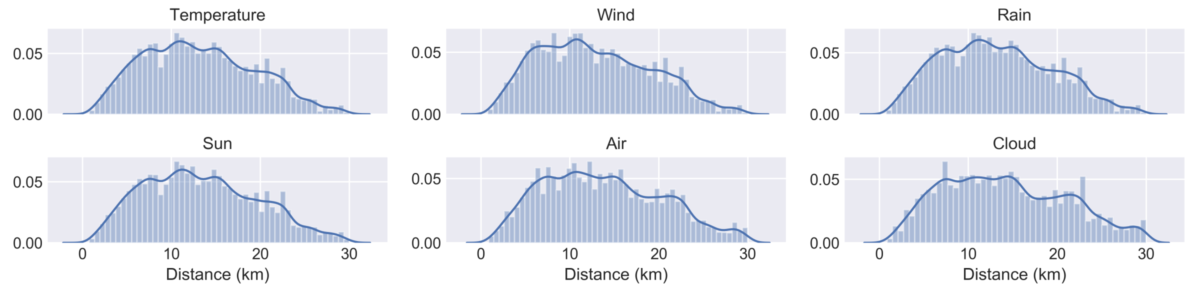


Figure 3.2: Distribution of distances from trips to weatherstations when a succesful connection is made.

30 km distance, then the trip was removed from the analysis. We could've remedied some of these missing data fields by interpolating from surrounding fields or correcting with other data sets from the same weather station. Even if this procedure is done flawlessly however there is no way of being absolutely sure of the semantic accuracy of these new values. Since we have more than enough data to provide reliable estimates we decided the trade-off between gaining completeness and losing accuracy was not worth the effort.

In contrast to the relatively raw state of the weather data, the travel behaviour data has been processed quite a bit already. Some trips' travel times were for example either added or modified later. Luckily these trips were marked, and all of them were removed from the analysis as matching with weather is based on precise travel times. Furthermore any trips where the origin or destination location was unknown were removed from the analysis, as again this location is used to determine the values for weather. We took a couple of extra measures, again valuing accuracy over completeness. Since our weather data is only available for the Netherlands we removed all trips that either departed from or arrived at a foreign location. Any remaining trips with a reported distance of greater than 300 km were also removed, with this number being determined by the maximum distance one can realistically travel within the country of the Netherlands which is roughly 300 km.

The last data preparation stage was estimating the choice set for each trip. This is a very complex issue, both theoretically and in practice and deserves its own section, which can be found below at Section 3.1.3.

3.1.3 Choice Set

A point of attention that falls within the data preparation is the choice set that is used to estimate the choice models. Choice models assume that a rational process underlies the decision making process based on which an alternative is chosen. This rational process weighs several options against each other. All options (often called alternatives) are part of a pre-specified choice set. The choice is thus made from alternatives within a choice-set that is both mutually exclusive (no two choices can be made at the same time) and collectively exhaustive (one of the alternatives must be chosen) [Ben-Akiva et al., 1985]. It should also reflect the underlying process as closely as possible, and thus exclude alternatives that were never considered in the first place. In the context of transportation mode choice, the choice set then consists of all the travel modes that could have been chosen for a certain trip.

Choice Sets in Revealed Preference data

A well-known problem when using revealed preference data to estimate choice models revolves around the fact that choice sets by their nature are mental processes, and are thus latent: it is impossible to determine the actual considered choice set using this type of survey data [Ben-Akiva et al., 1985]. This makes it impossible or very hard to retro-actively determine the choice set by querying the respondent for the alternatives that were considered.

This problem necessitates the approximation of the true considered choice set [Ton et al., 2018]. This choice set can often be approximated using so-called constraints, where proxies are used to determine whether or not an alternative was part of the choice set. Aside from using no constraints at all, three types of constraints can be identified: 1) spatial constraints, where only trips within a certain distance are used to estimate the model to ensure all modes could realistically be used for the trip; 2) deterministic constraints, which are rule-based constaints based on individual or trip characteristics; and 3) probabilistic constraints, which are constraints based on individual or trip characteristics that vary across individuals (in contrast to deterministic constraints, which simply apply the same rule to all individuals). This probabilistic nature allows one to include variables whose impact on choice set consideration is not universal, such as socio-demographics. An example where a probabilistic approach would improve on deterministic rules is the following: students are less likely to consider the car. Some students however will still travel by car, making it impossible to set a deterministic rule (If someone is a student, he/she does not consider the car). A probabilistic approach would enable inclusion of whether or not someone is a student to adjust the probability by which he/she considers the car. These different methods for specifying the choice set and the definitions of the associated choice set are shown in Figure 3.3.

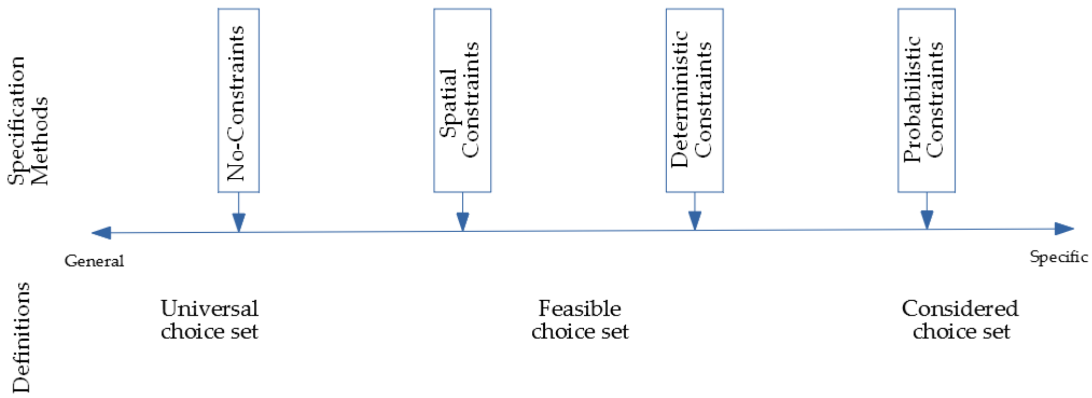


Figure 3.3: Mode choice set specification methods and their definitions, adapted from [Ton et al., 2019, p. 3]

As can be seen in Figure 3.3, the starting point is the universal choice-set, where all alternatives are available to all individuals for all trips. Spatial, deterministic, and probabilistic constraints are each getting closer to approximating the true considered choice-set, with the feasible choice-set being somewhere in-between these two extremes. A feasible choice-set is based on the feasibility of using a certain mode for the trip in question, based around characteristics of the trip and of the individual making the trip.

This research will make use of deterministic constraints, due to the fact that probabilistic constraints are both very difficult to incorporate into a running Latent Class model and its limited benefit compared to the deterministic choice-set.

Within the deterministic constraints, two different types of rules can be identified: rules based on availability of a travel mode, and rules based on consideration of a travel mode [Calastri et al., 2017]. Rules based on availability are made based on individual characteristics, such as the ownership of- or access to a car. Consideration rules however are based on trip characteristics, such as the distance of a trip or whether the trip can feasibly be made using public transport. Below the rules used in this research will be introduced, together with descriptive statistics and visualizations.

> A choice set is a collection of all alternatives from which a choice has been made. Here this entails all travel modes which could have been used by a traveller for their trip. We have no direct information about which modes were part of the choice set, so we need to approximate the choice-set using general rules. There are two types of rules. The first type is based on the availability of a travel mode. If people don't have a bicycle, they will not be able to travel using this mode. The second type is the consideration of a travel mode. For trips that are 100km long, people will not decide to go walking.

Availability

As said before, availability rules are based around whether or not a person has access to a travel mode and/or is able to use this travel mode. The rules will be specified per travel mode and applied to all individuals.

The first rules are based around access to specific travel modes. For walking no constraints are put in place, with the assumption that this mode is available to everyone. It is possible that for elderly or (temporarily) disabled people this mode is actually unavailable due to health concerns, but because of the limited size of this potential group and the difficulty of making hard rules that exclude the group this is not accounted for. Public Transport is also assumed to be available for everyone, as the option to buy a one-time ticket is always available. It is possible to define rules based on the ownership of public transport cards or subscriptions, but neither is able to fully determine that public transport is or is not available.

The availability rules do apply to both the car and bicycle modes, who share a similar rule: car/bicycle are only available if a person has access to (some type of) automobile or bicycle. A loose definition of 'Having access' is used: as long as there is a car or bicycle of some sort (including electric cars, electric bicycles, and folding bikes) in the household, people are said to have access to this specific mode. Stricter rules (for example determining that the person should own the mode his/herself) lead to a greater mis-specification, as both modes are used for travel quite often even when they are not owned by the travelling person him/herself.

The combination of these rules lead to a division of trips into trips where certain modes either were or were not available. The mode shares of trips under these various conditions are given in Table 3.1.

Table 3.1: Mode shares for different availability criteria

		Mode Share (%)			
Condition	N	Car	Transit	Bicycle	Walk
All trips	161 491	50	5	29	16
Car available	141 629	55	4	26	15
No car available	19 862	13	13	47	27
Bike Available	155 076	49	5	30	16
No Bike Available	6 415	58	7	8	27

Both availability criteria perform in a similar fashion: whilst they reduce the mode share of the mode in question, this mode share is not reduced down to (near-)zero. For the car it might for example be possible that people still travel with the car as a passenger, whilst bicycles might be borrowed from family/friends for short trips. Another option is that car- or bikesharing programs (like OV-fiets) are used. Table 3.1 clearly shows that these rules are not perfect, but a clear difference between available and non-available groups in mode share can be seen as well. For this reason the rules are applied during this research.

> We use two availability rules, one for the car and one for the bicycle. If a traveller does not have access to these modes (as indicated by the traveller him-/herself), then we say that they are not able to travel with the mode in question.

Consideration

Consideration is a trip-level characteristic, where the characteristics of the trip (purpose, distance, times) can be used to design rules that are able to exclude certain travel modes from the considered choice set for this trip. In other words one can say that the choice set is reduced, given the characteristics of a certain trip.

For the active modes, the most important characteristic is the distance of the trip. Long-distance trips are in general not travelled using either the bike or as a pedestrian. To set a boundary-value, the 99th percentile distance of trips travelled by bike or on foot are used. The distribution of trip distance using these modes is given in Figure 3.4, together with the 99th percentile value. To give some margin for error, the 99th percentile values are rounded up to the nearest 5 km value. The final boundary values are thus 10 km for pedestrian trips and 20 km for bicycling trips.

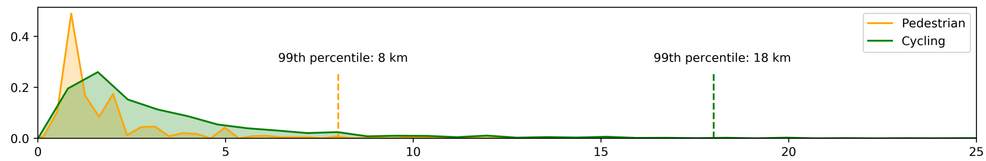


Figure 3.4: Distribution and 99th percentile of distances travelled using active modes

It is possible that there is an effective minimum distance for public transport and car trips. For this reason the distribution of trips is also analysed and visualized for these modes. The visualisation can be seen in Figure 3.5 and clearly shows that there is no effective minimum distance for either car or public transport. No distance-based rule is thus imposed on these two travel modes.

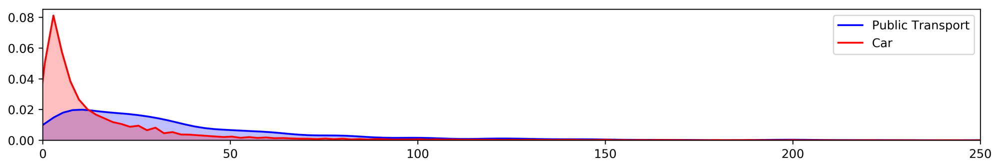


Figure 3.5: Distribution of distances travelled using car and public transport

Whilst this means that the car is assumed to always be considered, consideration of public transport depends on the existence and viability of trips from the origin to the destination at the time of the trip. Especially in the more rural parts of the Netherlands it is very possible that trips with public transport either don't exist or will take much longer than competing modes. Since there is no direct information about the existence or characteristics of public transport trips when other modes have been chosen this information needs to be approximated. This is done by GoudAppel for the first four waves of the MPN. From wave 5 onwards this information is approximated using the Google API. In both cases this provides information regarding the existence of a transit option for the trip and how much time this option would take. There is also information about the travel time of the same trip when a car is used instead. Using these estimated characteristics rules can be applied to exclude PT availability from certain trips. This is straightforward when the information indicates that no trip was available at all, as in these cases public transport is simply not included in the choice set. There is one detail that needs clarification: if there are trips available according to the API, but these trips are fully made on foot, then the same rule applies and public transport is excluded from the choice set.

The problem becomes slightly subjective when trips are available, but the duration might be unreasonable when compared to competing modes. Keep in mind that we're using deterministic rules that will be applied to everyone, whilst there might be considerable heterogeneity within the population regarding this issue. We deal with this heterogeneity by identifying individuals that might be 'transit-captives': individuals for whom competing modes are unavailable are unlikely to care about comparisons with these competing modes. For these individuals, transit is always considered provided a trip is actually possible.

For people who do own competing modes relatively lax rules are applied: PT is excluded from the choice set based on either its relative performance vs. the car or its absolute performance vs. the car. Comparisons with bicycles are not explicitly made, under the assumption that car travel is not slower than bicycle trips in almost all circumstances (although the bicycle can certainly be the more practical option). Absolute performance refers to the difference between trip duration using public transport and using the car. If this difference is greater than 90 minutes (so PT takes more than 1,5 hour longer than the car), PT is not included in the choice set. The relative performance is the trip duration of public transport divided by the trip duration when one would use the car. If this value exceeds 3 (so: if public transport takes more than 3 times as long than the car) then public transport is excluded from the choice-set. An exception for the very short trips is put in place, as the minimum absolute difference must be 10 minutes. These rules are specific to the context of the Netherlands, where no trips cover more than 300 kilometres and the public transport network is relatively advanced. In other more spacious countries the absolute 90 min difference might be (far) too little, as trips can take much more time to complete.

The mode shares under the varying consideration criteria are displayed in Table 3.2.

Table 3.2: Mode shares under various consideration criteria

		Mode Share (%)			
Condition	N	Car	Transit	Bicycle	Walk
All trips	161 491	50	5	29	16
Bike considered	137 642	45	2	34	19
Not considered	23 849	79	20	1	0
Walk considered	121 719	40	2	37	21
Not considered	39 772	79	16	5	0
PT considered	125 736	44	6	32	18
Not considered	35 755	71	1	20	8

Given the data-driven nature of the bicycle and pedestrian rules it is not a surprise that the mode share of these modes does drop down to near-zero when the respective modes are not being considered. The rules for public-transport are however devised in a similar fashion to the availability rules: by trying to approximate what reasonable choices would be. The results for this criterion are - also unsurprisingly - then very similar to those of the availability criteria, as PT share shows a marked decrease when PT is unavailable, but does not seem to tend to zero. Again the rule does show clear discriminatory ability, which is why it is used to determine the final deterministic choice-set.

> We use three consideration rules. For bicycling and walking we use a maximum reasonable distance at which these modes are used. For walking this is 10km, for cycling this is 20km. Finally public transport is not part of the choice set if a trip between the origin and destination is either not serviced by public transport or if its duration is very long in relation to alternative options for the traveller.

Home-based trips

The final choice set rule is not mode-specific, but has a drastic effect on the number of trips that are used in the final model. In short, this rule is that trips are only used in the model if they depart from the residential location of the respondents. The rationale behind this rule is that this is the place where all modes are both available (as in, where the cars are parked, bicycles are stored and PT subscription card is) and considered, as people will generally make a return journey with the same mode that they used to travel to their initial destination. When someone leaves their home using their car, it is very likely that they will return home with the car as well. In other words the assumption is that the actual mode choice is made when people depart their home, rather than when they leave for each new trip. If this assumption holds, estimating a choice model on all trips might lead to faulty estimates. However one loses a considerable amount of trips (both in terms of absolute numbers and relative to the overall number of trips) when only using trips that depart from the residence. If the above assumption doesn't hold then, it is very important not to implement the rule, as that will bias results and lead to lower standard-errors than necessary.

The key question thus becomes whether or not the assumption that the mode choice is made when people depart their home holds. To answer this question we look at how often respondents switch modes after they have left their residential location. This information is displayed in Table 3.3. The N/A values occur when it was impossible to match a trip with the mode that was chosen when respondents left their home. This occurs when the first trip of a day departs from another place than the residence. This happened for 12K trips, which is less than 10% of the total amount of trips.

Table 3.3: Mode shares compared to mode choice when people left their house

	Mode Share (%)			
Mode from Home	Car	Transit	Bicycle	Walk
Car	97	0	1	2
Transit	4	85	3	8
Bicycle	1	1	96	2
Walk	3	1	2	94
N/A	54	8	24	14

As can be seen in Table 3.3, more than 90% of the trips starting at non-residential locations use the same mode that was used to leave the residential location, which indicates that the assumption holds true for the vast majority of trips. This is reason to remove all trips that did not start at the residential location from the data set, as the actual decision to use a certain mode seems to be made when leaving the residence. This entails a drastic reduction of the size of the data-set, as a further 92K trips are removed (more than half of all trips). 68 thousand trips remain however, which should be more than enough to gather reliable and valid estimates.

We assume that a rational decision is only made at the residential location of the traveller. Trips departing from other locations are in almost all cases made using the same mode that was used to depart from the residential location earlier the same day.

Final Choice Set

The combination of the consideration and availability criteria leads to the final determination of the choice-set for each trip. The choice set should consist of a mutually exclusive and collectively exhaustive set of travel modes [Ben-Akiva et al., 1985], which means that a respondent can only use one mode for one trip and that a respondent needs to use a mode in the choice set. To evaluate to which extent the final choice set resulting from the deterministic procedure outlined above conforms to this ideal the observed mode shares of the four travel modes are calculated for each possible choice set. This information is displayed in Table 3.4. The choice sets are depicted using the first letter of modes that are included in the choice set. Choice set C thus only includes the car, whilst choice set TBW includes transit, bicycling, and walking. Some choice sets from the 16 possible combinations are missing, as these combinations did not exist.

Table 3.4: Mode shares under various choice-sets

Choice Set	N	Mode Share (%)			
		Car	Transit	Bike	Walk
C	1 197	96	3	1	0
T	1 012	28	68	4	0
CT	8 014	80	18	2	0
CW	186	65	1	4	31
CB	1 238	90	3	7	0
TW	805	22	9	17	51
TB	338	24	49	27	1
CTB	4 660	76	9	14	1
CTW	1 041	58	2	6	34
CBW	8 129	55	0	28	16
TBW	6 042	5	4	58	32
All	36 022	37	1	37	25

As said before, in the ideal case a choice set would be collectively exhaustive. This means that mode shares of the modes in the choice set should sum up to 100%. Another way to conceptualise the same principle: if a mode is not in the choice set, it should not be chosen. A quick scan of Table 3.4 shows that this condition does not hold: especially Car and Bicycle are often chosen even when these modes are not part of the choice-set. At the same time, most choice-sets show a discriminatory ability: the probability that an alternative is chosen is much higher if it is included in a choice set. The question thus becomes if the current rules are good enough to approximate true considered choice sets. Whilst the violation of the requirement of being collectively exhaustive can be mended easily (by adding a mode to the choice-set if it is chosen), this solution is applied on an ad-hoc basis: it only corrects an error if the mode is actually chosen, not if it is considered, but not chosen. Of course perfection is not achievable with deterministic rules for the reasons set out at the start of this chapter: choice sets are inherently latent variables that can only be approximated.

One way of estimating the error is determining how often a mode is chosen when it was not part of the choice set and compare this to the total number of times that a mode was chosen. This gives a measure of the relative frequency with which modes were excluded from the choice set when they should not have been. Table 3.5 displays this information: the total number of trips, number of trips when part of the choice set and number of trips when not part of the choice set are given, as well as the percentage of times that the chosen mode was not included in the choice set. It is important to keep in mind that this is only one of the two errors that might occur, the other being that an alternative is included in a choice set even when it's not considered.

Table 3.5: Percentage of times a mode is chosen, split per its inclusion in the choice set

Mode	Total	From choice set	Outside choice set	% Outside choice set
Car	31 798	30 926	826	2.74
Transit	3 554	3 455	99	2.79
Bike	20 245	19 844	401	1.98
Walk	13 087	13 043	44	0.34

The table indicates that the vast majority of times a chosen mode was in fact part of the choice set, with errors being smaller than 3% across the board. The errors for the active modes are noticeably smaller than the errors of public transport and car, which can be explained by the fact that the data set is used to determine the consideration rule for the active modes.

3.1.4 Data Description

In this section the data used to estimate the statistical models will be described and visualised, partly to get some idea about the underlying distributions and ranges and partly to check whether or not the values have changed drastically as a result of the reduction in size caused by using the choice set criteria. These descriptives for travel behaviour data thus use the choice-set, which means that for example only home-based trips are included.

The first kind of data is the travel behaviour data. Table 3.6 shows the mode shares, distances, and purpose shares of the trips in the choice set split out per year.

Table 3.6: Travel behaviour descriptives across the five waves of the MPN

		Mode Share (%)				Distance (km)		Purpose Share (%)		
Year	# trips	Car	Transit	Bicycle	Walk	Mean	St.D.	Work	Education	Leisure
2013	11 146	49	5	28	19	10.5	21.9	20	6	74
2014	15 002	44	5	32	19	10.5	21.8	20	8	72
2015	11 976	45	6	29	18	11.5	23.2	22	7	71
2016	13 061	47	6	29	18	12.1	23.3	23	6	70
2017	17 499	47	5	27	21	11.0	22.3	20	5	75

Some yearly changes can be seen in Table 3.6, although the cause of these changes is unknown (and might simply be due to sampling inconsistencies!). All changes are relatively small however, pointing to the lack of clear trend-breaks or drastic shifts during the years of collection which enables the pooling of all waves.

Figure 3.6 shows how the main weather variables develop during a day, by displaying mean values for the 144 ten minute intervals of all days from september to november (the period when MPN data is collected). A couple of interesting points can be made, the first being that four of the six kinds show the same trend: wind, temperature, rain, and optical range all show a similar pattern where the values peak in the afternoon and then slowly decline until the following morning before climbing again. The pattern of Wind, temperature, and optical range follow this pattern very well, whilst rain seems to show some small deviations (with more rain in the early morning). All of these variables follow from the radiation of the sun, which peaks slightly earlier. This solar radiation causes temperatures to rise, which in turn has an effect on the atmosphere that increases wind speed and rain fall

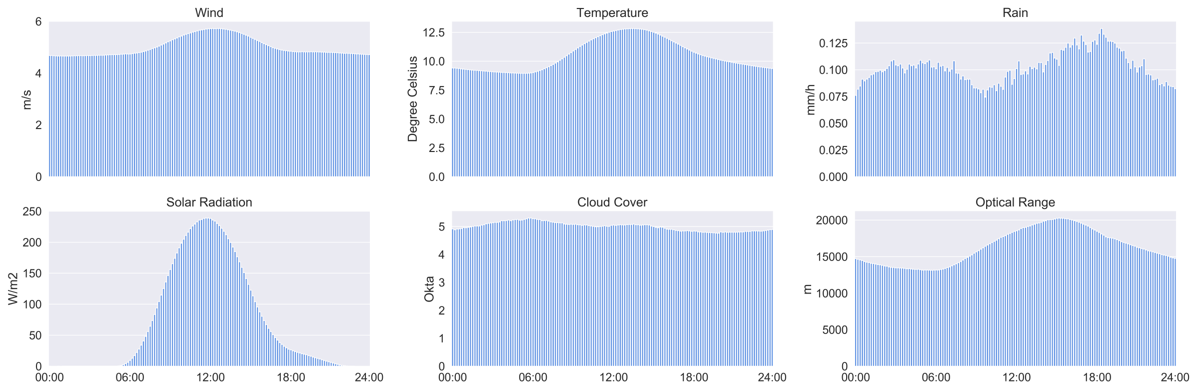


Figure 3.6: Mean values of main weather variables for different times of day in the months september through november

3.2 MODELS

When both the data combination and the selection of the variables are finished the models can be specified. As said before in the introduction a regression model is used to estimate trip generation and a choice model is used to estimate mode choice. Both modelling techniques and the tools needed to perform them are described in the section below.

3.2.1 Regression Models

Since the main dependent variable of the regression model is a count variable, ordinary least squares (OLS) regressions have serious drawbacks: the assumption of constant variance and normal distribution of the error terms are particularly unlikely to hold [Coxe et al., 2009]. For this reason Poisson regression models are often used as an improvement when estimating count variables. Rather than predicting the counts themselves, Poisson regressions estimate the natural log of the counts. Furthermore, the underlying distribution of the error term is assumed to be a Poisson distribution, rather than a normal distribution [Gardner et al., 1995]. This distribution is discrete (rather than the continuous normal distribution) and its variance is equal to its mean [Coxe et al., 2009]. The probability density function for the Poisson distribution then can be seen in Equation 3.1.

$$P(Y = y|\mu) = \frac{\mu^y}{y!}e^{-\mu} \tag{3.1}$$

Within a Poisson regression, the variance is assumed to be equal to the estimated mean (the mean conditional on the predictors), as in Equation 3.2

$$var(Y) = \bar{Y} \tag{3.2}$$

The mean and variance are thus estimated simultaneously, usually through a natural log link function as in Equation 3.3

$$ln(\bar{Y}) = b_0 + \sum_m b_m * x_m \tag{3.3}$$

where $\bar{Y}$ is the estimated variable, based on scores X and weights b of all m attributes and an intercept b_0.

However, it's possible that the actual variance is not equal to the mean: if this is the case the variance is usually larger, a condition known as overdispersion. There are multiple solutions to this problem, but in this thesis we have used negative binomial regressions, which can be seen as a generalization of the Poisson regression as they use an extra parameter to estimate the variance [Ver Hoef and Boveng, 2007]. The variance is often estimated as follows:

$$var(Y) = \bar{Y} + k\bar{Y}^2 \tag{3.4}$$

where $k \geq 0$. Since the Poisson regression is more parsimonious it is good practice to first estimate a Poisson regression and only estimate a Negative Binomial regression if the results indicate that there is overdispersion.

3.2.2 Choice Models

The quantitative method for determining mode choice is the discrete choice model. A choice model postulates that the decision to travel using a certain mode is the result of a rational process on the part of the decision maker. The choice model tries to approximate this underlying rational process, with the aim to gain a better understanding of how and why people made certain choices. Discrete choice models are models for choice sets that are exhaustive and whose alternatives are mutually exclusive [Ben-Akiva et al., 1985]. The general idea and process behind these discrete choice models will be explained using the linear-additive random utility maximization model in Section 3.2.2. Afterwards latent class models will be introduced and explained in Equation 3.2.2. The MNL specification will be used in Chapter 4, as it is simple to specify, quick to estimate and easy to interpret, making it ideal for estimating and comparing many models. The Latent Class specification will be used in Chapter 6, as it is able to uncover heterogeneities in the population and is able to provide better estimates for the effect of weather on mode choice by relaxing some of the assumptions made by the MNL model.

Multi-Nomial Logit model

This model assumes that decision makers associate each of the alternatives i with an utility U. This utility is modelled as both systemic utility V and an error term e, which can be conceptualised as the unobserved utility on the part of the modeller. This is denoted in equation 3.5.

$$U_i = V_i + e_i \tag{3.5}$$

The systemic utility that is associated with alternative i is the result of a rational calculation using specified attributes of the alternatives. Examples could be the cost or travel time associated with an alternative. In the linear-additive choice model, the systemic utility is the sum of attribute values X that are weighted by taste parameters β, as can be seen in equation 3.6.

$$V_i = \sum_m \beta_{im} * X_{im} \tag{3.6}$$

The probability that a decision maker chooses alternative i is based on the probability that the total utility for this alternative i is greater than the total utility of all other alternatives in the choice set j. If the error terms of these alternatives are independent and identically distributed (i.i.d.), this probability can be calculated using the canonical Multinomial logit (MNL) models. The equation used to calculate this probability is shown in equation 3.7.

$$P(i) = \frac{exp(V_i)}{\sum_{j=1..J} exp(V_j)} \tag{3.7}$$

Whilst this MNL discrete choice model is intuitive and easy to calculate, it has some drawbacks resulting from two assumptions: the first the assumption that the error terms are i.i.d., as mentioned above. The second potentially problematic assumption is that all decision makers use the same decision rule of utility maximization. These assumptions are relaxed for Chapter 6 by way of using a Latent Class choice model instead of a MNL model. The Latent Class choice model is described in section 3.2.2.

Latent Class Discrete Choice Model

As said before, latent class models are able to relax assumptions underlying the MNL model. This enables latent class models to elegantly capture taste heterogeneity [Hess et al., 2009]. Taste heterogeneity refers to different sensitivities in the population with respect to model attributes (differences in taste parameters β in the population). An example would be that one segment of the population is insensitive to weather, whilst another segment is very sensitive to weather. Latent classes can be used to estimate both decision rules at the same time, thus capturing heterogeneity with respect to the decision rule.

The starting point for the addition of latent class models to the MNL-model outlined above is the notion that the taste parameters β can be conditioned on the membership of discrete classes S, leading to S number of different β parameters. Some parameters might be fixed across classes, for example if there is limited heterogeneity or if the heterogeneity of these parameters is not the focal point of the model.

$$P_{i,n}(\beta_1, ..., \beta_s) = \sum_{s=1}^{S} \pi_{n,s} P_{i,n}(\beta_s) \tag{3.8}$$

where $\pi_{n,s}$ is the probability that individual n belongs to class S, which means that $0 \leq \pi_{n,s} \leq 1$, whilst $\sum_{s=1}^{S} \pi_{n,s} = 1$. This probability is estimated as part of the choice modelling procedure, and by convention also takes the form of a logit function to ensure that the sum of class-membership probabilities add up to 1.

$$\pi_{n,s} = \frac{exp(\delta_s)}{\sum_l^S exp(\delta_l)} \tag{3.9}$$

This gives the discrete mixture model, where the membership probability for the classes is constant across all respondents. This discrete mixture model can be turned into a true Latent Class model by adding predictors to the class-membership function, in effect allowing the membership probability to vary across individuals. These predictors z_n are typically person-level characteristics such as relevant socio-demographics or attitudes. The vector γ_s consists of the weights capturing the influence of the predictors on the class-membership function of each specified class, whilst δ_s is the class specific constant.

$$\pi_{n,s} = \frac{exp(\delta_s + g(\gamma_s, z_n))}{\sum_l^S exp(\delta_l + g(\gamma_l, z_n))} \tag{3.10}$$

We have to account for the panel structure of our data, where we have multiple choices for each individual. Estimating cross-sectional latent class models on panel data results in two errors: first it results in an over-estimation of the information available in the data, meaning that standard errors are lower than they should be. The second problem results from the classes specified in the latent class analysis: in Latent Class models we assign class-membership probabilities to the observations. Conceptually it makes sense that the class-membership probabilities are constant for each individual, rather than for each observation. This is accounted for by estimating a panel latent class model, where the likelihood of observing multiple choices for alternative i across t observations for each decision-maker n can be calculated as follows:

$$L_n(i_t, ..i_T|\beta) = \sum_{s=1}^{S} \pi_{ns}(\prod_{t=1}^{T} P_n(i_t|\beta)) \tag{3.11}$$

We thus calculate the likelihood of observing choices T given the taste parameters of each class and then calculate a sum across classes weighted by the class-membership probability assigned to decision-maker n. To get the likelihood across all observations, we multiply the likelihood calculated using Equation 3.11 across all n individuals, as in the formula given in Equation 3.12.

$$L(\beta) = \sum_{n=1}^{N} \sum_{s=1}^{S} \pi_{ns}(\prod_{t=1}^{T} P_n(i_t|\beta)) \tag{3.12}$$

The natural logarithm of this likelihood is maximised through a maximum likelihood estimation procedure. A difficulty here is that this log-likelihood function is not concave, which means that the maximum likelihood estimation might get stuck in an local optimum. This issue is typically dealt with by trying multiple starting values for the estimated parameters β and comparing their likelihood score. The highest of these estimations is more likely to be the global optimum, although there is no guarantee.

4 INCORPORATION OF WEATHER IN DECISION-MAKING

One of the knowledge gaps emerging from the literature review in Chapter 2 concerns the uncertainty of how weather impacts travel behaviour. This chapter aims to explore a facet of this topic by looking into *which* weather influences travel behaviour and thus how to operationalize weather in the context of travel behaviour. The difficulty here is that both weather and travel behaviour are distributed through spatial and time dimensions: to calculate a singular weather value for each trip is thus no trivial task. The question is first how weather **conceptually** influences travel behaviour and then how this conceptual information can be used in the process of **operationalization**, where the weather phenomena are reduced down to a single value for each day or trip.

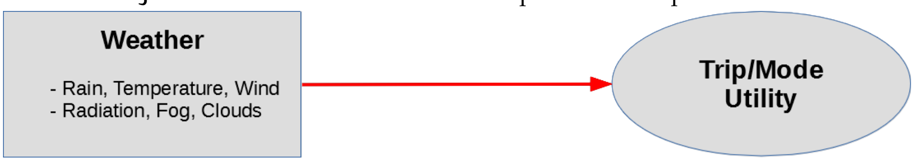
Starting with the conceptualisations then, one can think of many different ways in which people look at the weather and how this influences the trip or mode utility they associate to their travels. One intuitive way is the idea that you simply look out of the window when you are going to leave, observing the weather directly and then basing your decision (whether or not to travel or which mode to use) on that observation. A similar idea is that you observe the weather for some time preceding the trip and use those observations to make your decision. These are not the only possibilities however! Weather forecasts have been around for millennia and these forecasts have gotten more and more precise, especially in recent times. This enables people to base their decisions on forecasts of weather, with varying degrees of assumed precision. One option is simply looking outside and trying to determine whether or not it is likely to rain sometime this day. Alternatively people might use newspaper, television or online reports to give them information about the likely weather circumstances of the day.

More precise forecasts are also possible, where people look ahead at the weather circumstances at the time and location of arrival. The final conceptual idea discussed in this chapter is that where people look at certain times of day where they know they will travel: if forecasts predict rain during my return journey from work that information might affect my mode choice going to work as well.

These conceptual options are operationalized using different algorithms that produce one value for the weather concepts (such as temperature) for each trip. The varying conceptualisations thus result in different operationalizations, and thus in different values for each trips. These so-called metrics will then be compared by estimating models using the different metrics and comparing their ability to explain variation in travel behaviour, or more specifically, mode choice. The goal of this chapter is thus to compare different conceptualizations of how weather influences individual trips.

In the conceptual model, the chapter thus directly deals with the link between weather and trip or mode utility, highlighted in red in the simplified conceptual model below in Figure 4.1.

Figure 4.1: Where weather metrics impact the conceptual model



This chapter aims to provide empirical evidence from a revealed preference study to see which method of connecting weather to travel behaviour most closely resembles the true decision-making process. This is done by specifying multiple MNL models, one for each of the different metrics of weather when connected to travel behaviour. The MNL models all use the same weather variables of precipitation, temperature, and wind speed. However, each model uses different values for these variables due to the use of varying algorithms to connect weather to travel behaviour. The resulting weather data are compared and described in Section 4.1. The models and underlying assumptions are described in Section 4.2. Model results are given in Section 4.3 and conclusions are drawn in Section 4.4.

4.1 DATA DESCRIPTION

In this chapter we will describe conceptualisations that are compared in this research in Section 4.1.1, followed by a comparison of the values for the weather variables between the different conceptualisations in Section 4.1.2.

4.1.1 Conceptualisations and operationalization

Different conceptualisations of connection methods exist in the literature, where broadly speaking three main strands can be identified:

1. People use a snapshot of weather to inform their decision.
2. People use a window of weather to inform their decision.
3. People use daily predictions/values to inform their decision.

A distinction can be made between the first two and the third strand, as the first two are based on the time of the trip within a day, whilst the third is based only on the day of the trip. With the first two strands multiple trips within the same day would thus get different values, whilst the values are the same for the

Within these three main strands, multiple more specific approaches can be identified. A snapshot of weather for example can be made based on both the weather at the place of origin or the weather at the place of destination. Windows might be based on strictly the weather of the past few hours, but they might also include forecasts or more general knowledge about the state of future weather. Daily values can be based around averages for the entire day or on specific daily windows, like the traffic peak-hours. This leads to a total of nine different algorithms, which are shown in Table 4.1.

Table 4.1: Overview of methods of connecting weather to trips

Base unit	Name	Description	Location
Trip	Origin	Weather at time of departure	Origin
	Destination	Weather at time of departure	Destination
	2-hour centre	2-hour average around time of departure	Origin
	1-hour centre	1-hour average around time of departure	Origin
	1-hour before	1-hour average before time of departure	Origin
Day	Daily	Daily average weather	Origin
	Weighted Daily	Weighted daily average weather	Origin
	Morning	Average weather during morning rush-hour	Origin
	Afternoon	Average weather during afternoon rush-hour	Origin

The first two algorithms - origin and destination - are the easiest to explain: for each trip, one selects the location of interest (origin or destination) and time of departure. The weather station that is closest to this location is selected and the data at the time of origin/departure are read. If this data is missing, the next-closest weather station is used. This method most closely approximates the weather conditions at the location and time of departure/arrival.

Only slightly more complex are the algorithms concerning the windowed values: 2-hour centre, 1-hour centre, and 1-hour before. The location is still matched to the closest available weather station. However rolling averages are used instead of singular values. To illustrate with an example: if the time of departure is 10:30 and the connection algorithm is 1-hour centre the values from 10:00, 10:10, 10:20, 10:30, 10:40, 10:50, and 11:00 are used to calculate an average value, which is the final value used. This means that for the 1-hour windows a total of seven values are used to calculate the average. If one or two of these seven values are missing, the average is calculated with the remaining five or six values. If more than two values are missing the next-closest weather station is used. For the two-hour window, the same general procedure is used. If more than 4 values are missing the next closest weather station is used instead: for less than 4 missing values the average is simply calculated using the remaining values.

The previous connection methods use the specific time of the trip to connect weather to travel behaviour. The following four methods use a different approach: the connection is based on the date of the trip rather than a specific time of arrival or departure.

The first is the simple daily average, which is very straightforward: all values in the day (6 per hour * 24 hours per day = 144 values) are used to calculate a numerical mean for that date. The second is the weighted daily average, where weights corresponding to traffic intensity are used to calculate a weighted average. The weights are based on the times of departure and arrival of the trips in the full MPN data set. These distributions are displayed in Figure 4.2. In effect this means that weather during night-time hours is considered almost irrelevant, whilst weather from 7-18h counts is considered to be more important.

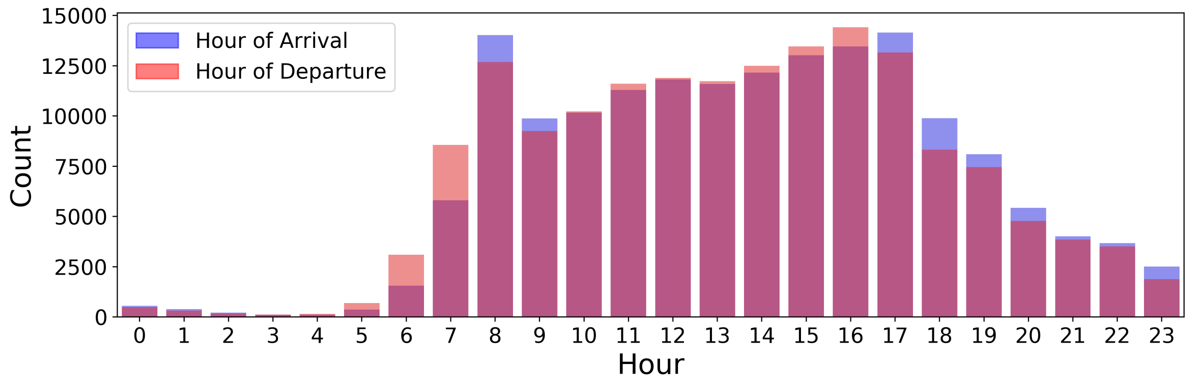


Figure 4.2: Distribution of arrival and departure times

The other two measures are daily windows surrounding the peak-hour times of the travel. These peak-hours are roughly centred around the working hours of 9-17h. This research uses the hours defined by the ANWB [ANWB, 2019a], the Dutch travellers' association. For the morning peak-hours the time-slot between 06.30 : 09.30 and for the afternoon this is the time-slot between 15.30 : 19.00. The values are again calculated as the mean of all 10-min interval values as reported by the KNMI.

4.1.2 Comparison of values

Before estimating models, it is useful to get an understanding of the actual differences between the varying operationalizations. This serves to answer the question whether the conceptually very different algorithms actually produce noticeably different values for the weather variables. To provide this answer, tables are used to give both the descriptives and correlations of the different values for temperature, wind, and rain, as operationalized through all of the algorithms described above in Section 4.1.1.

The first table gives the descriptives and correlations for the rain intensity, as can be seen in Table 4.2. The descriptives contain the mean, standard deviation, and maximum values. The minimum values for rain intensity were 0 across the board.

Table 4.2 shows some expected findings: the snapshot algorithms result in higher maximum values: the longer the time-period, the smaller the maximum value. Afternoons turn out to receive more rainfall on average than mornings, which is in line with expectations in the late-summer months [Verseput, 2009]. Furthermore correlations are relatively low for most combinations of methods. Exceptions are methods which are conceptually very similar, such as the windowed values of 2H centre, 1H centre, and 1H before which share correlation coefficients upwards of 0.8. This might also enable the inclusion of multiple methods in one model, as there are no multicollinearity issues. Especially combinations of methods that are conceptually very different (such as the 'snapshot' of the origin and the weighted

Table 4.2: Descriptives and correlations of rain across the connection methods

	Descriptives			Correlation (pearson's R)								
	Mean	St.D.	Max.	1	2	3	4	5	6	7	8	9
1. Origin	0.086	0.70	60.1	1	0.46	0.58	0.7	0.59	0.22	0.24	0.15	0.15
2. Destination	0.087	0.74	65.6		1	0.47	0.54	0.3	0.19	0.22	0.12	0.15
3. 2h centre	0.085	0.34	13.4			1	0.88	0.84	0.43	0.47	0.31	0.31
4. 1h centre	0.086	0.42	19.0				1	0.81	0.35	0.39	0.26	0.25
5. 1h before	0.083	0.41	25.5					1	0.35	0.38	0.27	0.25
6. Daily	0.093	0.17	1.56						1	0.84	0.5	0.56
7. W. Daily	0.096	0.19	1.96							1	0.38	0.79
8. Morning	0.061	0.23	4.31								1	0.05
9. Afternoon	0.142	0.47	8.22									1

daily average, whose correlation is 0.24) might be effective at capturing a multi-dimensional effect of rain on travel behaviour, for example when someone's decision is both influenced by rain at the point of origin, but also by the vaguer concept of whether or not the day overall is rainy.

The descriptives and correlations for temperature and wind are given in Table 4.3 and Table 4.4 respectively. The descriptives show similar patterns as could be seen inTable 4.2 for the rain variables: windowed or daily averages show smaller ranges, evidence of the smoothing that results from using mean values instead of single values. Some other expected, but noteworthy results are the higher temperatures for afternoon values compared to morning values and the fact that single-point values are closer to those higher afternoon values rather than the lower morning values.

The correlations however contrast with the values found for rain: they are in general much, much higher (ranging roughly from 0.85 and 0.8 upwards for temperature and wind respectively). This means that using more than one method in one model will lead to problems due to multicollinearity, effectively making it impossible to do so. For temperature and wind then one method needs to be chosen, whilst for rain some methods might be combined in a single model.

Table 4.3: Descriptives and correlations of temperature across the connection methods

	Descriptives				Correlation (pearson's R)								
	Mean	St.D.	Min.	Max.	1	2	3	4	5	6	7	8	9
1. Origin	11.3	4.24	-3.9	24.2	1	0.99	0.99	0.99	0.99	0.89	0.89	0.86	0.87
2. Destination	11.4	4.2	-4.3	24.2		1	0.99	0.99	0.97	0.89	0.9	0.86	0.88
3. 2h centre	11.3	4.2	-4.0	23.8			1	1	0.99	0.89	0.9	0.86	0.87
4. 1h centre	11.3	4.27	-4.14	24.0				1	0.99	0.89	0.89	0.86	0.87
5. 1h before	11.2	4.26	-3.89	23.9					1	0.88	0.88	0.85	0.86
6. Daily	10.6	3.64	0.01	19.3						1	0.98	0.97	0.96
7. W. Daily	11.5	3.78	0.98	20.9							1	0.93	0.98
8. Morning	10.1	3.89	-2.01	18.9								1	0.88
9. Afternoon	11.3	3.92	-1.14	21.2									1

Table 4.4: Descriptives and correlations of wind across the connection methods

	Descriptives				Correlation (pearson's R)								
	Mean	St. D.	Min.	Max.	1	2	3	4	5	6	7	8	9
Origin	4.03	2.4	0	22.7	1	0.92	0.99	0.99	0.97	0.86	0.87	0.82	0.79
Destination	4.05	2.39	0	26.9		1	0.92	0.92	0.9	0.81	0.83	0.78	0.83
2-hour centre	4.03	2.32	0.013	21.7			1	0.99	0.99	0.89	0.9	0.85	0.86
1-hour centre	4.04	2.35	0	22.1				1	0.98	0.88	0.89	0.84	0.81
1-hour before	4.01	2.35	0	21.6					1	0.88	0.88	0.85	0.8
Daily	3.76	1.99	0.22	18.1						1	0.98	0.92	0.9
Weighted Daily	4.03	2.07	0.28	18.5							1	0.9	0.94
Morning	3.82	2.27	0.09	22.7								1	0.74
Afternoon	3.63	2.17	0.19	16.9									1

4.2 MODEL DESCRIPTION

To test the different connection methods a simple MNL model structure is used. This model is estimated using different methods of connecting weather data to travel behaviour data. Since the aim of this part of the research is to enable comparisons between weather metrics rather than identifying the pure effect of weather on travel behaviour, these models are set-up in a relatively simple and straightforward manner. The simple model specification allows for a straightforward comparison between models, as the differences between models are captured by relatively few parameters that are very interpretable

The model consists of an alternative specific constant, which gives differences in utility when all other variables are zero, a distance parameter, and weather parameters for temperature, wind, and rain. All of these parameters are estimated for each mode separately, with the exception of one mode which serves as the reference mode. Since utility is latent all parameters are estimated relative to this reference mode. The reference mode in the models is the car, which means that its utility is always set to 0 and no parameters are estimated for this mode. The utility function of all other modes thus consists of five parameters: the alternative specific constants and parameters for distance, temperature, wind and rain. A total of 15 parameters are thus estimated per model. The one exception is the no weather model, which does not contain any weather parameters and thus only estimates six parameters.

The MNL models are estimated in the apollo choice modelling package for R [Hess and Palma, 2019]. The code for the estimation process can be found on my GitHub.

4.3 RESULTS

The results from estimating the 10 different models are displayed in Table 4.5. Since the N and Null Log-Likelihood (LL_0) are constant across models, these values are given once. The differences between the Log-Likelihoods of the models and the non-weather model are calculated and given as well. Statistical significance of the parameters is signalled by **bolding** the parameters that are significant at the 5% level.

The quantitative estimates described in Table 4.5 can roughly be divided in two parts: the model-fit descriptives and the parameter estimates. The two sections below (Section 4.3.1 and Section 4.3.2 respectively) will discuss each of these estimates separately. Finally another test is run to research the differences between the models: all models are used to predict choices for Wave 6 of the MPN, the newest wave that is not used during the estimation process. Seeing which model is able to best predict choices for this wave is a form of testing the external validity of the models and is discussed in Section 4.3.3.

Table 4.5: Comparison of models using standardized variables

	No Weather	Origin	Destin.	Daily	W. Daily	2H Centre	1H Centre	1H Before	Morning	Afternoon
Alternative Specific Constants										
PT	**-2.492**	**-2.511**	**-2.508**	**-2.503**	**-2.502**	**-2.512**	**-2.511**	**-2.514**	**-2.502**	**-2.501**
Bike	**-1.349**	**-1.353**	**-1.351**	**-1.359**	**-1.360**	**-1.356**	**-1.355**	**-1.356**	**-1.357**	**-1.357**
Walk	**-7.180**	**-7.180**	**-7.181**	**-7.180**	**-7.181**	**-7.181**	**-7.182**	**-7.179**	**-7.181**	**-7.180**
Distance										
PT	**0.343**	**0.341**	**0.346**	**0.347**	**0.346**	**0.341**	**0.341**	**0.340**	**0.345**	**0.345**
Bike	**-3.629**	**-3.629**	**-3.626**	**-3.651**	**-3.653**	**-3.636**	**-3.633**	**-3.638**	**-3.645**	**-3.646**
Walk	**-16.688**	**-16.683**	**-16.687**	**-16.695**	**-16.696**	**-16.687**	**-16.687**	**-16.680**	**-16.694**	**-16.692**
Temperature										
PT		**-0.155**	**-0.079**	-0.025	-0.025	**-0.160**	**-0.157**	**-0.180**	-0.027	-0.021
Bike		**0.093**	**0.114**	**0.136**	**0.147**	**0.090**	**0.093**	**0.075**	**0.117**	**0.149**
Walk		**0.087**	**0.095**	**0.131**	**0.138**	**0.086**	**0.087**	**0.089**	**0.120**	**0.138**
Wind										
PT		**0.082**	**0.103**	**0.113**	**0.099**	**0.089**	**0.085**	**0.075**	**0.086**	**0.084**
Bike		**-0.118**	**-0.108**	**-0.104**	**-0.100**	**-0.115**	**-0.118**	**-0.124**	**-0.123**	**-0.117**
Walk		**-0.102**	**-0.098**	**-0.080**	**-0.071**	**-0.100**	**-0.103**	**-0.100**	**-0.090**	**-0.077**
Rain										
PT		-0.017	**-0.157**	**-0.085**	**-0.063**	**-0.082**	-0.045	**-0.076**	**-0.121**	0.014
Bike		**-0.074**	**-0.074**	**-0.137**	**-0.127**	**-0.119**	**-0.097**	**-0.116**	**-0.109**	**-0.053**
Walk		**-0.027**	0.013	**-0.064**	**-0.047**	**-0.044**	**-0.039**	**-0.055**	**-0.032**	-0.016
Model Descriptives										
N	9303									
LL_0	-78250									
LL_β	-57896	-57744	-57743	-57722	-57649	-57700	-57722	-57703	-57707	-57713
$\delta(LL_\beta)$	-	152	153	174	247	196	174	193	189	183
r^2	0.260	0.262	0.262	0.262	0.263	0.262	0.262	0.262	0.262	0.262

Bold parameters are significant at the 5% level.

4.3.1 Model-fit results

Using the values related to model-fit two conclusions stand out: first that the inclusion of weather doesn't entail a drastic improvement of model-fit, as the pseudo-r^2 value increases by a maximum of 0.003 (comparing the 0.263 of the best weather-model with 0.260 of the model not incorporating weather parameters). Log-likelihood increases by 152-247 points. To compare whether this increase is due to chance, we can use the Likelihood Ratio Test. The Likelihood Ratio Test can be used for nested models, as is the case here: the non-weather model is extended by each of the nine models using different weather metrics. Since the added parameters will always cause an increase of the log-likelihood, the Likelihood Ratio Test can be used to determine whether this improvement could reasonably be due to peculiarities in the sample (thus making the smaller model, in this case the no weather model, the best model for the population) King [1998]. The Likelihood Ratio Test uses the Likelihood Ratio Statistic (LRS), which is chi-square distributed with q degrees of freedom, where q is the number of extra parameters estimated in the bigger model (here: 9 parameters). The LRS is calculated using Equation 4.1.

$$LRS = -2 * (LL_{no\_weather} - LL_{weather}) \tag{4.1}$$

To test whether the LRS is due to chance a χ^2 table is used, which gives the threshold value for significance levels at q degrees of freedom. This threshold value for the 0.1% significance level is 27.877, which is exceeded by all metrics by a substantial margin. The main conclusion here is then that including weather in the choice model does increase the ability of the model to explain mode choices, although the difference might be smaller than expected (judging by the pseudo-ρ^2).

Comparison of model-fit shows that all of the models using weather variables are statistically better at explaining mode shares than the model without weather variables.

The second conclusion concerns the goal of this chapter, which was to assess which of the conceptualizations (and corresponding operationalizations) would best be able to explain travel behaviour. Here again the Log-Likelihood and pseudo-ρ^2 of the nine different models are used to provide a comparison. Using the pseudo-ρ^2, we can see that there is no earth-shattering difference between the models, with a maximum increase of 0.001 point (from 0.262 to 0.263). The biggest difference in log-likelihood are found between the Origin and Weighted Daily models, with the former performing the worst and the latter the best, with a difference of 95 log-likelihood points between the two.

Notably this improvement does not come at the cost of estimating more parameters and the models are not nested versions of one-another. The Likelihood Ratio Test thus can not be used to determine whether or not such an improvement is due to chance/sampling peculiarities. To compare non-nested models different tests can be used. Here I use the Ben-Akiva & Swait test [Ben-Akiva and Swait, 2008], which generates an upper limit for the probability that the worse performing model is still the model that truly underlies mode choice decisions [Chorus, 2012, p. 55]. The equation for this test is given in Equation 4.2.

$$p = NormSDistr(-\sqrt{2 * N * ln(j) * (LL(B) - LL(A))/LL(0)}) \tag{4.2}$$

where N is the number of observations, j is the number of alternatives, A is the better-performing model and B is the worse-performing model.

For example, to determine the probability that the Destination model is statistically better than the Origin model one would enter the following:

$$p = NormSDistr(-\sqrt{2 * 9303 * ln(4) * (-1/-78250}) = 0.289 \tag{4.3}$$

The probability that the Destination model thus is the model that more closely approximates the true decision making rule is 0.289, which means that the performance of the Destination model is not statistically significant (at any reasonable cut-off anyway, such as 5%) and we can't conclude that this is the model that more closely approximates the decision making process in the population.

Since there are nine different models, there are a total of 36 unique values for this Ben-Akiva and Swait Test. The table containing all models and the result of the test is given in the appendix (Table B.1). Here I will highlight the probabilities for the combinations with the best performing model, the Weighted Daily model. The minimum difference in model-fit, when compared to the second-best model of 2H centre, is 49 Log-Likelihood points. Such a difference results in a probability that is smaller than 0.001, indicating that the Weighted Daily model does statistically perform better than all other models, even at the 1% threshold. This supports the conclusion that the Weighted Daily metric is best able to capture the effect of weather on mode choice, although the (very) small difference in McFadden's ρ^2 indicates that the differences between metrics are not ground-breaking.

The best model is the Weighted Daily model, which performs statistically better than all other models, even when using a 1% significance threshold. This provides evidence for the proposition that the population of decision makers use impressions of weather during travelling hours for their travel-related decisions.

4.3.2 Parameter Evaluation

Since the differences in model-fit are small (although statistically significant), it's interesting to inspect the differences in parameter estimates between the models. Stable parameter estimates indicate that the metrics don't really matter, whilst big changes across models would indicate the opposite. The parameter values that are discussed below are given in Table 4.5.

As expected the alternative specific constants and the distance parameters are highly stable across all models: of course, the distance data is the same for all models and the stability only indicates that the explanatory power of weather doesn't interfere with the explanatory power of distances.

The parameter values for the wind parameters are also highly stable across models. The signs are constant across all models(positive for train; negative for bicycle and walking). For nearly all models, the magnitude of the effect on biking is greatest, followed by the effect on public transport and then walking. The only dissonant here is the daily model, where the influence on public transport is a tiny fraction stronger than the effect on biking. Finally the range of parameter estimates is small for all modes, ranging from 0.082 to 0.113 for public transport, -0.124 to -0.100 for bicycling and -0.103 to -0.071 for walking. All of these observations support the conclusion that the wind parameters are very stable across the weather metrics and that the decision to choose one of them is unlikely to lead to different conclusions regarding the influence of wind speed on mode choice.

> All weather metrics give roughly the same estimates for the impact of wind speeds on mode shares, indicating that the conceptualisation of how weather is incorporated in decisions of travellers doesn't matter for the estimated effect of wind speeds.

Although the signs for temperature are constant, an interesting difference can be observed between the models that are based on the time of the trip (origin, destination, and the time-window models) and those based on the date of the trip (daily and rush-hour values). The most striking difference is that the parameter for public transport is much stronger for the trip-based models than for the day-based models. For all of the day-based models these parameters turn out to be insignificant! In contrast the parameters for biking and walking are stronger for the date-based models. This effect repeats for all of the date-based models compared to all of the time-based models, suggesting that it must be caused by the fundamental differences between these two overarching ways of connecting weather. I hypothesize that the difference can be explained by the fact that mode shares change during the day: in the morning most trips will be utilitarian and public transport mode share will be higher as a result (since mode share of public transport is higher for utilitarian trips than for leisure trips). Since the time of day is correlated to the average values for temperature (i.e. it is colder in the mornings and late afternoons than around noon), an effect that is not accounted for in these simple MNL models, the trip-based models actually over-estimate the influence of temperature on public transport trips as this influence is explained by the colder times of day when people tend to use public transport more often. This also leads to an under-estimation of the effect of temperature on the active modes of biking and walking, which are less used during the morning. The temperature parameters thus differ quite substantially between the date-based and trip-based operationalizations, but within these groups they are again very stable.

> There is a substantial difference between conceptualisations based on the time of the trip and conceptualisations based on the date of the trip. Time-based metrics estimate a sizeable negative effect of temperature on public transport mode shares, whilst date-based models all find insignificant effects. This can be explained by the higher mode shares for public transport during peak-hours, which occur during relatively colder hours of the day. Time-based models thus estimate a causal effect that probably does not exist.

This brings me to the rain parameters, where correlations between operationalizations were notably lower than for the other two weather parameters and the differences between conceptualisations are thus expected to be substantially larger. This expectation is indeed confirmed when looking at the results, where even a sign-change can be observed. All operationalizations lead to the estimation of a negative effect of rain on public transport use, except the operationalization of the afternoon rush-hour which indicates a (small and insignificant) positive effect. Here the cause is not so obvious, as this finding indicates that people are slightly more likely to use public transport on days when afternoons are affected by precipitation. The other date-based models do indicate a negative effect, with the morning operationalization indicating a substantial negative effect of precipitation. It could be that afternoon rain is simply not taken into account by public transport travellers, although there is no substantive explanation for this phenomenon as of yet. Evidence for the lack of attention paid to afternoon rain might also be found in the relatively small indicators for bicycling and rain of this operationalization. Another sign change can be observed for the walking parameters, where the destination model finds an insignificant positive impact of rain on mode share of walking whilst all others find a negative impact, which is more in-line with literature and intuition. Again it might be possible that pedestrians don't pay attention to potential rain at their destination, although this explanation is not very intuitive. The only consistent finding then is the negative effect of rain on the mode share of the bicycle, which is indeed expected in the literature. The effect seems slighlty stronger for the date-based models, indicating that rainy days in general dissuade bicycling use.

> Most models estimate a negative effect of rain on public transport, bicycling, and walking. There are some interesting outliers however. Results indicate that there is no effect of afternoon rain on public transport use, which could perhaps signify that public transport travellers do not look ahead at afternoon weather. The other outlier is the lack of an effect of rain at the destination on walking use, for which we have no good explanation at this time.

In conclusion the different conceptualisations often lead to very similar parameter estimates, which is expected when taking the similarity in overall model fit into account. The parameters for rain however can vary sizeably, which results from the more volatile nature of this weather type when compared to temperature and wind speed.

4.3.3 External Validity

Another way to test the implied differences between the models is by using the fitted model to predict choices in a data-set that was not used during the estimation process. In essence the available data sources are split into two data sets, of which one is used for estimating the model (the training data) and one is used for testing the estimated models against recorded choices (the test data). It's possible to repeat this procedure of estimation and testing on changing fractions of data, which is a way to reduce the chance that the test data is not an accuracte representation for the population and test results are biased by sampling peculiarities. This procedure is not used here, as instead the first five waves of the MPN are used for estimating the model (as described in Chapter 3) and the newest sixth wave is used for testing purposes.

This sixth wave of the MPN is recorded in autumn of 2018 and weather data is matched to it using the procedures outlined in this chapter and in Appendix A, which are the same procedures as were used for the earlier five waves. The 10 estimated models are then all used to calculate the probabilities that a mode is used for each trip. The mode with the highest probability is selected as the predicted mode for each trip and these predictions are compared to the recorded modes.

One numerical estimate for the performance of the prediction is the hit rate, which is the percentage of times when the chosen mode was the same as the mode that was predicted by the models. These hit-rates are given in Table 4.6

Table 4.6: The hit rates of the operationalizations on the external Wave 6 of the MPN

Model	No Weather	Origin	Destin.	Daily	W. Daily	2H Centre	1H Centre	1H Before	Morning	Afternoon
Hitrate	0.5695	0.5706	0.5707	0.5711	0.5700	0.5709	0.5710	0.5714	0.5719	0.5691

The hit-rates show a couple of surprising things: the first is that the choice of model and thus the conceptualisation and operationalization really don't matter very much, with the worst weather model scoring only 0.02 lower on this hitrate than the best weather model. Furthermore the difference with the no-weather model (incorporating only the distance!) is not very large at all, meaning that the predictive purposes of weather are either not very large or can not be found with the admittedly simplistic modelling done in this chapter. The third finding contradicts all previous findings, because the hit-rates do not show that the weighted daily conceptualization is the best. Rather this metric performs below average compared to the other weather metrics. The best metrics here are the daily average, the 1H before and the morning based predictions. Again, the difference between these models is very small but this was still a confusing surprise.

4.4 CONCLUSIONS

In this chapter we have studied how people incorporate the weather into their travel decision-making processes, more specifically looking at *which* weather is used during these processes: weather at the time and location of the trips origin, daily weather circumstances, or other ways?

We have systematically compared nine different methods of assigning weather values to a decision, based on the location and time of the trip. There are two fundamentally different groups of methods in terms of the time that is used to assign the value. The first group uses the specific time of the trip, for example using the weather at or around the time of departure. These methods thus assume that decision-makers change their decision based on variation in the weather during the day itself. The other methods use the date of the trip, for example calculating daily averages or values during specific hours of the day such as rush-hours (irrespective of the time when the trip is made). These methods assume that decision-makers do not base their decision on variation within the day, but rather only on variation between weather across different days. In general we find that the date-based models give more realistic parameter estimates, as the time-based models do not separate effects caused by weather changes throughout the day from effects caused by other daily patterns, such as the general habit of making utilitarian trips during colder hours of the day (the morning and late afternoon).

Our findings also show that the difference in explanatory performance between these various methods is only very modest. The differences between methods using the location of the origin or that of the destination to assign weather values is almost negligible. The difference between the use of various times throughout the day to measure weather is much larger and statistically significant, with the best performance being obtained by the method based on weighted daily average values of weather. This average weights the weather based on the number of trips made during certain times of day. Weather during the night (when few people travel) thus is almost not used when calculating this average, whilst weather during rush-hour is much more influential. This result is in contrast to the current state of the literature, where using the time and location of the trips' origin is the most common practice.

This result suggests that people's decisions are based on their perception of the weather during the entire day, such as whether or not it will rain, a general feel for the wind speed and the temperature of the day. It also suggests that travel decisions are based more on changes in weather between days than on changes in weather within a day. Since the perception of the entire day seems to matter, both past and future weather impact decision-making. The fact that people seem to partly base their decision on future weather prompts the question how they attain this information about the future state of the weather and whether different methods of gaining such knowledge lead to different incorporations of weather in the decision-making process. This question is left unanswered in this thesis, but would serve for interesting future research.

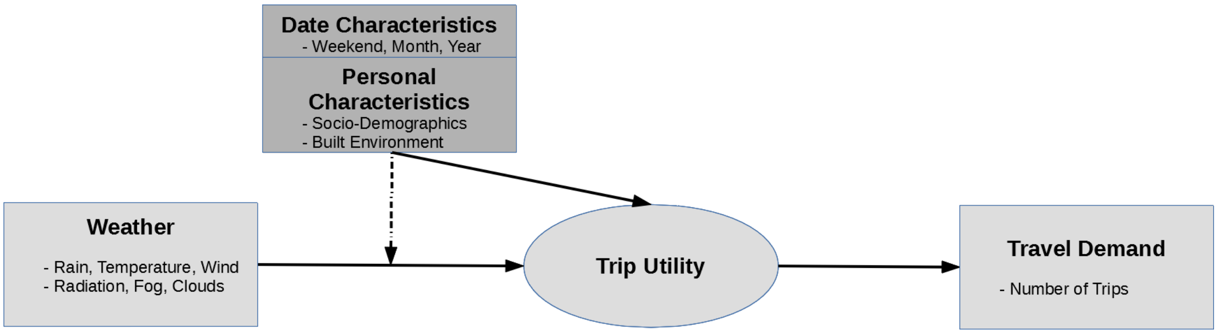
5 TRAVEL DEMAND

In this chapter we estimate the influence of weather on travel demand, whilst focusing on ensuring our knowledge is able to fill two existing gaps in the literature. The first gap relates to whether a specification where we estimate non-linear and interaction effects between weather variables lead to an improved understanding of the relation between weather and travel demand. The second gap relates to whether this relationship varies across urban and rural areas.

This is achieved by estimating count regression models that use daily weather data to explain the number of trips made by an individual on a certain day, whilst controlling for several other variables that influence the number of trips. As explained in Section 3.2.1 Poisson and negative-binomial regression models will be used for modelling. Within these model specifications we specify non-linear and interaction effects to find if they improve the performance of the model and if they lead to meaningfully different findings about the effect of weather on travel behaviour. We then also estimate models with interaction effects between the influence of weather and the spatial environment (rural or urban), which allows us to see if the weather affects travel demand differently in urban environments when compared to rural environments.

Considering the above, this chapter focuses on the following parts of the conceptual model:

Figure 5.1: Conceptual Model of the relationships that will be studied in this chapter



The unit of analysis of this chapter is the day, as we're looking into daily travel demand measured by the number of trips made per person per day. This unit of analysis is thus different than the one in Chapter 4 and Chapter 6 (where the unit of analysis is the trip itself), which has several repercussions. The first is the fact that we need to re-think the conceptualisation of how weather influences travel demand generally and the decision whether or not to travel specifically. Conclusions from Chapter 4 can help, but offer no solution that can directly be implemented. The second relates to other variables that might be included in the model, as trip-level characteristics (such as trip-purpose or trip-distance) are not available. The third is that the sample is now different: we're no longer sampling trips, but rather sampling unique person-day combinations. For each person in the MPN information about three days of travel is collected per year, meaning we get three observations per person per year.

Furthermore an important distinction must be made between the goal of this chapter and the goal of Chapter 4. In this chapter we try to estimate the pure influence of weather on a dimension of travel behaviour (in the form of travel demand), rather than just comparing different conceptualisations of the link between weather and mode utility. This means that the multi-variate analysis should control for other independent variables to ensure that possible confounding and mediating effects are controlled for. Consequently this analysis uses more variables during the modelling procedure. These new variables will be introduced in this chapter.

The data description and preparation process is described in Section 5.1. The model estimation procedure is given in Section 5.2 and the resulting estimates are described and briefly discussed in Section 5.3. These results lead to conclusions that can be found in Section 5.4.

5.1 DATA DESCRIPTION AND PREPARATION

As said above, the dependent variable of the analysis in this chapter is the number of trips made per day per person. The unit of analysis of this chapter is thus the day, rather than the trip (as in Chapter 4 and Chapter 6). This means that all independent variables should also be stable for a whole day (although they can of course vary across respondents and days).Furthermore this chapter will use variables that were not used in Chapter 4, due to the different goals of the chapters. There are both new weather variables (extending beyond rain, temperature, and precipitation) and non-weather variables, where variables related to both people and days are used to control for possible confounding effects.

But first we'll explain how the weather variables are operationalized for *days*, as opposed to *trips* (the discussion of this operationalization of course being central to Chapter 4). This explanation is given in Section 5.1.1. Then we're going to describe and visualize the dependent variables of the analysis, which is the number of trips made by a person per day. This can be found in Section 5.1.2. Next the weather variables will be discussed in Section 5.1.3, which is followed by a visualization of the relationship between weather variables and the number of trips in Section 5.1.4. Finally the non-weather independent variables will be discussed in Section 5.1.5.

5.1.1 Matching weather to travel demand

The weather that needs to be calculated and matched to the days is thus some kind of daily connection, such as the daily average, weighted daily average, or weather during certain peak-hours (irrespective of actual trips' origin/destination times). Based on the conclusions of Chapter 4 that weighted daily average is the best predictor of mode choice across the day the assumption here is that this connection method will also be the best predictor of travel demand during the day.

Another conclusion from Chapter 4 is that there is no discernible difference between using the locations of origin and those of destination. Conceptually however decisions whether or not to travel at all will primarily be based on the weather at the place where the trip potentially originates. One intuitive way of determining the location is to use all locations that have been used as the origin of a trip during the day and calculate their average. This approach presents some problems however. The first of these is that there is no clear location for people who have not made a trip; the second is that the location of the trip and the number of trips made are conceptually dependent on the weather. Using these dependent characteristics to calculate weather risks running into circular reasoning, biasing the results. The third results from the fact that we're interested in people deciding whether or not to travel. The decision whether or not to travel is however not made in all locations that are visited during the day: if someone decides to go shopping the weather at the shopping center will probably not compel or dissuade them into making fewer or more trips. If the weather at their residential location however is already very bad they might be dissuaded from going shopping at all. We assume that the decision to travel or not to travel is mostly made when at certain locations, such as the residential location.

The location(s) that is/are used should satisfy three conditions:

1. Be able to specify a location for people who have made no trips
2. Shouldn't depend on number of trips made
3. Correspond to locations where decisions whether or not to travel are made

By assuming that the residential and work locations are the two locations where the decision to travel or not is made, we're able to abide by these three conditions. This assumption is based on the fact that people spend most of their time during the day at both of these locations and that trips often depart from here. Of course the return journey from work to home is a given when people have gone to work earlier during the day, but at work decisions to make additional trips can be made based on the weather circumstances, for example when people go for a walk when it's nice outside.

Weighted daily average weather will be determined for both the residential and work locations. For people who did not visit their work location, the weighted daily average at the residential location is used for the analyses in this chapter. For people who did visit their work location the average between the two locations will be calculated and will be used during the analyses. Of course, to determine the weather at each location the weather as collected by the closest weather station is used as detailed in Appendix A.

Since we are now trying to find weather for a day, rather than for a trip, there is no obvious location to use for the collection of weather information, such as a trips' origin or destination. We decided to use the residential and work location of respondents under the assumption that people are able to make a decision to make additional trips at these locations. The work location is only used if people visited their work during the day.

5.1.2 Number of Trips

The dependent variable for the regression models are thus the number of trips made per person per day. Since the effects of weather on travel demand are mode specific (see Chapter 2), we've made the decision to differentiate between these modes and use the number of trips per mode per person per day as the dependent variable of regression models as well. As described in Chapter 3 four distinct modes are identified: the car, public transport, the bicycle, and simply walking. Each of these modes are thus the dependent variable for their own modelling procedure (more about this later on in Section 5.2). The fifth dependent variable is simply the total number of trips made per day, irrespective of modes used.

The descriptives and correlations for the number of trips made per mode are given in Table 5.1. The 'other' mode is included in this table as well, and contains modes such as mo-peds and steps. Since the number of trips made with other modes is relatively small and many very distinct modes are grouped together to form this 'other' modes they are not used as the dependent variable in the regression models.

Table 5.1: Descriptives of number of trips made per person per day for various travel modes

	Descriptives			Correlations					
Number of Trips	Mean	St. Dev.	Max.	1	2	3	4	5	6
1. Total	3.01	2.30	25	1	0.50	-0.03	0.35	0.37	0.09
2. Car	1.72	1.93	21		1	-0.22	-0.34	-0.13	-0.12
3. PT	0.19	0.61	8			1	-0.11	-0.02	-0.04
4. Bike	0.99	1.60	16				1	-0.08	-0.09
5. Walk	0.55	1.16	13					1	-0.04
6.Other	0.12	0.63	13						1

The mean trips show that people make an average of 3 trips per day, with most of these (1.72 on average) using the car. The relatively low maximum number of trips for public transport stands out as well: where all other modes have a maximum of 10+ trips per day the maximum for public transport trips is 8. Standard deviations are also noticeably high, with variances ($St.D.^2$) exceeding the mean values in all cases. Although this is an indication of possible over-dispersion, this is no conclusive evidence: it might be possible that the variance is equal to the mean conditional on the independent variables.

When turning our attention to the correlations, we can see that all trips using a specific mode are correlated negatively with each other: if someone uses one mode more, he/she is less likely to make numerous trips with other transport modes. The size of the correlations between Public Transport and the active modes are noticeably smaller than the size of the correlations between the car and the active modes, evidencing the fact that public transport and active mode trips are not just competitors, but also complementors. It's worth mentioning here again that we're looking at full trips rather than trip-legs, meaning that small trip-legs to and from public transport access points are not even included here. If they were, it stands to reason that there might be a positive correlation between PT and active modes' number of trips.

The overall number of trips are positively correlated with all modes, except for public transport. The reason might be found in the difference of the maximum number of trips: the maximum of 8 trips travelled using public transport are much less than the maxima for the other modes. Public Transport use perhaps locks the traveller into a specific origin/destination pair (for which the traveller is familiar with public transport). Another explanation could be that public transport trips are usually longer distance trips that take more time, thus meaning that the travelled distance is the same with relatively fewer trips.

To gain a further understanding of the distribution of the number of trips the distribution and a Poisson distribution with mean (and variance) equal to the mean of the number of total trips are given in Figure 5.2

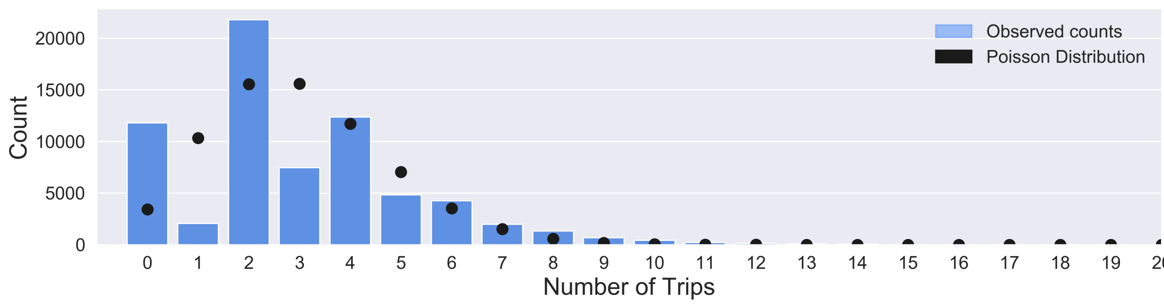


Figure 5.2: Distribution of the number of trips, compared to Poisson distribution

As can be seen in Figure 5.2, the number of trips contain an excess amount of zeros compared to a Poisson distribution. Furthermore the even trips seem to occur more often - and the uneven numbers less often - than would be predicted by a Poisson distribution. This can easily be explained by the fact that trips often go somewhere and then return back to the original location. Trips thus often come in pairs of two, which leads to the abundance of even number of trips that can be seen in the figure. High number of trips seem to occur slightly more often than predicted by the Poisson regression, possibly pointing to overdispersion.

The same figure made for all separate travel modes can be seen in Figure 5.3. The patterns from the total number of trips hold up for all modes: the even trips occur more often than uneven trips, which is to be expected. The mode distributions show that zero number of trips is by far the most often occurring number of trips per mode per day.

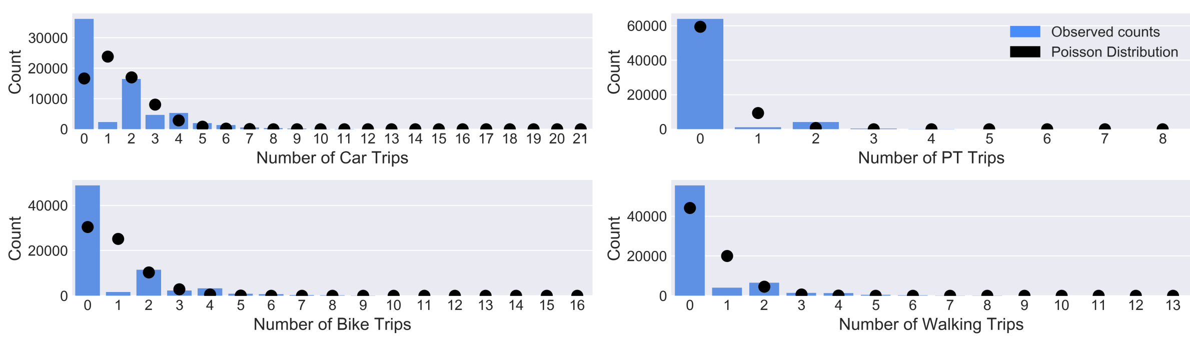


Figure 5.3: Distribution of the number of trips, compared to Poisson distribution

5.1.3 Weather Variables

Since the goal of this chapter is to estimate the effect of weather on the dependent variables described above we of course need to introduce the weather variables. Precipitation, temperature, and rain have already been used and introduced in Chapter 4. This chapter uses more variables however, the first of which is the humidity. There are two main ways of measuring humidity: either absolute humidity, which is the amount of water vapour present in the air (measured in grams per cubic metre or grams per kilogram) or relative humidity, which is the current absolute humidity divided by the maximum absolute humidity at the current temperature (higher temperatures allow for air to contain more water vapour). Here relative humidity is used, as this kind of humidity is more relevant for the perception of temperatures by humans: higher relative humidities increase the perceived temperatures as less sweat can be evaporated. Furthermore, absolute humidities would be highly correlated with temperature which would run into problems of multi-collinearity making it very difficult or impossible to estimate the effects of temperature and humidity.

The second addition is that of cloud cover, measured in okta. Cloud cover is simply the amount of sky that is covered by clouds, with measurements ranging from 0 (fully clear) to 8 (fully covered). Closely related is solar radiation or sunshine, which is the energy provided by the radiation from the sun measured in W/m^2. The final variable used here is Meteorological Optical Range, which is the distance in the atmosphere where light emitted from a source is reduced to 5% of its original intensity. This essentially corresponds to the level of visibility and serves to measure fog or other circumstances that reduce the visible range. Descriptives and correlations of the weather variables used in the regression analyses can be found in Table 5.2.

Table 5.2: Descriptives of weather variables

	Descriptives				Correlations						
Variable	Mean	St. Dev.	Min.	Max.	1	2	3	4	5	6	7
1. Temperature (C)	11.4	3.78	0.98	20.9	1	0.18	-0.01	-0.38	-0.10	0.50	0.11
2. Wind Speed (m/s)	4.05	2.1	0.28	18.5		1	0.19	-0.31	0.05	-0.02	0.21
3. Rain Intensity (mm/h)	0.1	0.19	0	1.96			1	0.26	0.27	-0.26	-0.12
4. Humidity (%)	83.1	8.79	53.3	100				1	0.54	-0.75	-0.63
5. Cloud Cover (okta)	5.75	2.27	0	8					1	-0.71	-0.27
6. Solar Radiation (W/m^2)	104.1	67.2	8.4	301						1	0.35
7. Optical Range (100m)	197	113	1.5	491							1

The correlations between weather variables is across the board more than low enough to avoid multicollinearity issues, with the exception of the correlation between humidity and solar radiation. Their correlation of -0.75 can become somewhat problematic. Furthermore the descriptives of the weather variables clearly show that the data set corresponds to autumn weather: no extreme temperatures (below 0 or above 25 degrees) exist, but the maximum wind speeds and to a lesser extent rain intensities clearly show that some stormy days were part of the days when the travel diaries were recorded.

5.1.4 Relationship weather and travel behaviour

We can perhaps get some valuable insights from the visualisation of the relationship between the number of trips and various weather characteristics. For the sake of visual clarity, this is done only for the three main weather parameters of temperature, wind, and rain. These continuous variables have been re-coded into two-level dummy variables, where one level is considered to correspond to inclement weather conditions. For temperature, days with a weighted average below 10 degrees C are classified as cold days. For wind speed, days with a weighted average value above 6 m/s are classified as being windy days. For rain, days with a weighted average value above 0.1 mm/h are classified as rainy days.

Figure 5.4 visualizes the overall number of trips under the varying combinations of these three conditions, as well as the percentage change of the mean value under one condition compared to the overall mean (irrespective of conditions).

The visualisations of the impact of the weather conditions on all four modes is visualized in Figure 5.5, which shows the percentage change of the mean number of trips using the different modes

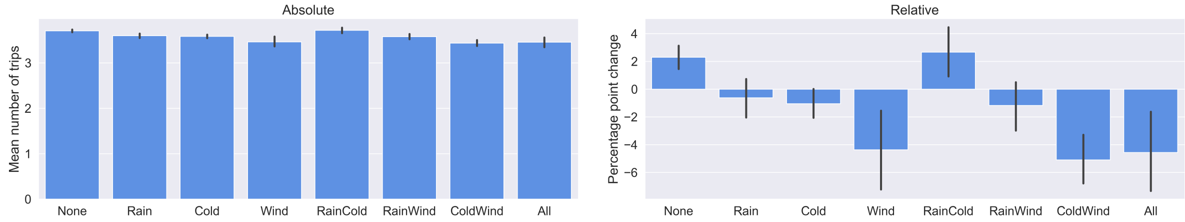


Figure 5.4: Changes in overall number of trips due to weather conditions

under varying conditions compared to the overall mean. The conditions are the same as were used for Figure 5.4.

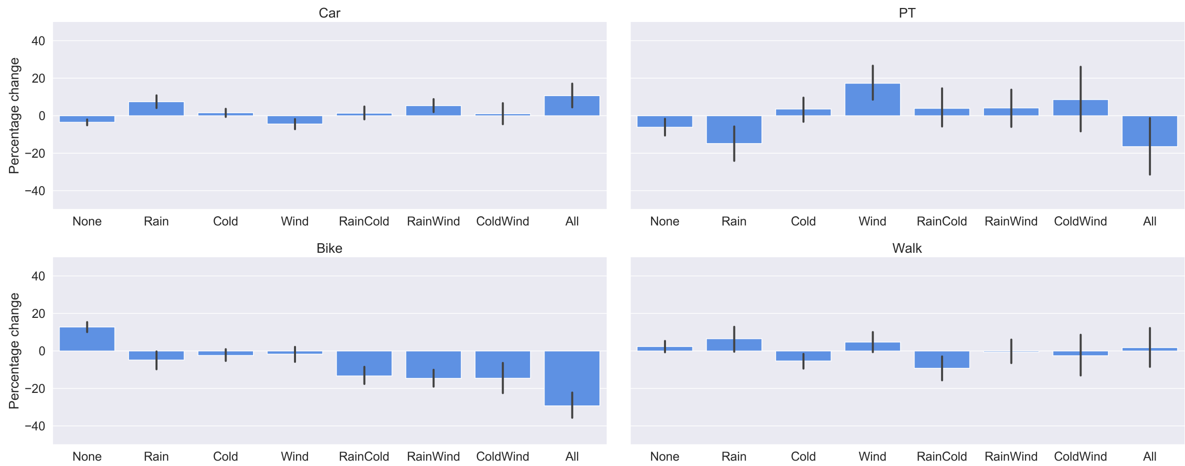


Figure 5.5: Changes in number of trips due to weather conditions, split per travel mode

When contrasting Figure 5.5 with Figure 5.4, one might conclude that the overall view presented in Figure 5.4 obscures very interesting mode-specific influences of weather. It is no surprise that the number of car trips is more robust to the different weather conditions (as can be surmised from the relatively small percentage changes).

One interesting tentative observation is the importance of the various interaction effects, perhaps most clearly seen in the figure pertaining to bicycle trips. Whilst rain, cold, and wind are all associated with a decrease in bike trips this association is very limited and falls within the confidence interval in all cases. The interaction terms however all show a much clearer decrease, indicating that the combinations of rain, wind, and cold are much more influential than the separate variables. The union of these circumstances shows a further decrease, indicating that the total interaction effect is also very important.

5.1.5 Other independent variables

There are two types of independent variables that are not related to the weather: variables that are dependent on the person making - or not making - the trips and variables that are dependent on the date when the number of trips are recorded. These two types will be discussed separately below.

Person characteristics

Seven person-level variables are used in this analysis, most of which are common socio-demographics. The first variable is the gender of the respondent, which is dummy-coded where a 1 represents that the respondent is of the male gender and a 0 that the respondent is female. Next someone's ethnicity is used as well, which has four levels and is also dummy-coded. The levels are Dutch ethnicity, foreign western ethnicity, foreign non-western ethnicity and unknown ethnicity. The next variable is education, a categorical variable with three levels: low-education, medium-education, and high-education. Low education is defined as having either not completed any high school level or having completed only the lower high school tier (vmbo). Medium education is defined as having completed higher-tier high schools (havo/vwo) or second vocational education (MBO). Finally high education is defined as having obtained at least a bachelor's degree from either an university of applied sciences (HBO) or a research university (WO). Age is the next variable, which is also coded categorically (due to privacy concerns). There are 10 different values, ranging from 1 to 10 where a 1 denotes people under 12 (which are not present in our sample) and a 10 denotes people aged older than 90. The first steps use increments of roughly five years. From the fifth step onwards (age 29) increments are 10 years instead.

The last two variables are determined based on the residential location of the respondent. The first of these variables is the density of the neighbourhood in which the residence is located, which is a categorical variable ranging from 1 to 5 where a 1 means 0-500 inhabitants per square kilometer and a 5 2500+ inhabitants per square kilometers. Not all steps of this scale are thus equally large: the first three steps contain a range of 500 inhabitants per square km, the fourth a range of 1000, and the fifth is everything above 2500 inhabitants. The second is related to the province of the residential location. In the Netherlands the Randstad is a conglomeration of major cities. The difference between Randstad and non-Randstad are often subject of public debate, which is why we decided to include this as a variable in the analysis. People living in the provinces of North-Holland, South-Holland, or Utrecht are said to live in the Randstad, whilst people living in other provinces are said not to live in the Randstad. Note that this is not perfect, as some parts of the mentioned provinces are not part of the Randstad (such as the northern part of North-Holland). This only serves as a rough first exploration to see if there are differences in travel patterns between Randstad and non-Randstad inhabitants.

Table 5.3: Overview of personal characteristic variables used in the analysis

Variable	Measurement Scale	Levels	Sample Percentage
Age	Ordinal	1-9	-*
Density	Ordinal	1-5	-*
Gender	Dichotomous	Male	46
		Female	54
Employment Status	Dichotomous	Employed	91
		Not Employed	9
Randstad	Dichotomous	In Randstad	58
		Outside Randstad	42
Ethnicity	Categorical	Native	91
		Western Foreign	6
		Non-Western Foreign	2
		Unknown	1
Education	Categorical	Low	34
		Med	37
		High	29

* There are too many levels to give the sample distribution in a table. See Figure 5.6 for sample distributions of these variables

In one of the models we try to estimate the difference between urban and rural areas. For these models we have re-coded the density variable into a dummy variable with two levels. THe first level ('rural') contains the first three levels of the density variable, whilst the second level ('urban') contains the highest two levels. From Figure 5.6 we can see that both levels contain roughly 50% of the data-points.

Date characteristics

Besides the personal characteristics, the date on which trips are made might also influence the number of trips. There are three variables related to the date, increasing in length of time (day, month, year).

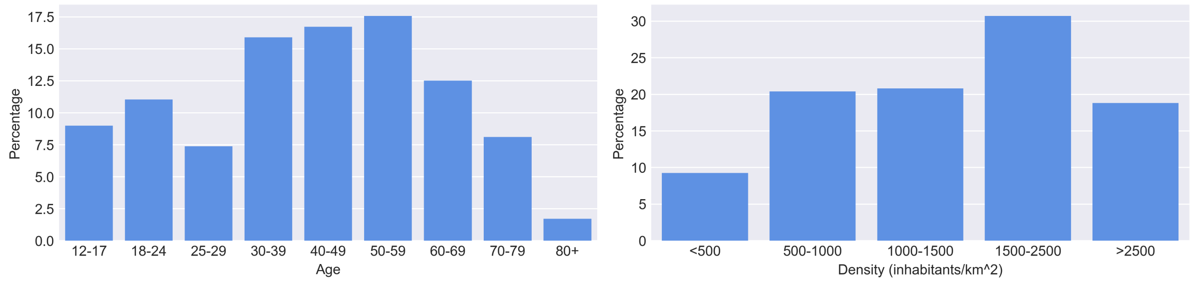


Figure 5.6: Sample distribution of ordinal personal variables

The first is the type of day, where we distinguish between week-days (monday through friday), saturday, and sunday. This categorical variable is dummy-coded, with week-day as the reference level. The second variable is the month. The MPN is collected during the months of september, october, and november. Again dummy-coding is used, where september is used as the reference. Finally we also include a variable for the year, again dummy coded with 2013 as the reference. The rationale for including variables related to months and years is to be able to separate non-weather influences of time on travel behaviour: there might be cultural reasons for travelling more in october for example; or the sampling might be slightly inconsistent from year to year. Finally we'd like to point out that we would have included a holiday variable for days that fall on a school- or national holiday. Since the MPN is sampled in the weeks before and after the Autumn holidays in the Netherlands and the absence of national holidays during this time of year no such days were present in the sample however.

Table 5.4: Overview of date characteristics used in the analysis

Variable	Measurement Scale	Levels	Sample Percentage
Month	Categorical	September	13
		October	35
		November	52
Year	Categorical	2013	17
		2014	23
		2015	17
		2016	19
		2017	24
Type of Day	Categorical	Weekday	70
		Saturday	15
		Sunday	15

5.2 MODEL DESCRIPTION

The visualisations of Section 5.1.4 show that the effects of weather on travel demand are mode-specific, as the change in the number of trips per weather condition is markedly different across the modes. It is thus necessary to estimate models with the number of trips per travel mode as the dependent variable, instead of just the total number of trips per day (irrespective of mode). This means that there will be five dependent variables: total trips, car trips, PT trips, bike trips, and pedestrian trips (all per day per person). The regression modelling procedure is thus repeated five times, once for each of the dependent variables.

The modelling procedure consists of estimating five iterative models, with each model estimating a more complex relationship between the number of trips and the weather. There are five of these iterations, with the first model estimating no weather relationship at all (by not using any weather variables). The second model adds a linear-in-parameters effect of weather variables. The third model adds quadratic effects to test the assumption of linearity. Our goal here is to examine whether non-linear transformations could add explanatory power to the model. We use quadratic transformations for two reasons, the first of which is that in the scarce literature researching non-linear effects some quadratic effects have been found (for example the existence of an ideal temperature for cycling), the second is the ease of specification and interpretation. The fourth model adds interaction effects between the weather variables and the fifth model finally adds both the quadratic and interaction effects. These models are first estimated using Poisson regressions: if these regressions point to overdispersion Negative Binomial regressions are estimated as well.

The iterations enable us to uncover the complexity of the relationship between weather and travel demand for each of the dependent variables: it could for example be true that a linear-in-parameters model best explains the number of total trips made per day, whereas the most complex model incorporating both quadratic and interaction effects is the best fit for the number of bike trips made per day. Estimating five models per dependent variable thus allows us to compare the models within the dependent variables and find which model is the best fit for each dependent variable.

> Since our visualisations have shown that the relationship between weather and travel demand is different per mode we estimate models for each mode separately. Furthermore we estimate five models per mode, each using an increasingly complex method of specifying the influence of the weather on travel behaviour. This allows us to see how complex we should model the relationship between weather and travel behaviour for each mode.

A further step is to compare the effects of weather across the different dependent variables, for example contrasting the effect of precipitation on the number of car trips and the number of bike trips. Effective comparison can be made in two ways: by displaying the same model for both modes and using the model parameters for both; and/or by selecting the best model for both modes and calculating the marginal effect of added precipitation on the number of trips for both models and comparing this number. Both approaches will be used in this thesis. Using values related to the trade-off between model-fit and the number of parameters used to estimate the models, the best model for each of the modes is picked. The model that performs the best across the modes is then used for the comparison of parameters between modes. Meanwhile the best model for each mode is used to calculate the marginal effect for each dependent variable.

During the model estimation procedure variables that were not significant in any of the 50 models were removed from further analysis. These variables were all weather variables: humidity, cloud cover, and optical range all proved to have no significant impact on travel demand for any travel mode.

> For all models the effects of humidity, cloud cover, and meteorological optical range were insignificant. For this reason these variables were removed from the models.

The models are all estimated using the statsmodels module [Seabold and Perktold, 2010] for Python. The code used during the analysis can be found on my GitHub

5.3 RESULTS

Since the procedure outlined in Section 5.2 leads to a total of 50 estimated models (5 models per dependent variable * 5 dependent variables * 2 methods (Poisson & Negative Binomial) most of the results are moved to Appendix C. Here we will first show the results of five negative binomial models where the number of bike trips per day are the dependent variable of the regression. In doing so we can show how progressively estimating more complex relations between weather and travel demand influences the model, its parameters, and its interpretability. This comparison is given in Section 5.3.1 and will be followed up by a comparison of the parameters estimated by one model with different modes as the dependent variable, which will be done in Section 5.3.2. Finally the best model will be selected for each dependent variable and this model will be interpreted using marginal effects and the calculated number of trips in Section 5.3.3.

5.3.1 Comparison of models per mode

Table 5.5 shows the results of the five negative binomial regressions, with bike trips as the dependent variable.

Using Table 5.5 we can see how estimating increasingly complex weather models affect the parameter estimates and significance of various parameters estimated by the model, as well as the model-fit as indicated by Chi-Square and Log-Likelihood values. The BIC and AIC values provide a reference for whether or not the increase in model-fit as indicated by the Log-Likelihood is due to a model that actually better represents the true underlying decisions or simply due to sampling peculiarities. Below we will first discuss the former effect (on parameters), and then the latter (on model fit).

In terms of parameters we see that some interesting changes occur in the non-weather variables when weather effects are estimated. To see this difference we can compare any model incorporating weather statistics with the model not incorporating weather variables. In the non-weather model the months have a strong significant effect on the number of bike trips, meaning that people use the bike less in october and november. When we incorporate weather effects these effects disappear entirely, indicating that this negative effect is caused by weather variation between the months (october and november are colder and more prone to high wind speeds than september). By incorporating weather variables we are thus able to separate the non-weather effects that cause variation between months (such as the existence of public holidays in a month) from the weather effects that cause variation between months. Here we see that there is no significant non-weather related variation between the months of september, october, and november (take into account that the MPN purposely does not sample during holiday weeks).

Whilst this disappearance of the effect of months was expected, the exact same effect can also be seen for the difference between years and even the Randstad vs. non-Randstad which is a surprising finding for us. In the years we see that in 2014 and 2015 the non-weather model indicates that significantly more bike trips have taken place: an effect that disappears when weather variables are added to the model. Furthermore we see that the difference between 2016 and 2013, which was at first not significant, becomes significant for various weather models. For the Randstad in our model without weather we would conclude that people in the Randstad bike less often, an effect that is as it turns out entirely explained by the weather. The Randstad area is placed in the western part of the country, closer to the sea which means that temperatures are slightly lower and wind speeds slightly higher than in the east of the country. Apparently this difference in weather causes the difference in number of bike trips, as the Randstad dummy estimate becomes negligible in all weather models.

> By including the effects of weather in the analysis some of the effects of other parameters change. The effects of months disappear, indicating that people bike less in November compared to September because the weather in November is less pleasant, rather than because of some other non-weather effect. Similarly the observation that people bike less in the Western part of the Netherlands seems to be caused by the fact that this part is wetter, colder, and windier than the rest of the country.

Turning our attention towards the effects of weather itself, we see that in almost all weather specifications the simple weather terms show the same direction: temperature and sunshine have a positive effect on the number of bike trips, whilst wind and rain have a negative effect which is in line with what we expected and is generally found in the literature. We see that sunshine and temperature have a significant squared relationship in the Quadratic model, indicating that their effect on bike trips is not linear-in-parameters: perhaps there is an optimal temperature or sunshine amount, which is also sometimes found in the literature. Interestingly almost all interaction parameters are significant as well, which means that the effect of an in- or decrease of one weather variable depends on the values for the other variables. One might expect that the negative effect of wind for example is stronger when temperatures are low and skies are overcast. The interpretation of these interaction effects is not straightforward, also due to the fact that so many of them are significant. When adding these interaction effects the linear effect of rain also becomes positive, which is a strange finding. A closer look into the effects of the weathervariables on the number of trips is given in Section 5.3.3.

Table 5.5: Negative Binomial Results, dependent variable are bike trips per day

	No Weather	Weather	Quadratic	Interaction	Quad + Int
Intercept	0.2769***	0.1374***	0.2033***	0.2060***	0.2303***
Age	-0.0599***	-0.0601***	-0.0601***	-0.0601***	-0.0601***
Male	-0.2030***	-0.1998***	-0.1998***	-0.1986***	-0.1987***
Employed	0.0963***	0.0888***	0.0887***	0.0898***	0.0894***
Randstad	-0.0347**	0.0087	0.0068	0.0058	0.0056
Density	0.0696***	0.0670***	0.0676***	0.0672***	0.0672***
Ethnicity (reference is Dutch native)					
Western Foreigner	-0.1898***	-0.1918***	-0.1922***	-0.1899***	-0.1908***
Non-Western Foreigner	-0.5551***	-0.5536***	-0.5542***	-0.5526***	-0.5523***
Unknown Ethnicity	0.1367**	0.1508**	0.1521**	0.1518**	0.1533**
Education (reference is low education)					
Education_Med	-0.1824***	-0.1810***	-0.1811***	-0.1830***	-0.1824***
Education_High	-0.0277	-0.0331*	-0.0333*	-0.0356*	-0.0347*
Month (reference is september)					
October	-0.0867***	0.0109	-0.0227	-0.0134	-0.0274
November	-0.2181***	0.0251	0.0011	0.0027	-0.0161
Year (reference is 2013)					
2014	0.0898***	0.0245	0.0201	0.0287	0.0300
2015	0.0703***	0.0054	0.0158	-0.0025	0.0116
2016	-0.0316	-0.0431*	-0.0403	-0.0534**	-0.0456*
2017	-0.0216	-0.0282	-0.0419	-0.0392	-0.0393
Type of day (reference is weekday)					
Saturday	-0.3709***	-0.3620***	-0.3607***	-0.3623***	-0.3615***
Sunday	-0.8936***	-0.8958***	-0.8941***	-0.8947***	-0.8947***
Weather variables					
Temperature		0.0680***	0.0581***	0.0448***	0.0392***
Wind		-0.0978***	-0.0904***	-0.1218***	-0.1132***
Rain		-0.0436***	-0.0530***	0.0221	0.0033
Solar Radiation		0.0587***	0.0834***	0.1067***	0.1085***
Temp Squared			-0.0125*		-0.0143
Wind Squared			-0.0046		-0.0048
Rain Squared			0.0039		0.0069*
Radiation Squared			-0.0280***		-0.0127
Wind:Sun				-0.0412***	-0.0416***
Wind:Rain				-0.0523***	-0.0499***
Wind:Rain:Sun				-0.0553***	-0.0544***
Rain:Sun				0.1023***	0.1011***
Temp:Wind				0.0294***	0.0343***
Temp:Sun				-0.0374***	-0.0209
Temp:Rain				-0.0416***	-0.0450***
Temp:Rain:Sun				-0.0303***	-0.0338***
Temp:Wind:Sun				0.0124	0.0066
Temp:Wind:Rain				0.0396***	0.0368***
All Interactions				0.0240*	0.0216
Model Statistics					
Chi-Square	68739.2	69433.4	69429.8	69545.6	69537.2
Log-Likelihood	-71329.4	-71247.4	-71240.1	-71228.5	-71226.2
AIC	142697	142541	142534	142525	142528
BIC	-605402	-605208	-605161	-605061	-605015

* variable is significant at 5% level
** variable is significant at 1% level
*** variable is significant at 0.1 % level

In the linear-in-parameters model, we see that the number of bicycle trips increases when temperatures and sunshine levels increase and decreases when precipitation and wind speed levels increase. Quadratic effects are observed for radiation and temperature, indicating a non-linear relationship between these variables and travel behaviour. Most interaction effects are significant, indicating that people use combined or holistic views of weather to make their travel-behaviour decisions.

Comparing the effects of the weather variables we see that the linear effect of wind speed is most important in all models, followed by temperature, rain, and sunshine who all have a more or less similar effect when compared across all models. Again it is difficult to assess which individual weather variable has the strongest effect on the number of bike trips in the more complicated models. We try to give an answer to this question in Section 5.3.3.

5.3.2 Comparison across modes

This section will display two tables, each comparing a model specification across modes. The goal of this comparison is twofold: first to see how different model specifications perform for each dependent variable, illustrating how some models might be better fit for one mode than for another and secondly to see how the effect of the weather varies across modes. To enable a direct comparison, we first display the simple linear-in-parameters weather model which enables direct comparisons of the parameters. Then we identify the best model by using the model with the best BIC/AIC values for each mode. When the BIC and AIC point to a different model, we use the model that contains the most parameters. This allows us to see how models perform across modes and enables a comparison of the best-fitting models effects of weather on that particular mode.

Linear-in-parameters model

The first model whose parameters will be compared across the range of travel modes is the simplest weather model, including only linear-in-parameters effect of rain, sun, wind, and temperature. This allows us to get a feel for the influence single weather variables have on each of the modes. The comparison of this simple model can be seen in Table 5.6.

We'll first quickly run through the various non-weather related results, before taking more time and space to compare how weather affects each mode and what the interesting differences are. Keep in mind that the following effects are controlled for possible weather circumstances. Looking at the socio-demographics, we see that older people are more likely to use the car or go for a walk, whilst they are less likely to use public transport or the bike. Men make less trips in total and for all modes except public transport, where the effect of gender is almost nil. Also in line with expectations is the finding that employed people travel more, especially with the car (whilst they walk less). Looking at the environment in which people travel, we see that in the Randstad area people make about the same number of total trips, but that the modal split is different: car is used less, whilst PT is used considerably more. This effect is even more pronounced in dense areas. We see that Dutch natives in general travel the most (as all other ethnicities travel less), and that western foreigners are more likely to use public transport. Non-western foreigners are much less likely to use the bicycle. When looking at the date when people travelled, we see that saturdays and sundays generally cause people to travel less, an effect that is particularly strong for public transport. There are also some monthly differences, with people travelling more by car in later months (after controlling for weather). Finally we see some relatively small differences across the years, which can be caused by a multitude of factors that are not controlled for here (chief amongst which is the sample difference between years).

Now let's turn our eyes towards the effects of weather. A general remark can be made, which is that (as expected) the effects of weather are very mode-specific. Some modes are used more when it starts to rain, whilst others are used less often. This points to substitution effects that are caused by the weather, where people decide their travel mode based on information about the weather. Based on previous literature these effects are definitely plausible, so it's good to see that our intuition and previous research is corroborated by the data. Furthermore we see that some modes are more affected by the weather than others, judging by the number of significant weather parameters and (more importantly)

Table 5.6: Comparison of negative binomial regression model with linear-in-parameters weather influence

	Total	Car	PT	Bike	Walk
Intercept	0.9245***	-0.2487***	-1.5330***	0.1374***	-1.7421***
Age	0.0220***	0.0781***	-0.2631***	-0.0601***	0.1433***
Male	-0.0685***	-0.0193*	0.0136	-0.1998***	-0.2990***
Employed	0.1319***	0.2893***	0.0247	0.0888***	-0.1272***
Density	-0.0052*	-0.0933***	0.2373***	0.0670***	0.0803***
Randstad	-0.0007	-0.0678***	0.4225***	0.0087	0.0462**
Ethnicity (Dutch native is reference)					
Western Foreign	-0.0759***	-0.0770***	0.3910***	-0.1918***	-0.1026**
Non-Western Foreign	-0.1927***	-0.0767**	0.1247	-0.5536***	-0.0948
Unknown Ethnicity	-0.0371	-0.2193***	-0.1447	0.1508**	-0.1330
Education (Low education is reference)					
Education_Med	0.0792***	0.2879***	0.1811***	-0.1810***	0.1278***
Education_High	0.1570***	0.3729***	0.3417***	-0.0331*	0.2429***
Type of day (weekday is reference)					
Saturday	-0.0946***	0.0955***	-0.6366***	-0.3620***	0.0793***
Sunday	-0.4651***	-0.2583***	-0.9986***	-0.8958***	-0.1178***
Month (September is reference)					
October	0.0307***	0.0754***	-0.1592***	0.0109	0.0726**
November	0.0218	0.0573**	-0.0485	0.0251	0.0283
Year (2013 is reference)					
2014	-0.0700***	-0.1099***	0.0217	0.0245	-0.1370***
2015	-0.1051***	-0.1540***	0.0832	0.0054	-0.2190***
2016	-0.0838***	-0.0924***	0.0568	-0.0431*	-0.1662***
2017	-0.0267**	-0.0001	-0.0322	-0.0282	-0.1051***
Weather variables					
Temperature	0.0275***	0.0034	0.0033	0.0680***	0.0422***
Wind	-0.0268***	0.0067	-0.0073	-0.0978***	-0.0241**
Rain	0.0013	0.0264***	-0.0192	-0.0436***	0.0002
Solar Radiation	0.0242***	0.0085	-0.0365*	0.0587***	0.0300**

* variable is significant at 5% level
** variable is significant at 1% level
*** variable is significant at 0.1 % level

the size of the estimated effects. We see that the weather parameters are mostly significant for the total number of trips and the number of trips by bike or walking. Public transport and the car are affected by only one variable: rain and sunshine respectively.

Now we will discuss the effects of weather per mode. For the total number of trips we see a positive effect for temperature and sunshine and a negative effect for wind speeds. We see that the positive effect of temperature is almost wholly caused by the increase in the use of active modes, whilst the use of the car and public transport are stable. This means that temperature increases cause additional demand, with no substitution taking place. People thus start making more trips with active modes, rather than using the active modes for trips they would have otherwise made using either the car of public transport. The story is almost the same for the effect of sunshine, but the (small) negative parameter for public transport trips shows that there are some substitution effects going on. For wind meanwhile we find the exact opposite story: people generally stop using the bicycle or going for walk, but they don't use the other modes much more. A completely different impact is found when it rains: here we see a positive impact on the number of car trips and a negative impact on the number of bicycle trips, whilst total trips remain stable. This indicates that people do make a substitution here, as they're inclined to use the car for trips where they normally use the bicycle. The opposite is true if there is no rain of course, but not to the same extent: since the distribution of rain is very skewed with a right-tail and a zero lower bound the minimum value of the standardized variable is much smaller than the maximum value.

Use of the car is only positively affected by rain, whilst the only weather effect on public transport use is the small negative effect of the amounts of sunshine. The number of bicycle and walking trips increase when temperatures and sunshine levels increase, whilst they decrease as a result of higher levels of wind speed. Rain has a negative effect on bicycle trips, but not on walking trips.
This indicates that increases in temperature and sunshine levels cause additional bicycle and walking demand, whilst rain causes a substitution from these active modes to the car.

Best model

One of the hypotheses posed in this work is that the influence of weather on travel behaviour can not be fully captured using the linear-in-parameters specification used in the models compared in Table 5.6. To test this hypothesis we estimated increasingly complex models for each mode and compared the model fit. The models that perform best based on AIC and BIC are displayed in Table 5.7.

We see that the the best model varies quite a bit, with 4 out of a maximum of 5 model specifications being the best fit for at least one mode. Interestingly the best model for public transport use is the one without weather, which means that within our observed range of weather values average use of this mode in the Netherlands seems to be stable. For the car the linear specification is best, probably due to the fact that the relationship between weather and car use is also not very complex: if it rains people seem to substitute active modes for the car, but no extra or fewer trips are made during non-extreme weather conditions.

Things are very different for the other three dependent variables: of course the total number of trips is highly dependent on the effect on the four constituent modes (and the 'other' mode, which we have not analysed here) so we will first discuss these specific modes. For cycling we see a very complex relationship, where all estimated parameters in the interaction model are significant. Quadratic effects were significant (as could be seen above in Section 5.3.1), but estimates were small and did not lead to a significant improvement of model fit. The more important point is thus the conclusion that the relationship is so complex and interrelated in the first place and that estimating linear-in-parameters effects only will lead to wrong conclusions about the effect the weather can have on the number of bike trips. The number of pedestrian trips meanwhile is estimated best by the Quadratic and Interaction model, which is the most complex model specification estimated. We see that the square parameters for temperature, wind, and sunshine are all significant and also very sizeable, indicating that a linear-in-parameters relationship will again lead to a wrong model specification. Less interaction prameters are significant than for the bicycle mode, with significant parameters almost all involving the effect of wind speed (the exception is the Rain:Sun parameter).We thus see that the effect of wind speed is

Table 5.7: Comparison of negative binomial regression with complex weather influence across modes

	Total	Car	PT	Bike	Walk
Best Model	QuadInt	Linear	None [1]	Interaction	QuadInt
Temperature	0.0240***	0.0040	0.0064	0.0423***	0.0367***
Wind	-0.0178***	0.0066	-0.0086	-0.1205***	0.0159
Rain	0.0149**	0.0260***	-0.0146	0.0206*	0.0146
Solar Radiation	0.0411***	0.0086	-0.0321*	0.1028***	0.0630***
Non-linear effects					
Temp Squared	-0.0052				-0.0304***
Wind Squared	-0.0091***				-0.0301***
Rain Squared	0.0007				-0.0021
Radiation Squared	-0.0040				-0.0308***
Interaction effects					
Wind:Sun	0.0019			-0.0439***	0.0285**
Wind:Rain	0.0029			-0.0515***	-0.0003
Wind:Rain:Sun	0.0005			-0.0549***	0.0035
Rain:Sun	0.0254***			0.1013***	0.0368**
Temp:Wind	0.0206***			0.0294***	0.0486***
Temp:Wind:Sun	-0.0089*			0.0144*	-0.0289***
Temp:Wind:Rain	0.0055			0.0418***	0.0131
Temp:Sun	-0.0156**			-0.0363***	0.0054
Temp:Rain	-0.0115**			-0.0408***	-0.0042
Temp:Rain:Sun	-0.0049			-0.0292***	0.0117
Temp:Rain:Sun:Wind	0.0079			0.0241**	0.0059

[1] We display the linear model here, so we can see the (lack of) influence of weather parameters
* variable is significant at 5% level
** variable is significant at 1% level
*** variable is significant at 0.1 % level

particularly dependent on the values of the other three weather variables involved, which makes sense intuitively: higher wind speeds become much more uncomfortable for walking when it is cold, wet, and/or overcast.

> Which specification of weather is best changes for each mode. The effect of weather on public transport use is so small that the model without weather effects is selected as the best model. For the car the linear-in-parameters model works the best, whilst for the bicycle the model estimating interaction effects (but no quadratic effects) is the best model. Finally for both the total number of trips and the number of walking trips the model including both quadratic and interaction effects is best able to capture the effect of weather on travel demand.

The effect on the total number of trips is mostly the result of the effects on the discrete modes: we see that temperature and sunshine improve the number of trips quite drastically, mostly due to the positive effects of these variables on the use of the active modes. Conversely wind has a negative effect, again mostly caused by the negative effect on especially bicycling and to a lesser extent walking. Some interaction parameters are significant, usually those that were significant for both of the active modes or were particularly sizeable for bicycling. The one surprising finding here is the positive estimate for the linear effect of rain, but perhaps this is the result of the negative effects being caught in the interaction terms.

5.3.3 Effect of weather changes

A returning problem when trying to see the effect weather can have on the number of trips is the fact that the more complex models are difficult to interpret. This is partly the result of the negative binomial model specification with its logarithmic transformation (see Equation 3.3) which causes all of the variables to be interacted to some extent by default and partly due to the non-linear and interacting model specifications. To improve the intuitive understanding of the estimates we've used two different methods. Both reduce the complexity of the model specification to some degree to provide easier and more interpretative findings. The first is the calculation of marginal effects, which are the percentage change caused in the number of trips per day per person for each mode as a result of a 1 unit change in weather variables. These marginal effects are calculated using both the simple and best model specifications as have been outlined in Section 5.3.2. The calculations are explained and results are given in Section 5.3.3. The second is the use of proto-typical days for autumn to see how various weather conditions influence the number of trips that have been made, which is explained in Table 5.3.3.

Marginal Effects

To get a better feel for the impact of changes in weather the marginal effects are calculated. Marginal effects here are the percentage changes in number of trips per mode due to a 1 unit increase in the weather variables. Since the weather variables are standardized this means an increase of 1 standard deviation for each weather variable. It is thus crucial to know what these standard deviations are for each of the weather variables. The standard deviations are thus given in Table 5.8.

Table 5.8: Standard deviations for the weather variables used in the regression analysis

Weather variable	St. D.	unit
Temperature	3.78	Degrees Celsius
Wind Speed	2.12	m/s
Rain	0.19	mm/h
Sunshine	67.2	W/m2

The marginal effects are calculated for both the simple model specification and the best model specification, as determined above in Section 5.3.2. Before we give the calculated marginal effects, first we will outline the procedure used for the calculations below.

The fitted model is used to predict the number of trips for all roughly sixty thousand person/day combinations used to estimate the model. The mean value is calculated, and is the baseline value for the number of trips. Then for each separate weather variable the original data is modified by adding one standard deviation to all values of this variable. The model is used to predict the new number of trips for each row: the mean value across all rows is the new modified value. The marginal effect then is the difference between the modified and the baseline values, divided over the baseline value and multiplied by a 100 to get a percentage change. This procedure is repeated for each of the weather variables and for all modes. As said, we have repeated this procedure for two models per mode, the first is the simple linear-in-parameters model and the second is the best model. First we will present the marginal effects for the simple model in Table 5.9.

Table 5.9: Percentage change in number of trips caused by a 1 standard deviation in-/decrease of weather variables in the linear-in-parameters weather model

	Total	Car	PT	Bike	Walk
Temperature	2.79	0.34	0.33	7.04	4.31
Wind Speed	-2.65	0.68	-0.73	-9.32	-2.38
Rain	0.13	2.67	-1.90	-4.27	0.02
Sunshine	2.45	0.85	-3.59	6.04	3.05

The marginal effects are more or less as expected, based on the parameter estimates of the model. We see that the total number of trips is hardly affected by rain at all. A decrease in total trips is caused by wind increases, whilst higher temperatures and sunshine levels cause increased total demand. Car trips are not affected very much by either temperature, wind speed, or solar radiation. Increasing rain does cause an increase of car trips, probably due to a corresponding decrease in bicycle trips when it rains. The bicycle is affected strongly by all four weather variables, positively by temperature and solar radiation and negatively by wind and rain. Finally walking is not affected by rain intensity, whilst the effects of the other weather variables show the same direction but a slightly smaller size compared to the bicycle.

> The marginal effects indicate that the bicycle is the most strongly affected mode, with positive effects from temperature and sunshine and negative effects from wind speed and rain. The largest effect is that of wind speed. Rain seems to mostly cause substitutions to take place between the car and the bicycle, whilst temperature, sunshine, and wind speed have an effect on total demand.

We have also calculated marginal effects with the best models. The calculation is done in exactly the same way as before, but now with the model that is selected as the best one per mode. The values for the weather variable in question are thus increased by one standard deviation. These new values are used to calculate new interaction and quadratic variables as well. The results can be seen below in Table 5.10.

Table 5.10: Percentage change in number of trips caused by a 1 standard deviation increase of weather variables in the best models for each mode

	Total	Car	PT[1]	Bike	Walk
Best Model	QuadInt	Linear	None [1]	Interaction	QuadInt
Temperature	1.99	0.34	0.0	5.68	0.26
Wind Speed	-2.98	0.68	0.0	-9.68	-2.10
Rain	1.38	2.67	0.0	2.69	2.38
Sunshine	3.54	0.85	0.0	10.17	3.28

[1] Since the best model is that without weather influence the marginal effects of all weather variables are zero.

Although most marginal effects have not changed much, we can see some interesting differences between the calculations using more complex models and the calculations with the linear-in-parameters models. Obviously the zeros across the board for public transport stand out first. They are the result of the fact that the best model for public transport use does not include any influence of weather, meaning that there are no marginal effects of weather variables.

The second very surprising change is the positive sign for the influence of rain on bicycle and pedestrian trips, which is both contra-intuitive and contrasts with the results from the linear-in-parameters models. We think that this is caused by the procedure to calculate the effects, which increases the amount of rain for every single trip. This will cause outliers where rain is introduced in otherwise very sunny days, which does not happen in reality. This results in model predictions for values that are not within the observed range, which can cause mistakes to occur. So where the interactions are meant to capture the fact that weather parameters naturally co-occur the predictions are partly made for parameter values that never co-occur. This issue is mended in Table 5.3.3, where we predict the number of trips for plausible days.

> We find puzzling positive effects of rain on the use of the bicycle within the complex model. This might be caused by the calculation procedure, where we introduce rain in all observed days. Perhaps this causes unrealistic combinations of weather patterns (rain during a sunny, warm, and windstill day for example). The predictions for these unrealistic combinations are likely to be wrong, causing mistaken marginal effects.

Proto-typical days

The second way to enable more intuitive insight into the effects of weather we've composed several typical days for the months of autumn. These days are based on our intuitive understanding of the weather, ranging from days with warm, dry weather to rainstorms. These days thus do use realistic combinations of weather variables, allowing our models to give predictions for weather that could be expected to occur during the observation period.

For each of these days we calculate the predicted number of trips using our models, first with the linear-in-parameters model and then with the best model. The labels and values for the weather parameters for these typical days are given in Table 5.11. To enable a comparison of the number of trips during the selected typical days to more normal weather we have calculated trips for a day with average weather as well.

Table 5.11: Labels and weather values for typical days

Label	Temperature Degrees C	Wind Speed m/s	Rain Intensity mm/h	Solar Radiation $W/m2$
Mean	11.39	4.06	0.1	104.28
Rainstorm	10	15	1.5	25
Rain, Overcast	10	3	0	10
Wind, Rain	10	7	1	100
Near-freezing	2	4	0	100
Good	15	2	0	150
Great	25	2	0	250

For these proto-typical days we use the fitted regression models to predict the number of trips, giving us an intuitive understanding of the effect weather differences can have on travel demand for each travel mode. Importantly for the non-weather variables, we'll use the reference alternative for the dummy variables and the mean value for continuous variables. This means that the predicted number of trips are calculated for female low-educated and unemployed people, aged somewhere in their forties, who lives outside the Randstad and travel on a week-day. We are thus not trying to give mean predictions for these days, rather we're looking at the differences in number of trips within a mode across the full observed range of weather circumstances.

Table 5.12: Predicted trips per day of linear-in-parameters model for prototypical days

	Total	Car	PT	Bike	Walk
Mean weather	2.79	0.90	0.10	1.01	0.54
Rainstorm	2.33	1.12	0.09	0.40	0.45
Rain, Overcast	2.71	0.87	0.10	0.97	0.52
Wind, Rain	2.66	1.03	0.09	0.70	0.51
Near-freezing	2.60	0.88	0.09	0.87	0.48
Good	3.00	0.88	0.10	1.26	0.58
Great	3.36	0.90	0.09	1.64	0.68

The difference in the total number of trips is quite sizeable, with a deviation of a full trip (roughly one-third of the maximum number of trips) between the worst and best days selected here for the total number of trips. This would mean that good or bad weather can cause and explain a significant amount of daily travel variation. The effect is less pronounced for car trips, where 0.87 trips is the minimum, and 1.12 the maximum, and public transport trips which seem to be hardly affected by the weather at all. For bicycle trips however the effect is even more sizeable, varying between 0.4 trips per person per day during stormy conditions to 1.64 trips during great conditions. The variation that can be observed for walking trips meanwhile is relatively similar to that caused in car trips.

We can also see that car trips are maximized by rainstorm conditions, probably due to substitution from bicycle and to a lesser extent pedestrian trips to the car. Both bicycle and walking trips are at their maximum during great weather conditions, probably due only in small part to substitution: instead these favourable conditions create extra travel demand, m in the form of increased leisure cycling and walking. Zooming in on the active modes, we can see that windy conditions are in general very detrimental to cycling demand (looking at both the stormy and windy/rainy days). There is an effect of temperature, as indicated by the lesser amount of trips for the near-freezing day (0.87 vs. mean 1.01), but this effect is not quite as large as the effect of increased wind speeds.

We have run the same calculations using the best model (as indicated in Table 5.3.2 for each mode instead. For the total number of trips and bicycling and walking these estimates differ from the numbers calculated above. For the car the linear-in-parameters model is used, whilst for public transport the predicted number of trips are not affected by weather at all.

Table 5.13: Predicted trips per day for prototypical days, predictions using best-fitting models

	Total	Car	PT	Bike	Walk
Best Model	QuadInt	Linear	None [1]	Interaction	QuadInt
Mean	2.94	0.91	0.10	1.08	0.60
Rainstorm	1.98	1.12	0.10	0.42	0.11
Rain, Overcast	2.77	0.88	0.10	0.99	0.52
Wind, Rain	3.05	1.03	0.10	0.67	0.55
Near-freezing	2.62	0.89	0.10	0.90	0.44
Good	3.02	0.90	0.10	1.27	0.56
Great	2.98	0.92	0.10	1.40	0.42

[1] Since the best model is that without weather influence the predicted number of trips are constant

In comparison to the estimation by the simple linear-in-parameters model some interesting differences can be observed. Below we'll highlight these differences, which are important and show that the linear-in-parameters model can lead to mistaken predictions, especially at values at the extreme end of the observed range. These differences of course can only be observed for the active modes and the total number of trips, as only these models use more complex specifications.

The differences between the predictions of these complex specifications and the linear-in-parameter specifications get bigger for days with values towards the extreme end of the observed range, such as the rainstorm and great days. For the rainstorm day we see that the predictions for the number of total and the number of pedestrian trips are much lower with the complex specification, probably due in part to the quadratic specification. These results are more in-line with what we would expect during such an extreme weather event compared to the relatively smaller effect estimated by the linear-in-parameters model.

For the great day we see similarly sizeable differences, although in a different way. The linear-in-parameters model estimates that the increase in weather conditions from the good to great day causes a fairly large further increase in the number of total, bicycle, and walking trips. In reality we expect this increase to have a smaller effect, as the weather during the good day will prompt most people to already increase their number of trips. This expectation is reflected better by the complex model specification, especially for the bicycle mode. For the number of pedestrian trips we even see a decrease in the number of trips compared to both mean and good conditions, indicating that ideal walking conditions might be at temperatures below 20 degrees Celsius. Whereas the linear-in-parameters models thus assume that the positive effect of increases in temperature are the same across the entire observed range our complex models are able to replicate reality better, as the positive effect of increases level off after a certain point.

The estimated number of trips when calculated by the more complex model give more realistic predictions, especially for the days with relatively extreme weather (rainstorms and very sunny, warm days). For rainstorms these models are able to predict a larger decline in the number of trips than the linear-in-parameters models, which seems to be more realistic. For great days these models are able to predict a smaller increase in the number of trips compared to better than normal weather circumstances. The models thus estimate that the increase in weather circumstances from good to great will not have a large impact on travel behaviour, which is more realistic as well.

5.3.4 Spatial Differences

In the models before we haven't investigated the difference in reaction to weather between people living in cities or urbanized regions and people living in rural regions. To investigate this difference the linear-in-parameters model is re-estimated for all dependent variables with interaction effects between weather parameters and a dummy-variable with two levels, rural and urban. This results in two different linear weather parameters for urban and rural environment. This allows us to reflect on these differences and determine whether urban environments do in fact prompt different responses to weather than rural environments. We chose to use the linear-in-parameters models as we are primarily interested in the differences between the two spatial environments, rather than a most accurate appraisal of the effect of weather on travel behaviour. The linear-in-parameters models are much easier to interpret than the complexer interaction and/or quadratic models.

To allow for a comparison we give both the estimated parameters in Table 5.14 and the predicted number of trips per mode during several weather circumstances in Table 5.15. Parameter estimates are given for the effect of the four temperature variables in each of the five models. Furthermore the estimate for the urban dummy on the number of trips directly is given as well. Please note that the significance of the urban variable is calculated with respect to the deviation from its rural counterpart, rather than the deviation from zero.

Table 5.14: Weather parameters for urban and rural environments

		Total	Car	PT	Bike	Walk
Urban Constant		-0.0283***	-0.1822***	0.4680***	0.0651***	0.1463***
Rural	Temperature	0.0260***	0.0009	0.0395	0.0802***	0.0222
	Wind	-0.0179***	0.0080	-0.0178	-0.0563***	-0.0273*
	Rain	0.0056	0.0246***	-0.0206	-0.0500***	0.0212
	Solar	0.0275***	-0.0019	-0.0746**	0.0783***	0.0613***
Urban [1]	Temperature	0.0293	0.0014	-0.0161*	0.0601	0.0633*
	Wind	-0.0355***	0.0093	-0.0062	-0.1381***	-0.0251
	Rain	-0.0023	0.0274	-0.0172	-0.0385	-0.0163*
	Solar	0.0210	0.0186	-0.0129*	0.0381**	0.0021***

[1] We display calculated effects of the weather variables in urban areas. The significance however is indicated for the interaction parameter itself.
* variable is significant at 5% level
** variable is significant at 1% level
*** variable is significant at 0.1 % level

In terms of the number of trips during average weather conditions we can see that slightly less total trips are made in urban environments, mostly due to a sharp decrease in the number of car trips. There are more bicycle, public transport, and pedestrian trips in urban environments compared to rural environments, with especially the increase in public transport trips being noticeable. Of course public transport networks are in general more developed for urban areas, so this result is conform our expectations. More exciting and innovative results are found when we compare the effect of weather for each travel mode.

Table 5.15: Predicted number of trips for urban and rural environments, as calculated by linear-in-parameters models

	Total	Car	PT	Bike	Walk
Rural Environment					
Mean	2.85	1.01	0.08	0.96	0.48
Rainstorm	2.60	1.26	0.07	0.44	0.45
Rain, Overcast	2.73	0.99	0.09	0.88	0.44
Wind, Rain	2.82	1.14	0.07	0.68	0.51
Near-freezing	2.66	0.99	0.07	0.80	0.45
Good	3.02	0.99	0.08	1.18	0.52
Great	3.37	0.99	0.08	1.64	0.61
Urban Environment					
Mean	2.77	0.84	0.12	1.02	0.56
Rainstorm	2.19	1.05	0.11	0.35	0.43
Rain, Overcast	2.71	0.80	0.13	1.03	0.56
Wind, Rain	2.58	0.97	0.11	0.69	0.49
Near-freezing	2.58	0.82	0.13	0.90	0.48
Good	3.00	0.83	0.12	1.29	0.61
Great	3.34	0.86	0.12	1.61	0.73

People travel less by car in urban areas compared to rural areas. There are however more public transport, bicycle, and pedestrian trips in urban areas.

For the total number of trips we see effects of temperature, wind, and sunshine. The effect of rain is very small and statistically insignificant for both urban and rural environments, with only a small difference between the two environments. There is also only a small difference between the two environments for both temperature and sunshine, where temperature has a bigger effect in rural areas and sunshine in urban areas. The biggest difference between the two environments however can be observed as a result of wind speed. Unexpectedly the effect of wind speed on the total number of trips is much stronger in urban environments than in rural environments, indicating that higher wind speeds are more likely to cause people to cancel their trips in urban areas than they would be in rural areas. We can see that the range of predicted trips is thus also much larger in urban environments, resulting from fewer trips being made during inclement weather conditions (2.19 trips in rainstorm conditions) when compared to rural environments (2.60 trips). The difference caused by good or great conditions meanwhile is next to nothing.

Use of the car is affected only by rain in both environments, with this positive effect being roughly equal in size for both environments. Sunshine levels have no effect in rural environments, but a noticeable positive effect in urban environments can be observed. In rural environments then there is no difference in predicted number of trips for the days without rain, whilst rain causes fairly large increases in car use. This effect of rain is almost similar in size in urban environments, but some small additional variation can be observed resulting from different sunshine levels.

Public transport use is not affected by the weather to a large, or even any, extent as has been discussed previously in this chapter. We do see some differences between urban and rural environments however. There is a negative effect of solar radiation on public transport use in rural environments, which is much smaller in urban environments. In urban environments however we have an additional negative effect of temperature, whereas temperature has no significant effect in rural environments. In both environments we see only small deviations in the predicted number of public transport trips, which mostly increase during cold and overcast days for both areas.

The influence of the weather on the use of the bicycle is much stronger in general and we can also see more interesting differences between the two environments. In both areas all weather variables have a sizeable effect on cycling behaviour. In rural environments we see that the size of the effect of all variables is roughly similar, whilst in urban areas the effect of wind is much stronger than the effect of the other variables. The directions are always as expected: positive for temperature and solar radiation and negative for wind and rain. The effect of temperature, rain, and sunshine however is thus less strong in urban areas whilst the effect of wind is much stronger in urban areas than in a more rural environment. This difference in the effect of wind speeds on bicycling use could also be seen in the total number of trips, where wind speed was much more influential in urban environments as well.

> The number of cyclist trips are strongly affected by weather in both environments. In an urban environment however the effect of wind is much stronger than in the rural environment, whilst the effects of the other three variables (temperature, rain, and sunshine) are stronger in rural environments.

For pedestrians we can also see interesting differences between the two environments. The effect of wind speed is relatively similar, which contrasts with the big difference found for bicycle trips. Furthermore in rural environments we see a positive effect of sunshine which completely disappears in urban environments, whilst in urban environments the positive effect of temperature is much larger. Finally we see no effect of rain in rural environments, whilst in urban environments a sizeable negative effect of rain can be found. This also means that in urban environments we expect slightly larger fluctuations, especially when weather gets cold, wet, and is paired with higher wind speeds.

In summary the effects of weather differ quite a lot between the two spatial environments, but we can't directly say that one or the other environment is more sensitive to weather. We must also keep in mind that our measurements of weather often take place outside the city, meaning it is impossible to account for the difference in actual weather between cities and rural areas. Cities often create small micro-climates with lower wind speeds and higher temperatures: perhaps people in cities are more sensitive to higher wind speeds because they occur less often.

5.4 CONCLUSION

There are many conclusions that can be drawn based on the results from the multitude of models we estimated in this chapter. First of all we can confidently say that the effect of weather on the number of trips is very mode-specific, which is in line with our expectations. Only estimating a model for the total number of trips would hide many interesting effects that are specific to one mode. For these total number of trips we see that total demand is increased by rising temperatures and increases in sunshine levels, whilst it is decreased by higher wind speeds.

When turning our attention to the mode-specific results, we see that in general the active modes of bicycling and walking are most affected by the weather. Use of the bicycle is increased by increases in temperature and sunshine, whilst it is decreased as a result of increases in wind speed and rain. The effects on the number of pedestrian trips is generally somewhat smaller, with these trips being positively affected by temperature and sunshine as well, whilst only wind speed is found be a negative factor. For the car we only see a positive effect of rain, whilst public transport use is hardly affected by weather changes at all. We thus see substitution effects, where people swap from one mode to another, most clearly during rainy conditions when people switch from the active modes to the car. Increases in temperature and sunshine levels meanwhile mostly cause increases in total travel demand, driven by more use of the active modes.

The differences can be sizeable, especially for the use of the bicycle, with the models predicting up to 4x as many trips using the bicycle during great weather conditions compared to rainstorm conditions and a 60% increase from mean days to these great days. Effects of weather on the other modes are smaller, although we still see changes in car use in the range of 5% during relatively normal weather conditions. When weather gets particularly nasty, such as windy and rainy conditions, car use is even predicted to increase by around 15%. These increases are likely to cause significant problems on the road with respect to congestion and road safety, especially as in these conditions the performance of the infrastructure is affected negatively [van Stralen et al., 2015].

The fact that our models find significant interaction effects gives more evidence to suggest that linear-in-parameters models are not the best specifications of weather effects. Conceptually this means that estimating single effects of weather variables is not a good reproduction of the decision-making process of people. Weather phenomena always co-occur and are probably perceived in relation to one another. Estimating singular effects would reduce this holistic view of the weather to more or less separate dimensions, when the effect of rain on travel behaviour for example probably depends on the co-occuring values for temperature and wind. Whilst estimating interaction effects is one way of dealing with this problem we have found that this makes interpretation of the results more difficult. Perhaps other approaches could be more successful, as long as they account for the fact that the effect of one variable depends on the values of other variables.

We also find interesting differences between urban and rural areas with respect to the effect of weather on travel demand. We find that especially interesting differences for cyclists: in urban areas cyclists are more sensitive to wind speeds but less sensitive to temperature, rain, and sunshine. These differences might be the result of varying travel patterns, different availability of alternatives, and the fact that people need to travel for longer distances in rural areas. But they might also be (partially) caused by the protection that the urban environment offers its travellers in the form of buildings reducing the effect of the wind and trapping some of the heat, thereby creating a more pleasant environment for cycling and walking during autumn.

The final finding that we want to highlight is the fact that the weather is able to explain away some of the effects of other variables on the number of trips. If we do not account for weather we would for example see that people travel less in the months of October and November compared to September. We find that this effect is entirely caused by the weather variation between these two months. Whilst this effect was expected for the influence of months, the exact same phenomenon can be observed for differences between the highly urbanized Randstad provinces and the non-Randstad provinces. The Randstad is situated close to the sea, which causes consistent differences in weather between the Randstad and the rest of the country. When accounting for weather we see that the difference in the number of bicycle trips between these two regions completely disappears. We are also able to explain some of the variance between years. By incorporating and controlling for weather effects on the number of trips we could thus more reliably estimate long-term trends in the number of trips that are driven by other factors than the weather.

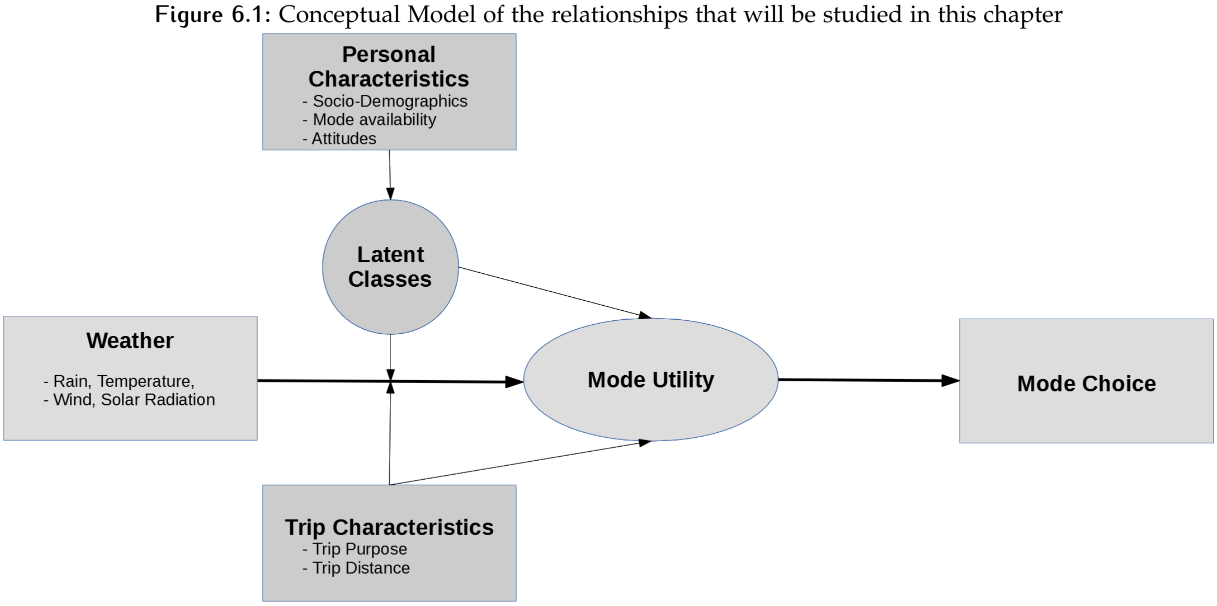
6 MODE CHOICE

This chapter will investigate the effect of weather on mode choice, with a specific focus to uncover the heterogeneity within the population with respect to this relationship. In terms of heterogeneities, we expect that people will have different reactions to weather as a result of their current travel patterns, different perceptions of weather, and different spatial environments. To give examples, we expect that people who are more used to riding a bicycle will not be as sensitive to increasing wind speeds than people who have less experience on the bicycle.

To estimate the effect of weather on mode choices we use trips made in the mobility panel Netherlands as the dependent variable. We have connected this data-set with weather data from weather stations, which allows us to estimate the effect by using discrete choice modelling techniques.

To uncover heterogeneities we could estimate model where we specify interactions between socio-demographics and the reaction to weather, which would give us different sensitivities to weather for multiple segments of the population based on the socio-demographics. Such an approach would however result in an extreme amount of parameters. An easier way of uncovering the heterogeneities can be accomplished by using a latent class discrete choice model. Here we use a discrete and pre-specified number of latent classes. Each latent class will respond differently to weather variations. Membership of the latent classes is affected by personal characteristics, such as socio-demographics, mode availabilities, and attitudes. Since the classes have different responses to weather variables these personal characteristics then serve as variables mediating the relationship between weather and travel behaviour. This leads to the following conceptual model visualized in Figure 6.1.

Figure 6.1: Conceptual Model of the relationships that will be studied in this chapter



The chapter builds on the findings from Chapter 4 for its connection method and incorporates some of the findings of Chapter 5 to aid parameter selection with the aim of using a parsimonious but 'complete' model. The structure of the chapter is similar to the earlier results chapters, with a description of the data in Section 6.1, followed by a model description in Section 6.2. Then we give the results in Section 6.3 and finally give conclusions from this chapter in Section 6.4.

6.1 DATA DESCRIPTION

The data set that is used for the modelling is the same as the one used in Chapter 4 which was described in Chapter 3. Below follows a very quick summary: the data set consists of a travel behaviour part, which is made up of the first five waves of the MPN, and a weather part, which is supplied by the KNMI. The data set is reduced to those trips that originate from the residential location (as this is where mode decisions are made) and for each trip a choice set is approximated using information about the trip and the traveller. The travel diaries are sampled in autumn, mostly during October and the first weeks of November with the occasional day falling at the end of September.

This chapter introduces some new variables. There are two types of variables, with the first being variables that vary for each trip and the second being variables that vary for each traveller. The sections Section 6.1.1 and Section 6.1.2 describe these kinds separately.

6.1.1 Trip Variables

Trip variables can roughly be divided into travel behaviour related variables (trip distance and travel purpose) and weather variables. The travel behaviour variables will be discussed first, followed by a discussion of the weather variables.

As said before two travel variables are used in the model: trip distance and travel purpose. Travel time is not included as this variable is highly related to trip distance. Travel costs would have been an interesting inclusion, but for practical reasons they were not included in the model. This would be an important omission if the goal of the model would be to most accurately predict mode shares. That is however not the goal of this thesis, as the main objective is to uncover the influence of weather on mode choice. As such the omission of travel costs is not very important.

The travelled distance is simply operationalized as the number of kilometres travelled during the trip. Two operationalizations of travel purpose were considered: the first operationalization would use dummy-coding with three levels: trips for work, trips for education, and leisure trips. The second operationalization would create an ordinal scale, ranging from pure leisure trips (tours/hiking/biking for fun) to utilitarian trips (for work) with other purposes ranked somewhere in between. We have chosen for the first operationalization, as the effects of work and education trips would be too different to combine into one scale variable. Estimating two dumy-coded variables will allow us to more accurately see the differences between not only leisure and utilitarian trips, but also between work and educational trips.

The other trip variables are the weather circumstances that were connected to the trip. As a metric again the weighted daily average is used, following the conclusions from Chapter 4 that this metric is best able to capture the influence of the weather on travel behaviour. Many weather variables are collected by the KNMI and including all of them would make the estimation and interpretation very complex. For this reason the conclusions from Chapter 5 were used and variables that had no significant impact on the number of trips with any mode were not used for the modelling procedure. This means that four weather variables are used: temperature, wind speed, precipitation, and solar radiation.

> In addition to the weather, we use the distance and purpose of the trip as explanatory variables in our models.

6.1.2 Person Variables

As discussed in Section 6.2.2 latent class discrete choice models use a class-membership model to assign probabilities of belonging to a specific latent class to individuals. By including personal variables in the class-membership model of the Latent Class choice model we're able to estimate the probabilities that an individual belongs to each of the classes, dependent on their personal characteristics. This is thus where the LC model used in this chapter extends on the simple MNL-models used in Chapter 4 the most and thus were the most new variables are introduced compared to this model, although some have been seen previously in Chapter 5. Three groups of such variables can be distuingished: (1) socio-demographics, (2) attitudes, and (3) mode availability.

Socio-Demographics

The socio-demographics that are used in the class-membership model are sex, employment status, education and the urban density at the residential location. Three attitude variables are used, which are the attitudes of the respondents to travel modes. The socio-demographic variables that are used in the model are sex, age, education, and employment status. Sex is dummy-coded, where a 0 is used for female respondents and a 1 for male respondents. Age is an ordinal variable, ranging from 1 to 10 where a 1 consists of people aged below 12 and a 10 of people aged above 80. The steps are 5 years for the first few levels, then growing to be 10 years for the last levels. Education is also dummy-coded with low-education as the reference alternative. Low-education is defined as having not completed at least secondary vocational education (MBO). This means people who have either not graduated high-school (yet) and people who have only graduated lower-tier high school. Medium-education is defined as having completed at least higher-tier high school (havo/vwo) or secondary vocational education (MBO). Finally high-education is defined as having attained at least a Bachelor's degree from a University of Applied Sciences (HBO) or Research University (Universiteit). Employment status then is again just a dummy-coded variable, where a 1 means that the person in question is employed and a 0 that the person is unemployed.

Attitudes

Attitudes were collected as part of a post-questionnaire distributed in both Wave 2 and Wave 4 of the MPN. In this post-questionnaire attitudes towards travel modes were collected by using a scale of seven questions for each mode. Under the assumption that the response to these individual questions is caused by a latent attitude towards the travel mode a factor analysis is used to determine the value of this underlying attitude for each individual.

The factor analysis is performed using SPSS, using a data-set where the post-questionnaires from Wave 2 and Wave 4 are added together resulting in a total sample size of 12,027. For each of the modes, only one factor has an Eigenvalue bigger than 1 and thus only one factor is extracted. The questions and factor loadings on this sole factor are given in Table 6.1.

Table 6.1: Questions related to Mode Attitudes and calculations of factor components

	Modes' Factor Loadings			
Questions	Car	Train	BTM	Bicycle
Travelling by (mode) is comfortable	0.859	0.871	0.891	0.828
Travelling by (mode) is relaxing	0.771	0.853	0.876	0.864
Travelling by (mode) saves time	0.765	0.688	0.788	0.662
Travelling by (mode) is safe	0.742	0.576	0.553	0.688
Travelling by (mode) is flexible	0.789	0.735	0.820	0.783
Travelling by (mode) is satisfying	0.852	0.882	0.897	0.877

As can be seen in Table 6.1, almost all factor loadings were higher than 0.7, whilst all loadings are higher than 0.5. The factor scores for all attitudes are thus calculated using all six items. The factor scores were calculated using a weighted regression that was then standardized, resulting in factor scores with a mean of 0 and a standard deviation of 1. A cronbach's alpha is calculated to give an estimate of the reliability of the items. For all factors this reliability score was higher than 0.8, indicating a reliable scale. The factor labels, statistics, and correlations between factors are given in Table 6.2.

Questions related to attitudes were only distributed in Wave 2 and Wave 4 of the five total waves of the MPN. This means that we have not recorded attitude data for all respondents within the entire MPN. People who were not part of either Wave 2 or Wave 4 were thus removed from the analysis, resulting in a loss of roughly 10,000 trips (from slightly more than 68,000 trips). People who were a part of both waves also present a problem for the latent class model, as the class-membership model

Table 6.2: Summary of attitudinal factor scores

Factor Label	α	Explained Variance(%)	Correlations 1.	2.	3.	4.
1. Car Attitude	0.881	60	1	-0.102	-0.113	0.052
2. Train Attitude	0.864	64		1	0.667	0.240
3. BTM Attitude	0.894	66			1	0.170
4. Bike Attitude	0.873	62				1

supposes that variables are constant for all individuals. Whilst it would be interesting to see how longitudinal changes in attitudes within the same individual affect the sensitivity to weather, such an analysis is not conducted here in the interest of time. For this reason the most recent questionnaire was used to determine the attitude scores throughout all years.

Mode Availability

The availability of travel modes form the last group of variables that are part of the analysis. We use two direct mode availability variables, one for the car and one for the electronic bicycle. These variables are dummy-coded, with a 0 meaning that the respondent doesn't personally own a car or e-bicycle and a 1 signifying that the person does own a car and/or bicycle. Ownership of public transport is not directly coded, as no clear variable can be used with the same purpose (as one doesn't own public transport in the same way one can own a car or a bicycle). Of course the same is true for the pedestrian mode, which is why this mode is also not included. We chose to use the e-bicycle as we posit that ownership of this mode could potentially lead to more robust travel behaviour with respect to weather: since the e-bicycle offers all of the advantages of the bicycle and combines them with a relatively easy effort that needs to be made by the bicyclist it's quite likely that he or she will keep using the e-bicycle during windy conditions that might prompt users of a normal bicycle to switch to public transport or cancel their trip altogether. The third variable related to mode availability is whether or not people own a driver's license and are thus able to drive a car. Again, this variable is included as we think that being able to drive a car might mean that people switch their mode choice earlier: for a trip where they normally cycle they can also just take a car if weather conditions become unfavourable.

> We use three kinds of variables related to the persons making the mode-choice decisions: socio-demographics, attitudes, and mode availability. The attitudes respondents' attitudes towards four travel modes: the car, the train, bus, tram and metro (btm), and finally the bicycle. In terms of mode availability we include whether or not someone owns an electronic bicycle, a driver's license, and/or a car.

6.2 MODEL DESCRIPTION

As is described in Equation 3.2.2 this chapter uses latent classes to uncover the heterogeneity in the population with respect to the influence of weather on travel behaviour, ideally uncovering groups that react differently to the same weather circumstances. The model consists of both a traditional utility function (here specified as a MNL model) where the unit of analysis is the trip itself and a class-membership model, where the unit of analysis is the person making trips. This entails that both model components use different variables, as the utility function uses variables that are unique for each trip whilst the class-membership model uses variables that are unique for each person.

A latent class model consists of two parts. The first part is the utility function, which assigns utility to an alternative based on the value on attributes of the trip (its distance, its purpose, and the weather circumstances). For each discrete class the calculated utilities are different, meaning that the same circumstances can lead to different estimated choice probabilities for each mode. The second part is the class-membership function, which estimates the probability that a traveller belongs to each discrete class based on personal characteristics. The final choice probabilities are thus dependent both on the attributes of the trip and on the attributes of the traveller.

6.2.1 Utility Function

The utility function for the models is linear-additive, which is in-line with conventional RUM-MNL specifications. This contrasts with the recommendations from Chapter 5 to consider the combined nature of weather, where all individual types co-occur. We choose for the linear-additive model due to two reasons. The first is the fact that latent class log-likelihood functions are already very complex, making estimation very difficult due to local optima. If we make the utility function of each class more complex by adding many interaction parameters estimation is likely to become even more difficult, if not outright impossible. Furthermore the interpretation of the interaction estimates is a non-trivial task, which is a problem in latent class models as we have estimates for each class that we need to be able to interpret in relation to one-another.

We use a linear-additive utility function with no interaction effects, despite the recommendations following from Chapter 5. We made this decision due to the fact that interacting utility functions would be very difficult if not impossible to interpret for the multiple discrete classes, whilst estimation would become more complex as well.

A technical detail is that since no variables are mode-specific, we need to set one mode's utility to zero. This will be the reference alternative and the utilities of all other modes can be estimated in comparison with this reference alternative. We have decided to use the car as this reference alternative. For the other modes the utility function consists of three parts: a trip-specific part, a weather-specific part, and a part containing an interaction between trip and weather variables. We have also added alternative specific constants since the alternatives are labelled.

For the trip-specific part we estimate the effect of trip distance and trip purpose. Since trip purpose is dummy coded with three levels (leisure, work, education) we use two variables to estimate the effect of the latter two, thereby automatically uncovering the effect of the first purpose. For the weather-specific part we estimate the linear-additive effect of the four weather variables identified above (temperature, wind, rain, sunshine). It's worth pointing out again that these weather variables are standardized so that the mean is 0 and the standard deviation is 1. For three of the four modes the combination of trip-specific and weather-specific variables is the complete utility function, but for the bicycle we have also estimated some interaction effects between trip purpose and the weather.

The goal here is to explore whether such a specification will enable us to see if different trip purposes are affected differently by weather conditions. To do so we have included eight interaction variables in total within the bicycle utility function: four variables each for the two purpose variables (work and education), where each of these four variables correspond to one weather variable. This means that we can estimate and hopefully gain insight into for example how increasing temperatures will affect leisure, work, and educational trips separately for the bicycle. The decision to estimate terms with the bicycle mode only is informed by the fact that this mode is affected by weather the most (as evidenced in Chapter 4 and Chapter 5), which makes this the most fertile ground for the exploration of possible interaction effects.

Finally for the latent class models we fix some variables across classes, ensuring that the classes pick up on variation within other variables. We allow variation for the weather-related parameters, such as the influence of temperature, and the interaction parameters between trip purpose and weather influence for the bicycle mode. Finally we also allow for variation between the alternative specific constants. This means that the distance and purpose parameters are fixed. Table 6.3 contains all parameters and indicates whether or not they have been varied across classes.

Table 6.3: Overview of all estimated parameters in the utility function

Parameter	Description	Varied across classes
Alternative Specific Constant	Estimates mode shares when other variables are set to zero	Yes
Distance	The distance travelled during the trip	No
Purpose work	Dummy variable, set to 1 if the purpose of the trip is work-related	No
Purpose edu	Dummy variable, set to 1 if the purpose of the trip is school-related	No
Temperature	Weighted daily average temperature	Yes
Rain	Weighted daily average rain	Yes
Wind	Weighted daily average wind	Yes
Sunshine	Weighted daily average sunshine	Yes
Purpose work * weather [1]	Interaction between weather and purpose for work related trips	Yes
Purpose edu * weather [1]	Interaction between weather and purpose for school related trips	Yes

[1] These are four variables each, one for each of the weather parameters. They are only estimated for the bicycle mode.

6.2.2 Class-Membership Model

The class-membership model determines the probability that an individual person belongs to one of the classes estimated in the modelling process. Since we assign this class-membership probability to each individual we can only use variables that use the person as the unit of analysis, which are described in Section 6.1.2. This also means that we are not able to use trip-related variables, such as the distance or purpose of a trip, as these variables change for each trip made by the traveller in question.

To estimate the class-membership probabilities a typical MNL structure is used, as described in Equation 3.2.2. The model estimates the influence of each of the variables on the class-membership probability of all but one of the classes. This one class is the reference class, in relation to which the probability of all other classes can be estimated. For each variable we thus calculate one parameter for each additional class we estimate. If we estimate a three-class model, two different parameters are estimated for each variable.

Table 6.4: Overview of all estimated parameters in the class-membership model

Parameter	Description
Delta	Estimates probabilities when all other variables are set to zero
Gender	The gender of the respondent, either male or female
Age	The age of the respondent
Employment	Whether or not the respondent is employed
Density	The urban density at the residential location of the respondent
Education	The highest attained education level of the respondent
Mode Attitude [1]	Attitude of the respondent to the travel modes
Car ownership	Whether the respondent personally owns a car
License ownership	Whether the respondent is in possession of a driver's license
E-bike ownership	Whether the respondent personally owns an e-bike

[1] Attitudes to four modes are used: car, train, bus, tram, metro (btm) and the bicycle.

6.2.3 Estimation Procedure

The latent class models are estimated using Apollo [Hess and Palma, 2019], a package for the statistical programming language R [R Core Team, 2017], which is developed to facilitate choice modelling. The code used for estimation can be found on my GitHub. Since the number of latent classes that need to be estimated are not known in advance multiple models with varying number of classes are estimated. Four models will be estimated: one MNL model and then four Latent Class models with 2 to 4 latent classes. The utility function in the MNL model will be the same as the utility functions used in the latent class models, but of course the MNL model will lack a class-membership model and thus will not use person-level variables.

The Maximum Likelihood Estimation searches the parameter space for solutions that maximise the Log-Likelihood, iteratively converging on a solution when an optimum is located. Whilst this procedure works well with the globally concave LL-functions of MNL models, the algorithm tends to get stuck in local optima for Latent Class models due to the fact that the LL-function for LC models is very complex, meaning it is not globally concave. To solve this problem one has to run the model using different starting values and select the model with the best fit, which is more likely to be the global optimum. Apollo facilitates this procedure by providing a function that automatically generates multiple starting points and uses an algorithm by Bierlaire et al. [2010] to select the best set of starting values. This algorithm is run using 20 sets of candidates and the selected set is used to estimate the final models.

e have used a software package called Apollo to estimate the models. Due to the complexity of the utility function, Latent Class models tend to get stuck in local optima. We tried 20 different starting values and selected the most promising one to alleviate this problem.

6.3 RESULTS

As said above, the estimation procedure involved estimating a MNL model and then progressed into estimating multiple Latent Class models. Before we discuss the Latent Class extension of the MNL model, we will first give the results from the MNL model estimation itself, which will facilitate the interpretation of the Latent Class models by providing baseline results. Furthermore the MNL outcomes will give some interesting information about the mean effects of weather on travel behaviour in their own right. The MNL model will thus be discussed first in Section 6.3.1, followed by the selection of the number of classes in Section 6.3.2 and the discussion and interpretation of the classes in Section 6.3.3. The class-membership model is then used see which individual factors influence the class-membership probabilities and thus which of these variables moderate the relationship between weather and travel behaviour.

6.3.1 MNL Analysis

Before we look into the parameters as estimated by the MNL model, we will first provide some model statistics and information. The model is estimated using nearly 60,000 trips made by 7,054 individuals (people who participated in either wave 2 or wave 4 of the MPN). More information can be found in Table 6.5.

Table 6.5: Model statistics about the MNL model

MNL model statistics	
Number of individuals	7054
Number of observations	59820
Estimated parameters	32
LL(final)	-47069.29
Adj.Rho-square (0)	0.3103
AIC	94202.57
BIC	94490.54

As discussed above in Section 6.1, the full Latent Class models contain trip-level variables and variables for the people making the trips. This last category is only included in the class-membership model, which is not present in the MNL model. The MNL model thus only contains trip-related variables. The utility function of the MNL model includes interaction effects between trip purpose and weather for bicycles. For this reason the weather parameters are given in a separate table where these interaction effects have already been calculated. The parameters and the robust T-ratio's of travel distance and trip purpose can be found in Table 6.6.

Table 6.6: Parameters estimated by MNL model

Parameter	Estimate	Rob.t-ratio
Alternative Specific Constants		
PT	-3.3297	-62.3
Bike	-1.9939	-26.86
Walk	-7.0416	-23.13
Distance		
PT	0.3006	15.86
Bike	-4.6291	-23.93
Walk	-16.3772	-23.54
Purpose		
Work-PT	1.1395	14.12
Work-Bike	0.7763	14.21
Work-Walk	-0.7747	-7.84
Edu-PT	3.4623	39.31
Edu-Bike	3.0754	31.41
Edu-Walk	0.6357	3.52

The alternative specific constants show that at the average distance and during average weather circumstance (since these are standardized variables, the value for the average is 0) the car is by far the most used mode. As expected the distance parameters for bicycling and walking are very negative, indicating that these modes are mostly chosen at shorter distances. The parameter for public transport is positive, although much smaller in magnitude than the parameters for bicycling and walking. This indicates that public transport use increases with increasing distances, a finding that doesn't surprise too much.

The dummy variables of trip purpose show some interesting patterns: public transport is used much more often for trips to and from work, whilst walking is used less often. For education the parameters are much larger in size and show that public transport and the bicycle are much more commonly used for trips to and from educational institutions. Students, pupils, and scholars will often be relatively young and lack the money to own a car making the alternatives more prevalent. Students who live far away from their institutions are likely to use public transport, in part due to the free public transport passes offered to Dutch students by the Dutch government. For closer trips the bike and pedestrian modes take preference above the car.

> At average distance and during average weather conditions for leisure trips the car is used the most, followed by the bicycle. For commute trips (to and from work) public transport is used used slightly more often, whilst for trips to and from educational institutions public transport and the bicycle are used much more often than for leisure trips.

The weather parameters are given in Table 6.7. For public transport and walking, these weather variables are directly estimated by the model. For bicycling interaction parameters between the weather variables and trip purpose are estimated. These interaction parameters mean that there are different weather parameters for bicycle trips for each of the three purpose categories (leisure, work, and education). All three parameters are given in Table 6.7 as well.

When observing the weather parameters, one can see that public transport share is hardly affected by the weather: the parameters for this mode are in general (much) smaller than the parameters estimated for either the bicycle or pedestrians with only the parameter for wind being statistically significant. The signs of the parameters are as expected: sunny, dry, warm, and wind still days see more use of the active modes as indicated by the negative parameters for rain and wind and the positive parameters for temperature and sunshine for these modes. The sole exception here is the negative sign for the

Table 6.7: Coefficients for the weather parameters estimated by the MNL model[1]

	PT	Walk	Bike (leisure)	Bike (work)	Bike (edu)
Temperature	-0.02	0.1009***	0.1418***	0.0336***	-0.0573***
Wind	0.0871***	-0.0639***	-0.1105***	-0.1258	-0.0339
Rain	-0.0085	-0.0236	-0.0873***	-0.0586	-0.075
Sunshine	-0.0217	0.1037***	0.0737***	0.0244	0.0275

[1] We display calculated effects of the weather variables in urban areas. The significance however is indicated for the interaction parameter itself.
* variable is significant at 5% level
** variable is significant at 1% level
*** variable is significant at 0.1 % level

effect of temperature on the mode share of bicycling for educational trips, suggesting that students are more likely to use the bicycle if temperatures decrease.

For wind, temperature, and sunshine the parameter sign for public transport is the opposite of the sign for the active modes, indicating that either people substitute public transport for trips that would otherwise have been made using an active mode or that travellers with public transport are not affected by the weather circumstances to the same extent and thus keep travelling when bicyclists and pedestrians choose not to travel. For rain however the signs of public transport and the active modes are in the same direction, indicating that the car's share increases at the expense of all other modes during precipitation.

> Mode shares of the active modes are negatively affected by wind and rain, whilst they are positively affected by temperatures and sunshine.

Zooming in on the bicycle trips, we can see that leisure trips are in general affected more strongly by the weather than commuting trips (for both trips to work and to educational institutions). This is especially true for the temperature and sunshine parameters, which are much larger for leisure trips than for utilitarian trips (indicating a higher sensitivity of leisure trips). There is one interesting exception to this rule however: the sensitivity of work trips to wind speeds (-0.1258) is greater than the sensitivity of leisure trips (-0.1105). Interestingly, educational trips' sensitivity to wind does follow the overall pattern and is much smaller compared to both other purposes. Finally the differences in sensitivity are smallest for rain, where the leisure trips are only slightly more sensitive (-0.0837) than work trips (-0.0586) and especially educational trips (-0.075).

6.3.2 Number of Classes

One key step in latent class analysis is determining how many classes need to be used in the model. There are two ways of specifying this number: the first uses statistical properties that assess the trade-off between the number of parameters in the model and model-fit. Two methods that can always be used are AIC and BIC, which need to be minimized. Since additional classes extend models with fewer classes the models are nested (a situation when model X uses all parameters used in model Y and some additional parameters) the Likelihood Ratio Test can also be used. The other way of deciding the number of classes uses the more subjective interpretability of the model: additional classes are only used if they are distinctly different from other classes and can be interpreted clearly. Since our goal is not to most accurately predict mode choices but rather to understand and explain the influence of weather on mode choice the second criterion is used as the primary decision making criterion here.

Four models are estimated: a simple RUM-MNL model and then four Latent Class models with increasing number of classes (from 2 to 4), where the underlying specification for the utility function of each separate class follows RUM-MNL specification as described in Section 6.2.1. The latent class models also contain parameters that are used in the class-membership model. The model-fit statistics for each model are given in Table 6.8.

Table 6.8: Model-level statistics from the models using up to 4 classes

	MNL	2 classes	3 classes	4 classes
Number of individuals	7054	7054	7054	7054
Number of observations	59820	59820	59820	59820
Estimated parameters	32	68	104	140
LL(final)	-47069.3	-42069.49	-39890.88	39212.32
Adj.Rho-square	0.3103	0.383	0.4143	0.4237
AIC	94202.57	84272.98	79989.76	78704.64
BIC	94490.54	84875.92	80925.66	79964.52

The statistical properties of the models evaluating the trade-off between number of parameters and model-fit improve for each added class: the 4-class model performs best on this criterion. As specified above however, the interpretability of the classes is more important. To interpret the classes we've tried to interpret each class as estimated by the latent class models. If all classes within a latent class model are both clearly distinguishable and internally consistent the model is easier to interpret than if either or both of these conditions are not satisfied. For the interpretation the parameters that are varied across classes are used to create tables like Table 6.7 for each class. The parameters are then compared to one-another to determine whether classes are distinguishable and internally consistent. Based on these tables we have decided to use three distinct latent classes, as all three classes show highly distinguishable behaviour that is consistent.

We have selected the model with three discrete latent classes. The three classes were highly distinguishable from one-another as opposed to the four-class model, whilst a significant improvement in model-fit has been obtained when compared to the two-class model.

6.3.3 Interpretation of Classes

For the interpretation of classes we'll look at the differences in parameter estimates of the parameters that are varied across the classes. There are two types of these parameters: the alternative specific constants and the weather parameters (see also Table 6.3). We'll first look at the alternative specific constants to interpret the differences between classes at average weather conditions. Then we'll look to the weather parameters to interpret the differences in behavioural response to weather between the two classes.

Differences during average weather conditions

Before we interpret the differences in the effect of weather on travel behaviour we want to see if there are differences between the classes during average weather conditions. These differences might arise from the flexible alternative specific constants, which basically give the probabilities during average conditions. To accurately assess the differences between classes implied by the variation in alternative specific constants, we give each classes' predicted choice probabilities for trips with average weather conditions in Table 6.9. We give the choice probabilities for the three different trip purposes (leisure, work, and education) and for two distances, the mean distance of 11 km and a shorter distance of 5 km.

A couple of observations can be made. We see substantial differences between leisure, work, and educational trips. We see that public transport and the bicycle are used more often for work trips than for leisure trips and that this effect is even stronger for educational trips. This last observation makes sense considering the fact that these types of trips are often made by young people who lack the money to use a car. Since students get free access to public transport in the Netherlands using a car is very expensive compared to both other modes. We also see that the predicted choice probability for walking is very low for all classes, especially for the longer trips. This again makes sense, as even the shorter 5 km trip is at the upper end of reasonable walking distances.

Table 6.9: Choice probabilities for modes in average weather conditions

	Leisure Trips				Work Trips				Educational Trips			
	Car	PT	Bicycle	Walking	Car	PT	Bicycle	Walking	Car	PT	Bicycle	Walking
Class	Distance: 11 km											
1	0.81	0.10	0.08	0.01	0.57	0.27	0.16	0	0.14	0.62	0.24	0
2	0.67	0.03	0.30	0	0.41	0.07	0.52	0	0.11	0.14	0.75	0
3	0.97	0.01	0.02	0	0.93	0.02	0.06	0	0.66	0.11	0.23	0
Class	Distance: 5 km											
1	0.58	0.07	0.19	0.16	0.41	0.18	0.36	0.05	0.09	0.37	0.49	0.04
2	0.4	0.02	0.57	0.02	0.20	0.03	0.77	0	0.04	0.09	0.87	0
3	0.92	0	0.06	0.01	0.82	0.01	0.16	0	0.43	0.07	0.49	0.01

There are also substantial differences between the classes. We see extremely high odds of using the car for the third class for both distances. Only for educational trips at shorter distances do we see that the car is not the mode with the highest choice probability by a big margin, but for most other trips the probability exceeds 0.8. We can thus say that people in this class use the car for the (vast) majority of their trips, with some use of the bicycle for work and educational trips as well.

Travellers in the second class almost exclusively use either the bicycle or the car, with high probabilities for either one of these modes under all conditions. For longer distance trips made either for leisure or work the car is preferred, whilst for all short distance trips the probability of using the bike is highest. The bicycle is also the predicted mode for longer distance trips made for educational purposes.

Finally for the first class we see much more multi-modal travel behaviour. We see much higher probabilities for both public transport (for the longer trips) and walking (for the shorter trips). Educational trips meanwhile see especially high public transport probabilities, especially compared to the other two classes. The traveller in this class is thus to a much greater extent than the other two classes able and willing to use all four modes.

> Based on the differences in choice probabilities during average weather conditions we can make a first distinction between our classes. The first class consists of multi-modal travellers. The second class uses either the car or the bicycle, depending on trip conditions and finally the last class uses the car for most, if not all, of their trips.

Reactions to weather

Of course there are also interesting differences between the classes in terms of the reactions to weather. To enable this comparison we use the estimated parameters for the effect of weather on the three classes. These estimates can be found in Table 6.10. Since these parameters can be difficult to interpret we also use calculated choice probabilities. We calculate these probabilities for trips of two distances (11 km and 5 km) during several days with different weather. These days are the same as those used in Chapter 5 and are supposed to be typical days that might feasibly occur within the months of September through November when the data was collected. They thus serve as typical examples of specific weather, ranging from very bad rainstorm days to days with very nice and enjoyable weather. More information about these days can be found in Table 5.11. The predicted choice probabilities for leisure trips are given in Table 6.11. The choice probabilities for trips with a work- or educational purpose are displayed in the appendix, in Table D.4 and Table D.5 respectively.

When looking at both the parameters and the calculated choice probabilities we see interesting differences between the classes with respect to how the weather influences travel behaviour. We will discuss these effects below for each class.

Table 6.10: Coefficients for the weather parameters and alternative specific constants for each class[1]

Class	Variable	PT	Walk	Bike (leisure)	Bike (work)	Bike (edu)
Class 1	ASC	-2.0423	-5.6348	-2.3266		
	Temperature	-0.1293	0.1479	0.1531	0.1127	0.0346
	Wind	0.1196	-0.1747	-0.0996	-0.3986	-0.2681
	Rain	-0.1438	-0.0591	-0.2328	-0.0246	-0.1396
	Sunshine	-0.0178	0.0444	0.0965	0.0629	0.1618
Class 2	ASC	-3.1446	-7.5217	-0.8161		
	Temperature	0.0571	0.1019	0.1332	-0.0509	0.0173
	Wind	0.0479	-0.005	-0.0696	-0.0003	-0.0062
	Rain	0.0886	-0.0863	-0.1354	-0.1079	-0.1018
	Sunshine	0.0435	0.1019	0.1119	0.075	0.0372
Class 3	ASC	-5.284	-8.6348	-3.8601		
	Temperature	0.1229	0.0061	0.1912	0.0242	-0.2291
	Wind	0.0128	-0.0194	-0.1971	-0.0634	0.1578
	Rain	-0.0559	-0.0116	-0.0884	-0.07	0.0031
	Sunshine	-0.3972	0.1079	0.0773	0.0875	0.0385

[1] We display calculated effects of the weather variables in urban areas. The significance however is indicated for the interaction parameter itself.
* variable is significant at 5% level
** variable is significant at 1% level
*** variable is significant at 0.1 % level

Class 1: multi-modal travellers

We have seen that people in class 1 use multiple modes for their travel, with the final decision partly being informed by the distance and purpose of the trip. We also expect that the choice is informed by the weather conditions during the day of the trip. To illustrate the effect of the weather we use parameter estimates and calculated probabilities during days with specified weather patterns.

From the estimated parameters we see that the use of public transport is affected more strongly by weather variables relative to the other two classes. Perhaps for travellers in the other two classes public transport is never actually considered, meaning that there is no effect of weather on public transport use for these classes. The public transport mode share for this class is especially affected by temperature (-), wind (+), and rain(-). Use of public transport thus increases during very cold days, whilst it decreases during warmer, rainy, and wind-still days. We can see possible substitutions both with the active modes and with the car. For the active modes the effects of temperature and wind are opposite to that of public transport, indicating that people switch from active modes to public transport as a result of low temperatures and hard wind speeds (and vice versa during opposite conditions). We see substitutions with the car during rainy days, probably caused by the exposure to precipitation during the trip-leg to and from the public transport access/egress points. The mode share of the car is thus at its maximum during the worst weather conditions, similar to a rainstorm. The shelter provided by the car is appreciated by travellers and people are fairly likely to switch to the car during inclement conditions.

We thus see relatively complex behavioural changes within this class, caused by the fact that multiple modes are considered and used often. This creates more possible substitutions between modes, perhaps all the way from biking to work if weather is nice, to using public transport during colder and windier conditions, to using the car during cold and rainy conditions. We thus see relatively large effects of weather on choice probabilities, especially for the bicycle. The probability of the bicycle varies between 0.05 and 0.59 for work trips at a distance of 5 km, where the mean probability is 0.36. These are very large swings, nearly up to 100% in either direction. The car and public transport are both competitors of the bicycle, with people more likely to choose the car during rainy conditions and more likely to turn to public transport during colder or windier conditions. There is also an effect on the number of pedestrians, but it's smaller than the effect on the number of cyclists with relatively smaller swings either way. The substition patterns are however probably similar to those of the bicycle.

Table 6.11: Predicted choice probabilities for leisure trips by all three classes given certain weather conditions

	Class 1				Class 2				Class 3			
Distance: 11 km												
Type of weather	Car	PT	BC	WK	Car	PT	BC	WK	Car	PT	BC	WK
Mean	0.81	0.1	0.08	0	0.67	0.03	0.3	0	0.97	0.01	0.02	0
Rainstorm	0.92	0.07	0.01	0	0.83	0.1	0.07	0	0.99	0	0	0
Rain, Overcast	0.8	0.12	0.08	0	0.7	0.02	0.28	0	0.97	0.01	0.02	0
Wind, Rain	0.91	0.07	0.02	0	0.79	0.06	0.15	0	0.99	0	0.01	0
Near-freezing	0.79	0.15	0.06	0	0.73	0.03	0.24	0	0.98	0	0.01	0
Good	0.79	0.09	0.12	0	0.6	0.03	0.37	0	0.96	0	0.03	0
Great	0.75	0.06	0.19	0.01	0.48	0.03	0.5	0	0.94	0	0.06	0
Distance: 5 km												
Type of weather	Car	PT	BC	WK	Car	PT	BC	WK	Car	PT	BC	WK
Mean	0.59	0.07	0.19	0.16	0.4	0.02	0.57	0.02	0.92	0	0.06	0.01
Rainstorm	0.86	0.06	0.02	0.05	0.71	0.08	0.2	0.01	0.98	0	0.01	0.01
Rain, Overcast	0.59	0.08	0.18	0.16	0.43	0.01	0.54	0.01	0.92	0.01	0.06	0.01
Wind, Rain	0.77	0.05	0.06	0.11	0.58	0.04	0.36	0.01	0.96	0	0.03	0.01
Near-freezing	0.62	0.11	0.15	0.12	0.47	0.02	0.5	0.01	0.94	0	0.04	0.01
Good	0.51	0.05	0.24	0.19	0.33	0.01	0.64	0.02	0.89	0	0.1	0.01
Great	0.4	0.03	0.33	0.24	0.23	0.01	0.75	0.02	0.82	0	0.16	0.01

Class 2: Bicycle + car

Class 2 consists of travellers who use either the car or the bicycle for the vast majority of their trips. The decision which of these modes to use is partly informed by the weather, although the effect of weather seems to be smaller than for the first class.

We see that the effect of weather on bicycle use in particular is much less strong than for the first class, especially for non-leisure trips. For both work and educational trips cycling use is almost constant across most weather conditions, with no real positive effect of great days or negative effect from most weather circumstances. Even during the rainstorm day the mode share of the bicycle is only cut roughly in half, whilst such conditions are fairly rare and definitely on the extreme end of the observed range of weather circumstances. There is thus a small negative effect of especially wind and rain on the mode share of the bicycle, which only becomes noticeable if there is a lot of both wind and rain. For leisure trips the effects are stronger, but still nowhere near as influential as for the first class. We could thus say that the more experienced and habitual cyclists of the second class are affected to a smaller extent than the other two classes.

We see that people in this class also use public transport, especially for educational and work trips. The main substitution mode is different for each of these two types of trips. For work trips we see that both car and public transport use increase during inclement conditions, probably as cyclists decide to use either one of these modes for the trip. For shorter distance trips the car is slightly more often preferred, whilst for longer distance trips the public transport is slightly more likely to be the substitution. For educational trips however we see that car use is much more stable, with only a small increase during inclement conditions. Instead public transport use increases during inclement conditions, as it serves as the substitute mode for bicycling during such circumstances.

Class 3: Car mostly

From the choice probabilities during average weather conditions we can see that travellers in class 3 use the car for the vast majority of their trips, especially those made with either a leisure or work purpose.

For these trips we see only small effects of weather on the choice probabilities, with dry, sunny, and hot conditions causing a small decrease in car use and a corresponding increase in bicycle use whilst inclement conditions have the opposite effect. The choice probabilities for both public transport and pedestrian trips are for all intents and purposes negligible during all weather conditions for trips with these purposes.

For educational trips meanwhile we do see some more multi-modal behaviour, with especially use of the bicycle increasing dramatically. If we look at both the parameters and the calculated choice probabilities we see some contra-intuitive effects of weather however: predictions for the use of the bicycle are highest during rainstorm conditions and lowest during great weather conditions. The parameters for the effects of temperature, wind, and rain on bicycle utility have swapped signs compared to the other classes and to trips with other purposes. We have two possible explanations for this finding. The first is that simply not enough students are part of class 3, causing the estimator to be unable to find reliable estimates. The second is that students are more likely to go to lectures during the latter weeks of the sampling time, which causes bike rides to be correlated with the worse weather during this time-period. More research is needed into educational trips specifically to see whether the findings are indeed caused by some kind of sampling inadequacy or whether they truly represent the underlying decision-making.

The choice probability for walking are very low under all circumstances, although they are slightly improved during dry, sunny, and wind-still weather.

Final Interpretation

The information described above about the differences amongst the classes between both general travel patterns under average weather conditions and the response of travel behaviour to weather are summarised below.

- Class 1: These are people who actively consider and use all four modes, with choice probabilities varying both as a result of the trips purpose and distance and the weather conditions. Bicycle and pedestrian behaviour are both relatively sensitive to inclement weather conditions, with the effect being slightly larger for bicycle behaviour. Especially utilitarian trips are relatively more sensitive to weather when compared to the other classes. We see that people substitute the bicycle for both public transport and the car, where public transport is favoured under cold but dry conditions and car is favoured under rainy conditions, especially if they're coupled with higher wind speeds. We call this class 'Multi-modal' and in general can say that behaviour changes quite a lot as a result of shifts in weather conditions.

- Class 2: Here we find travellers who use either the bicycle or the car for the vast majority of their trips. These travellers are thus more experienced with the bicycle compared to both other classes. The utilitarian bicycle trips are really only affected by rain, which causes a mode shift to the car for leisure trips and the car or public transport. The effect of weather in general is much less sizeable compared to class 1. People in this class basically never walk, preferring to use the bicycle for shorter distance trips. This class is called 'Bicycle + car', as public transport use is really only sizeable for longer distance educational trips. Their behaviour is more robust to weather conditions, especially for non-leisure trips.

- Class 3: These people use the car for almost all of their trips, particularly those made for leisure or work. Other modes are used for educational trips, especially at shorter distance where use of the bicycle is sizeable. The effects of weather are small, although we see some increases in bicycling mode share under good weather conditions, mostly for leisure trips. For work related trips we see some effect of weather, more-so than for class 2 but not as much as for class 1. Again we see almost no pedestrians within this class. We find strange effects of weather on mode shares for educational trips, as inclement weather conditions seem to cause an increase of bicycling mode share. This is a very surprising finding that might be caused by sampling peculiarities and should be researched more. We label this class 'Car mostly', and the effect of the weather is relatively small, especially for leisure trips.

Since these classes can be categorized based on their travel patterns during average weather conditions we can compare our findings with previous latent class (clustering) analyses of Dutch travellers which have not accounted for the influence of the weather. We know of two such analyses, Kroesen [2014] and Molin et al. [2016]. Both use latent class clustering analysis (LCCA) techniques as opposed to the latent class choice model (LCCM) used in our analyses. Within LCCM we're trying to capture taste heterogeneity within the population, essentially estimating separate effects of one explanatory variable on choice probabilities, whilst LCCA classification is based on differences in indicators [Molin et al., 2016, p. 5]. Since we find that the taste heterogeneity with respect to weather is mostly based on travel patterns our classes can be compared to the ones found with LCCA techniques.

Both Kroesen [2014] and Molin et al. [2016] find five different classes, as opposed to our three classes. This means that our classes are expected to be aggregations of the five classes identified by these previous papers. We show how our three classes can be seen as aggregations for the classes identified by both papers in Table 6.12.

Table 6.12: Comparison of our classes to the classes identified by previous papers

Our Study		Other studies	
Class	Label	Molin et al. [2016]	Kroesen [2014]
Class 1	Multi-modal	Bike multi-modal PT	PT Light Traveler
Class 2	Bicycle + car	Car + Bike Car multi-modal	Strict bicycle Joint car + bicycle
Class 3	Car mostly	Car mostly	Strict car

In this table we can see that our results are generally in-line with the results from the previous papers: two of our three classes are also found by both previous researchers (Bike + Car and Car mostly), whereas the other multi-modal class can be seen as an aggregation of the PT and multi-modal classes found by Kroesen [2014] and Molin et al. [2016]. The additional classes estimated by the two papers (car multi-modal and strict bicycle respectively) can be part of the bigger bicycle + car class found in our research, although one could also argue that the car multi-modal class is a subsection of our multi-modal class. Since our classes are comparable we can also see whether the class-membership functions show similar covariates for the latent class structure. To do so we'll first establish how membership of our classes is affected by covariates related to socio-demographics, mode-availability and attitudes and then compare our findings to the two mentioned papers.

6.3.4 Class Membership Model

With the interpretations and implications of the three classes clear in our mind we can look to the class-membership model. The variables in the class-membership model give us information about which type of people are part of each class and thus which variables can mediate the relationship between weather and travel behaviour. These mediators in turn offer us an avenue for policy-making, as some of them can be influenced (to varying extents) by policy makers.

We have included three groups of variables in the class-membership model: 'normal' socio-demographics, mode availability, and attitudes. In Table 6.13 we see the values for these parameters as estimated for classes 2 and 3. Class 1 is the reference, which means that its value is always 0.

As discussed earlier, there are three types of variables included: socio-demographics, mode availabity, and attitudes. There is one last variable which is the delta, which can be interpreted as the constants added to the membership model to give the sizes of the classes when all other variables are set to 0. Below we will first discuss these deltas and then the groups of variables in the order given above.

We see very high values for both the class 2 and class 3 delta, indicating that these classes are much larger than class 1 when all other variables are set to zero. Whilst some other variables drastically reduce this difference there is still a sizeable difference in size between the classes in the population, as 19% of people are grouped into class 1, contrasting with 44% and 37% for classes 2 and 3 respectively.

In terms of socio-demographics we see that gender doesn't really play a role: the dummy variable has very low values that are not statistically significant. The opposite can be said for employment: the

Table 6.13: Class-membership function estimates

	Class 2	Class 3
delta	**5.2089**	**5.1777**
Male	0.0015	0.0281
Age	**-0.1441**	**-0.1145**
Employed	**-2.9688**	**-3.4724**
Education	**-0.0573**	**-0.1107**
Density	**-0.088**	**-0.1347**
E-bike	**0.4308**	-0.1757
Car	**-0.3486**	**0.9348**
License	-0.1728	**2.1134**
Car Attitude	**-0.1389**	**0.3407**
Train Attitude	-0.0477	-0.1066
BTM Attitude	-0.0304	0.0586
Bike Attitude	**0.5118**	**-0.2694**

Bold parameters are significant at the 5% level.

sizeable negative estimates show that employed people are much more likely to be part of class 1 than unemployed people. Similarly increasing levels of education also increase the odds of being part of class 1, which comes at the cost of the probability of being part of class 3. The probability of being a member of class 2 are also increased, although only to a limited extent. Age also has a fairly sizeable effect, where older people are more likely to be part of class 1, coming at the cost of class 2. The effect however is not very large. Finally the urban density of the residential location also has an impact increasing the probability of being part of class 1 and to a lesser extent also of class 2, at the cost of class 3. Based on these socio-demographics we could thus say that class 1 consists mostly of urban, employed and highly educated people, class 2 of a relatively representative sample with slightly more younger and unemployed people and class 3 of rural, sometimes unemployed and lower educated people.

Turning our attention to mode availability variables, we see that these variables have a fairly strong impact on class-membership probability. People who personally own a car are much more likely to be part of class 3, coming at the cost of both class 2 and to a lesser extent class 1. Having a driver's license is also very influential, again increasing the odds of being part of class 3 dramatically, coming at the cost of classes 2 and 1. Finally personally owning an e-bike increases the odds of being part of class 2, at the cost of mostly the probability of being part of class 3. All of these effects of course are somewhat to be expected: class 3 is the car enthusiast group which means that people need to be able to travel by car for most of their trips, necessitating a license and a car. Class 2 meanwhile is the group of experienced cyclists: investing in an e-bike is probably mostly a sensible decision for this group that uses the bike often.

The final group of variables concerns the attitudes. Here we see a strong and significant effect of car and bike attitudes, whilst the effect of the two public transport attitudes (for train and bus, tram, and metro) are both not significant. As expected, increases in Car attitude strongly increase the probability of being part of the car enthusiast class 3, mostly at the cost of the more bicycle oriented class 2. The effect of bike attitudes is of course the reverse, with a strong increase of the odds of being in class 2 mostly at the expense of class 3. Increasing PT attitudes, whilst not significant, both increase the likelihood of being part of class 1. For train attitudes this comes mostly at the cost of class 3, whilst BTM attitudes actually cause an increase for the likelihood of being part of class 3 at the cost of class 2.

Based on all of these variables we can thus make a typology of the three classes. When interpreting this typology we have to keep in mind that the effects of socio-demographics in particular are not that large, so we should avoid hasty generalizations or far-reaching conclusions based on these variables. Furthermore these are only the effects of variables that we have been able to include in the analysis: it's possible that the effect of other variables is large and explains some of the effects we observe. In particular our lack of any income variable means that both age and high-education effects are probably partly an income effect in reality. We find that slightly older and higher educated people are more likely to be in the multi-modal class for example, but it's very possible that the effect is mostly resulting from higher income levels within older and higher educated individuals. The typology is given in Table 6.14.

Table 6.14: A typology of the three classes, based on the class-membership model

Class	Socio-Demographics	Mode Availability	Attitudes
Class 1 Multi-modal	Slightly older, employed and higher-educated living in urban areas	Less likely to own car or license	Slightly positive to train
Class 2 Bicycle + car	Relatively even sample. Slightly younger, slightly more often unemployed.	More likely to own an e-bike Less likely to own car or license	Positive towards bicycle Negative towards car
Class 3 Car mostly	More often unemployed, living in rural-areas. Slightly lower educated.	Less likely to own an e-bike Always owns car and license.	Positive towards car Negative towards bicycle

Again we might compare our finding to earlier research into latent classes of travel behaviour in the Netherlands, in particular the studies by Kroesen [2014] and Molin et al. [2016]. We find that attitudes are in general congruent with travel behaviour, in accordance with expectations from theory and the findings by Molin et al. [2016]. The multi-modal travellers of the first class are more likely to live in more urbanized areas and have enjoyed a relatively high education, whereas the car only class consists of more rural and lower-educated travellers. Both results are consistent with the findings by the other two papers. We do find some inconsistent results with respect to age, as we find older people in the multi-modal class and younger people in the bike + car class. These classes are difficult to compare one on one with classes found by the previous papers, but they seem to contradict their findings. Molin et al. [2016] for example finds that people in the Car multi-modal and Car + bike groups are older than average. Perhaps this difference is caused by our exclusion of an income variable. As income is correlated to age, our findings with respect to age might be spurious.

6.4 CONCLUSION

In this chapter we have estimated the effect of weather on mode choices. Simultaneously we have tried to find heterogeneous groups within the population: groups of travellers that react differently to changes in weather circumstances. We have thus investigated the impact of weather on mode shares.

For the whole population, we find roughly the effects we expected based on the results from Chapter 5. The mode share of cycling is most strongly affected by weather circumstances, with particularly strong effects being measured for temperature (positive) and wind (negative). Both rain (negative) and sunshine (positive) also have a sizeable impact on the mode share of the bicycle however. The share of pedestrian trips is influenced by the same variables and in the same direction, but to a smaller degree. The effect on public transport is relatively minor for rain, but its share increases modestly as a result of increases in wind speed and decreases modestly when there is more solar radiation. The effects on the mode share of the car are the net result of the effects on the mode share of the other modes: it's affected positively by wind speed and rain, and negatively by sunshine and temperature.

We tried to find whether trip purpose moderates the effect between weather and bicycle mode shares. We find that in general utilitarian trips are much less sensitive to the effects of weather, particularly in the case of temperature and sunshine. This indicates that whilst increasing temperatures and sunshine levels increase the number of leisure trips made with the bicycle, the number of utilitarian trips are much less strongly affected. For rain and wind speed the differences in reaction are much smaller. This indicates that rain and wind speed are the main barriers for adopting commute bicycling habits.

One of the main extensions of the literature is achieved by our attempt to find heterogeneous groups within the population that differ in their travel behaviour reaction to weather circumstances. We find three such groups, defined mostly by their normal travel patterns. The first group has a multi-modal travel patterns where all four modes are considered and used relatively often. The second group uses either the bicycle or the car for the vast majority of the trips and the third group uses just the car for nearly all of their trips. The first group is the smallest, consisting of about 20% of all individuals whilst both other groups consist of about 40% of the sample.

Effectively we find that people with different current travel patterns (in terms of their mode shares during average weather conditions) will have different reactions to variations in weather. If people only use the car during average conditions (the third group) they are mostly affected by very pleasant conditions, which increase the mode share of the bicycle. People who live in rural areas, have a lower education and are unemployed are more likely to be part of this group of people. If people are prone to use the bicycle or the car, then weather circumstances will have a moderate effect on the decision to use either one of these modes. During inclement conditions these people keep using the bicycle to a much greater extent than people in other classes. The final group uses all four modes during average conditions. Reactions are qualitatively different than for the other two groups due to the fact that public transport and walking are alternatives that are actually considered often. This means for example that during very cold weather not only car, but also PT shares are increased and that pleasant conditions do the opposite. People in this group are more likely to be higher educated, employed, and living in urban environments.

Another findings is that people who own an e-bike are much more likely to be part of class 2, whilst people who own a car and a driver's license are much more likely to be part of the third class. These results could indicate that buying an e-bicycle means one is much more likely to cycle a lot and keep cycling during inclement conditions. Alternatively the causality might be reversed, where people who cycle a lot and keep cycling during inclement conditions are more likely to own an e-bicycle. More research is needed to establish this direction of the causal chain. We also see that people who own a car are very likely to use it, even during good weather conditions. Finally we also find that people's travel patterns are congruent with their attitudes towards the travel modes: if they are more positive towards a mode they are also more likely to use it. Again, causality might run both ways in this relationship.

7 CONCLUSION AND DISCUSSION

In this research we have used multiple methods to investigate the relationship between weather and travel behaviour, with a specific focus on possible factors that mediate this relationship. Uncovering these mediating factors and the extent to which they mediate the relationship between the weather and travel behaviour has been our research objective. To attain this objective we have stated multiple research questions, which we will answer in this chapter. Using these answers we can then provide an assessment of our research objective and find possible policy implications that might follow from the knowledge gained during this research.

The chapter will thus start by answering the research questions in Section 7.1. We will then connect these answers to the research objective and then discuss how our findings might lead to policy implications in Section 7.2, where we use the possible avenues described in the introduction. We will then discuss our findings in the context of previous research, both globally and in the context of the Netherlands in Section 7.3, which will also include the limitations of this research. Finally we will give recommendations for future research in Section 7.4.

7.1 RESEARCH QUESTIONS AND RESEARCH OBJECTIVE

Each of the research questions have been the subject of a chapter of this thesis, where we have tried to gather evidence in order to answer the questions. The answers to these questions will be given below, followed by a discussion to which extent the research objective has been attained.

1. *In which ways can weather affect travel behaviour and what is known about the existence and strength of these effects?*

We have found multiple ways through which weather can influence travel behaviour through our literature review. Behaviour is found to be the result of some mental processes, which we assume to be at least somewhat rational. Within these mental processes the perceptions of weather lead to changes in the utility associated with the different alternatives (whether or not to take a trip for travel demand and the different modes for mode choice). These perceptions of weather in turn are influenced by the objective state of the weather, although we don't know exactly how this relationship works . Perhaps the weather circumstances that can be directly perceived outside are important, but perhaps precise forecasts for the near future or informed expectations about daily weather play a role. In general we know that the effect of perceived weather is thus stronger than the effect of observed weather, as observed weather is prior to perceived weather in the causal chain. Unfortunately due to data limitations we have been unable to study this perceived weather effect. A second effect runs through the infrastructure and how this infrastructure changes during various weather conditions. This is for the most part a completely different research area that again can not be studied effectively within the scope of this thesis. In this thesis we thus focus on the effect of objective weather on the attractiveness or utility of alternatives of the traveller.

The causal paths discussed above can be moderated by characteristics of the traveller, such as socio-demographics, availabilities of travel modes, and attitudes. They can also be moderated by trip characteristics, such as its purpose or distance, and built-environment characteristics, such as the levels of shelter provided by public transport stations. These effects would typically moderate the relationship between objective weather and the attractiveness of the alternatives.

Some general findings and patterns emerge from the literature studied within this paper. The effects of weather on cycling use are the strongest, with most research finding a negative relationship between cycling use and increasing wind speeds and rain and a positive relationship with temperature. Some studies also find a negative effect of humidity and fog levels. The sign of the effects of these weather variables on walking behaviour are often found to be similar, but the size of the effects is smaller. These findings are mostly similar across many different geographic areas and are also replicated in the Netherlands.

For the use of the car and public transport more disagreement can be found in the literature. General findings would be that car use is affected negatively when temperatures near zero degrees Celsius, probably due to the risk of freezing roads. Rain conditions are found to increase car use to some extent, probably due to substitution from more exposed modes such as the bicycle or public transport. For public transport the results are very contradictory indeed, probably due to greater variation in both the performance and availability of public transport across geographic regions. Within the context of the Netherlands previous studies find a negative effect of temperature and a positive effect of precipitation on the use and mode share of public transport.

2. *Which conceptual connection between weather and travel behaviour is best able to capture the influence of weather on travel behaviour?*

To answer this question we have compared nine different conceptual ways of incorporating weather into the travel demand decision making. These ways relate to the spatial and temporal dimensions which are used to connect weather to a trip, resulting in the value of weather for this trip. For example, we use daily average values and values at the time of departure. In total we compare nine such ways of determining the value for the weather. We use these values to estimate nine models, enabling us to compare the performance of these models.

When comparing the parameter estimates between models we find some differences between conceptualisations that are based on the date of the trip and conceptualisations that are based on the time of the trip. The date-based trips conceptually assume that differences between days are driving mode choice decisions, when in contrast time-based trips assume that intra-day weather variations are also important for mode choice decisions. The time-based models find a negative effect of temperature on public transport mode share, the date-based models do not find this effect. We postulate that the finding by the time-based models (higher temperatures cause lower public transport mode share) is actually partly or wholly incorrect. Public transport trips are used for utilitarian purposes relatively often compared to the other modes, and utilitarian trips are usually during the morning and early evening when it is relatively cold. People do not use the public transport during these relatively cold times of the day because they're cold, but because they're the de-facto standard times to travel to and from work. The time-based models thus presuppose a causal effect between temperature and public transport use that doesn't exist.

We find that the weighted daily metric performs best based on the log-likelihood of the models, although the predictive capabilities of all models on an out-of-sample data set were very similar. The weighted daily metric entails that for each trip we take the average weather for the entire day at the weather station closest to the origin of the trip, weighted by the number of trips people on average make in any given hour. Essentially the weights ensure that weather at night has almost no effect on the weighted average and that the weather during rush-hour has the most effect. The conceptual conclusion then would be that people use information about the weather during the entire day when making their decisions. They thus make their decisions based on both past- and future weather.

This is the first revealed preference study which explicitly compares different conceptual ways of how weather affects travel behaviour and as such contributes to the literature in view of the call for this type of research by the most recent literature review by Liu et al. [2017]. We show that the most common practice up until now, where the trips' time and location of departure are used to measure weather, is both not the best option and could potentially lead to wrong conclusions regarding the influence of weather variables on mode choice, in particular the influence of temperature on public transport mode shares.

3. *What are the effects of combined weather on travel demand and how is this effect different between urban and rural areas?*

To determine the combined effects of weather on travel demand we have estimated regression models. We use five different dependent variables, namely four travel modes (car, public transport, bicycle, and walking) and the total number of trips made during a day. For each of these dependent variables we estimate five models, each specifying an increasingly complex relationship between weather and travel behaviour. We find that the effects of weather on travel behaviour are mode-specific and that for some modes effects estimated by linear-in-parameters models are not accurate.

Use of the active modes (cycling and walking) is in particular strongly affected by weather circumstances: in general there is a positive influence of sunshine and temperature and a negative influence of wind speed and rain on these variables. The effect is stronger for bicycling use than it is for pedestrian use. All of these findings are in accordance with previous literature [Creemers, 2010; Sabir, 2011; Böcker et al., 2016]. Our contribution lies in the finding that the demand for these modes is explained much better when we specify interaction effects between weather variables, in essence accounting for the co-occurring nature of weather. For walking we also find improvements when specifying non-linear effects, pointing to an ideal walking temperature. When using these models to predict the number of trips made during certain weather types we find that the estimation of non-linear and interaction effects lead to substantial differences, in particular for days at the extreme end of the observed range such as days with rainstorms or particularly high temperature and sunshine levels.

A further interesting finding is the fact that including weather variables in the multi-variate analysis changes the estimates for other variables. When we don't estimate any weather parameters we see that the dummy-variables of month are becoming significant instead, indicating that there are less bicycle trips in October and November than in September. By including weather effects, these month estimates become insignificant. The fact that the effects of month are in fact effects of weather is of course expected, but as it turns out the same notion can be found for the differences between the Randstad and non-Randstad regions and those between consecutive years. Variation in weather patterns thus explains why people cycle less in the Randstad area when accounting for the higher density of this area. The Randstad enjoys slightly colder weather with higher wind speeds, which is less enjoyable for cyclists. Accounting for variations in weather also explains some of the variation in observed travel behaviour from year to year. By including weather we are thus able to separate the variation between years into two parts: one that is caused by weather variation and one that is caused by other changes. Accounting for the influence of the weather thus enables a better identification of causal effects of other measures that are taken to change travel behaviour.

We also see interesting differences in the effect of weather on travel behaviour between rural and urban environments. Increasing temperatures increase the number of public transport trips in a rural context, but decrease the number of public transport trips in an urban environment. Perhaps this is due to the different use of public transport between rural and urban environments: in rural environments public transport could be used more often for longer-range trips when the alternative is the car, whilst in the urban environment public transport trips could be used more often for short-distance trips where the competitor is the bicycle. We see that the effect of wind speed on bicycle use is much stronger in urban environments, whilst the effects of the other weather variables (temperature, rain, sunshine) are slightly stronger in rural environments. Perhaps this is again caused by the different travel options available to urban and rural travellers, although the different trip characteristics might also play a role.

Our findings that linear-in-parameters models are unable to accurately capture the influence of weather on travel behaviour, leading to reduced predictive capabilities, show that the practice of estimating linear-in-parameters models by default can only lead to limited conclusions. We furthermore show that there is significant spatial variation in terms of the effects of weather, which means that it is difficult to estimate a singular effect of weather for large research areas. Finally the fact that weather is able to explain some of the spatial and temporal variation in the number of trips is a very important finding as well and means that researchers focusing on the differences in travel behaviour between say coastal and in-land regions should consider the impact of weather.

4. *What are the effects of weather on mode choice decisions of heterogeneous parts of the population?*

We identified three different classes within the population that have a different travel pattern and respond differently to variations in weather. The three classes are the following: 1) Multi-modal, 2) Bicycle + car and 3) Car mostly. The first group is the smallest, with about 19% of respondents being most likely to be part of this group. 44% of travellers are part of the second class and 37% belong to the third class.

The car mostly class is mostly unresponsive to weather variation, with the exception of increases in temperature and sunshine that see a relatively small increase in cycling use amongst this group, probably caused by the group making more leisure cycling trips. The second group uses the bicycle more often and will decide between these two modes based in large part on variations in weather. Noticeably even in very bad conditions, with very high wind speeds and precipitation, this group still cycles for a considerable number of trips especially at shorter distance trips. The first class have a more multi-modal perspective, seemingly deciding for a certain mode based to some extent on weather circumstances. We see that this group walks considerably more than both other groups at a distance of 5 km, with the modal share of walking going up to a quarter of all trips at this distance (which is pretty far for walking trips). Again we see that car use of this group is relatively stable at a longer distance, with some substitutions made for bicycling and public transport. Public transport benefits from cold and rainy weather conditions, whilst cycling benefits from warm and dry conditions. Car use is greatest when wind speeds are high, combined with relatively low temperatures.

The multi-modal class consists of relatively more people with a high education, living in urban areas and a job, who do not own a car or a driver's license and have positive attitudes towards the train. The bicycle + car class consists of a relatively representative sample of the population, with slightly more younger and unemployed people. These people are more likely to have an e-bike and have positive attitudes towards the bicycle, whilst having negative attitudes towards the car. The final class of car mostly travellers meanwhile consists of relatively more lower-educated rural people who own a car and a driver's license. They furthermore have positive attitudes towards the car and negative attitudes towards the bicycle.

Using the answers from these research questions, we can turn our attention to the research objective:

Determine the moderating effects of individual, trip, and spatial characteristics on the effect of the weather on travel behaviour in the Netherlands.

To determine the moderating effects of individual, trip, and spatial characteristics we need to identify the general effects of weather on travel behaviour first. Of the four modes investigated here bicycle and walking are most affected by weather characteristics. Higher temperatures and sunshine cause more trips with these modes, whilst increasing amounts of rain and wind cause decreases in the number of trips. Public transport use is hardly affected by the weather at all, whilst car use mostly sees an increase as a result of precipitation. Whilst general effects of weather on these two modes are small, they are also the modes where supply and demand are most closely matched causing full roads or trains at times. The small variation in use caused by weather patterns might thus still be very important, as a small increase of car usage during peak-hours can have sizeable effects on the amount of congestion on the roads.

In terms of personal characteristics we can say that the most influential socio-demographics are age, employment, and education. The urban density of the residential location moderates the relationship as well, but will be discussed below. Older people are more likely to either use the car during most circumstances or to consider all four modes, choosing between them partly on the basis of weather. People who are employed are much less likely to only use the car during all circumstances, instead again basing their mode choice partly on weather circumstances whilst they consider all modes. Finally people who have finished higher educational levels are also less likely to just use the car, instead again being much more likely to consider all four modes depending partly on weather variation.

We find that the purpose of the trip moderates the relationship between weather and travel behaviour as well. For the bicycle we see that the effect of weather on trips with utilitarian trip purposes (trips made for work or educational purposes) is less strong than on trips with leisure purposes. The only exception is the effect of wind, which is stronger for work trips than for leisure trips.

Finally the moderating effects of spatial variation has been investigated in two ways: first by trying to find the difference between rural and urban environments in the effect of weather on the number of trips and then by trying to find if the spatial environment has an effect on the class membership probability. Using the first method we find that rural trips are in general more sensitive to weather circumstances than urban trips, a fact that is especially true for bicycle use. We postulate that this could be explained by the greater distances per trip in rural areas and/or the fact that the car is a better alternative in rural than in urban areas, but perhaps other factors are relevant as well. Our second estimations point in the first place to the fact that people in dense areas are much more likely to use public transport, which automatically means that they respond differently to weather as well compared to people for whom public transport simply is not an actual alternative. In response to good weather conditions (dry, warm, relatively little wind and some sunshine) these people are more likely to switch modes from the car and public transport to cycling and walking, whilst in response to inclement weather conditions they are more prone to start using the car. People in rural areas are more likely to rely on the car for almost all trips, which makes them almost insensitive to weather changes: even during great weather they are more likely to keep using the car for trips.

With the research of this thesis we have been able to attain our research objective, having found both important effects of the weather on travel behaviour and having identified multiple moderating factors for this relationship. We've shown that these moderating factors need to be considered both by the scientific community when trying to determine the relationship between weather and travel behaviour and by policy makers when trying to control the effect the weather might have on their mobility related goals. However, to really gain an understanding that could lead to concrete policies we would need to include lower-level concepts in the analysis. Our recommendations for such future research and the implications for policy makers of the current findings can be found in the remainder of this chapter below.

7.2 POLICY IMPLICATIONS

In this section we will describe the policy implications that follow from our research on the effects of weather on travel behaviour. We will describe the policy implications using the two avenues identified earlier in Section 1.1. The first section will thus discuss policy implications following from the knowledge gained about the direct effects of weather on travel behaviour. The second section meanwhile will contain the implications following from possible measures and facilities that moderate the relationship between weather and travel behaviour.

7.2.1 Using the effect of weather

Knowledge about the direct effect of weather on travel behaviour can be used in three ways. The first of these is the ability to more accurately forecast future travel behaviour in the short-term. The second is the ability to explain the influence of the weather on variation in observed travel patterns across both the spatial and temporal dimensions and the third is the ability to use our knowledge about climate change to predict long-term shifts in travel behaviour.

Short-term predictions of future travel behaviour can for example be used to provide more accurate information the public, as is currently done in congestion forecasts given by the ANWB and Rijkswaterstaat. Since the effect of weather on the bicycle is particularly large (swings of up to 50% in bicycle use from mean weather conditions to inclement and great conditions), the greatest benefit can probably be achieved by policy makers and other actors that provide services aimed at bicyclists. Bicycle parking facilities might be able to pro-actively determine whether or not their facility will be stretched to capacity, increasing staff and taking other measures to avoid the facility being overcrowded by bicycles. Similarly public transport companies that ferry bicyclists across waterways might be able to increase or decrease the planned frequency of the ferry service depending on the predicted weather. The information could also be used to adaptively the frequency of other public transport options, but our research found that there are no substantial impacts of the weather on overall public transport use. It's possible (and probably even likely) that weather does impact the use of some parts of the overall public transport system, so this could still be relevant advice for some public transport routes.

The second way is the ability to control for the effect of weather, allowing researchers to establish the non-weather related differences in travel behaviour across both time and space. We'll explain how this information can be used by policy makers for both of these dimensions below.

Policies are often based on long-term trends of travel behaviour, which are supposed to show a steady signal of travel behaviour change which is expected to continue into the future. To estimate these long-term trends researchers have to account for factors that introduce noise. Such factors could be different sampling techniques or yearly anomalies that have changed the recorded travel behaviour, but are not expected to influence future travel behaviour. One of these noise factors is the weather. In an especially cold year with many snowy days travel behaviour will radically change, but this change is not expected to continue into the next year. We thus need to control for weather circumstances to calculate the overall behavioural trend. We can also control for the effect of weather in before- and after measurements of travel behaviour, which enables a more accurate assessment of the impact of various policy instruments.

Similarly we can observe many differences in travel behaviour between regions within the country of the Netherlands, such as urban vs. rural areas or coastal vs. in-land areas. Policy makers might try to change the travel behaviour within specific regions, for example by investing money into better public transport networks in rural areas under the assumption that the difference between rural and urban regions is partly caused by the performance of the public transport network. To know which policies could get we thus need to establish which factors cause the structural differences in travel behaviour between regions. Our research finds that the weather is one of these factors: it is structurally different for coastal regions when compared to in-land regions. Colder temperatures, higher wind-speeds, and increased precipitation can be observed in coastal regions, making them less attractive for especially bicycling and walking. Policy makers are advised to consider this effect when trying to find policies aimed at changing travel behaviour within regions.

Finally we can use our knowledge about climate change to improve our forecasts of long-term travel behaviour. We know that temperatures will increase in the next decades by up to 2 degrees. The Royal Netherlands Meteorological Institute (KNMI) has produced multiple scenario's for the effects of climate change in the Netherlands [Klein Tank et al., 2014]. Briefly summarised temperatures are expected to increase by 1 to 2.5 degrees Celsius, leading to hotter summers and milder winters. Precipitation is expected to increase in the colder months especially, whilst solar radiation levels are expected to increase slightly for the entire year. With this information and the results of our research we can link the expected changes in weather circumstances to behavioural shifts.

The effects of climate change are expected to vary for each season. Since our data is collected mostly in autumn we unfortunately have no insights into how increased drought and very hot temperatures in summer will affect travel behaviour, but we can give an impression of the effects in autumn conditions. We expect leisure bicycle usage to increase, caused by higher temperatures and solar radiation levels. Precipitation increases meanwhile might cause commuters to use the car more often, which could raise demand for car usage at peak-hour times and thus increase congestion problems. These effects could be included in strategic travel forecasting models such as the Dutch national models LMS and NRM [Rijkswaterstaat, 2019], which are used in strategic decision-making with respect to travel infrastructure.

The expected effects are relatively small however, within the range of 5% for bicycle use and 0-2% for car use. This begs the question if the pure effects of weather will be noticeable within the many societal and mobility-related changes expected in the next decades in the face of the transition to a society that no longer emits greenhouse gasses. Rietveld et al. [2012] seem to think that the effects of weather itself are interesting, but probably not large enough to warrant specific adaptation measures in the grand scheme of changes that will be brought about in the future. Our analyses show slightly larger effects of weather, but we are inclined to agree on a national level. On a local level it's possible that weather changes might cause certain destinations to become more attractive as well (beaches for example), which might exacerbate the effects of weather on travel demand on this local level. Local authorities near such places are advised to investigate the strength of the expected effects and take adaptation measures based on the results.

7.2.2 Changing the effect of weather

Policy makers are also able to change the effects of the weather on travel behaviour, in effect allowing them to steer travel behaviour into desirable directions. This can be done by making changes in measures, facilities, or other factors that moderate the relationship between weather and travel behaviour. We have investigated the influence of two different types of moderators: those based on the individual making the trip and those based on the characteristics of the trip itself, such as its location or purpose. Implications following from these two types are described below.

In our analyses we have identified groups of travellers that react differently to the same weather circumstances. Membership of a group is based on multiple factors related to the individual, such as his/her socio-demographic background, attitudes, and the availabilities of travel modes. By changing these factors policy makers would be able to steer people into a higher probability of being a member of desirable groups. Our findings show that people with different travel patterns will display different reactions to weather circumstances, finding that people who use the bicycle more often during average weather conditions are also relatively more likely to keep using the bicycle during inclement weather conditions. This would indicate that the barrier of inclement conditions is less imposing for experienced cyclists. Policies that try to increase the number of commuting bicyclists could thus use relatively pleasant seasons to allow people to gain experience on their bicycle, knowing that they are then also more likely to keep using the bicycle during inclement conditions. Another option would be to accommodate cycling during inclement conditions, for example by stimulating the provision of shower facilities at the workplace.

We also find that people with positive attitudes towards cycling are more likely to keep cycling during inclement weather conditions, whilst people with positive attitudes towards the car are less likely to keep cycling. Policies aimed at changing these attitudes, for example by increasing or decreasing the overall attractiveness of these modes, might thus cause people overall to become less sensitive to inclement weather conditions. Of course such policies will be designed first and foremost to change the overall travel behaviour, but increased weather resilience could be used as a supportive argument. There are also some examples of policies that specifically try to alter the usefulness of a travel mode during inclement conditions. One such policy is to reduce waiting times for cyclists at traffic lights during inclement conditions. Currently tests are done where this is done if rain is detected. Based on our findings these tests could be extended to include days with high wind speeds, as wind speeds seem to have strong negative effects on the utility of cycling as well.

The characteristics of the trip can also moderate the relationship between weather and travel behaviour. We have investigated if trip purpose and the residential location of the trip change the sensitivity of travel behaviour to weather influence. We find that conform our expectations trips with a more utilitarian purpose (commute trips for example) are less sensitive to weather influence. This would indicate that established cycling habits for a commute are more robust, meaning that changing those habits is a promising way of ensuring people use their bicycle more often during inclement weather conditions. Furthermore we find that the bicycle is used very often to cycle to and from educational institutions: it seems that people form these robust bicycling habits early in life only to transition to more car-dependent usage later on. This presents a promising window for policies, where young professionals could be incentivized to keep using the bicycle for work. Temporary financial incentives (as those used as part of current bike stimulation packages) could be used to allow young professionals to start the habit of cycling to work.

We have also found substantial differences in reactions to weather between urban and rural environments, finding some interesting differences. Before any policies can be designed to decrease te overall weather sensitivity however we need to know which factors cause these differences. Based on theory we think that there are multiple possibilities, for example that travel distances in rural areas are on average longer or that buildings within urban environments provide protection against the elements. To give policy recommendations related to this topic we would thus need to carry out future research focusing on explaining the difference in weather sensitivity between regions.

7.3 DISCUSSION

Whilst the conclusion and policy recommendations follow directly from the results we want to reflect on our research in this discussion. We do so in two different ways, first focusing on the relation between our research and the existing literature and then on the limitations of this research.

7.3.1 Other literature

With respect to the directions of the effects of weather variables on the use or mode shares of transport modes most of our results are in line with previous research: use of the active modes increases at higher temperatures and decreases at higher levels of wind speed or precipitation. Most research however either hasn't included the effect of sunshine levels in their research or has found there to be no effect, whilst we find a sizeable effect of this variable on the use of public transport (negative) and especially the active modes (positive for both cycling and walking).

As has been discussed multiple times prior the influence of weather is very specific to a location, partly due to the local climate and partly due to local customs with respect to travel patterns, travel infrastructure, and weather perceptions. As such we will compare our results to previous research in the Netherlands and Flanders, the upper half of Belgium.

Chapter 3 of the doctoral dissertation by Sabir [2011] is very similar to our travel demand chapter, as it also estimates regression models to estimate the influence of weather on travel demand. Within the observable range of our research, which uses data collected in the Fall season, the results are in general similar with respect to the quantitative estimates of temperature and precipitation. We find a very sizeable effect of wind speeds on the use of the bicycle, whilst Sabir finds an insignificant effect. A further contrast is the fact that Sabir has not estimated the effect of sunshine.

Mode choice findings are estimated by [Creemers et al., 2015], who find significant effects of precipitation, sunshine and Physiologically Equivalent Temperature (PET), which is calculated based on values for temperature, wind speed, and humidity on mode shares with the same directions as found in this research. Effects on use of the non-motorised modes are more sizeable than effects on public transport.

Finally our conclusions are also generally in-line with other previous research in Dutch context whose design is less similar to ours. The stated preference approach by [van Stralen et al., 2015] finds that light rain conditions prompt cyclists to switch to the car, a finding that is replicated in our revealed preference study. The research into possible effects of climate change by [Böcker et al., 2013b] shows seasonal variations, which we can't replicate. In terms of the influence of weather variations, they also find increases in cycling use as a result of higher temperatures in autumn. Research into perceptions and valuations of travel show that active modes are affected strongly by cold, windy, and wet weather as well.

In sum, the qualitative and quantitative findings produce results with respect to the influence of weather on travel behaviour that are in-line with previous research. We add to that knowledge by estimating interaction effects of weather variables, the identification of heterogeneous groups and the finding that the reactions to weather are very different for urban and rural areas.

7.3.2 Limitations

As with any research, and perhaps master's theses specifically, there are improvements that could be made upon this research. We specify the most important limitations below, roughly going from limitations resulting from the data to methodological limitations to limitations following from some arguable decisions made when doing the work required for this research.

The first two limitation of this research follows from the spatial resolution of the weather data. This data is gathered in roughly 50 weather stations distributed fairly evenly across the Netherlands. Most stations however are located near cities, which means that the average distance to a weather station in general is smaller for trips made within cities compared to trips in rural areas. We used 30 km as the maximum distance between the origin of a trip and the destination, but this number is fairly arbitrary. Some of the differences in effect between urban and rural areas might thus be caused by the better measurements for urban areas. In general the precision of the method can be put in doubt, especially with regards to the influence of rain. Rain can be a very local phenomenon that is also very time-specific and there is zero doubt that some trips that are estimated to have been during rainy conditions were actually completely dry and vice versa. This might have a pretty severe impact on our findings with regards to the incorporation of weather into the decision-making process. The date-based methods assume less precision on the part of the decision-maker. Perhaps these are found to be the best methods because of the mistakes in our connection method, rather than because people actually use weighted daily values for their decision making. Finally the matching to the weather stations could have been improved. Two examples are the use of interpolation for locations that are in-between two or more weather stations and using the travel route to select the weather station more accurately.

Our travel behaviour data also has some small flaws for the goals of this research. We'd have liked to estimate the effects of fairly detailed built-environment variables, for example wind-screens on bicycle paths or so called 'snelfietspaden', which are fast cycling lanes. The limited number of respondents within such specific regions make this impossible however, even if we would have been able to gather such precise data from other sources. Furthermore we use self-reported data, which is always a little bit tricky as people are found to forget trips/trip details or they get fatigued during the registration of trips and leave some out.

In terms of methods it would have been a strong addition to this research to make more use of the panel structure of our travel behaviour data set, for example enabling us to estimate some causal effects or transitions between our latent classes, using longitudinal methods. Our methods do incorporate panel effects, but they don't make use of the longitudinal structure of the data. Turning to another point, we debated whether or not to estimate zero-inflated count models. Especially for the modes with one mode as the dependent variable a zero was recorded more than would be expected, which gives the empirical basis for estimating such a model. Theoretically the zero-inflated model postulates that these excess zeros are generated by a different process than the other zeros. For example this would be if people were unable, rather than unwilling, to use a certain mode. There is definitely some theoretical and empirical justification for estimating these zero-inflated models, but in the end we decided not to mostly due to practical reasons: they would've required a significant time investment for relatively limited return compared to our normal negative binomial models.

Finally the mediating variables considered in this thesis, especially with regards to spatial differences, were very high-level concepts that will not lead to concrete policy recommendations. This is partly due to the fact that this data was included in the MPN and thus easy to implement, and partly due to the fact that the reach of this research (the entire Netherlands) combined with its relatively small sample size if we're zooming in on specific locations means that more specific variables, which are also very localized, are not a good fit to analyse the impact of such specific built environment characteristics.

7.4 RECOMMENDATIONS FOR FUTURE RESEARCH

During the six months it took to write this thesis we have continuously had to make decisions of whether or not to include factors, relations, and other complications. Six months of course is by far not enough time to investigate everything there is to know about the relationship between weather and travel behaviour relationship. Therefore we have had to scrap many factors that seemed interesting, many relations that seemed promising and many complications that seemed curious and definitely deserve attention. Below we will try to highlight the ones that stood out to us the most.

The first recommendations concern the data. In terms of weather data the approach that uses weather stations and then connection techniques such as the one used in Chapter 4 to connect this to all forms of travel behaviour data is the most common. The technique can be improved by adding more weather stations, such as the ones from weather underground, a platform where hobbyists can record and upload data with their own weather stations. Communication with weather experts (for example from national meteorological institutes) to decide which weather underground stations are reliable and valid for which circumstances could prove a worthwhile enrichment, reducing average distance to weather stations and improving overall accuracy. Another approach would be to use radar or satellite data, which is already available for the Netherlands, although accessibility is more complicated compared to weather station data. Again, communication with experts could facilitate the use of this data whose spatial resolution is much better than the weather station approach.

For travel behaviour data it would be very interesting to use data gathered from smartphone applications that not only record locations and start/stop-times, but also routes and speeds. This would help to avoid self-reporting limitations, but would also enable us to investigate the influence of weather on route choice and travel speeds. Other forms of data could be used as well, such as loop-counter data before and after an infrastructural project to determine how reactions to weather have changed as a result of such a project. In an ideal world these data sources are combined: smart phone applications could be used to gather more in-depth and person-specific information (such as the socio-demographics, other mode alternatives, and attitudes), whilst the loops or counters are used to gather population data for the use of the bicycle path. By linking the smart phone application data to certain routes and using this data as a sample for the route, we might be able to determine who bikes where and how different weather circumstances affect these different groups. Finally we would recommend surveyors to incorporate questions pertaining to perceived weather. Of course the more detailed and numerous these questions are, the more interesting research could be done. By asking respondents to indicate perceived weather for both the day and for their trips we could see which of these perceived variables is a better predictor of travel behaviour for example. However in the context of travel diary surveys we have to account for the fatigue that people undergo when they respond to such elaborate surveys. Even with little data, such as one question per day concerning the perceptions of weather, we would gain interesting insights into how weather circumstances are perceived and how these perceptions affect travel behaviour.

Using such perceived weather data the investigation into the incorporation of weather into the decision-making process could be improved, for example by using perceived weather as the dependent variable. Such an addition would enable us to see how people perceive weather and determine if there is considerable heterogeneity here as well. Such research could be complemented by more qualitative interview-style research where we try to determine how people perceive the weather in the first place, especially with regards to forecasts of weather and how these impact their decision-making.

BIBLIOGRAPHY

Al Hassan, Y. and 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.

ANWB (2018). 20 procent meer files op de Nederlandse wegen.

ANWB (2019a). Dagelijkse dukke trajecten ochtend en avondspits.

ANWB (2019b). Verkeersverwachting.

Arana, P., Cabezudo, S., and Peñalba, M. (2014). Influence of weather conditions on transit ridership: A statistical study using data from Smartcards. *Transportation Research Part A: Policy and Practice*, 59:1–12.

Barry, R. G. and Chorley, R. J. (2009). *Atmosphere, weather and climate*. Routledge, New York, 8th edition.

Ben-Akiva, M. and Bierlaire, M. (1999). Discrete Choice Methods and their Applications to Short Term Travel Decisions. In Hall, R., editor, *Handbook of Transportation Science*, pages 5–33. Spinger, Boston, Massachussets, USA.

Ben-Akiva, M. and Swait, J. (2008). The Akaike Likelihood Ratio Index. *Transportation Science*, 20(2):133–136.

Ben-Akiva, M. E., Lerman, S. R., Ben-Aakiva, M., and Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*. MIT Press, Cambridge, Massachussets, 1 edition.

Bierlaire, M., Themans, M., and Zufferey, N. (2010). A heuristic for nonlinear global optimization. *INFORMS Journal on Computing*, 22(1):59–70.

Boöcker, L. (2015). *Climate, weather and daily mobility : transport mode choices and travel experiences in the randstad Holland*. PhD thesis.

Böcker, L., Dijst, M., and Faber, J. (2016). Weather, transport mode choices and emotional travel experiences. *Transportation Research Part A: Policy and Practice*, 94:360–373.

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

Böcker, L., Dijst, M., and Prillwitz, J. (2013a). Impact of Everyday Weather on Individual Daily Travel Behaviours in Perspective: A Literature Review. *Transport Reviews*, 33(1):71–91.

Böcker, L., Prillwitz, J., and Dijst, M. (2013b). 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.

Boonstra, H., Van den Brakel, J., and Das, S. (2019). Multilevel time-series modeling of mobility trends. Technical report.

Calastri, C., Hess, S., Choudhury, C., Daly, A., and Gabrielli, L. (2017). Mode choice with latent availability and consideration: Theory and a case study.

Centraal Bureau voor de Statistiek (2018). Meer kilometers dan ooit door Nederlandse wegvoertuigen.

Chorus, C. G. (2012). Spijtminimalisatie als discrete-keuzemodel: Een overzicht van empirische vergelijkingen met het nutsmaximalisatie-model. *Tijdschrift Vervoerswetenschap*, 70(3):50–67.

Clifton, K. J. and Handy, S. L. (2003). Qualitative Methods in Travel Behaviour Research. In Jones, P. and Stopher, P., editors, *Transport Survey Quality and Innovation*, pages 283–302. Emerald Publishing Ltd.

Cools, M. and Creemers, L. (2013). The dual role of weather forecasts on changes in activity-travel behavior. *Journal of Transport Geography*, 28:167–175.

Cools, M., Moons, E., Creemers, L., and Wets, G. (2010). Changes in Travel Behavior in Response to Weather Conditions. *Transportation Research Record: Journal of the Transportation Research Board*, 2157(1):22–28.

Coxe, S., West, S. G., and Aiken, L. S. (2009). The analysis of count data: A gentle introduction to poisson regression and its alternatives. *Journal of Personality Assessment*, 91(2):121–136.

Creemers, L. (2010). *Do weather conditions and weather forecasts trigger changes in our daily travel behaviour*. Master's thesis, Universiteit Hasselt.

Creemers, L., Wets, G., and Cools, M. (2015). Meteorological variation in daily travel behaviour: evidence from revealed preference data from the Netherlands. *Theoretical and Applied Climatology*, 120(1-2):183–194.

de Nazelle, A., Morton, B. J., Jerrett, M., and Crawford-Brown, D. (2010). Short trips: An opportunity for reducing mobile-source emissions? *Transportation Research Part D: Transport and Environment*, 15(8):451–457.

Dijksterhuis, A. and Van Baaren, R. (2017). Fietsen door weer en wind: fietsen in de regen stimuleren met psychologisch onderbouwde feedback. Technical report, D&B, Nijmegen.

Fietsersbond (2019). 6 Leuke en slimme verkeerslichten - Fietsersbond.

Gallop, C., Tse, C., and Zhao, J. (2018). A Seasonal Autoregressive Model Of Vancouver Bicycle Traffic Using Weather Variables. *i-manager's Journal on Civil Engineering*, 1(4):9–18.

Gardner, W., Mulvey, E. P., and Shaw, E. C. (1995). Regression Analyses of Counts and Rates: Poisson, Overdispersed Poisson, and Negative Binomial Models. *Psychological Bulletin*, 118(3):392–404.

Gärling, T. and Schuitema, G. (2007). Travel demand management targeting reduced private car use: Effectiveness, public acceptability and political feasibility. *Journal of Social Issues*, 63(1):139–153.

Gebhart, K. and Noland, R. B. (2014). The impact of weather conditions on bikeshare trips in Washington, DC. *Transportation*, 41(6):1205–1225.

Golob, T. F. (2003). Structural equation modeling for travel behavior research. *Transportation Research Part B: Methodological*, 37(1):1–25.

Goudappel Coffeng (2017). Tussenevaluatie fietsstimuleringsprojecten. Technical report.

Grabow, M. L., Stone, B., Patz, J. A., Spak, S. N., Holloway, T., and Mednick, A. C. (2011). Air Quality and Exercise-Related Health Benefits from Reduced Car Travel in the Midwestern United States. *Environmental Health Perspectives*, 120(1):68–76.

Guo, Z., Wilson, N., and Rahbee, A. (2008). Impact of Weather on Transit Ridership in Chicago, Illinois. *Transportation Research Record: Journal of the Transportation Research Board*, 2034:3–10.

Heinen, E. (2011). *Bicycle Commuting*. PhD thesis.

Heinen, E. and Handy, S. (2012). Similarities in Attitudes and Norms and the Effect on Bicycle Commuting: Evidence from the Bicycle Cities Davis and Delft. *International Journal of Sustainable Transportation*, 6(5):257–281.

Heinen, E., Maat, K., and van Wee, B. (2011). Day-to-Day Choice to Commute or Not by Bicycle. *Transportation Research Record: Journal of the Transportation Research Board*, 2230(1):9–18.

Heinen, E., van Wee, B., and Maat, K. (2010). Commuting by bicycle: An overview of the literature. *Transport Reviews*, 30(1):59–96.

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

Hess, S., Ben-Akiva, M., Gopinath, D., and Walker, J. (2009). Advantages of latent class choice models over continuous mixed logit models. In *12th International Conference on Travel Behaviour Research*.

Hess, S. and Palma, D. (2019). Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*, In Press.

Hjorthol, R. (2013). Winter weather – an obstacle to older people's activities? *Journal of Transport Geography*, 28:186–191.

Hoogendoorn-Lanser, S., Schaap, N. T., and Oldekalter, M. J. (2015). The netherlands mobility panel: An innovative design approach for web-based longitudinal travel data collection. In *Transportation Research Procedia*, volume 11, pages 311–329. Elsevier.

Hyland, M., Frei, C., Frei, A., and Mahmassani, H. S. (2018). Riders on the storm: Exploring weather and seasonality effects on commute mode choice in Chicago. *Travel Behaviour and Society*, 13:44–60.

IPCC (2019). Climate change and land. Technical report.

Kashfi, S. A., Bunker, J. M., and Yigitcanlar, T. (2016). Modelling and analysing effects of complex seasonality and weather on an area's daily transit ridership rate. *Journal of Transport Geography*, 54:310–324.

Keay, K. and Simmonds, I. (2005). The association of rainfall and other weather variables with road traffic volume in Melbourne, Australia. *Accident Analysis and Prevention*, 37(1):109–124.

Kennisinstituut voor Mobiliteitsbeleid (2017). Mobiliteitsbeeld 2017. Technical report, KiM Netherlands Institute for Transport Policy Analysis.

Kennisinstituut voor Mobiliteitsbeleid (2018). Kerncijfers Mobiliteit 2018. Technical report, KiM Netherlands Institute for Transport Policy Analysis.

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: Policy and Practice*, 31(3):181–203.

Kilpeläinen, M. and Summala, H. (2007). Effects of weather and weather forecasts on driver behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour*, 10(4):288–299.

King, G. (1998). *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.*, volume 53. University of Michigan Press.

Klein Tank, A., Beersma, J., Bessembinder, J., Van den Hurk, B., and Lenderink, G. (2014). KNMI14: Klimaatscenario's voor Nederland. *KNMI publicatie*.

Koetse, M. J. and 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.

Kroesen, M. (2014). Modeling the behavioral determinants of travel behavior: An application of latent transition analysis. *Transportation Research Part A: Policy and Practice*, 65:56–67.

Lawless, J. F. (2006). Negative binomial and mixed poisson regression. *Canadian Journal of Statistics*, 15(3):209–225.

Li, J., Li, X., Chen, D., and Godding, L. (2017). Assessment of metro ridership fluctuation caused by weather conditions in Asian context: Using archived weather and ridership data in Nanjing. *Journal of Transport Geography*, 66:356–368.

Ligterink, N. and Smokers, R. (2016). Uitstoot van auto's bij snelheden hoger dan 120 km/u.

Liu, C., Susilo, Y. O., and Karlström, A. (2015a). 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., and Karlström, A. (2015b). 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., and Karlström, A. (2017). Weather variability and travel behaviour–what we know and what we do not know. *Transport Reviews*, 37(6):715–741.

Ma, L., Xiong, H., Wang, Z., and Xie, K. (2018). Impact of weather conditions on middle school students' commute mode choices: Empirical findings from Beijing, China. *Transportation Research Part D: Transport and Environment*.

Markolf, S. A., Hoehne, C., Fraser, A., Chester, M. V., and Underwood, B. S. (2018). Transportation resilience to climate change and extreme weather events – Beyond risk and robustness. *Transport Policy*, 74:174–186.

Mathisen, T. A., Annema, J. A., and Kroesen, M. (2015). The effects of weather and climate change on cycling in northern Norway. *EJTIR Issue*, 15(2):261–273.

McConnell, K. E. (1995). Consumer surplus from discrete choice models. *Journal of Environmental Economics and Management*, 29(3):263–270.

McKinney, W. (2010). Data Structures for Statistical Computing in Python.

McNally, M. G. (2000). The Four Step Model, Handbook of Transport Modeling. In *Institute of Transportation Studies and Department of Civil & Environmental Engineering*.

Meng, M., Zhang, J., Wong, Y. D., and Au, P. H. (2016). Effect of weather conditions and weather forecast on cycling travel behavior in Singapore. *International Journal of Sustainable Transportation*, 10(9):773–780.

Meyer, M. D. (1999). Demand management as an element of transportation policy: Using carrots and sticks to influence travel behavior. *Transportation Research Part A: Policy and Practice*, 33(7-8):575–599.

Miao, Q., Welch, E. W., and Sriraj, P. (2019). Extreme weather, public transport ridership and moderating effect of bus stop shelters. *Journal of Transport Geography*, 74:125–133.

Miranda-Moreno, L. and Nosal, T. (2011). Weather or Not to Cycle. *Transportation Research Record: Journal of the Transportation Research Board*, 2247(1):42–52.

Molin, E., Mokhtarian, P., and Kroesen, M. (2016). Multimodal travel groups and attitudes: A latent class cluster analysis of Dutch travelers. *Transportation Research Part A: Policy and Practice*, 83:14–29.

Motoaki, Y. and Daziano, R. A. (2015). A hybrid-choice latent-class model for the analysis of the effects of weather on cycling demand. *Transportation Research Part A: Policy and Practice*, 75:217–230.

Nankervis, M. (1999). The effect of weather and climate on bicycle commuting. *Transportation Research Part A: Policy and Practice*, 33(6):417–431.

Nkurunziza, A., Zuidgeest, M., Brussel, M., and Van Maarseveen, M. (2012). Examining the potential for modal change: Motivators and barriers for bicycle commuting in Dar-es-salaam. *Transport Policy*, 24:249–259.

Nordbakke, S. T. D. and Olsen, S. (2019). Who are most likely to adapt their travel behaviour to changes in weather conditions? A study of weather tolerance and travel behaviour in Norway. *European Journal of Sustainable Development*, 8(1):69–81.

Nosal, T. and Miranda-Moreno, L. F. (2014). The effect of weather on the use of North American bicycle facilities: A multi-city analysis using automatic counts. *Transportation Research Part A: Policy and Practice*, 66(1):213–225.

R Core Team (2017). R: A Language and Environment for Statistical Computing.

Rérat, P. (2018). Cycling to work: Meanings and experiences of a sustainable practice. *Transportation Research Part A: Policy and Practice*.

Rietveld, P., Sabir, M., and Van Ommeren, J. (2012). Fietsen door weer en wind:. *TVW Tijdschrift Vervoerswetenschap Jaargang*, 48(4):46–59.

Rijksoverheid (2018a). IenW maakt werk van veilige en duurzame mobiliteit — Nieuwsbericht — Rijksoverheid.nl.

Rijksoverheid (2018b). Klimaatbeleid — Klimaatverandering — Rijksoverheid.nl.

Rijkswaterstaat (1995). Wind in de Schermen: beschrijving van het proefproject Beschut Fietsen. Technical report.

Rijkswaterstaat (2019). Nederlands Regionaal Model (NRM) en Landelijk Model Systeem (LMS).

Rutte, M., van Haersma Buma, S., Pechtold, A., and Seegers, G. (2017). Vertrouwen in de Toekomst: regeerakkoord 2017-2021. Technical report, Rijksoverheid, The Hague.

Sabir, M. (2011). *Weather and Travel Behaviour*. PhD thesis, Vrije Universiteit Amsterdam.

Sabir, M., Van Ommeren, J., and Rietveld, P. (2013). Weather to travel to the beach. *Transportation Research Part A: Policy and Practice*, 58:79–86.

Saneinejad, S., Roorda, M. J., and Kennedy, C. (2012). Modelling the impact of weather conditions on active transportation travel behaviour. *Transportation Research Part D: Transport and Environment*, 17(2):129–137.

Schmiedeskamp, P. and Zhao, W. (2016). Estimating Daily Bicycle Counts in Seattle, Washington, from Seasonal and Weather Factors. *Transportation Research Record: Journal of the Transportation Research Board*, 2593(1):94–102.

Seabold, S. and Perktold, J. (2010). Statsmodels: Econometric and Statistical Modeling with Python Quantitative histology of aorta View project Statsmodels: Econometric and Statistical Modeling with Python. In *Proceedings of the 9th Python in Science conference*, pages 57–61.

Shaaban, K. and Muley, D. (2016). Investigation of Weather Impacts on Pedestrian Volumes. In *Transportation Research Procedia*, volume 14, pages 115–122. Elsevier.

Singhal, A., Kamga, C., and Yazici, A. (2014). Impact of weather on urban transit ridership. *Transportation Research Part A: Policy and Practice*, 69:379–391.

Stover, V. and McCormack, E. (2012). The Impact of Weather on Bus Ridership in Pierce County, Washington. *Journal of Public Transportation*, 15(1):95–110.

StudioSK (2017). Trekfietstracé Fietsviaduct A4: beeldkwaliteitsplan. Technical report.

Tao, S., Corcoran, J., Hickman, M., and Stimson, R. (2016). The influence of weather on local geographical patterns of bus usage. *Journal of Transport Geography*, 54:66–80.

Tao, S., Corcoran, J., Rowe, F., and Hickman, M. (2018). To travel or not to travel: 'Weather' is the question. Modelling the effect of local weather conditions on bus ridership. *Transportation Research Part C: Emerging Technologies*, 86:147–167.

Thomas, T., Jaarsma, R., and Tutert, B. (2009). Temporal variations of bicycle demand in the Netherlands : The influence of weather on cycling. *Transportation Research Board 88th Annual Meeting*.

Thomas, T., Jaarsma, R., and 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.

Ton, D., Bekhor, S., Cats, O., Duives, D. C., Hoogendoorn-Lanser, S., and Hoogendoorn, S. P. (2019). The experienced mode choice set and its determinants: Commuting trips in the Netherlands. *In Press*.

Ton, D., Duives, D. C., Cats, O., Hoogendoorn-Lanser, S., and Hoogendoorn, S. P. (2018). Cycling or walking? Determinants of mode choice in the Netherlands. *Transportation Research Part A: Policy and Practice*.

Van Baaren, R. and Dijksterhuis, A. (2015). Resultaten onderzoek fietsstimuleringsprojecten. Technical report.

Van Der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering*, 13(2):22–30.

van Nieuwenhuizen Wijbenga, C. and van Veldhoven-van der Meer, S. (2019). Kamerbrief over voortgang vorming mobiliteitsfonds.

van Stralen, W. J., Calvert, S. C., and Molin, E. J. (2015). The influence of adverse weather conditions on probability of congestion on dutch motorways. *European Journal of Transport and Infrastructure Research*, 15(4):482–500.

van Veldhoven-van der Meer, S. (2018). Ambities Fiets.

Van Wee, B. and Banister, D. (2016). How to Write a Literature Review Paper? *Transport Reviews*, 36(2):278–288.

Ver Hoef, J. M. and Boveng, P. L. (2007). Quasi-poisson vs. negative binomial regression: How should we model overdispersed count data? *Ecology*, 88(11):2766–2772.

Verseput, S. (2009). Zijn er uren waarin het meer of minder regent?

Wadud, Z. (2014). Cycling in a changed climate. *Journal of Transport Geography*, 35:12–20.

Webster, J. and Watson, T. (2004). R., 2002. Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, 1(2):0–10.

Wei, M., Corcoran, J., Sigler, T., and Liu, Y. (2018). Modeling the Influence of Weather on Transit Ridership: A Case Study from Brisbane, Australia. *Transportation Research Record*, page 036119811877707.

Wickham, H. (2017). tidyverse: Easily Install and Load ' Tidyverse' packages.

Wiebes, E. (2019). Kamerbrief voorstel voor een Klimaatakkoord.

Winters, M., Davidson, G., Kao, D., and Teschke, K. (2011). Motivators and deterrents of bicycling: Comparing influences on decisions to ride. *Transportation*, 38(1):153–168.

WMO (2019). WMO Statement on the State of the Global Climate in 2018. Technical report.

Zanni, A. M. and Ryley, T. J. (2015). The impact of extreme weather conditions on long distance travel behaviour. *Transportation Research Part A: Policy and Practice*, 77:305–319.

Zhao, J., Wang, J., Xing, Z., Luan, X., and 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.

A DATA PREPARATION

This appendix is a technical report of the data preparation process used to clean and combine the weather- and travel behaviour datasets that have been used in this thesis. The report will describe the data-preparation process and the assumptions/simplifications made in the process. It is meant to enable reproduction of my work and could perhaps be used as an aid for similar scientific work by others. The code will be available on my github. I have used Python and its data science libraries Pandas [McKinney, 2010] and NumPy [Van Der Walt et al., 2011] as well as R [R Core Team, 2017], where I have mostly used the tidyverse package [Wickham, 2017].

The overall process consists of multiple steps, which can be somewhat complicated to replicate. Most of this complexity is the result of the privacy-sensitive nature of the trip diary data, which necessitated that part of the code is executed in a safe digital environment. However this also means that most of this complexity can be avoided if a different (anonymous/private by design) set of travel data is used.

The main steps of the process are the following:

1. Retrieve the weather data
2. Retrieve & combine the travel behaviour data sets
3. Connect weather to trips
4. Connect weather to days

Each of these steps will be explained in some detail in the following sections.

A.1 WEATHERDATA

The weather data used in this thesis is collected and supplied by the KNMI, the Dutch Meteorological Institute. From the many types and forms of data that they provide as open data I have used the 10-min interval data, which is automatically updated every month. These datasets are unvalidated and missing values are not imputed or otherwise dealt with in any way: it is truly 'raw' data. This does mean that some extra work needs to be done to deal with the missing values, but it also means that no unknown operations, possibly influencing results, have been performed on the data. The data can be downloaded online from https://data.knmi.nl/datasets?q=bergman and can also be accessed by an API through the KNMI open data portal: ftp://data.knmi.nl/download/.

For this research six distinct datasets have been downloaded from the portal. The datasets are distinguished by their main weather variable(s), and were the following: precitipation, humidity & temperature, atmospheric pressure, wind, clouds, and sunshine. Each dataset contains multiple variables related to the overall theme. The temperature & humidity set for example contains variables like air temperature, dew point temperature, humidity, max. temperature, and more. Each set also contains a datetime column and information about the weather station where the data was collected, such as its name and location. A simplified tabular view of the datasets can be found in Table A.1. The data is collected by weather stations spread throughout the country of the Netherlands. Some weather stations collect all of the above weather themes, whilst others only collect information about one theme, for example a station that only measures wind speed. The location of the weather stations and an indication for the amount of weather types that are measured by them can be seen in Figure A.1.

One difficulty in preparing the data is its decentralized structure, as the individual data files contain just one month of data. All data files belonging to a specific year are bundled in a folder. There are thus 6 (datasets)* 5 (years) = 30 folders, each containing 12 monthly data files for a total of 360 files. It is more efficient and convenient to use a centralized structure instead, where all data files

Figure A.1: Location of weatherstations in the Netherlands

Table A.1: Weather dataframes

Datetime	Weatherstation	Location	Weathervariables
2013-11-14 09:00:00	De Kooy	lat=52.9...	T=6.0 °C ...
2013-11-14 09:10:00	De Kooy	lat=52.9...	T=6.2 °C ..
etc.			

belonging to one dataset are combined into a single file. This is achieved by use of a Python script that processes the data files and creates a single Pandas dataframe that is written away to a single data file in multiple formats (.csv for its universal compatibility and .pkl for its much faster read/write performance in pandas). This process is very memory-sensitive, which necessitated the creation of another function that can be used to reduce the memory requirements of the dataframe within python. The final datafiles are still multiple GB's in size and a computer with enough memory to process these files (16GB+ of RAM) is necessary to run the program and transform the data into a single file.

Some of the data could be removed, further reducing both processing time- and memory-requirements. The datasets contained records for weather stations placed in the North-Sea, which of course can't be coupled to on-land travel behaviour. These weather stations were removed. Furthermore the main source of Travel Behaviour data, the MPN, only records data within the last four months of the year (sept. - dec.). Data from the remaining eight months can thus be removed without consequence. The reduced data sets are saved as new files to ensure both the raw and unprocessed files are accessible at all times.

Missing values are not dealt with or imputed in any way at this stage. They will be dealt with during the matching stage, where trips or days are matched to weather data. This reduces processing time spent on cleaning parts of the data that will not be connected to trips and ensures that this critical step in data processing (and the underlying assumptions) are carried out in a single place, increasing both flexibility and transparency.

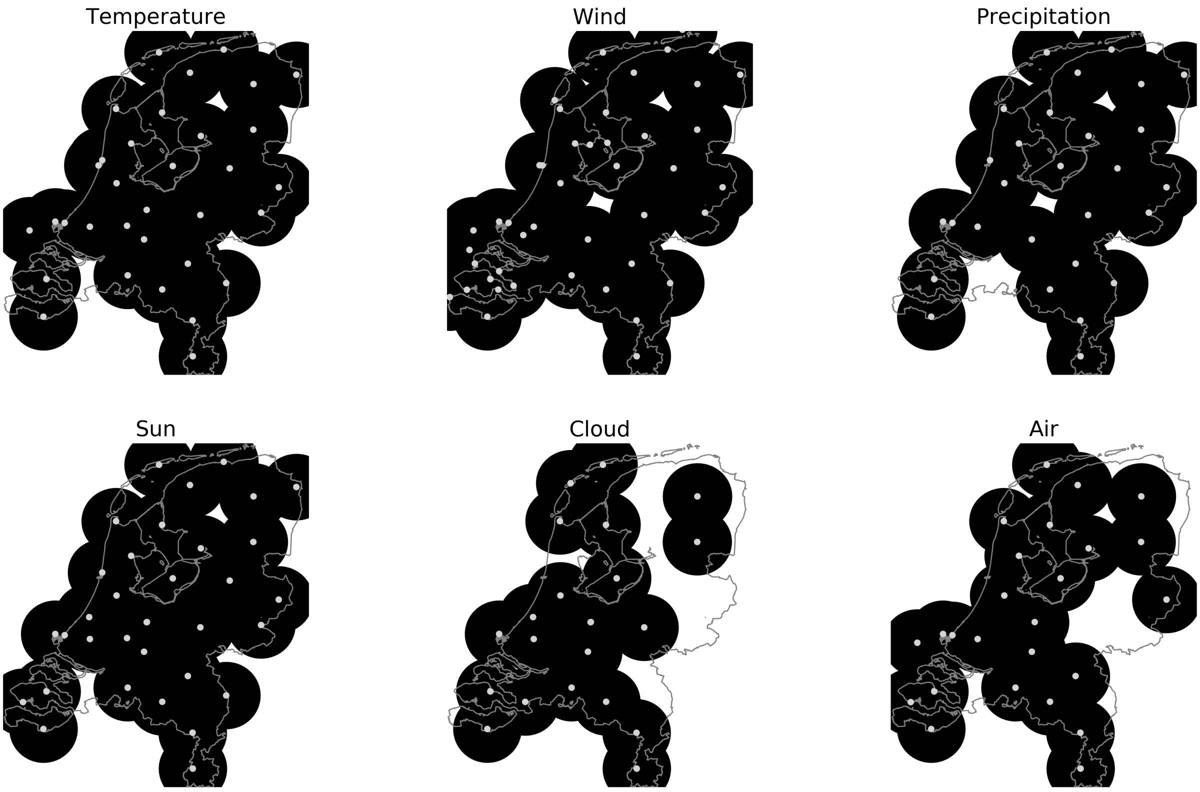


Figure A.2: Coverage of 30km zone around weatherstations, split per category

A.2 TRAVEL BEHAVIOUR DATA

This report uses the Mobility Panel Netherlands (MPN) as its source of travel behaviour data. The MPN is a longitudinal panel data set collected by the KiM Netherlands Institute for Transport Policy Analysis (KiM) [Hoogendoorn-Lanser et al., 2015]. Collection of the MPN started in 2013 and as of april 2019 five yearly waves are processed and available (2013 through 2017). The data set contains two principal parts: the first is a survey consisting of multiple questionnaires that contain information regarding socio-demographics, attitudes, locations, and travel behaviour of respondents. The second is a travel diary, where each respondent reports their travel behaviour for three consecutive day. For each year this results in four different data files. Furthermore all years apart from the first one contain a fifth data file with additional information. For odd waves (2015 and 2017) this data file contains geo-spatial information such as distances from the home location to the city centre, shops, or transit stops. For the even waves (2014 and 2016) this data file consists of information regarding attitudes of the respondents . The four data files are described in table A.2.

These individual data files can be combined by using the primary keys within each set. Primary keys are values that are unique within the data set: for the Person dataset this is the personal ID value that is unique for each respondent. The personal information can then be combined with household

Table A.2: The different files within the MPN

Name	Description
Travel Diary	Information regarding the trips made by a person each day
Daily	Information about the days included in the diary
Person	Personal characteristics
Household	Household characteristics
HH distance	Distance from location of home to shops, city centre, etc.
Special Person	Information about personal attitudes

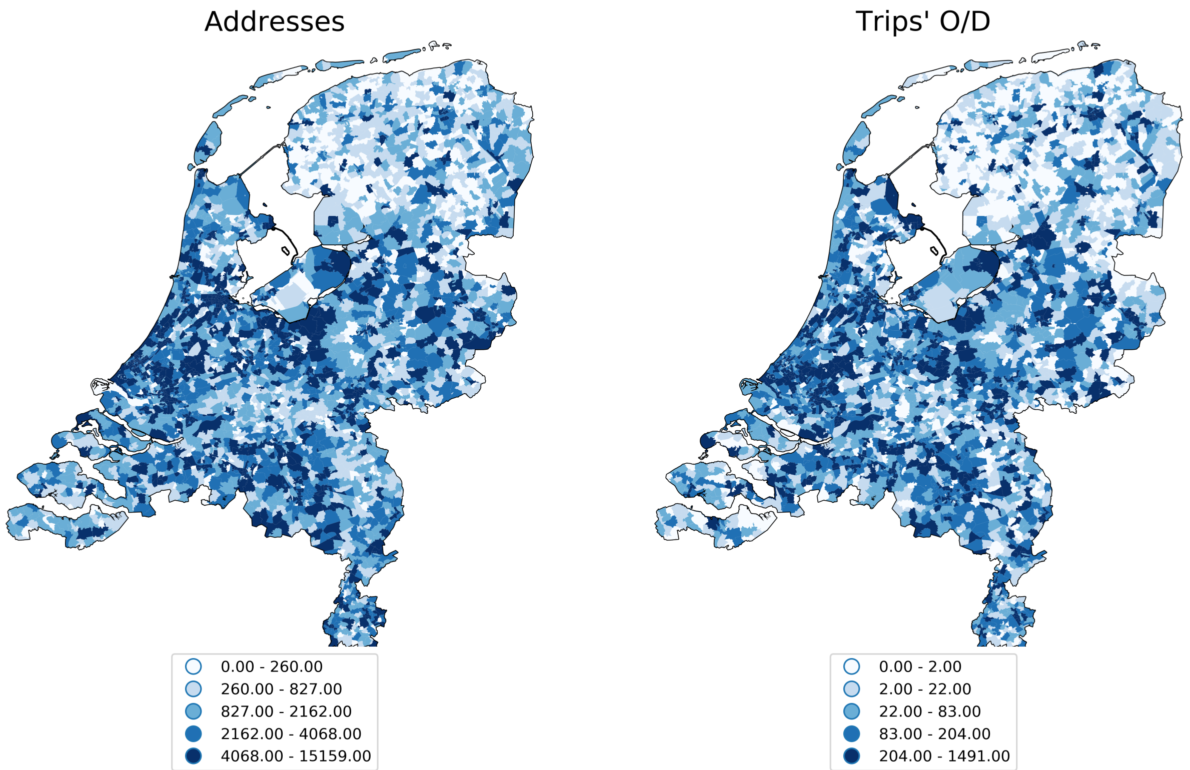


Figure A.3: Comparison of number of addresses per PC4 area and how often it is a trips' origin or destination

information by matching the household ID key values to the person ID values. This combining of data files is necessary to create a final data set that contains all the information available within the MPN that we need to estimate the models. This however also means that the combined data set is more privacy sensitive: by connecting multiple sources of information this one file could be directly traced back to individuals if compromised. The connections were thus done in the governmental digital environment to ensure that the data file was safe and secure. All personal information was stripped before the files were exported to my private laptop where they could be analysed in more detail. For the combination and preparation of the MPN data I made use of R code, written in Rstudio and using the tidyverse paradigm.

There are two final datasets resulting from this combination: one for each of the analyses performed in this thesis. One set contains records of individual trips, combined with relevant information about the person making the trip and the household that they are a a part of. The second set contains records on a daily level, again combined with relevant information about the person and household.

A.3 MATCHING TO TRIPS

Two parts of this research use the trip as the unit of analysis: Chapter 4 tries to determine which way of connecting weather to trips is best and Chapter 6 uses this method and a latent class approach to be able to determine the influence of weather on mode choice whilst providing possible levers for policy makers to use. This section will explain how travel behaviour data pertaining to trips is connected to locally collected weather data from weather stations.

There are two dimensions across which weather needs to be matched to a specific trip: space and time. Time-wise many different possibilities exist, the most prominent of which have been compared in Chapter 4. In all variations either the exact time of destination/arrival or at least the date on which the trip has been made are used to collect weather data. I will illustrate the matching process using two example algorithms, the first of which uses the time of origin and the second of which uses the weighted daily value.

For the first algorithm the original data as available after the downloads from KNMI and pre-processing described in Section A.1 can be used. For the second algorithm the weighted daily value has to be calculated for each day for all weather stations. This calculation uses 24 weights, each corresponding to one hour of the day and based on the amount of trips departing/originating during this hour, to calculate an average of all values. This gives a new dataframe, indexed by the combination of date and weatherstation. An example is shown in Table A.3.

Table A.3: Example of weighted daily average dataframe

Datetime	Weatherstation	Location	Weathervariables
2013-11-14	De Kooy	lat=52.9...	T=7.5 °C ...
2013-11-15	De Kooy	lat=52.9...	T=5.4 °C ..
etc.			

The other dimension across which trips need to be matched is the spatial dimension. Based on the location of the trip I need to determine which (combination of) weather station needs to be used to gather data that is relevant to this trip. The most intuitive location is that of the trips' origin or destination. For both the MPN contains details up to the specific address level, but due to privacy concerns (as the data needs to be processed outside the secure environment) the postcode-4 (pc4) level is used instead, consisting of a unique four digit combination corresponding to an area that is a couple of km^2 in size. These pc4 areas are then reduced to a single point by using open data from the Dutch Kadaster, the Basisregistratie Adressen en Gebouwen (BAG). Each address within the pc4 area is given, together with its coordinates. The average latitude and longitude of the addresses within each pc4 is used to determine a point-location for each pc4. Keeping in mind that pc4 areas change each travel behaviour wave prompted the use of BAG data corresponding to the years where these waves were collected. This results in five different data sets containing pc4 as the index and the average point-location as the value.

Table A.4: Example of dataset with pc4 locations

pc4	latitude	longitude
1011	52.371...	4.904...
1012	52.373..	4.895...
etc.		

To match these point values with weather stations, one needs to determine how close each of the weather stations are to the pc4 locations. Distance matrices between the weather stations and pc4 locations are made for each wave. A distance matrix contains all pc4 locations on the index, whilst all weather stations make up the columns. A combination of index and column would then give you the straight-line distance between the two.

These distance matrices can be used to look-up the closest weather station to any given PC 4 area and how close this weatherstation is. For this research the maximum distance between pc4 and weather station was set to 30km: any PC 4 areas that are further away from the closest possible station can not be matched to weather data. A pandas series (similar to an ordered dictionary) is used to store all the weather stations that are closer than 30 km to each PC 4 area, sorted by distance so that the closest station is considered first. As not all weather stations collect all sorts of weather data, these series were made for each of the six types of weather data that were collected.

Finally these series are used to create a mapping from all trips to all weather stations that can be used to read the data for this trip. These mappings are then used to look-up the data of the closest weather station at the time or date of the trip. If no data exists, the next closest weather station is used (if it is not farther away than 30km, and is thus part of the series mentioned above).

A.4 MATCHING TO DAYS

To match the weather to days the procedure is overall relatively similar: the algorithm above is almost entirely re-used, with only some small adjustments necessary to deal with the complications caused by the fact that we now don't have a time or location of a trip. Instead we can only match based on the date in question. In the end we used a weighted daily approach, where we calculated the average weather for each day where the weights are the number of trips made during this specific hour. In doing so hours during the night are valued much less, whilst peak-hours are valued much more. We calculate these weighted daily averages for each weather station and for all days within the range from the 1st of September to the 31st of December.

The only difficulty is then to decide which weather stations' data to use for each person-date combination. We assume that people make a decision to travel or not to travel in a select few locations: their residential location and their work location. As such we can use one or both of these locations. If we use both locations we average the calculated weighted daily averages for the two different weather stations. The weighted daily average thus averages across the temporal dimension, whilst the average of these values serves as a rudimentary interpolation across the spatial dimension.

Obviously we must ensure we do not use the values for the work location if people never visited their work in the first place. To do so we had to search all trips within the day for a trip that would have the purpose of going to work. If such a trip was found the destination location of this trip was used as this days' work location. If such a trip was not found, we simply used the residential location.

B INCORPORATION OF WEATHER IN DECISION MAKING APPENDIX

This is the appendix for Chapter 4, which includes material that we have decided not to include in the main-text of this thesis. In this case we display a single table here, which gives the outcomes of the Ben-Akiva & Swait test for all combinations of models. This test is used to compare the model fit of multiple non-nested models and tries to compare whether or not this improvement could be due to chance. The results are displayed in Table B.1.

Table B.1: Results of the Ben-Akiva & Swait test for each unique combination of models

	LL	1. -57744	2. -57743	3. -57722	4. -57649	5. -57700	6. -57722	7. -57703	8. -57707	9. -57713
1. Origin	-57744		0.283	0.004	<0.001	0	0.004	0	0	0.001
2. Destination	-57743			0.004	<0.001	0	0.004	0	0	0.001
3. Daily	-57722				<0.001	0.004	0.5	0.006	0.013	0.042
4. W. Daily	-57649									
5. 2H Centre	-57700				<0.001					
6. 1H Centre	-57722			0.5	<0.001	0.004		0.006	0.013	0.042
7. 1H Before	-57703				<0.001	0.16				
8. Morning	-57707				<0.001	0.064		0.125		
9. Afternoon	-57713				<0.001	0.019		0.035	0.08	

From the table we can conclude that the improvement in model fit obtained by the weighted daily model in comparison with all other models is statistically significant: it is thus very unlikely to be solely due to chance.

C REGRESSION APPENDIX

This is the appendix for Chapter 5 of my thesis. This appendix will contain results from the regression models we have estimated as part of this chapter. The appendix is divided into two sections. In Section C.1 the tables consist of comparisons of the same model across five different dependent variables (total trips, car trips, PT trips, bicycle trips, and pedestrian trips). These tables thus serve to illustrate the differences in the effects of weather between the various modes. Section C.2 displays the parameter estimates for all 50 models that were estimated as part of the modelling procedure (5 dependent variables * 5 different model specifications * 2 regression methods). We display these models for each mode, enabling a comparison between the performance of increasingly complex models with the same dependent variable.

C.1 COMPARISON ACROSS MODES

Table C.1: Comparison of the selected negative binomial regression models

Best Model	Total QuadInt	Car Linear	PT None	Bike Interaction	Walk QuadInt
Intercept	0.9653***	-0.2487***	-1.5330***	0.2060***	-1.6186***
Age	0.0222***	0.0781***	-0.2631***	-0.0601***	0.1440***
Male	-0.0690***	-0.0193*	0.0136	-0.1986***	-0.3003***
Employed	0.1322***	0.2893***	0.0247	0.0898***	-0.1265***
Density	-0.0053**	-0.0933***	0.2373***	0.0672***	0.0806***
Randstad	-0.0011	-0.0678***	0.4225***	0.0058	0.0432**
Ethnicity (Dutch native is reference)					
Western	-0.0759***	-0.0770***	0.3910***	-0.1899***	-0.1012**
Non-Western	-0.1940***	-0.0767**	0.1247	-0.5526***	-0.0988
Unknown Ethnicity	-0.0372	-0.2193***	-0.1447	0.1518**	-0.1374
Education (Low education is reference)					
Education_Med	0.0788***	0.2879***	0.1811***	-0.1830***	0.1293***
Education_High	0.1559***	0.3729***	0.3417***	-0.0356*	0.2431***
Day of week (weekday is reference)					
Saturday	-0.0923***	0.0955***	-0.6366***	-0.3623***	0.0919***
Sunday	-0.4624***	-0.2583***	-0.9986***	-0.8947***	-0.1115***
Year (2013 is reference)					
October	0.0066	0.0754***	-0.1592***	-0.0134	0.0114
November	0.0038	0.0573**	-0.0485	0.0027	-0.0225
Year (2013 is reference)					
2014	-0.0644***	-0.1099***	0.0217	0.0287	-0.1319***
2015	-0.1039***	-0.1540***	0.0832	-0.0025	-0.1925***
2016	-0.0772***	-0.0924***	0.0568	-0.0534**	-0.1509***
2017	-0.0299***	-0.0001	-0.0322	-0.0392	-0.1261***
Weather Variables					
Temperature	0.0240***	0.0034	0.0033	0.0448***	0.0375**
Wind	-0.0178***	0.0067	-0.0073	-0.1218***	0.0153
Rain	0.0149**	0.0264***	-0.0192	0.0221	0.0147
Solar Radiation	0.0411***	0.0085	-0.0365*	0.1067***	0.0660***
Quadratic Parameters					
Temp Squared	-0.0052				-0.0324***
Wind Squared	-0.0091***				-0.0296***
Rain Squared	0.0007				-0.0024
Radiation Squared	-0.0040				-0.0312**
Interaction Parameters					
Temp:Sun	-0.0156**			-0.0374***	0.0044
Temp:Rain	-0.0115**			-0.0416***	-0.0012
Temp:Wind	0.0206***			0.0294***	0.0511***
Wind:Sun	0.0019			-0.0412***	0.0279*
Wind:Rain	0.0029			-0.0523***	0.0021
Rain:Sun	0.0254***			0.1023***	0.0336
Wind:Rain:Sun	0.0005			-0.0553***	0.0071
Temp:Wind:Sun	-0.0089*			0.0124	-0.0275**
Temp:Wind:Rain	0.0055			0.0396***	0.0128
Temp:Rain:Sun	-0.0049			-0.0303***	0.0140
All Interactions	0.0079			0.0240*	0.0069
Model Statistics					
Chi-Square	61359.3	64819.3	89246.5	69545.6	66038.2
Log-Likelihood	-127758	-98002.9	-23104.1	-71228.5	-50144.8
AIC	255593	196052	46254.2	142525	100366
BIC	-588522	-588904	-640633	-605061	-622749

Table C.2: Comparison of negative binomial regression with complex weather influence across modes

	Total	Car	PT	Bike	Walk
Intercept	0.9653***	-0.2503***	-1.4925***	0.2303***	-1.6186***
Age	0.0222***	0.0784***	-0.2633***	-0.0601***	0.1440***
Male	-0.0690***	-0.0200*	0.0110	-0.1987***	-0.3003***
Employed	0.1322***	0.2890***	0.0199	0.0894***	-0.1265***
Density	-0.0053**	-0.0936***	0.2361***	0.0672***	0.0806***
Randstad	-0.0011	-0.0663***	0.4194***	0.0056	0.0432**
Ethnicity (Dutch native is reference)					
Western	-0.0759***	-0.0765***	0.3861***	-0.1908***	-0.1012**
Non-Western	-0.1940***	-0.0781**	0.1256	-0.5523***	-0.0988
Unknown Ethnicity	-0.0372	-0.2189***	-0.1556	0.1533**	-0.1374
Education (Low education is reference)					
Education_Med	0.0788***	0.2873***	0.1821***	-0.1824***	0.1293***
Education_High	0.1559***	0.3714***	0.3421***	-0.0347*	0.2431***
Type of day (weekday is reference)					
Saturday	-0.0923***	0.0951***	-0.6241***	-0.3615***	0.0919***
Sunday	-0.4624***	-0.2549***	-0.9795***	-0.8947***	-0.1115***
Month (September is reference)					
October	0.0066	0.0729	-0.1123**	-0.0241	0.0114
November	0.0013	0.0640	-0.0221	-0.0106	-0.0225
Year (2013 is reference)					
2014	-0.0644***	-0.1059***	0.0069	0.0300	-0.1319***
2015	-0.1039***	-0.1676***	0.0941*	0.0116	-0.1925***
2016	-0.0772***	-0.0826***	0.0208	-0.0456*	-0.1509***
2017	-0.0299***	0.0018	-0.0489	-0.0393	-0.1261***
Weather variables					
Temperature	0.0240***	0.0134	-0.0043	0.0392***	0.0375**
Wind	-0.0178***	0.0185**	0.0325	-0.1132***	0.0153
Rain	0.0149**	0.0377***	-0.0607*	0.0033	0.0147
Solar Radiation	0.0411***	0.0070	-0.0321	0.1085***	0.0660***
Temp Squared	-0.0052	0.0073	-0.0112	-0.0143	-0.0324***
Wind Squared	-0.0091***	-0.0053	-0.0393***	-0.0048	-0.0296***
Rain Squared	0.0007	-0.0059**	0.0144*	0.0069*	-0.0024
Radiation Squared	-0.0040	0.0108	-0.0174	-0.0127	-0.0312**
Wind:Sun	0.0019	0.0179**	-0.0128	-0.0416***	0.0279*
Wind:Rain	0.0029	0.0314***	-0.0268	-0.0499***	0.0021
Wind:Rain:Sun	0.0005	0.0315***	-0.0179	-0.0544***	0.0071
All	0.0079	-0.0023	0.0072	0.0216	0.0069
Rain:Sun	0.0254***	-0.0074	0.0245	0.1011***	0.0336
Temp:Wind	0.0206***	0.0009	0.0628***	0.0343***	0.0511***
Temp:Wind:Sun	-0.0089*	-0.0115	0.0079	0.0066	-0.0275**
Temp:Wind:Rain	0.0055	-0.0193**	0.0658**	0.0368***	0.0128
Temp:Sun	-0.0156**	-0.0272***	0.0370	-0.0209	0.0044
Temp:Rain	-0.0115**	0.0081	0.0008	-0.0450***	-0.0012
Temp:Rain:Sun	-0.0049	0.0037	-0.0135	-0.0338***	0.0140

Table C.3: Full results from the linear-in-parameters model investigating the difference between urban and rural environments

	Total	Car	PT	Bike	Walk
Intercept	0.9191***	-0.4458***	-1.0296***	0.3052***	-1.5710***
Age	0.0220***	0.0774***	-0.2605***	-0.0595***	0.1437***
Male	-0.0687***	-0.0185*	0.0054	-0.2013***	-0.3016***
Employed	0.1319***	0.2919***	0.0214	0.0866***	-0.1268***
Randstad	0.0039	-0.0952***	0.5041***	0.0462***	0.0742***
Urban	-0.0283***	-0.1822***	0.4680***	0.0651***	0.1463***
Ethnicity (Dutch native is reference)					
Western Foreigner	-0.0742***	-0.0843***	0.4054***	-0.1803***	-0.0937**
Non-Western Foreigner	-0.1901***	-0.0902**	0.1590*	-0.5331***	-0.0798
Unknown Ethnicity	-0.0359	-0.2185***	-0.1210	0.1581**	-0.1290
Education (Low education is reference)					
Education_Med	0.0788***	0.2910***	0.1844***	-0.1817***	0.1274***
Education_High	0.1577***	0.3666***	0.3743***	-0.0202	0.2538***
Day of week (Weekday is reference)					
Saturday	-0.0946***	0.0972***	-0.6335***	-0.3636***	0.0796***
Sunday	-0.4653***	-0.2586***	-0.9908***	-0.8986***	-0.1174***
Month (September is reference)					
October	0.0321***	0.0684***	-0.1552***	0.0176	0.0782**
November	0.0229	0.0489*	-0.0405	0.0329	0.0347
Year (2013 is reference)					
2014	-0.0706***	-0.1061***	0.0074	0.0194	-0.1362***
2015	-0.1061***	-0.1514***	0.0779	0.0010	-0.2187***
2016	-0.0833***	-0.0887***	0.0450	-0.0441*	-0.1643***
2017	-0.0267**	0.0012	-0.0327	-0.0279	-0.1042***
Weather variables					
Temperature	0.0260***	0.0009	0.0395	0.0802***	0.0222
Wind	-0.0179***	0.0080	-0.0178	-0.0563***	-0.0273*
Rain	0.0056	0.0246***	-0.0206	-0.0500***	0.0212
Solar Radiation	0.0275***	-0.0019	-0.0746**	0.0783***	0.0613***
Interaction parameters					
Temp:Urban	0.0033	0.0005	-0.0556*	-0.0201	0.0411*
Wind:Urban	-0.0176***	0.0013	0.0116	-0.0818***	0.0022
Rain:Urban	-0.0079	0.0028	0.0034	0.0115	-0.0375*
Sun:Urban	-0.0065	0.0205	0.0617*	-0.0402**	-0.0592***
Model Statistics					
Chi-Square	61322	64548.2	89755.7	69249.8	65478.6
Log-Likelihood	-127789	-98046.5	-23136.1	-71259.7	-50178.8
AIC	255631	196147	46326.2	142573	100412
BIC	-588625	-588957	-640630	-605195	-622967

C.2 COMPARING MODEL ESTIMATES

Table C.4: Poisson Results, dependent variable are total trips

	No Weather	Weather	Quadratic	Interaction	Quad + Int
Intercept	0.9825***	0.9304***	0.9625***	0.9535***	0.9689***
Age	0.0218***	0.0219***	0.0220***	0.0221***	0.0221***
Male	-0.0712***	-0.0710***	-0.0711***	-0.0711***	-0.0714***
Western	-0.0727***	-0.0738***	-0.0741***	-0.0732***	-0.0739***
Non-Western	-0.1930***	-0.1917***	-0.1920***	-0.1925***	-0.1930***
Unknown Ethnicity	-0.0442*	-0.0410*	-0.0418*	-0.0403	-0.0412*
Education_Med	0.0793***	0.0791***	0.0790***	0.0784***	0.0785***
Education_High	0.1578***	0.1554***	0.1554***	0.1540***	0.1541***
Employed	0.1314***	0.1296***	0.1296***	0.1301***	0.1301***
Density	-0.0045**	-0.0054***	-0.0053**	-0.0054***	-0.0055***
Randstad	-0.0110**	-0.0016	-0.0023	-0.0023	-0.0021
Saturday	-0.0940***	-0.0938***	-0.0942***	-0.0917***	-0.0916***
Sunday	-0.4637***	-0.4645***	-0.4617***	-0.4647***	-0.4617***
October	-0.0120	0.0291***	0.0155*	0.0131	0.0067
November	-0.0744***	0.0216*	0.0118	0.0156	0.0054
2014	-0.0467***	-0.0699***	-0.0708***	-0.0670***	-0.0648***
2015	-0.0791***	-0.1043***	-0.1006***	-0.1076***	-0.1038***
2016	-0.0795***	-0.0841***	-0.0832***	-0.0797***	-0.0773***
2017	-0.0221***	-0.0276***	-0.0325***	-0.0310***	-0.0302***
Temperature		0.0270***	0.0228***	0.0249***	0.0237***
Wind		-0.0256***	-0.0180***	-0.0291***	-0.0184***
Rain		0.0013	0.0024	0.0202***	0.0150***
Solar Radiation		0.0240***	0.0336***	0.0420***	0.0407***
Temp Squared			-0.0057***		-0.0052*
Wind Squared			-0.0065***		-0.0088***
Rain Squared			0.0000		0.0009
Radiation Squared			-0.0105***		-0.0031
Wind:Sun				0.0025	0.0008
Wind:Rain				-0.0038	0.0012
Wind:Rain:Sun				-0.0046	-0.0020
All				0.0093**	0.0088**
Rain:Sun				0.0286***	0.0266***
Temp:Wind				0.0146***	0.0201***
Temp:Wind:Sun				-0.0049	-0.0078**
Temp:Wind:Rain				0.0096**	0.0070
Temp:Sun				-0.0223***	-0.0158***
Temp:Rain				-0.0125***	-0.0123***
Temp:Rain:Sun				-0.0066*	-0.0054
Chi-Square	101561	101454	101411	101390	101368
Log-Likelihood	-132572	-132481	-132458	-132429	-132413
AIC	265182	265008	264971	264926	264903
BIC	-550397	-550535	-550536	-550518	-550506

Table C.5: Negative Binomial Results, dependent variable are total trips

	No Weather	Weather	Quadratic	Interaction	Quad + Int
Intercept	0.9775***	0.9245***	0.9587***	0.9481***	0.9653***
Age	0.0219***	0.0220***	0.0221***	0.0221***	0.0222***
Male	-0.0690***	-0.0685***	-0.0687***	-0.0687***	-0.0690***
Western	-0.0745***	-0.0759***	-0.0760***	-0.0751***	-0.0759***
Non-Western	-0.1933***	-0.1927***	-0.1931***	-0.1936***	-0.1940***
Unknown Ethnicity	-0.0403	-0.0371	-0.0380	-0.0363	-0.0372
Education_Med	0.0792***	0.0792***	0.0791***	0.0786***	0.0788***
Education_High	0.1593***	0.1570***	0.1570***	0.1557***	0.1559***
Employed	0.1336***	0.1319***	0.1317***	0.1325***	0.1322***
Density	-0.0043	-0.0052*	-0.0050*	-0.0052*	-0.0053**
Randstad	-0.0105	-0.0007	-0.0014	-0.0014	-0.0011
Saturday	-0.0946***	-0.0946***	-0.0950***	-0.0924***	-0.0923***
Sunday	-0.4641***	-0.4651***	-0.4622***	-0.4654***	-0.4624***
October	-0.0108	0.0307***	0.0162	0.0137	0.0066
November	-0.0757***	0.0218	0.0109	0.0147	0.0038
2014	-0.0466***	-0.0700***	-0.0711***	-0.0667***	-0.0644***
2015	-0.0801***	-0.1051***	-0.1009***	-0.1083***	-0.1039***
2016	-0.0796***	-0.0838***	-0.0838***	-0.0795***	-0.0772***
2017	-0.0215**	-0.0267**	-0.0323***	-0.0302***	-0.0299***
Temperature		0.0275***	0.0229***	0.0255***	0.0240***
Wind		-0.0268***	-0.0186***	-0.0292***	-0.0178***
Rain		0.0013	0.0030	0.0199***	0.0149**
Solar Radiation		0.0242***	0.0344***	0.0420***	0.0411***
Temp Squared			-0.0052*		-0.0052
Wind Squared			-0.0067***		-0.0091***
Rain Squared			-0.0002		0.0007
Radiation Squared			-0.0113***		-0.0040
Wind:Sun				0.0038	0.0019
Wind:Rain				-0.0020	0.0029
Wind:Rain:Sun				-0.0018	0.0005
All				0.0081	0.0079
Rain:Sun				0.0280***	0.0254***
Temp:Wind				0.0150***	0.0206***
Temp:Wind:Sun				-0.0059	-0.0089*
Temp:Wind:Rain				0.0081	0.0055
Temp:Sun				-0.0227***	-0.0156**
Temp:Rain				-0.0121***	-0.0115**
Temp:Rain:Sun				-0.0063	-0.0049
Chi-Square	61230.5	61312.4	61314.6	61353	61359.3
Log-Likelihood	-127856	-127801	-127787	-127768	-127758
AIC	255750	255648	255627	255605	255593
BIC	-588721	-588678	-588638	-588563	-588522

Table C.6: Poisson Results, dependent variable are car trips

	No Weather	Weather	Quadratic	Interaction	Quad + Int
Intercept	-0.1549***	-0.1637***	-0.1528***	-0.1571***	-0.1631***
Age	0.0671***	0.0670***	0.0671***	0.0672***	0.0673***
Male	-0.0159**	-0.0160**	-0.0161**	-0.0164**	-0.0166**
Western	-0.0735***	-0.0743***	-0.0739***	-0.0739***	-0.0740***
Non-Western	-0.0754***	-0.0765***	-0.0762***	-0.0768***	-0.0776***
Unknown Ethnicity	-0.1928***	-0.1932***	-0.1938***	-0.1920***	-0.1935***
Education_Med	0.2804***	0.2805***	0.2803***	0.2802***	0.2799***
Education_High	0.3657***	0.3666***	0.3664***	0.3657***	0.3651***
Employed	0.2802***	0.2797***	0.2797***	0.2794***	0.2792***
Density	-0.0928***	-0.0928***	-0.0929***	-0.0927***	-0.0929***
Randstad	-0.0590***	-0.0674***	-0.0670***	-0.0672***	-0.0666***
Saturday	0.0886***	0.0805***	0.0785***	0.0809***	0.0803***
Sunday	-0.2724***	-0.2770***	-0.2769***	-0.2767***	-0.2736***
October	0.0584***	0.0713***	0.0673***	0.0566***	0.0622***
November	0.0337***	0.0588***	0.0566***	0.0563***	0.0614***
2014	-0.1107***	-0.1118***	-0.1101***	-0.1091***	-0.1074***
2015	-0.1500***	-0.1549***	-0.1547***	-0.1584***	-0.1687***
2016	-0.0937***	-0.0958***	-0.0951***	-0.0824***	-0.0856***
2017	0.0005	-0.0054	-0.0047	-0.0048	-0.0029
Temperature		0.0057	0.0047	0.0108*	0.0147**
Wind		0.0058	0.0059	0.0111**	0.0143**
Rain		0.0249***	0.0398***	0.0220***	0.0340***
Solar Radiation		0.0084*	0.0114*	0.0124**	0.0071
Temp Squared			-0.0026		0.0063
Wind Squared			-0.0019		-0.0051**
Rain Squared			-0.0043***		-0.0051***
Radiation Squared			-0.0004		0.0107**
Wind:Sun				0.0160***	0.0144***
Wind:Rain				0.0225***	0.0272***
Wind:Rain:Sun				0.0210***	0.0246***
All				0.0007	0.0012
Rain:Sun				-0.0058	-0.0070
Temp:Wind				0.0021	0.0033
Temp:Wind:Sun				-0.0108**	-0.0089*
Temp:Wind:Rain				-0.0125**	-0.0147**
Temp:Sun				-0.0184***	-0.0272***
Temp:Rain				0.0030	0.0065
Temp:Rain:Sun				0.0001	0.0045
Chi-Square	143533	143499	143497	143458	143433
Log-Likelihood	-112857	-112828	-112823	-112806	-112796
AIC	225751	225702	225701	225680	225667
BIC	-520754	-520768	-520733	-520691	-520667

Table C.7: Negative Binomial Results, dependent variable are car trips

	No Weather	Weather	Quadratic	Interaction	Quad + Int
Intercept	-0.2410***	-0.2487***	-0.2379***	-0.2437***	-0.2503***
Age	0.0781***	0.0781***	0.0782***	0.0783***	0.0784***
Male	-0.0193*	-0.0193*	-0.0195*	-0.0198*	-0.0200*
Western	-0.0762***	-0.0770***	-0.0765***	-0.0764***	-0.0765***
Non-Western	-0.0769**	-0.0767**	-0.0769**	-0.0775**	-0.0781**
Unknown Ethnicity	-0.2193***	-0.2193***	-0.2200***	-0.2173***	-0.2189***
Education_Med	0.2878***	0.2879***	0.2879***	0.2876***	0.2873***
Education_High	0.3721***	0.3729***	0.3728***	0.3719***	0.3714***
Employed	0.2895***	0.2893***	0.2893***	0.2893***	0.2890***
Density	-0.0934***	-0.0933***	-0.0934***	-0.0933***	-0.0936***
Randstad	-0.0583***	-0.0678***	-0.0673***	-0.0668***	-0.0663***
Saturday	0.1038***	0.0955***	0.0932***	0.0958***	0.0951***
Sunday	-0.2529***	-0.2583***	-0.2582***	-0.2580***	-0.2549***
October	0.0645***	0.0754***	0.0723***	0.0607***	0.0676***
November	0.0369**	0.0573**	0.0560**	0.0557**	0.0624**
2014	-0.1110***	-0.1099***	-0.1082***	-0.1070***	-0.1059***
2015	-0.1515***	-0.1540***	-0.1537***	-0.1563***	-0.1676***
2016	-0.0910***	-0.0924***	-0.0929***	-0.0782***	-0.0826***
2017	0.0047	-0.0001	0.0002	0.0006	0.0018
Temperature		0.0034	0.0025	0.0087	0.0134
Wind		0.0067	0.0074	0.0151**	0.0185**
Rain		0.0264***	0.0438***	0.0246**	0.0377***
Solar Radiation		0.0085	0.0119	0.0121	0.0070
Temp Squared			-0.0015		0.0073
Wind Squared			-0.0024		-0.0053
Rain Squared			-0.0051**		-0.0059**
Radiation Squared			-0.0006		0.0108
Wind:Sun				0.0196**	0.0179**
Wind:Rain				0.0267***	0.0314***
Wind:Rain:Sun				0.0281***	0.0315***
All				-0.0032	-0.0023
Rain:Sun				-0.0052	-0.0074
Temp:Wind				-0.0004	0.0009
Temp:Wind:Sun				-0.0138*	-0.0115
Temp:Wind:Rain				-0.0173*	-0.0193**
Temp:Sun				-0.0180**	-0.0272***
Temp:Rain				0.0037	0.0081
Temp:Rain:Sun				-0.0018	0.0037
Chi-Square	64766.5	64819.3	64829.2	64848	64857.4
Log-Likelihood	-98012.5	-98002.9	-98000.9	-97994.9	-97990.9
AIC	196063	196052	196056	196058	196058
BIC	-588960	-588904	-588860	-588776	-588730

Table C.8: Poisson Results, dependent variable are PT trips

	No Weather	Weather	Quadratic	Interaction	Quad + Int
Intercept	-1.4895***	-1.4676***	-1.4531***	-1.4852***	-1.4273***
Age	-0.2737***	-0.2738***	-0.2737***	-0.2735***	-0.2732***
Male	-0.0079	-0.0078	-0.0091	-0.0080	-0.0093
Western	0.3896***	0.3898***	0.3853***	0.3898***	0.3858***
Non-Western	0.1029	0.1003	0.0975	0.1005	0.0992
Unknown Ethnicity	-0.2632**	-0.2656**	-0.2676**	-0.2618**	-0.2647**
Education_Med	0.2750***	0.2740***	0.2735***	0.2753***	0.2745***
Education_High	0.3866***	0.3843***	0.3844***	0.3853***	0.3848***
Employed	0.0245	0.0246	0.0251	0.0246	0.0254
Density	0.2123***	0.2122***	0.2116***	0.2118***	0.2108***
Randstad	0.4078***	0.4141***	0.4109***	0.4130***	0.4119***
Saturday	-0.6736***	-0.6717***	-0.6648***	-0.6659***	-0.6591***
Sunday	-1.0215***	-1.0172***	-0.9933***	-1.0151***	-0.9993***
October	-0.0894***	-0.1015***	-0.0940**	-0.0652	-0.0875**
November	0.0339	-0.0028	-0.0054	0.0297	-0.0054
2014	0.0159	0.0076	0.0011	-0.0091	-0.0057
2015	0.0810**	0.0855**	0.0853**	0.0808*	0.0973**
2016	0.0718**	0.0781**	0.0657*	0.0414	0.0448
2017	0.0059	0.0039	-0.0008	-0.0140	-0.0118
Temperature		0.0077	0.0055	0.0084	0.0003
Wind		-0.0093	0.0441***	-0.0242*	0.0347*
Rain		-0.0121	-0.0617***	-0.0054	-0.0509*
Solar Radiation		-0.0296*	-0.0403**	-0.0244	-0.0254
Temp Squared			0.0121		-0.0145
Wind Squared			-0.0367***		-0.0422***
Rain Squared			0.0128***		0.0127**
Radiation Squared			-0.0019		-0.0193
Wind:Sun				-0.0079	-0.0098
Wind:Rain				-0.0558**	-0.0242
Wind:Rain:Sun				-0.0416*	-0.0210
All				0.0225	0.0094
Rain:Sun				0.0260	0.0206
Temp:Wind				0.0418***	0.0644***
Temp:Wind:Sun				0.0168	0.0025
Temp:Wind:Rain				0.0722***	0.0564***
Temp:Sun				0.0181	0.0402*
Temp:Rain				0.0077	-0.0030
Temp:Rain:Sun				-0.0058	-0.0095
Chi-Square	126419	126355	126228	126069	126215
Log-Likelihood	-27008.8	-27006.4	-26989.2	-26989	-26971.8
AIC	54055.5	54058.8	54032.5	54046	54019.5
BIC	-621525	-621486	-621476	-621400	-621390

Table C.9: Negative Binomial Results, dependent variable are PT trips

	No Weather	Weather	Quadratic	Interaction	Quad + Int
Intercept	-1.5618***	-1.5330***	-1.5213***	-1.5439***	-1.4925***
Age	-0.2630***	-0.2631***	-0.2634***	-0.2634***	-0.2633***
Male	0.0134	0.0136	0.0118	0.0128	0.0110
Western	0.3911***	0.3910***	0.3871***	0.3898***	0.3861***
Non-Western	0.1246	0.1247	0.1227	0.1248	0.1256
Unknown Ethnicity	-0.1408	-0.1447	-0.1513	-0.1478	-0.1556
Education_Med	0.1821***	0.1811***	0.1813***	0.1817***	0.1821***
Education_High	0.3436***	0.3417***	0.3421***	0.3418***	0.3421***
Employed	0.0221	0.0247	0.0207	0.0207	0.0199
Density	0.2372***	0.2373***	0.2367***	0.2371***	0.2361***
Randstad	0.4153***	0.4225***	0.4193***	0.4211***	0.4194***
Saturday	-0.6403***	-0.6366***	-0.6297***	-0.6285***	-0.6241***
Sunday	-1.0040***	-0.9986***	-0.9757***	-0.9949***	-0.9795***
October	-0.1381***	-0.1592***	-0.1470***	-0.1202**	-0.1402**
November	0.0084	-0.0485	-0.0484	-0.0125	-0.0451
2014	0.0265	0.0217	0.0163	0.0022	0.0069
2015	0.0743	0.0832	0.0857	0.0770	0.0941*
2016	0.0484	0.0568	0.0453	0.0162	0.0208
2017	-0.0323	-0.0322	-0.0369	-0.0531	-0.0489
Temperature		0.0033	0.0015	0.0035	-0.0043
Wind		-0.0073	0.0448**	-0.0230	0.0325
Rain		-0.0192	-0.0758***	-0.0124	-0.0607*
Solar Radiation		-0.0365*	-0.0484*	-0.0299	-0.0321
Temp Squared			0.0140		-0.0112
Wind Squared			-0.0353***		-0.0393***
Rain Squared			0.0152**		0.0144*
Radiation Squared			-0.0023		-0.0174
Wind:Sun				-0.0098	-0.0128
Wind:Rain				-0.0551**	-0.0268
Wind:Rain:Sun				-0.0376	-0.0179
All				0.0167	0.0072
Rain:Sun				0.0288	0.0245
Temp:Wind				0.0433**	0.0628***
Temp:Wind:Sun				0.0212	0.0079
Temp:Wind:Rain				0.0798***	0.0658**
Temp:Sun				0.0186	0.0370
Temp:Rain				0.0123	0.0008
Temp:Rain:Sun				-0.0078	-0.0135
Chi-Square	89289.8	89246.5	89347.2	89248.5	89581.9
Log-Likelihood	-23105.5	-23104.1	-23099	-23097	-23093.1
AIC	46248.9	46254.2	46252	46262	46262.1
BIC	-640678	-640633	-640571	-640498	-640433

Table C.10: Poisson Results, dependent variable are bicycle trips

	No Weather	Weather	Quadratic	Interaction	Quad + Int
Intercept	0.3528***	0.2245***	0.2771***	0.2869***	0.3010***
Age	-0.0646***	-0.0639***	-0.0638***	-0.0638***	-0.0637***
Male	-0.2033***	-0.2019***	-0.2017***	-0.2012***	-0.2012***
Western	-0.1931***	-0.1935***	-0.1933***	-0.1917***	-0.1919***
Non-Western	-0.5549***	-0.5506***	-0.5504***	-0.5519***	-0.5518***
Unknown Ethnicity	0.1652***	0.1744***	0.1758***	0.1746***	0.1758***
Education_Med	-0.1886***	-0.1899***	-0.1902***	-0.1916***	-0.1911***
Education_High	-0.0418***	-0.0507***	-0.0511***	-0.0535***	-0.0528***
Employed	0.0658***	0.0592***	0.0591***	0.0602***	0.0600***
Density	0.0684***	0.0662***	0.0667***	0.0663***	0.0663***
Randstad	-0.0489***	-0.0019	-0.0041	-0.0041	-0.0047
Saturday	-0.3777***	-0.3669***	-0.3658***	-0.3667***	-0.3657***
Sunday	-0.8919***	-0.8886***	-0.8885***	-0.8863***	-0.8871***
October	-0.0894***	-0.0034	-0.0312*	-0.0241	-0.0334**
November	-0.2092***	0.0007	-0.0151	-0.0139	-0.0280
2014	0.0864***	0.0258	0.0233	0.0272*	0.0288*
2015	0.0642***	0.0081	0.0146	-0.0033	0.0054
2016	-0.0370**	-0.0458***	-0.0392**	-0.0552***	-0.0464***
2017	-0.0362**	-0.0423***	-0.0530***	-0.0538***	-0.0522***
Temperature		0.0596***	0.0528***	0.0387***	0.0350***
Wind		-0.0948***	-0.0924***	-0.1193***	-0.1170***
Rain		-0.0412***	-0.0502***	0.0182*	0.0034
Solar Radiation		0.0488***	0.0716***	0.0975***	0.0958***
Temp Squared			-0.0161***		-0.0154***
Wind Squared			-0.0013		-0.0009
Rain Squared			0.0040*		0.0061**
Radiation Squared			-0.0225***		-0.0062
Wind:Sun				-0.0467***	-0.0474***
Wind:Rain				-0.0498***	-0.0482***
Wind:Rain:Sun				-0.0533***	-0.0520***
All				0.0234***	0.0203**
Rain:Sun				0.0993***	0.1010***
Temp:Wind				0.0294***	0.0333***
Temp:Wind:Sun				0.0166**	0.0125*
Temp:Wind:Rain				0.0441***	0.0415***
Temp:Sun				-0.0357***	-0.0220**
Temp:Rain				-0.0398***	-0.0435***
Temp:Rain:Sun				-0.0276***	-0.0315***
Chi-Square	159841	159260	159159	159087	159049
Log-Likelihood	-87941.5	-87642.8	-87620.7	-87573.2	-87566.5
AIC	175921	175332	175295	175214	175209
BIC	-538150	-538704	-538704	-538722	-538691

Table C.11: Negative Binomial Results, dependent variable are bicycle trips

	No Weather	Weather	Quadratic	Interaction	Quad + Int
Intercept	0.2769***	0.1374***	0.2033***	0.2060***	0.2303***
Age	-0.0599***	-0.0601***	-0.0601***	-0.0601***	-0.0601***
Male	-0.2030***	-0.1998***	-0.1998***	-0.1986***	-0.1987***
Western	-0.1898***	-0.1918***	-0.1922***	-0.1899***	-0.1908***
Non-Western	-0.5551***	-0.5536***	-0.5542***	-0.5526***	-0.5523***
Unknown Ethnicity	0.1367**	0.1508**	0.1521**	0.1518**	0.1533**
Education_Med	-0.1824***	-0.1810***	-0.1811***	-0.1830***	-0.1824***
Education_High	-0.0277	-0.0331*	-0.0333*	-0.0356*	-0.0347*
Employed	0.0963***	0.0888***	0.0887***	0.0898***	0.0894***
Density	0.0696***	0.0670***	0.0676***	0.0672***	0.0672***
Randstad	-0.0347**	0.0087	0.0068	0.0058	0.0056
Saturday	-0.3709***	-0.3620***	-0.3607***	-0.3623***	-0.3615***
Sunday	-0.8936***	-0.8958***	-0.8941***	-0.8947***	-0.8947***
October	-0.0867***	0.0109	-0.0227	-0.0134	-0.0274
November	-0.2181***	0.0251	0.0011	0.0027	-0.0161
2014	0.0898***	0.0245	0.0201	0.0287	0.0300
2015	0.0703***	0.0054	0.0158	-0.0025	0.0116
2016	-0.0316	-0.0431*	-0.0403	-0.0534**	-0.0456*
2017	-0.0216	-0.0282	-0.0419	-0.0392	-0.0393
Temperature		0.0680***	0.0581***	0.0448***	0.0392***
Wind		-0.0978***	-0.0904***	-0.1218***	-0.1132***
Rain		-0.0436***	-0.0530***	0.0221	0.0033
Solar Radiation		0.0587***	0.0834***	0.1067***	0.1085***
Temp Squared			-0.0125*		-0.0143
Wind Squared			-0.0046		-0.0048
Rain Squared			0.0039		0.0069*
Radiation Squared			-0.0280***		-0.0127
Wind:Sun				-0.0412***	-0.0416***
Wind:Rain				-0.0523***	-0.0499***
Wind:Rain:Sun				-0.0553***	-0.0544***
All				0.0240*	0.0216
Rain:Sun				0.1023***	0.1011***
Temp:Wind				0.0294***	0.0343***
Temp:Wind:Sun				0.0124	0.0066
Temp:Wind:Rain				0.0396***	0.0368***
Temp:Sun				-0.0374***	-0.0209
Temp:Rain				-0.0416***	-0.0450***
Temp:Rain:Sun				-0.0303***	-0.0338***
Chi-Square	68739.2	69433.4	69429.8	69545.6	69537.2
Log-Likelihood	-71329.4	-71247.4	-71240.1	-71228.5	-71226.2
AIC	142697	142541	142534	142525	142528
BIC	-605402	-605208	-605161	-605061	-605015

Table C.12: Poisson Results, dependent variable are walking trips

	No Weather	Weather	Quadratic	Interaction	Quad + Int
Intercept	-1.6095***	-1.6720***	-1.5805***	-1.6393***	-1.5575***
Age	0.1354***	0.1354***	0.1355***	0.1359***	0.1361***
Male	-0.2880***	-0.2878***	-0.2883***	-0.2885***	-0.2893***
Western	-0.0884***	-0.0897***	-0.0903***	-0.0889***	-0.0907***
Non-Western	-0.1160**	-0.1160**	-0.1178**	-0.1169**	-0.1180**
Unknown Ethnicity	-0.1630**	-0.1601**	-0.1649**	-0.1585**	-0.1627**
Education_Med	0.1009***	0.1010***	0.1012***	0.1001***	0.1013***
Education_High	0.2173***	0.2153***	0.2159***	0.2130***	0.2145***
Employed	-0.1286***	-0.1299***	-0.1299***	-0.1287***	-0.1287***
Density	0.0847***	0.0839***	0.0840***	0.0846***	0.0840***
Randstad	0.0286**	0.0358***	0.0336**	0.0320**	0.0331**
Saturday	0.0577***	0.0611***	0.0603***	0.0759***	0.0738***
Sunday	-0.1445***	-0.1434***	-0.1316***	-0.1418***	-0.1356***
October	0.0158	0.0672***	0.0327	0.0455**	0.0114
November	-0.1043***	0.0170	-0.0108	0.0144	-0.0300
2014	-0.1043***	-0.1387***	-0.1425***	-0.1399***	-0.1324***
2015	-0.1704***	-0.2085***	-0.1970***	-0.2044***	-0.1763***
2016	-0.1575***	-0.1645***	-0.1670***	-0.1568***	-0.1473***
2017	-0.0940***	-0.1037***	-0.1190***	-0.1165***	-0.1205***
Temperature		0.0401***	0.0288***	0.0422***	0.0354***
Wind		-0.0223***	0.0124	-0.0213**	0.0167
Rain		-0.0037	-0.0067	0.0218*	0.0129
Solar Radiation		0.0248***	0.0483***	0.0536***	0.0597***
Temp Squared			-0.0094*		-0.0284***
Wind Squared			-0.0276***		-0.0303***
Rain Squared			0.0014		-0.0016
Radiation Squared			-0.0296***		-0.0300***
Wind:Sun				0.0316***	0.0289***
Wind:Rain				-0.0133	-0.0010
Wind:Rain:Sun				-0.0024	0.0026
All				0.0027	0.0030
Rain:Sun				0.0595***	0.0389***
Temp:Wind				0.0235***	0.0454***
Temp:Wind:Sun				-0.0158*	-0.0304***
Temp:Wind:Rain				0.0193*	0.0112
Temp:Sun				-0.0363***	0.0052
Temp:Rain				-0.0176**	-0.0069
Temp:Rain:Sun				0.0012	0.0101
Chi-Square	148855	148741	148583	148553	148388
Log-Likelihood	-62227.8	-62210	-62172.9	-62154.7	-62122
AIC	124494	124466	124400	124377	124320
BIC	-571074	-571066	-571096	-571056	-571077

Table C.13: Negative Binomial Results, dependent variable are walking trips

	No Weather	Weather	Quadratic	Interaction	Quad + Int
Intercept	-1.6744***	-1.7421***	-1.6385***	-1.7063***	-1.6186***
Age	0.1434***	0.1433***	0.1434***	0.1438***	0.1440***
Male	-0.2993***	-0.2990***	-0.2992***	-0.2999***	-0.3003***
Western	-0.0997**	-0.1026**	-0.1005**	-0.1007**	-0.1012**
Non-Western	-0.0934	-0.0948	-0.0949	-0.0991	-0.0988
Unknown Ethnicity	-0.1375	-0.1330	-0.1397	-0.1307	-0.1374
Education_Med	0.1273***	0.1278***	0.1282***	0.1282***	0.1293***
Education_High	0.2437***	0.2429***	0.2443***	0.2415***	0.2431***
Employed	-0.1252***	-0.1272***	-0.1295***	-0.1251***	-0.1265***
Density	0.0810***	0.0803***	0.0803***	0.0810***	0.0806***
Randstad	0.0402*	0.0462**	0.0445**	0.0422**	0.0432**
Saturday	0.0773***	0.0793***	0.0781***	0.0939***	0.0919***
Sunday	-0.1176***	-0.1178***	-0.1074***	-0.1166***	-0.1115***
October	0.0162	0.0726**	0.0314	0.0477	0.0114
November	-0.1051***	0.0283	-0.0056	0.0249	-0.0225
2014	-0.1017***	-0.1370***	-0.1416***	-0.1400***	-0.1319***
2015	-0.1780***	-0.2190***	-0.2077***	-0.2216***	-0.1925***
2016	-0.1576***	-0.1662***	-0.1690***	-0.1605***	-0.1509***
2017	-0.0933***	-0.1051***	-0.1223***	-0.1212***	-0.1261***
Temperature		0.0422***	0.0282*	0.0455***	0.0375**
Wind		-0.0241**	0.0083	-0.0215*	0.0153
Rain		0.0002	-0.0032	0.0222	0.0147
Solar Radiation		0.0300**	0.0548***	0.0595***	0.0660***
Temp Squared			-0.0134		-0.0324***
Wind Squared			-0.0259***		-0.0296***
Rain Squared			0.0015		-0.0024
Radiation Squared			-0.0313***		-0.0312**
Wind:Sun				0.0329**	0.0279*
Wind:Rain				-0.0085	0.0021
Wind:Rain:Sun				0.0044	0.0071
All				0.0048	0.0069
Rain:Sun				0.0545***	0.0336
Temp:Wind				0.0274**	0.0511***
Temp:Wind:Sun				-0.0143	-0.0275**
Temp:Wind:Rain				0.0191	0.0128
Temp:Sun				-0.0418***	0.0044
Temp:Rain				-0.0129	-0.0012
Temp:Rain:Sun				0.0040	0.0140
Chi-Square	65802.3	65814.2	65880.3	65986.3	66038.2
Log-Likelihood	-50175.7	-50170.2	-50160.2	-50154.3	-50144.8
AIC	100389	100386	100374	100377	100366
BIC	-623031	-622976	-622904	-622814	-622749

D MODE CHOICE APPENDIX

This is the appendix for Chapter 6, where we estimated a latent class choice model. Within this appendix we'll give the full results of the estimation and a couple of additional tables that did not fit within the main text. We display the tables in three main sections, somewhat following the order of the chapter in the main text. We first give the full results of the MNL analysis in Section D.1, followed by all of the estimated parameters of the latent class analysis in Section D.2. Finally we give predictions for choice probabilities that were made using this latent class analysis in Section D.3.

The following sections are present:

D.1 MNL RESULTS

In Table D.1 we give the parameter estimations from the MNL model. These estimations are used in Section 6.3.1.

Table D.1: Parameter estimations from the MNL model.

	PT	BC_leisure	WK	BC_work	BC_edu
ASC	-3.3297***	-1.9939***	-7.0416***		
Distance	0.3006***	-4.6291***	-16.3772***		
Purp_work	1.1395***	0.7763***	-0.7747***		
Purp_edu	3.4623***	3.0754***	0.6357***		
Temp	-0.0200	0.1418***	0.1009***	-0.1082**	-0.1991**
Wind	0.0871**	-0.1105***	-0.0639**	-0.0153	0.0766
Rain	-0.0085	-0.0873***	-0.0236	0.0287	0.0123
Sunshine	-0.0217	0.1037***	0.0737**	-0.0493	-0.0462

* variable is significant at 5% level
** variable is significant at 1% level
*** variable is significant at 0.1 % level

D.2 LATENT CLASS RESULTS

All parameters estimated by all three of the latent class models are given in Table D.2. The estimations are used to interpret each class for all of the latent class models. These interpretations lead to the final selection of the model with three latent classes.

Table D.2: Results from estimating models with 1-5 latent classes

	MNL model		2 Classes		3 Classes		4 Classes	
	estimate	T ratio	estimate	T ratio	estimate	T ratio	estimate	T ratio
asc_pt_1	-3.33	-62.3	-2.51	-35.02	-2.04	-12.68	-5.28	-9.36
asc_bc_1	-1.99	-26.86	-0.89	-9.97	-2.33	-14.2	-1.88	-11.28
asc_wk_1	-7.04	-23.13	-7.03	-22.5	-5.63	-16.76	-8.2	-13.4
asc_pt_2			-4.49	-32.35	-3.14	-23.23	-1.58	-4.12

Continued on next page

	MNL model		2 Classes		3 Classes		4 Classes	
	estimate	T ratio	estimate	T ratio	estimate	T ratio	estimate	T ratio
asc_bc_2			-3.7	-31.54	-0.82	-8.29	0.1	0.66
asc_wk_2			-7.05	-22.37	-7.52	-20.26	-6.65	-9.3
asc_pt_3					-5.28	-27.76	-4.84	-6.01
asc_bc_3					-3.86	-30.94	-4.77	-21.27
asc_wk_3					-8.63	-24.75	-8.85	-14.28
asc_pt_4							-2.45	-3.3
asc_bc_4							-2.56	-8.99
asc_wk_4							-5.74	-10.87
b_dst_pt	0.3	15.86	0.29	13.73	0.3	13.61	0.26	9.84
b_dst_bc	-4.63	-23.93	-4.52	-20.73	-4.63	-20.18	-4.98	-18.71
b_dst_wk	-16.38	-23.54	-16.47	-23.12	-17.22	-20.62	-17.46	-12.42
b_purp_work_pt	1.14	14.12	1.35	14.92	1.31	13.66	1.34	8.64
b_purp_work_bc	0.78	14.21	1.03	15.85	1.04	15.4	1.06	11.93
b_purp_work_wk	-0.77	-7.84	-0.71	-7.07	-0.8	-7.64	-0.8	-6.03
b_purp_edu_pt	3.46	39.31	3.28	34.84	3.49	30.78	3.26	24.74
b_purp_edu_bc	3.08	31.41	2.79	27.51	2.82	26.22	2.79	23.44
b_purp_edu_wk	0.64	3.52	0.39	2.23	0.5	2.79	0.41	2.1
b_temp_pt_1	-0.02	-0.59	-0.01	-0.29	-0.13	-1.82	0.11	0.58
b_temp_bc_1	0.14	6.44	0.16	5.17	0.15	1.99	0.23	5.78
b_temp_wk_1	0.1	3.92	0.11	2.47	0.15	2.49	0.13	2.34
b_wind_pt_1	0.09	2.9	0.1	2.35	0.12	1.77	0.1	0.54
b_wind_bc_1	-0.11	-5.11	-0.1	-3.41	-0.1	-1.18	-0.15	-3.5
b_wind_wk_1	-0.06	-2.86	-0.03	-0.74	-0.17	-3.76	-0.09	-1.44
b_ri_pt_1	-0.01	-0.31	0	0.01	-0.14	-2.6	0.32	3.04
b_ri_bc_1	-0.09	-4.93	-0.13	-5.71	-0.23	-3.51	-0.09	-2.73
b_ri_wk_1	-0.02	-1.26	-0.08	-2.47	-0.06	-1.39	-0.08	-1.76
b_ss_pt_1	-0.02	-0.62	0.02	0.33	-0.02	-0.25	0.25	0.93
b_ss_bc_1	0.1	4.66	0.11	3.37	0.1	1.37	0.1	2.57
b_ss_wk_1	0.07	3.03	0.11	2.4	0.04	0.76	-0.02	-0.43
b_work_temp_bc_1	-0.11	-2.61	-0.2	-3.06	-0.04	-0.27	-0.11	-1.08
b_work_wind_bc_1	-0.02	-0.38	0.02	0.27	-0.3	-2.04	0.04	0.39
b_work_rain_bc_1	0.03	0.84	0.03	0.63	0.21	1.92	-0.01	-0.14
b_work_sun_bc_1	-0.05	-1.2	-0.06	-0.84	-0.03	-0.25	0	0.01
b_edu_temp_bc_1	-0.2	-2.75	-0.19	-1.99	-0.12	-0.47	-0.62	-2.91
b_edu_wind_bc_1	0.08	1.18	0.07	0.8	-0.17	-1.02	0.32	1.49
b_edu_rain_bc_1	0.01	0.22	0.03	0.43	0.09	0.55	0.26	1.5
b_edu_sun_bc_1	-0.05	-0.69	-0.05	-0.63	0.07	0.37	0.13	0.78
b_temp_pt_2		-0.03	-0.38	0.06	0.89	0.11	1.19	
b_temp_bc_2		0.2	4.06	0.13	3.91	0.14	1.78	
b_temp_wk_2		0.09	2.36	0.1	2.33	0.21	1.89	
b_wind_pt_2		0.09	1.39	0.05	0.82	0.05	0.47	
b_wind_bc_2		-0.22	-3.82	-0.07	-2.23	-0.09	-1.36	
b_wind_wk_2		-0.08	-2.38	-0.01	-0.12	-0.04	-0.37	
b_ri_pt_2			-0.1	-1.31	0.09	1.87	-0.04	-0.49
b_ri_bc_2			-0.12	-2.41	-0.14	-5.38	-0.21	-4.49
b_ri_wk_2			0.01	0.23	-0.09	-2.59	-0.1	-1.68
b_ss_pt_2			-0.16	-2.02	0.04	0.69	0.09	1.04
b_ss_bc_2			0.09	1.71	0.11	3.1	0.3	3.39
b_ss_wk_2			0.06	1.59	0.1	2.13	0.38	3.28
b_work_temp_bc_2		-0.07	-0.78	-0.18	-2.37	-0.24	-1.64	
b_work_wind_bc_2		0.06	0.71	0.07	0.95	0.07	0.58	
b_work_rain_bc_2		0.08	1.04	0.03	0.47	0.08	0.98	

Continued on next page

	MNL model		2 Classes		3 Classes		4 Classes	
	estimate	T ratio	estimate	T ratio	estimate	T ratio	estimate	T ratio
b_work_sun_bc_2		-0.05	-0.5	-0.04	-0.48	-0.18	-1.42	
b_edu_temp_bc_2		-0.37	-2.35	-0.12	-1.04	0.07	0.35	
b_edu_wind_bc_2		0.26	1.97	0.06	0.64	0.05	0.39	
b_edu_rain_bc_2		0.13	1.11	0.03	0.41	-0.04	-0.27	
b_edu_sun_bc_2		0.03	0.25	-0.07	-0.76	-0.29	-1.58	
b_temp_pt_3				0.12	1.07	0.09	0.75	
b_temp_bc_3				0.19	3.17	0.27	2.1	
b_temp_wk_3				0.01	0.12	-0.09	-0.99	
b_wind_pt_3				0.01	0.12	-0.02	-0.12	
b_wind_bc_3				-0.2	-2.49	-0.33	-2.88	
b_wind_wk_3				-0.02	-0.31	0.01	0.05	
b_ri_pt_3					-0.06	-0.57	-0.1	-0.47
b_ri_bc_3					-0.09	-1.57	-0.06	-0.7
b_ri_wk_3					-0.01	-0.28	0.01	0.05
b_ss_pt_3					-0.4	-2.72	-0.27	-1.39
b_ss_bc_3					0.08	1.2	0.06	0.52
b_ss_wk_3					0.11	2.05	0.14	1.14
b_work_wind_bc_3				0.13	1.09	0.25	1.6	
b_work_rain_bc_3				0.02	0.22	-0.02	-0.17	
b_work_sun_bc_3				0.01	0.09	0.05	0.33	
b_edu_temp_bc_3				-0.42	-2.27	-0.45	-2.01	
b_edu_wind_bc_3				0.35	1.71	0.45	1.42	
b_edu_rain_bc_3				0.09	0.65	-0.02	-0.07	
b_edu_sun_bc_3				-0.04	-0.2	-0.08	-0.34	
b_temp_pt_4						-0.21	-2.14	
b_temp_bc_4						0.13	1.24	
b_temp_wk_4						0.13	1.86	
b_wind_pt_4						0.13	0.74	
b_wind_bc_4						-0.1	-1.05	
b_wind_wk_4						-0.18	-3.03	
b_ri_pt_4							-0.14	-0.82
b_ri_bc_4							-0.22	-3.26
b_ri_wk_4							-0.06	-1.11
b_ss_pt_4							-0.12	-0.75
b_ss_bc_4							0.09	1.08
b_ss_wk_4							0.02	0.33
b_work_temp_bc_4						-0.03	-0.15	
b_work_wind_bc_4						-0.32	-2.15	
b_work_rain_bc_4						0.17	1.42	
b_work_sun_bc_4						-0.06	-0.39	
b_edu_temp_bc_4						-0.03	-0.08	
b_edu_wind_bc_4						-0.22	-1.26	
b_edu_rain_bc_4						0.03	0.17	
b_edu_sun_bc_4						0.01	0.03	

Table D.3: Class-membership function estimates for models with 2-5 latent classes

	2 Classes		3 Classes		4 Classes	
Variable	Estimate	T ratio	Estimate	T ratio	Estimate	T ratio
delta_1	0		0		0	
delta_2	-0.6321	-1.25	5.1507	2.43	8.8348	0.38
delta_3			5.0726	2.36	9.0867	0.39
delta_4					-3.66	-0.01
gamma_ebike_2	-0.3422	-3.04	0.2937	2.08	0.2428	1.39
gamma_male_2	0.0241	0.33	-0.0183	-0.2	0.333	2.52
gamma_age_2	0.129	5.42	-0.1355	-5.01	-0.1834	-4.96
gamma_employed_2	-0.4624	-0.99	-2.8884	-1.39	-8.8994	-0.38
gamma_education_2	-0.0212	-0.91	-0.0919	-3.26	-0.0106	-0.17
gamma_density_2	-0.0697	-2.24	-0.0784	-1.8	0.2167	3.09
gamma_license_2	-1.0934	-5.87	0.2487	1.39	0.3513	1.6
gamma_CarAtt_2	0.4009	9.02	-0.1801	-3.1	-0.2975	-3.63
gamma_TrainAtt_2	-0.2225	-5.45	0.1202	1.71	0.1207	1.53
gamma_BikeAtt_2	-0.8765	-16.35	0.7246	10.41	0.269	1.34
gamma_Car_2	1.2401	10.05	-0.3665	-2.63	-0.6779	-3.04
gamma_BTMAtt_2	-0.2832	-6.15	-0.2424	-3.45	0.0697	0.58
gamma_ebike_3			-0.1797	-1.18	-0.4346	-2.48
gamma_male_3			-0.0088	-0.09	0.1236	1.01
gamma_age_3			-0.0774	-2.45	-0.0207	-0.32
gamma_employed_3			-3.6157	-1.72	-9.8343	-0.42
gamma_education_3			-0.1096	-3.38	-0.0495	-0.63
gamma_density_3			-0.1304	-2.81	0.0592	0.73
gamma_license_3			-1.8512	-5.73	-2.4509	-1.78
gamma_CarAtt_3			0.3473	5.72	0.4212	5.91
gamma_TrainAtt_3			-0.1512	-2.11	-0.2268	-3.05
gamma_BTMAtt_3			-0.1256	-1.69	0.1676	1.44
gamma_BikeAtt_3			-0.2943	-4	-1.0687	-15.2
gamma_Car_3			0.9208	4.51	1.1088	1.44
gamma_ebike_4					-0.1205	-0.71
gamma_male_4					0.1125	1.04
gamma_age_4					0.0665	1.62
gamma_employed_4					1.9811	0
gamma_education_4					0.0568	1.46
gamma_density_4					0.1402	1.71
gamma_license_4					0.0665	0.24
gamma_CarAtt_4					0.0376	0.37
gamma_TrainAtt_4					-0.078	-1.02
gamma_BTMAtt_4					0.2153	2.11
gamma_BikeAtt_4					-0.5735	-5.07
gamma_Car_4					0.1182	0.41

D.3 LATENT CLASS POST-ESTIMATION

For each of the three latent classes we calculated predicted choice probabilities for trips given certain weather condition. These probabilities are used in the interpretation of the classes in Section 6.3.3.

Table D.4: Predicted choice probabilities for work trips by all three classes given certain weather conditions

	Class 1				Class 2				Class 3			
Distance: 11 km												
Type of weather	Car	PT	BC	WK	Car	PT	BC	WK	Car	PT	BC	WK
Mean	0.57	0.27	0.16	0	0.41	0.07	0.52	0	0.93	0.02	0.06	0
Rainstorm	0.74	0.24	0.02	0	0.53	0.2	0.27	0	0.96	0.02	0.02	0
Rain, Overcast	0.55	0.29	0.16	0	0.43	0.06	0.52	0	0.92	0.03	0.05	0
Wind, Rain	0.7	0.21	0.09	0	0.5	0.13	0.37	0	0.95	0.01	0.04	0
Near-freezing	0.52	0.37	0.11	0	0.38	0.05	0.57	0	0.93	0.01	0.05	0
Good	0.53	0.21	0.25	0	0.4	0.06	0.53	0	0.92	0.02	0.07	0
Great	0.5	0.14	0.35	0	0.4	0.08	0.52	0	0.91	0.01	0.08	0
Distance: 5 km												
Type of weather	Car	PT	BC	WK	Car	PT	BC	WK	Car	PT	BC	WK
Mean	0.41	0.18	0.36	0.05	0.19	0.03	0.77	0	0.82	0.01	0.16	0
Rainstorm	0.71	0.21	0.05	0.02	0.33	0.12	0.55	0	0.91	0.02	0.06	0
Rain, Overcast	0.39	0.19	0.37	0.05	0.2	0.03	0.77	0	0.82	0.02	0.15	0
Wind, Rain	0.57	0.16	0.24	0.04	0.28	0.07	0.66	0	0.88	0.01	0.11	0
Near-freezing	0.41	0.27	0.28	0.04	0.17	0.02	0.81	0	0.83	0.01	0.15	0
Good	0.33	0.12	0.5	0.06	0.19	0.03	0.78	0	0.8	0.01	0.18	0.01
Great	0.27	0.07	0.59	0.07	0.19	0.03	0.77	0.01	0.77	0.01	0.21	0.01

Table D.5: Predicted choice probabilities for educational trips by all three classes given certain weather conditions

	Class 1				Class 2				Class 3			
Distance: 11 km												
Type of weather	Car	PT	BC	WK	Car	PT	BC	WK	Car	PT	BC	WK
Mean	0.14	0.62	0.24	0	0.1	0.14	0.75	0	0.66	0.11	0.23	0
Rainstorm	0.25	0.72	0.03	0	0.15	0.38	0.47	0	0.49	0.09	0.43	0
Rain, Overcast	0.14	0.64	0.22	0	0.09	0.22	0.69	0	0.62	0.18	0.21	0
Wind, Rain	0.24	0.63	0.13	0	0.13	0.31	0.56	0	0.62	0.08	0.3	0
Near-freezing	0.11	0.71	0.18	0	0.09	0.22	0.69	0	0.57	0.08	0.35	0
Good	0.14	0.5	0.36	0	0.09	0.2	0.71	0	0.72	0.11	0.18	0
Great	0.14	0.35	0.51	0	0.08	0.19	0.73	0	0.79	0.09	0.11	0
Distance: 5 km												
Type of weather	Car	PT	BC	WK	Car	PT	BC	WK	Car	PT	BC	WK
Mean	0.09	0.37	0.49	0.04	0.04	0.09	0.87	0	0.43	0.07	0.49	0.01
Rainstorm	0.24	0.65	0.08	0.02	0.07	0.19	0.74	0	0.25	0.04	0.7	0
Rain, Overcast	0.09	0.4	0.47	0.04	0.04	0.09	0.87	0	0.42	0.11	0.46	0.01
Wind, Rain	0.18	0.45	0.32	0.04	0.06	0.14	0.8	0	0.37	0.04	0.58	0.01
Near-freezing	0.08	0.48	0.41	0.03	0.04	0.09	0.88	0	0.32	0.04	0.63	0.01
Good	0.08	0.25	0.63	0.05	0.03	0.08	0.88	0	0.51	0.07	0.41	0.01
Great	0.06	0.15	0.73	0.06	0.03	0.07	0.89	0	0.63	0.07	0.29	0.02

COLOPHON

MSc Thesis: The influence of weather on travel behaviour - a multi-method analysis

Master Engineering & Policy Analysis
Faculty of Technology, Policy, and Management
Delft University of Technology
Jaffalaan 5, 2628 BX
Delft, the Netherlands

The result of a thesis internship at
KiM Netherlands Institute for Transport Policy Analysis
Bezuidenhoutseweg 20, 2594 AV
The Hague, the Netherlands.

This document was typeset using LaTeX. The document layout was generated using the arsclassica package by Lorenzo Pantieri, which is an adaption of the original classicthesis package from André Miede.