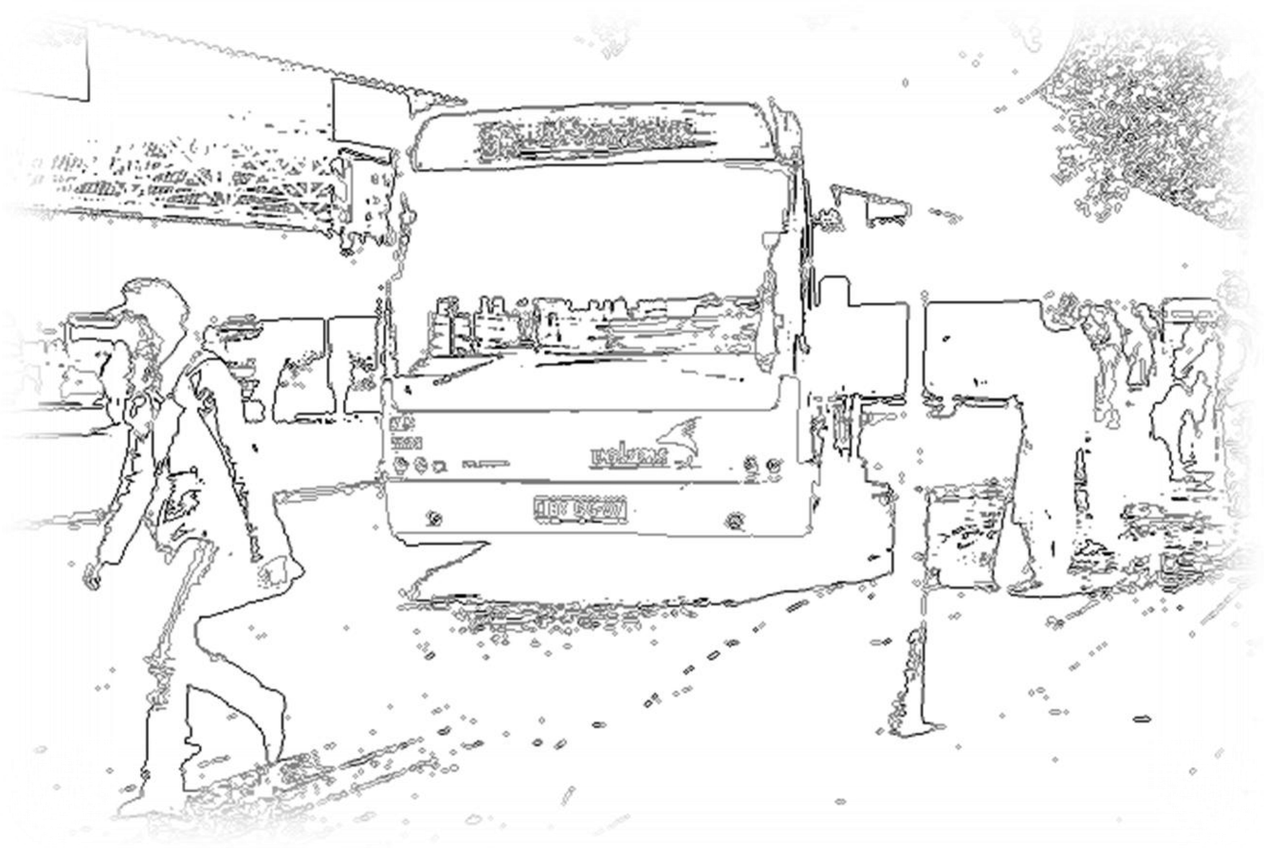
Exploring the demand for bus transport

An exploration of the factors that determine the demand for bus transport and the development of a bus demand model



Master thesis

[Final report]

Marijn Smit

June, 2014

Exploring the demand for bus transport

An exploration of the factors that determine the demand for bus transport and the development of a bus demand model

M.W. Smit BSc
Civil Engineering and Management
University of Twente
marijnsmit@gmail.com

Supervisors:

Prof. Dr. Ing. K.T. Geurs
University of Twente

Ing. K.M. van Zuilekom
University of Twente

Dr. Ir. M.E. Kraan
Grontmij

Ing. C. Doeser
Grontmij

UNIVERSITY OF TWENTE.

Grontmij

Illustration title page: Bus station Breda, adapted from Mensonides (2008).

Abstract

Introduction

The challenge for bus operators is to optimize the supply of bus transport to demand as cost effective as possible. This demand is however not static; it depends on the supply of bus transport, population characteristics, urban developments, etc. In order to optimize the supply of bus transport, detailed information is needed about the impact of bus transport, population and spatial characteristics on demand. This thesis explores the factors that determine the demand for bus transport and uses those factors to develop a model which is able to predict the number of boarders and alighters at a bus stop and assigns them to the bus network.

Methodology

A four step model approach is followed to create bus demand models, including the stages trip generation, distribution and assignment. A literature study is performed to identify the factors that determine the demand for bus transport. Data from different sources are used to represent those factors. The data of these variables are collected for the catchment area of each bus stop in Breda. Veolia Transport supplied chipcard data of Breda and Tilburg which are used to create the trip generation models via regression analyses. Also the variation of the number of passengers over time is analyzed. For the distribution step, distribution functions are calibrated for weekdays (Monday to Friday), Saturdays and Sundays.

Results

The factors that determine the demand for bus transport can be categorized in the categories population characteristics, spatial characteristics, bus service characteristics and trip specific characteristics, although many factors are related to multiple categories. The research showed that the demand for bus transport is mainly explained by the presence of the central station, the number of stops that can be reached from a stop without transfer, the floor area of offices and shops and the number of students at higher educational institutes.

Six trip generation models are developed for Breda: boarders and alighters weekday (Monday to Friday), boarders and alighters Saturday and boarders and alighters Sunday. The models perform quite well for Breda, but worse for Tilburg. The calibrated top log-normal distribution function is well able to reproduce the observed trip length distribution of both Breda and Tilburg. The combination of the trip generation and distribution models result in an average difference of 16% with the observed number of passengers at the busiest line segments in Breda, while this is 34% in Tilburg.

Conclusion

The demand for bus transport can mainly be explained by spatial factors, like the presence of the central train station, offices, shops and higher education institutes. These spatial factors represent facilities that attract a relatively large number of passengers. Since the models perform worse for Tilburg than for Breda, it can be concluded that bus demand models developed for one area cannot simply be applied to another area. This is emphasized by the high sensitivity of the models for the sample of bus stops which are used to create the models.

Summary

Introduction

The challenge for bus operators is to optimize the supply of bus transport to demand as cost effective as possible. This demand is however not static; it depends on the supply of bus transport, population characteristics, urban developments, etc. In order to optimize the supply of bus transport, detailed information is needed about the impact of bus transport, population and spatial characteristics on demand. The recent introduction of the public transport chipcard and the availability of more detailed socio-economic and spatial data offer opportunities to gain more insight in the factors that determine the demand for bus transport. This thesis explores the factors that determine the demand for bus transport and uses those factors to develop a model which is able to predict the number of boarders and alighters at a bus stop and assigns them to the bus network.

Research design

The objective of this research is to explore the factors that determine the demand for bus transport and implement these characteristics in a model that is able to predict the number of boarders and alighters at a bus stop and use those to predict the number of bus passengers in the bus network.

The developed model is a four step model with the stages trip generation, trip distribution, modal split and assignment. The third stage -the modal split- is not considered, since it is a unimodal demand model. Figure 1 shows the procedure of the development of the four step model.

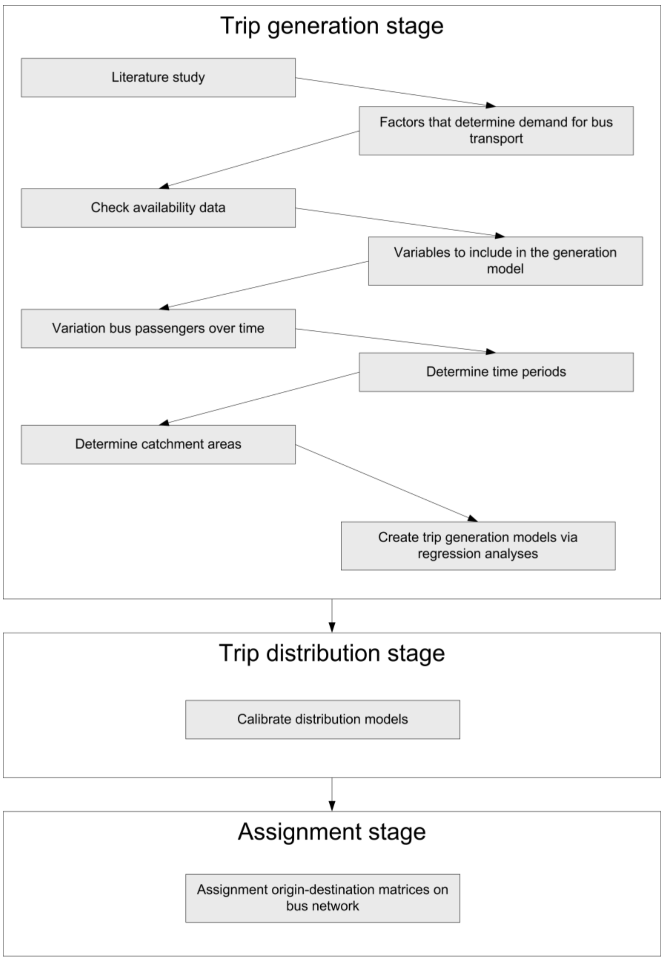


Figure 1: Procedure of the development of the four step model

Literature
The factors that influence the demand for bus transport which were found in literature can be categorized in the categories spatial characteristics, population characteristics trip specific characteristics and bus service characteristics. Examples of factors are age, population (size), ethnicity (population characteristics), college enrolments, number of jobs, density (spatial characteristics), trip purpose (trip specific characteristics), accessibility, fare and quality of service (bus service characteristics).

Data assessment
Data from several sources are used to represent the factors that were found in literature. The main sources are the Centraal Bureau voor de Statistiek (CBS) for population characteristics, Dienst Uitvoering Onderwijs (DUO) for education related indicators, OpenStreetMap for the roads, buildings and bus stop locations, Kadaster for the data about addresses (Basisregistratie Adressen en Gebouwen) and Veolia Transport for the chipcard data.

Most of the data is available at the spatial levels 4-position postcode, neighbourhood, squares of 500m and squares of 100m. All spatial data which is used is converted to the squares of 100m, since this is the spatially most detailed level. The last stage of data processing is the collection of the data within the catchment area of 200 meters of each bus stop.

Veolia Transport provided the chipcard data of Breda and Tilburg of 2012. This data contains the hourly number of boarders and alighters at each bus stop in Breda and Tilburg in 2012. These datasets were used to create the trip generation models. Veolia Transport also provided data about the yearly number of passengers between each bus stop pair in Breda and Tilburg in 2012, divided in yearly number of passengers on weekdays (Monday to Friday), yearly number of passengers on Saturdays and yearly number of passengers on Sundays. None of the available chipcard data contains information about transfers.

Variation over time
The percentage of boarders during each month is approximately 7.7-10.8% of the yearly number of boarders, except for July and August with only 3.9 and 5.3%. Of all bus passengers, 90% travel on weekdays with an equal distribution over the weekdays (Monday to Friday), 6.3% travel on Saturdays and the remaining 3.7% travels on Sundays. Holidays have a quite large impact on the number of bus passengers: during the holiday weeks in 2012 the weekly number of boarders is about half of the average number of boarders. Although there is a morning peak on weekdays between 8:00 and 9:00h, during a period of four hours in the afternoon the hourly number of passengers is equal to or higher than during the morning peak.

It seems that there are some weather characteristics that influence the number of boarders. The temperature and sun hours seems to have a negative correlation with the number of boarders, while the correlation of precipitation is positive. Also the effect of large events is rather limited. The effect of some events is visible in the evening hours, like Carnival and Dancetour, but the change of the number of passengers is negligible.

Generation models

Six trip generation models are developed: boarders and alighters weekday (Monday to Friday), boarders and alighters Saturday and boarders and alighters Sunday. First, from all 67 explanatory variables the variables are selected which have a significant correlation with the dependent variable (e.g. boarders on weekdays) at both Breda and Tilburg. Via an iterative process the variables are selected which are included in the final generation models. Table 1 shows the variables which are included in the six trip generation models.

Table 1: Variables of the six trip generation models

Variable	Boarders			Alighters		
	weekday	Saturday	Sunday	Weekday	Saturday	Sunday
Number of stops without transfer	✓	✓	✓	✓	✓	✓
Dummy for Central station	✓	✓	✓	✓	✓	✓
Percentage low income households	✓	✓	✓	✓	✓	✓
Floor area of offices	✓	✓	✓	✓	✓	✓
Floor area retail, social gathering and accommodation	✓	✓	✓	✓	✓	✓
Address density	✓	✓	✓	✓	✓	✓
Total number of cars	✓	✓	✓	✓	✓	✓
% of floor area with residential function	✓			✓		
Number of students at higher education institutes	✓			✓		
Dummy for University (of Applied Science)	✓			✓		
Dummy for shopping centre (>5000 m^2 floor area retail, social gathering and accommodation)		✓	✓		✓	✓

Distribution models

While the trip generation models determine the number of boarders and alighters at each bus stop, distribution models are used to determine how many passengers there are at each bus stop pair with a direct connection. Three distribution models have been created: one for weekdays, one for Saturdays and one for Saturdays. For each time period three distribution functions are calibrated from which the best one is selected. The top log-normal function is selected, since it is almost perfectly able to reproduce the observed trip length distribution.

Validation

Nine models have been developed for Breda: six trip generation models and three distribution models. The generation models are validated for Breda and Tilburg, the distribution models are validated for Tilburg and the combination of the generation and distribution models are validated for both Breda and Tilburg.

The differences between the observed and modelled number of boarders and alighters at the weekdays summed over all bus stops in the municipality are very limited. An overestimation at some bus stops is compensated by underestimations at other bus stops. The over- and underestimations are however larger at Tilburg than at Breda, which indicates a worse performance of the models. At Saturdays and Sundays, the differences between the observed and modelled numbers are much larger. The distribution models perform slightly worse when applied to Tilburg, but they also match the observed trip length distributions very well.

The combination of the trip generation and trip distribution models perform very well for Breda. The cumulative average difference between the observed and modelled number of passengers is only 16% at the line segments with at least 1,000 passengers and increase to 40% at the busiest half of the line segments. For Tilburg the differences are 34% at the line segments with at least 1,000 observed passengers and 66% at the busiest half of the line segments.

Conclusion

The demand for bus transport can mainly be explained by spatial factors, like the presence of the central train station, offices, shops and higher education institutes. These spatial factors represent facilities that attract a relatively large number of passengers. Other factors that are important are the accessibility indicator (number of stops that can be reached without transfer), the address density, percentage of low income households and the total number of cars.

The trip generation and distribution models are developed for Breda and perform quite well for Breda, but worse when applied to Tilburg. The worse result in Tilburg is mainly caused by the generation models. Since the models perform worse for Tilburg than for Breda, it can be concluded that bus demand models developed for one area cannot simply be applied to another area. Very specific combinations of circumstances have a large influence on the developed models, which is emphasized by the high sensitivity of the models for the sample of bus stops.

Recommendations

Larger study area

Very specific combinations of circumstances appeared to have a large influence on the trip generation models. Therefore more research is necessary to identify the variables that determine the demand for bus transport in general. An analysis with more municipalities might show why a variable is correlated with the number of passengers in one municipality and not in the other.

Use more detailed chipcard data

The chipcard data contains more detailed data than used in this study, like the number of passengers per bus line, the travel product people use to pay for their bus trip and the daily number of passengers between each bus stop pair. Using this more detailed data could provide more insight in improve the developed models.

Add explanatory variables

Some variables that could be relevant for the demand for bus transport are not included in this study but could be examined in further research. Examples of such variables are the directness of the bus route, frequency of the bus service and accessibility of the bus stops.

Preface

This Master's thesis concludes my study Civil Engineering and Management at the University of Twente in Enschede. It is the result of eight months of internship at the mobility department of Grontmij in De Bilt.

The start of the graduation process was quite a challenge. Finding subjects was not really an issue: there were plenty of interesting topics. After a selection process, a few topics related to bus transport remained. These topics raised an issue: finding data of public transport in general is difficult, but for bus transport this is even harder. Attempts of myself to gain access to chipcard data came to nothing, but luckily Cees Doeser found Veolia Transport willing to supply chipcard data. None of my supervisors had experience with the possibilities and limitations of this relatively new data source however, so also the design of the research was a challenge.

Finally, the research could begin. At some point, the research went so smooth, that Kasper van Zuilekom challenged me to extend the scope of the thesis. Until that moment I only considered the number of boarders at each bus stop, but now I also considered the distribution of the bus passengers over the bus stops and the number of passengers at the bus lines. Although some extra months were necessary because of the extended scope, I am happy with the added value it has to the thesis.

There are some people who I would like to thank. First I would like to thank Rob Kooloos and Piet van den Bosch of Veolia Transport for the supply of the chipcard data. I would also like to thank them for sharing their knowledge of the chipcard data and the bus system in general. I would like to thank my supervisors at Grontmij Mariëtte Kraan and Cees Doeser for their feedback and assistance. I also would like to thank Kasper van Zuilekom, my daily supervisor, for the joined-up thinking about the research and the discussions about the reports and Karst Geurs for the feedback on the reports. Finally, I would like to thank my family for supporting me.

De Bilt, June 2014

Marijn Smit

Table of contents

1 Introduction

People that want or need to travel have several transport modes to choose from. People living in large urban areas can choose for the bicycle, car and several types of public transport like train, metro, tram or bus. People living in rural areas are usually restricted to bicycle, car and sometimes bus. Travelling is quite easy for people that choose for the bicycle or the car: the road network is always available and offers only limited restrictions. For people that choose for public transport the story becomes more complicated, since the supply of public transport is both time and space restricted. If the supply of public transport is not properly aligned with the transport demand, it will be more attractive for people to use the bicycle or car.

Train, tram and metro operators only have limited tools to align supply with demand, since they are bounded to fixed infrastructure. The bus system however is more flexible, because it is only to a limited extent bounded to fixed infrastructure. The challenge for bus operators, municipalities and concession holders is to optimize the supply of bus transport to meet the demand as cost-efficient as possible. This demand is however not static, it depends, among others, on the supply of public transport, population specific characteristics and urban developments. Because of changing population characteristics in neighbourhoods it might be that other bus routes are more effective.

Before the national introduction of the public transport chipcard (finished in 2012), the NVS counts were the main source of information about the number of passengers in the buses. These counts were only spot checks performed during two weeks in March or November, which resulted in very rough estimations. The estimations were rough because of potential counting (or estimation) errors and because of the limited number of counting days. These counts are therefore not suitable to optimize the supply of bus transport. Because of the lack of detailed data, it was not known which effect urban and demographic developments have on the number of bus passengers. The chipcard data however offers detailed information about the number of boarders and alighters per hour at each bus stop and the origin and destination stops.

This detailed chipcard data and more detailed data from institutions like the Dutch Statistics offers opportunities to gain a better insight in the factors that influence the demand for bus transport. More insight in those factors offers opportunities to develop more detailed models that provide a more accurate prediction of the number of bus passengers and can be used to optimize the exploitation of the bus system by bus operators. It also enables predictions about the number of boarders at planned bus stops to align the supply of bus transport with the demand. This study therefore explores the factors that determine the demand for bus transport and use those factors to develop a model that is able to predict the number of boarders at a bus stop and assign those to the bus network to support decision-making.

1.1 Public transport in The Netherlands

1.1.1 Concession areas

Since 2001 the 19 regional public transport authorities are obliged to tender the public transport in their areas. In that year the new law that manages public transport in the Netherlands came into force: the “Wet Personenvervoer 2000”. The transport company that wins the tender is granted the exclusive right to provide the regional public transport in the concession area for a certain period of time. Figure 2 shows the 45 concession regions, which are granted to 13 transport companies (Kennisplatform Verkeer en Vervoer, 2013).

Figure 2: Concession areas in the Netherlands at January 1st, 2014. Adapted from Kennisplatform Verkeer en Vervoer, 2013.

1.1.2 Chipcard system

Since 1980 passengers could pay for their journey with public transport with the *strippenkaart*. Although it was an easy and cheap system, there were some disadvantages. One of the major disadvantages was that the sales of the *strippenkaart* was centrally organized and thus only provided information about the sales of the *strippenkaart* and not about how they were used. The *strippenkaart* did not provide any information about the number of passengers nor about their travel behaviour (e.g. board and alight stops). To gain some insight in the number of passengers each year during a few weeks the NVS counts (Normeringssysteem Voorzieningenniveau Streekvervoer) were conducted in March and/or November. At each bus line some counting stops were appointed. At those bus stops the bus driver counted the number of bus passengers in the bus (Directoraat Generaal van het Verkeer, 1981). This system gave some insight in the number of passengers at that moment, yet only a rough indication. It gave however no information about the origins and destinations of those passengers nor about the number of passengers in other periods.

The central organisation of the sales of the *strippenkaart* also resulted in difficulties with the distribution of the turnover over the operators. A complex system, the WROOV system (Werkgroep Reizigers Omvang en Omvang Verkopen), was needed to determine which fraction of the sales belonged to each operator (Centrum Vernieuwing Openbaar Vervoer, 2004).

Another disadvantage of the *strippenkaart* was the tariff structure, namely through zones. Since the tariff depended on the number of zones a passenger passed, the size and location of zone borders had much influence on the costs of travelling with public transport. This way a short trip could be more expensive than a long trip, just because of the zone layout.

From 2007 the new chipcard system was introduced gradually. In this new system the *strippenkaart* was replaced by a system with chipcards, requiring passengers to check in and check out (Van der Zwan, 2011). This provides detailed information about the boarding and alighting stops, travelled distance and allows spatial and temporal tariff differentiation. The chipcard system is maintained by Trans Link Systems (TLS), a joint venture of public transport operators. All transactions between passengers and the transport operators go through TLS. Since the chipcard system enables detailed information collection, there is a strict privacy policy. Because of this policy, not all information that is collected by the chipcard system can be accessed by the transport operators. Because of this, TLS does not provide information about the transfers of public transport passengers to the operators, creating an empty space in their data.

Data which is available are the number of boarders and alighters per bus line per bus stop, aggregated per hour. It also shows information about the travel product that is used to pay (e.g. credit, student subscription, age discount, etc.) and the distances travelled. It is however not possible to identify individual passengers to see for example how their travel behaviour over multiple days is.

1.2 Study area

The study focuses on the municipalities of Breda and Tilburg since the data of these municipalities is made available by Veolia. Breda is a municipality in the province of Noord-Brabant with a population of 178.000 of which 147.000 in the city itself (breda.incijfers.nl). Figure 3 shows the bus lines in Breda. The bus network consists of 10 urban and 18 regional bus lines. Those lines serve 185 bus stops in the municipality. Special facilities in Breda that are relevant for bus transport are the hospital Amphia with two locations, the Universities of Applied Sciences Avans and NHTV, the Royal Military Academy (KMA) in the city centre and the stadium of Breda's football club NAC. Breda has two train stations: Central station (intercity station) and Breda-Prinsenbeek (regional train station).

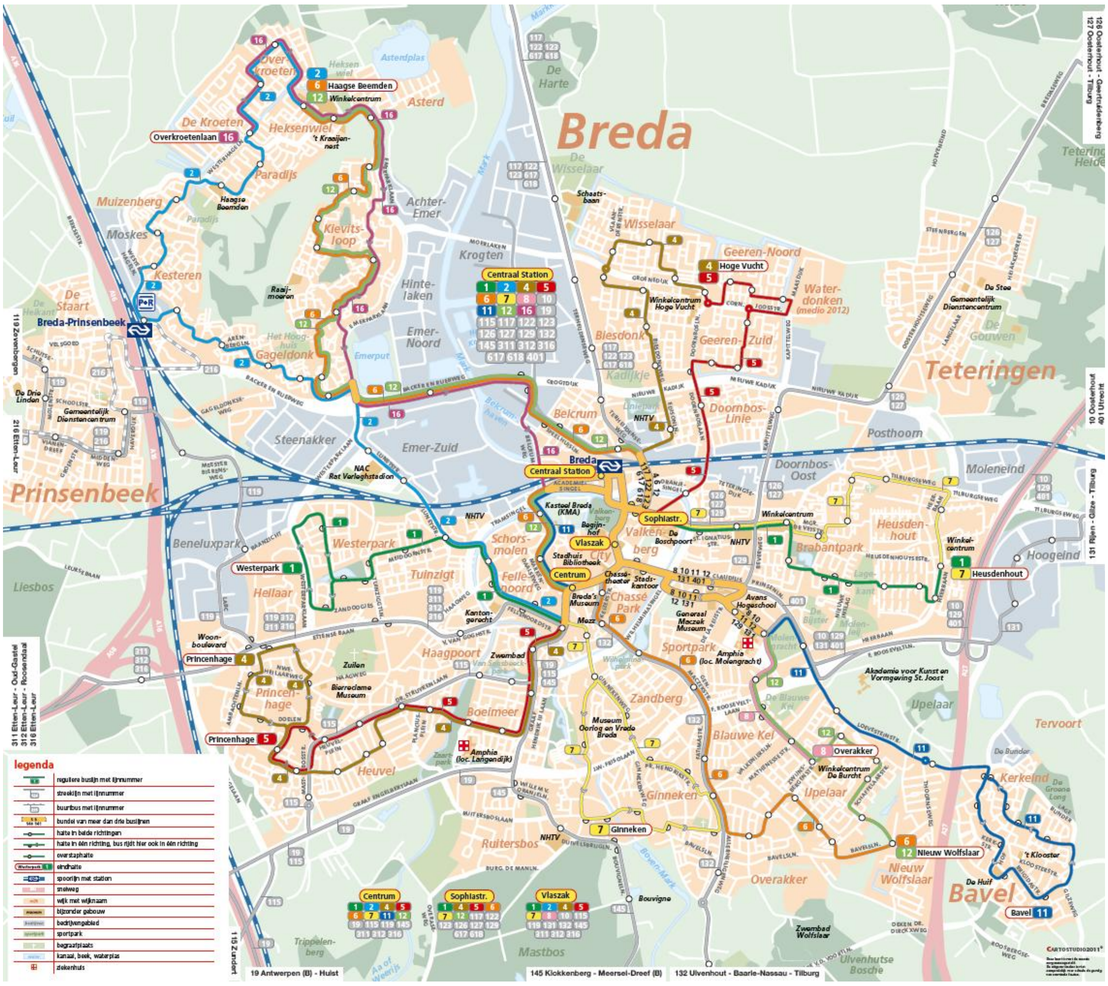


Figure 3: Bus lines in the municipality of Breda. Reprinted from Veolia Transport (2012).

The municipality of Tilburg has a population of 208.000 of which 190.000 in the city itself (Tilburg-stadsmonitor.buurtmonitor.nl). Figure 4 shows the bus network of Tilburg. In total there are 10 urban and 15 regional bus lines. The bus network in the municipality consists of 191 bus stops; 172 in Tilburg and 19 Udenhout and Berkel-Enschot. Special facilities that are relevant for bus transport are the university with 13.000 students, the universities of Applied Science Avans and Fontys with in total 15.000 students, ROC Tilburg with 11.000 students and the two hospitals. Tilburg has three train stations: Tilburg Central Station, Tilburg University and Tilburg Reeshof.

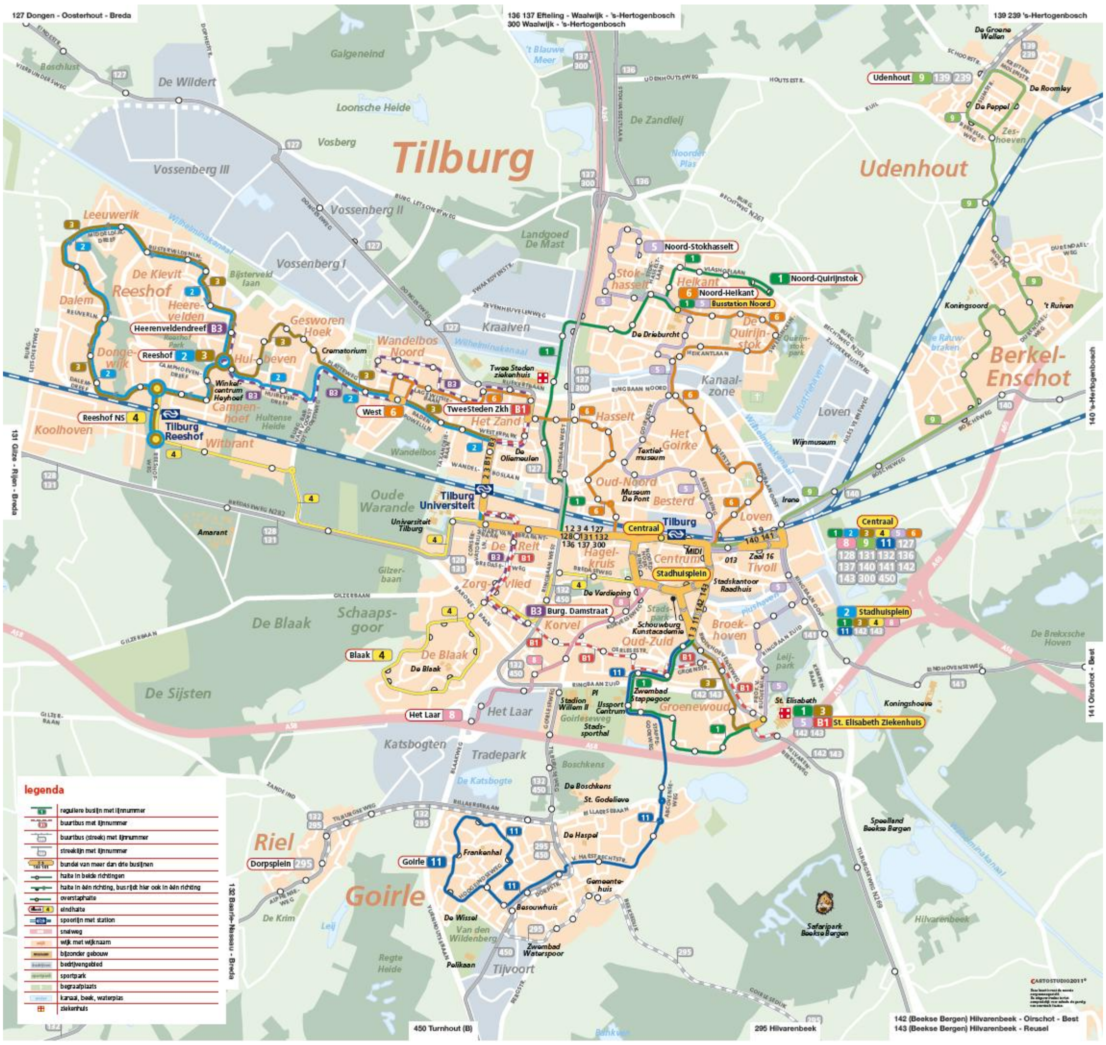


Figure 4: Bus lines in the municipality of Tilburg. Reprinted from Veolia Transport (2012).

2 Research design

This chapter elaborates on the design of the research. It first gives the research objective. The second section shows the research questions and the third section explains the methodology that is used to answer the research questions.

2.1 Research objective

With more insight in the factors that influence the demand for bus transport, transport planners are able to optimize the exploitation of their bus system and adapt the system on changes in urban development. Therefore the influences of the factors that determine the number of boarders and alighters and their impact on the number of boarders and alighters at a bus stop need to be explored. The aim of this study is therefore to explore the factors that determine the demand for bus transport and implement these characteristics in a model that is able to predict the number of boarders and alighters at a bus stop and use those to predict the number of passengers in the bus network.

2.2 Research questions

The main research question of this study is:

Which factors determine the demand for bus transport and how can these factors be implemented in an explanatory model that is able to predict the number of boarders and alighters at a bus stop and used to predict the spatial distribution of those passengers?

In order to answer the main research question and meet the research objective, the following sub questions are composed:

1. How can bus demand be modelled according to the theory and what are the requirements of the model for this study?
2. Which factors determine the demand for bus transport according to the literature?
3. Which indicators can represent the factors that were found in the second research question and which data is available about these indicators?
4. How does the number of bus passengers vary over time?
5. How can the number of boarders and alighters at bus stops be predicted using a model?
6. How can the distribution of bus passengers over the network be modelled?
7. How valid are the developed models?

2.3 Operationalisation

1. *How can bus demand be modelled according to the theory and what are the requirements of the model for this study?*

   A literature review has been performed about some examples of different types of models that are used to model the demand for public transport. With these model types in mind, the setup of the models that are developed in this study is given.

2. *Which factors determine the demand for bus transport according to the literature?*

   A literature study has been conducted in order to gain insight in the factors that determine the demand for transport, with a focus on bus transport. This has resulted in an overview with the most important factors that determine demand for bus transport.

3. *Which indicators can represent the factors that were found in the first research question and which data is available about these indicators?*

   In order to include the factors that influence the demand for bus transport that were found in research question 2, indicators are needed that represent those factors. The availability of data about those indicators and suitability to consider in this study were checked.

4. *How does the number of bus passengers vary over time?*

   The variation of the number of bus passengers over time are explored, as well as the influence of certain weather characteristics and the effect of large events in the study area.

5. *How can the number of boarders and alighters at bus stops be predicted using a model?*

   The indicators that influence the demand for bus transport and the chipcard data are combined in a regression model which is able to predict the number of boarders and alighters at bus stops.

6. *How can the distribution of bus passengers over the network be modelled?*

   For this research question a Gravity model is produced by calibrating distribution functions. The distribution functions should be able to produce an origin-destination matrix with a given number of boarders and alighters at bus stops.

7. *How valid are the developed models?*

   The models which are developed for Breda are validated using data of Tilburg. By comparing the results of the models and the observed data, the validity of the models is assessed. Also the sensitivity of the models for the sample of bus stops is checked.

3 Model development

In order to support decision making, a bus demand model is developed. This model should be able to forecast the number of boarders and alighters at the bus stops and determine the destinations of the bus passengers. Socio-economic data is used to explain the number of boarders and alighters which is retrieved from the chipcard data. There are two types of aggregated models that are potentially suitable for these purposes. The first is the classic four step model, which is actually a collection of four models: trip generation model, distribution model, modal split model and the assignment model. The second model type is the direct demand model. The direct demand model combines the trip generation and distribution and modal split models in one model: the demand of each origin-destination pair is determined per mode. The two types of models are discussed in the following three sections. In the final section a choice is made for the most suitable model type with more elaboration on the model development.

3.1 Four step model

The four step model is a general, often used approach to determine the amount of traffic on links in the transport network. Although it origins from the 1960's, it is still a common approach. Figure 5 shows the four stages of the model. The first step, the generation stage, determines the number of trips that start and end in each zone. In this study it determines the number of boarders and alighters at each bus stop in a particular time period. The second stage determines how the boarders and alighters are distributed over the bus network, so how many people travel between each pair of bus stops. This results in an origin-destination matrix. (Ortúzar and Willumsen, 2001)

The modal split stage determines which mode people use, but since only bus transport is considered in this study, this stage is not explored. The final stage is the assignment of the origin-destination matrix to the network. In this study, this is the final stage since the capacity of the network is not considered. In some situations the four stage model is an iterative approach, since the assignment of the traffic to the network might influence travel times and thus results in a different distribution in the next iteration. In this study however, it is assumed the assignment of the passengers to the network does not influence the distribution.

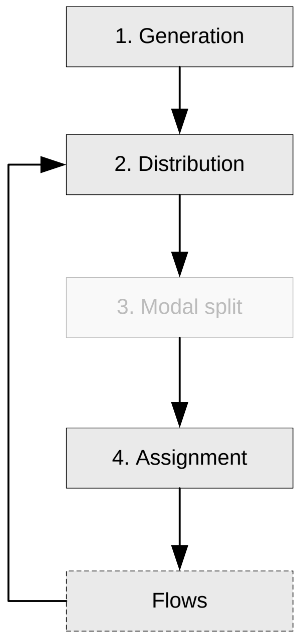


Figure 5: The classic four step model

3.2 Direct demand model

The direct demand model is somewhat similar to the four step model, but it combines the trip generation, distribution and modal split stages in one model. Socio-economic characteristics of the catchment areas of both the boarding and alighting stop are combined in the model, as well as trip specific characteristics. The result is an origin-destination matrix with the number of passengers on the bus stop pairs with a direct bus connection. Once the origin-destination matrix is created, the passengers can be assigned to the bus network. (Ortúzar & Willumsen, 2001)

3.3 Model choice

From the four step model and the direct demand model the four step model is assumed to be most appropriate for this study. The separate models for the trip generation and distribution stage allow more insight in the functioning of the stages. At the start of the study it was not yet known how well the demand for bus transport could be estimated by demand models, so the separate models offers some flexibility. The four step model offers more manual control, e.g. if the trip generation model estimates a number of boarders at a bus stop that is obviously much too low, it is possible to manually change the number of boarders before proceeding to the distribution stage. This kind of flexibility is not offered in the direct demand model.

Next to the lack of flexibility of the direct demand model, the available data is more suitable for the development of the four step model. Data about the number of passengers on origin-destination pairs is only available in aggregated form on year-level, while the data about boarders and alighters at each bus stop is available on hour-level. In the four step model both data levels can be used, while only the year-data is usable for the direct demand model.

The aggregated basis of both models has some disadvantages. It is for example hard to implement trip specific characteristics, such as travel time and accessibility. Both models do not model behaviour and are therefore unable to model the effect of policy and changing travel behaviour (Domencich & McFadden, 1975).

Because of the advantages of the four step model over the direct demand models, a four step model is developed. The following subsections elaborate the development of the four step model.

3.3.1 Generation

The most common approach of determining the number of trips generated by and attracted to each zone is by using socio-economic data. The generation models can be produced with a regression analysis. Some aspects need to be addressed in order to produce the generation models: The factors that determine the demand for bus transport, the variation of the number of bus passengers over time, the number of boarders and alighters at the bus stops and the catchment area of bus stops.

1. *The factors that determine the demand for bus transport*

In order to identify the factors that determine the demand for bus transport a literature study is performed. The literature study provides a list of variables that are important for the demand for bus transport. The second step is to explore the availability of data about these factors and explore the availability of additional data that might be relevant. The final stage is to process and prepare the data for the trip generation models.

2. *The variation of the number of boarders and alighters over time*

The variation of the number of bus passengers over time is explored by creating figures of the number of boarders per time period. The time periods that are assessed are months of the year, weeks of the year, days of the week and hours of the day. Next to these time periods the influence of holidays and events are explored as well.

3. *The number of boarders and alighters at the bus stops*

The observed number of boarders and alighters at the bus stops per time period is necessary to create the trip generation models. This data is available through the chipcard data.

4. *The catchment area of bus stops*

The literature study also addresses the catchment area of bus stops that is found in other studies. Apart from this, stepwise regression models are created for several catchment areas to see which catchment area gives the best regression model. The catchment area with the best fitted regression model is assumed to be most suitable.

Once these four subjects are addressed, the generation models can be produced. The following methodology will be followed for each trip generation model:

1. Create a correlation matrix of all selected variables with the number of boarders in Breda and a correlation matrix with the number of boarders in Tilburg.

2. Select the variables that are significant in both Breda and Tilburg and identify other variables that are either significant in Breda or Tilburg or variables that are considered to be important.

3. Create a correlation matrix with these selected variables for both Breda and Tilburg

4. Remove the variables that are strongly correlated (≥ 0,5) with multiple other variables

5. Create regression models for both Breda and Tilburg, remove variables that are insignificant in both Breda and Tilburg

6. Once all models are created, try to uniformize the models by adding and/or removing variables resulting in models with similar combinations of variables. E.g. there are five models with a variable 'floor area of offices' and one with a variable 'number of offices'. Both variables are strongly correlated, so if the influence on the model is limited, the variable 'number of offices' is replaced with 'floor area of offices'.

7. The developed models for Breda are the final models.

3.3.2 Distribution

There are several methods to determine the distribution of the passengers between the bus stops. The focus is either on the trip itself, using generalized costs, or on the activity that is the reason to travel. In this study, the focus is on the trip itself, since no information is available about the activities that are performed by the bus passengers. The most used method for trip distribution is the Gravity model, which uses travel costs to determine the impedance to travel between each pair of bus stops. The model assumes the willingness to travel to a particular bus stop only depends on those travel costs. Examples of travel costs are generalized costs (a combination of several costs), distance and travel time. The gravity model uses a distribution function (also known as deterrence function) that needs to be calibrated. The result of the distribution model is an origin-destination matrix.

An alternative for the gravity model is distribution with an intervening-opportunity model. The basic assumption of intervening-opportunity models is that people will travel to the destination which satisfies the aim of the trip which is closest by or best accessible. It works with the probability a destination is able to satisfy the objective. If there is a bus line which serves bus stops A to E (in order of serving) and a passenger travels from A to E, then apparently stops B, C and D cannot fulfil the objective of the journey although they are closer by. Calibrating the model determines the likelihood of each bus stop to satisfy the trip. This model is however quite complicated to compute and more difficult to understand than the gravity model. (Ortúzar & Willumsen, 2001)

In literature various distribution functions can be found, but since it is not feasible to examine all those functions, a few different functions need to be assessed. The function that gives the best results of the distribution is selected. The functions that are assessed are:

- Exponential function: $f\left(c_{ij}\right) = \mathrm{e}^{-\beta c_{ij}}$

- Tanner function: $f(c_{ij}) = c_{ij}^{\alpha} e^{-\beta c_{ij}}$
- Top log-normal function: $f(c_{ij}) = e^{-\alpha(\log(\frac{c_{ij}}{\beta}))^2}$

3.3.3 Assignment

The final stage of the four step model is the assignment of the distribution matrix to the bus network. In situations where the assignment of the traffic to the network influences the distribution (e.g. because travel times increase making other routes more attractive) an iterative procedure is recommended. In this study however, it is assumed people will take the shortest route (measured by number of bus stops). In reality there are several factors that determine which bus line people take, like frequency and directness of the route, but these are ignored in this thesis.

3.4 Conclusion

Two types of aggregated models are identified which are potentially suitable for the purposes of this thesis: the classic four step model and the direct demand model. Since the four step model gives more insight and flexibility in the performance of the separate models, the four step model is chosen. The four step model consists of the stages trip generation, distribution, modal split and assignment. Since this study only considers bus transport, the modal split stage is skipped. The subjects that need to be addressed in order to develop the generation model are the factors that determine the demand for bus transport, the availability of data to represent those factors, data about the number of boarders and alighters at bus stops and determining the catchment area of bus stops. For the distribution stage the distribution function that is most suitable to reproduce the observed trip length distribution needs to be identified. Figure 6 summarizes the procedure of the development of the four step model.

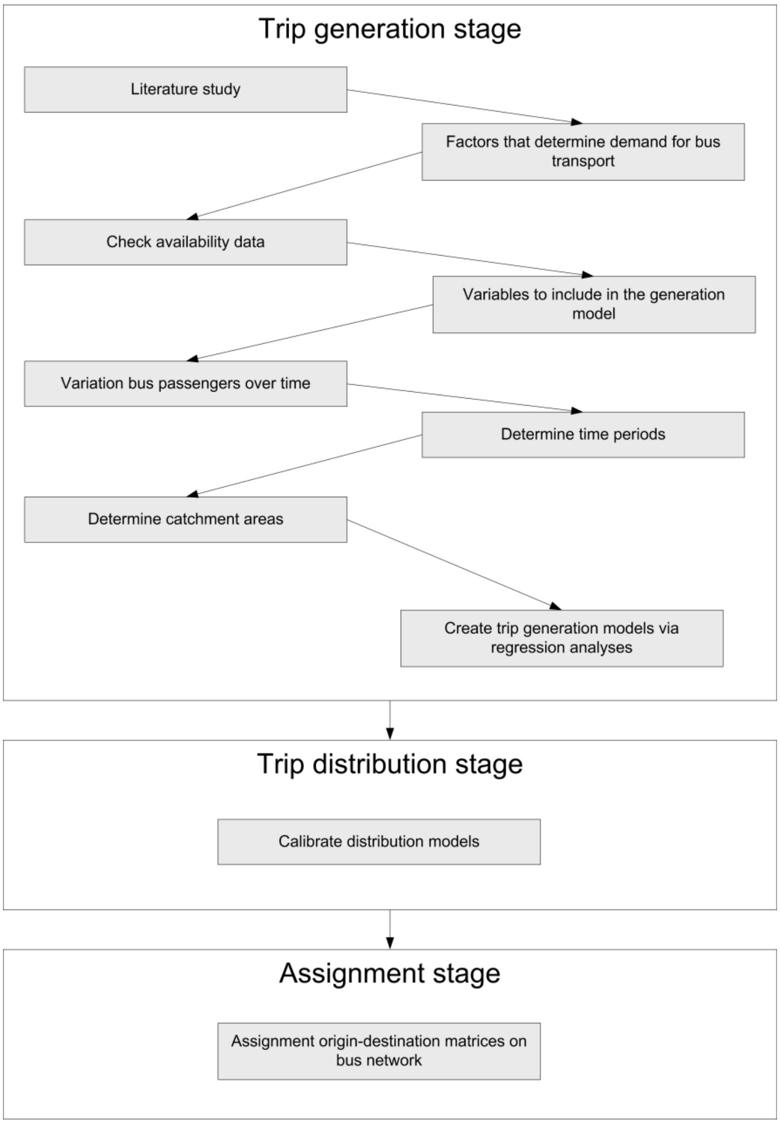


Figure 6: Procedure of the development of the four step model

4 Literature review

The studies that examine the demand for bus transport on aggregated level can be categorized in two categories: those that focus on factors that explain the number of bus passengers and those that focus on factors that cause a change in demand for bus transport. Although there is an overlap between the factors in those categories, this study focuses on the first category and therefore the focus will be on studies from the first category. Those factors can be classified in the categories population characteristics, spatial characteristics, bus service characteristics and trip specific characteristics.

4.1 Spatial characteristics

Spatial characteristics are important factors that determine travel time and travel distances. Because of suburbanisation in the Netherlands the distances between residential areas and commercial and industrial areas are relatively large, which has a negative effect on bus transport demand (Steg and Kalfs, 2000). The spatial characteristics that influence the demand for bus transport can be discussed using the five D's of development:

1. Density
2. Diversity
3. Design
4. Destination accessibility
5. Distance to transit

Density
Density is usually the number of people or employees per square kilometre. In high density areas more people live within the catchment area of public transport stops or stations, which allows more efficient public transport. According to several studies more trips are undertaken by public transport in high density areas (Campoli and MacLean, 2002; Lee and Cervero, 2007; Limtanakool, Dijst and Schwanen, 2006).

Diversity
Diversity of land use, like mixed land use, especially influences demand for transport for non-work related trips (Lee and Cervero, 2007). Mixed land use allows people to walk or cycle towards their destination instead of having to travel a longer distance (Cervero and Kockelman, 1997), although this might have a negative effect on the demand for bus transport. Lee and Cervero (2007) found that a good mix of residents and jobs reduces vehicle trips.

Urban design
The urban design of the environment of a bus stop needs to facilitate the access routes for bus passengers. Most bus passengers travel to the bus stop by walking or cycling, which can be supported by offering direct and save routes between the origins and destinations of the passengers and the bus stops, in other words: offering a pedestrian and cyclist friendly environment (Lee and Cervero, 2007).

Destination accessibility
Public transport rarely brings passengers to their desired location; usually an additional trip is needed to reach the final location. Therefore the accessibility of the destination is an

important factor. When the destination is well accessible from the bus stop (e.g. at a walkable distance), the bus will become more attractive to use. Destination accessibility can be described to be the number of activities like jobs, parks, shops that can be reached within some amount of time (Ewing, Meakins, Bjarnson and Hilton, 2011).

Distance to transit

The circle theory assumes there is a maximum distance people are willing to travel to access a transit stop or station. The theory is that people living closer to a bus stop are more likely to travel via that bus stop than people living further away. In the Netherlands, this radius is usually presumed to be about 500 meters, because this is found to be the maximum distance people are willing to walk to a stop (Van der Blij, Veger and Slebos, 2010). There is however not an unambiguous distance that can be used in this study, so this needs to be investigated.

4.2 Population characteristics

Ethnicity

Ethnic minorities make other choices in travelling than Dutch natives. The bicycle ownership among Moroccans, Surinamers, Antilleans and Turks is far lower than that of Dutch natives, namely a quarter of these minority households do not own a bicycle, while this is 3% for natives. The minorities travel less by bicycle and car and more with public transport. In similar situations, when natives choose to use the bicycle, minorities choose for public transport. Possible explanations are cultural differences (differences between men and women, low social status of bicycle) and a relatively large number of people who never learnt to cycle (Harms, 2006). It might also be influenced by the location where ethnic minorities live, e.g. in higher density neighbourhoods in the vicinity of well served bus stops. The number of ethnic minorities might therefore have a significant influence on the demand for bus transport. Also Taylor, Miller, Isekia and Fink (2008) showed the number of recent immigrants have a large influence on the demand for bus transport.

Age

For certain age categories age has a large influence on demand for bus transport. Other than one might expect, elderly (65+) do not use the bus more often than other people, yet the category 80+ make 3% of their trips with the bus compared to 2% for the whole population (Bakker, Zwaneveld, Berveling, Korteweg & Visser, 2009). Balcombe et al. (2004) however found that the bus accounts for 12% and 13% of the trips of respectively the age groups 70+ years and 17-20 years. Since the group of 70+ is small, the total influence of this group on the demand for bus transport is limited. Pupils and students in the Netherlands make about 29% of their trips with public transport and are therewith an important age group to consider.

Income & car ownership

Income mainly has an indirect effect on bus transport demand. Income has a very large impact on car ownership and availability, which is a major factor in bus transport demand (Balcombe et al., 2004). People that do not own or have access to a car depend on other ways of transport, like public transport, cycling or walking. Taylor et al. (2008) showed the number of households without a car influences the demand for bus transport.

Balcombe et al. (2004) found that people in the UK that own a car are 66% less likely to travel by bus. Figure 7 shows the modal distribution of the number of passenger kilometres for people that do and do not have a car available. The figure shows that the group of people in the Netherlands that do not have a car available travel significantly more with public transport. For this group, public transport can be considered to be important, but their share in the total number of passenger kilometres is fairly limited (Bakker et al., 2009).

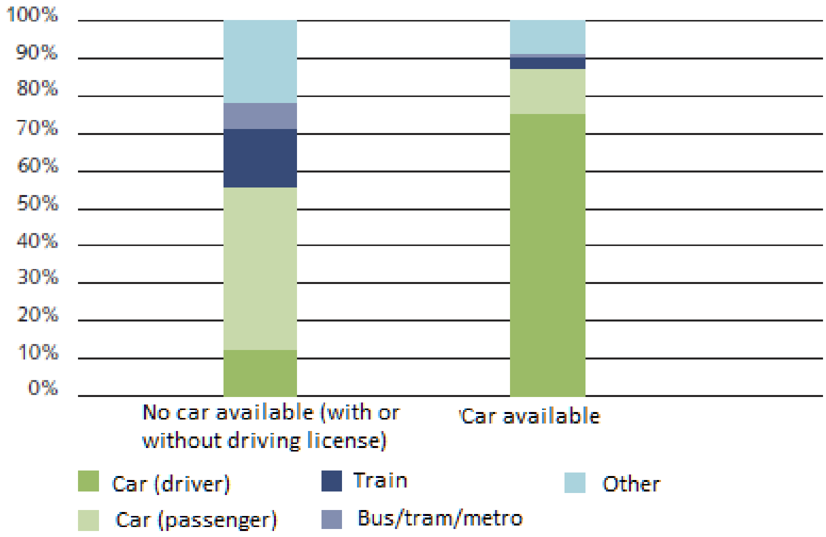


Figure 7: Modal distribution of the number of passenger kilometres for people that do and do not have a car available in the Netherlands. Adapted from Bakker et al., 2009.

Bicycle ownership

Approximately 84% of the Dutch people own at least one bicycle (Fietsersbond, 2011). Bicycles are however also an important competitor for bus transport, since both operate on relatively short distances. The average bus, tram and metro journey in the Netherlands is about nine kilometres from stop to stop, while figure 08 shows that the bus, tram and metro (BTM) have a very small share of trips shorter than 7.5 km.

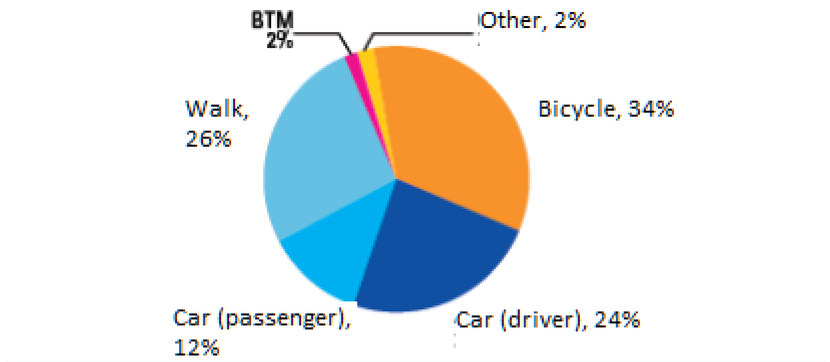


Figure 8: Modal choice for trips shorter than 7.5km in the Netherlands. BTM = Bus, tram and metro. Adapted from Fietsersbond, 2011.

4.2.1 Trip specific characteristics

Trip purpose

The trip purpose is the reason why people travel; the activity they are going to perform. The main motives for bus transport in the Netherlands are school, work, shopping and recreation. 43% of the passenger kilometres that is made for educational purposes is made by public transport, making public transport an important mode for this purpose. The share of public transport is significantly lower for commuter trips, viz. 14%. Of the total number of vehicle kilometres of bus, tram and metro, 33% is made for educational purposes and 30% is commuter traffic. Other purposes are visiting people, shopping and other social activities (Bakker et al., 2009). Although the trip purpose is hard to operationalize for this study, there are some indicators that are related to those purposes, like population (home based trips), college enrolments (educational trips) and number of jobs (commuter trips). Taylor et al. (2008) found the latter two to have a significance influence on the demand for bus transport.

Travel time
The travel time is the time that is needed to travel from the origin to the destination. The travel time of public transport is often longer than that of private vehicles, since the public transport is usually not from door to door, other than for example cars or bicycles. The travel time of bus transport not only consists of the in-vehicle travel time, but also the access- and egress time and the waiting time. According to Balcombe et al. (2004), the amount of time that people spent travelling is more or less constant at a value of about one hour. This means that the longer the access- and egress time and waiting time are, the less time people are willing to spend in a vehicle. This means that the actual area that can be reached with the bus is smaller than with the car.

Factors that are important in travel time are the directness of the route, the necessity to interchange or not and the distance between the origin and the boarding stop and the destination and the alighting stop.

4.3 Bus service characteristics

Fare
Fare has to be paid to make usage of a bus service, with exception of cases where bus transport is free for particular groups. Taylor et al. (2008) found that the fare levels of bus services have a large influence in the number of passengers. The fare levels do however not differentiate within the study area and is therefore not included in this analysis.

Quality of service
The quality of service is used to measure the performance of a public transport service. A service with a good performance has a better quality of service than services with a low performance. The quality of service can be determined using different factors that all together determine the quality of service. It can be measured using objective and subjective indicators and some indicators have both an objective and a subjective value, for example the punctuality (Eboli and Mazzulla, 2011). The objective indicators are measurable and unambiguous (assuming objectivity exists), while the subjective indicators reflect the judgments of travellers about those indicators. If only subjective indicators are used, the indicators together result in a perceived quality of service.

The first indicator that can be used to measure the quality of service is the travel time, which is discussed above. It is a subjective indicator, since people can perceive the travel differently, even if the travel time is the same. It depends for example on how people perceive interruptions of the bus journey by stopping at bus stops. A factor that is closely related to the travel time is the frequency or headway of the bus service, which is an objective indicator. The frequency is usually the number of busses per hour, while the headway is the number of minutes between two consecutive buses. A higher frequency results in a higher bus demand, since waiting time is reduced resulting in a lower total travel time.

Rojo, Gonzalo-Orden, dell'Olio and Ibeas (2012) found that bus characteristics (i.e. the quality of the bus) only influence demand for bus transport for long distances (> 15km), although it does influence user satisfactory for all trips. According to Paulley et al. (2006), the conversion of high-floor buses to low-floor buses results in a demand growth of about 5%. This increased demand would come from people in a wheelchair, people with small children and people with a lot of luggage. Another example of a bus characteristic that is important to some people is the presence of a functioning air-conditioning system (Eboli and Mazzulla, 2011).

Another factor that impacts the quality of service is whether people have to interchange or not. People do not like to interchange on their journey, although the degree in which it influences the demand for bus transport depends on the waiting time during transfer (Balcombe et al., 2004). Another major factor is the total length of the journey; people are less dissatisfied about transferring at longer journeys than at short ones (Wardman, 2001). The value of the transfer is found to be between 4.5 and 30 in-vehicle-minutes (Guo, 2003). The value of a transfer in the Netherlands is approximately 11 minutes (Oostra, 2004 via Goudappel Coffeng, n.d.).

The reliability of a bus service is the ability to hold on to the schedule and to depart and arrive at the desired locations at the planned moments (Eboli and Mazzulla, 2011). An unreliable bus service results in longer waiting and travel times for passengers and hence will dissatisfy them and decrease demand for bus transport. Since busses usually don't stop at stops where no passengers desire to leave or board the bus, this causes some unreliability in bus services.

The final factor related to the quality of service discussed here is the provision of information. Information can be provided online, at the bus stops and in the bus itself. Usually the information includes the timetable, but the availability of information about delays is increasing. According to Grotenhuis the most desired pre-trip information is the total travel time, interchanges and real time delay information. The most desired roadside information is real time delay information, trip advice and the departure platform (Grotenhuis as cited by Grotenhuis, Wiegmans and Rietveld, 2007). Nearly all bus stops in the Netherlands are fitted with a time table and the number of digital passenger information systems with current departure times is growing.

The factors related to the quality of service discussed above are the travel time, frequency, quality of the bus, necessity to interchange, reliability and provision of information. Most factors are difficult to include in this study because they are trip specific, because of a lack of data or because of insufficient relevance. The travel time and necessity to interchange are trip specific and therefore not included in this study. Veolia Brabant has bought 200 new buses in 2007, so it can be assumed that the quality of the buses is overall of a high standard (Omroep Brabant, 2007). All buses in Brabant are low-floor busses (Reizigersoverleg Brabant, n.d.). The quality of buses is therefore not included in the study. No information is available about the reliability of the different bus lines and the presence of digital passenger information systems at bus stops in 2012, so those factors are not considered.

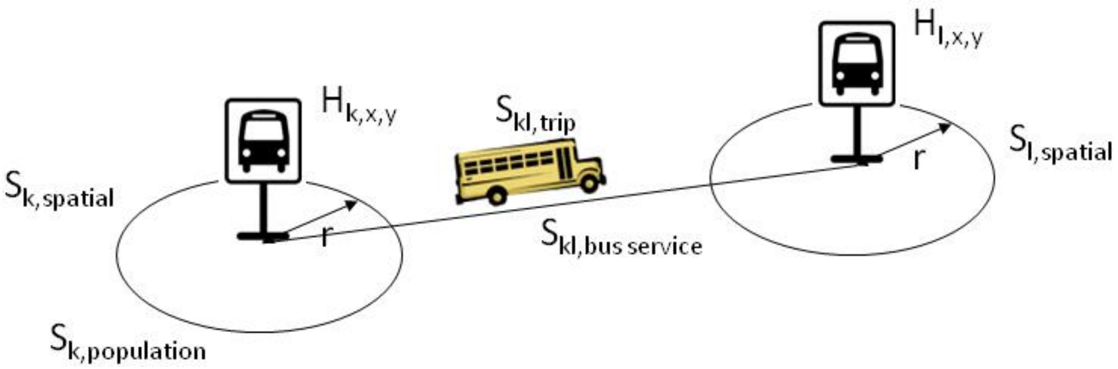
Reputation

Several studies have shown that the bus has a rather bad reputation among people (Pommer, Van Kempen and Eggink, 2008; Tertoolen and Van Uum, 2004). It is remarkably that frequent bus users rate the bus higher than not-users. The bus services in the Netherlands received ratings of 7.1 to 8.3, which show that bus users actually are quite satisfied with the bus services (Kennisplatform Verkeer en Vervoer, 2013). Because of a lack of appropriate data this factors is not considered in this study.

4.4 Conclusion

The literature study came up with several factors that influence the demand for bus transport. These factors can be divided in four categories: spatial characteristics, population characteristics, trip specific characteristics and bus service characteristics. Figure 9 shows the four categories in relation to the bus system. It shows the boarding and alighting stop, each with a catchment area. The catchment area of the boarding stop has spatial and population characteristics while the alighting stop only has spatial characteristics. The bus service and trip characteristics are related to the connection between the two stops.

Figure 9: The four categories of factors that influence the demand for bus transport in relation to the bus system.



Where

$H_{k,x,y}$ = boarding stop k with coordinates x and y
r = radius of the catchment area of the bus stops
$S_{k,spatial}$ = spatial characteristics of bus stop k
$S_{k,population}$ = population characteristics of boarding stop k
$S_{k,trip}$ = Trip specific characteristics from boarding stop k
$S_{k,bus\ service}$ = bus service characteristics from boarding stop k
$H_{l,x,y}$ = alighting stop l with coordinates x and y
$S_{l,spatial}$ = spatial characteristics of alighting stop l

Table 2 shows the characteristics in the four categories spatial characteristics, population characteristics, trip characteristics and bus service characteristics. Some factors can be categorized in multiple categories, but are assigned to one category. There are two factors which are only applied to the alighting stop: college enrolments and number of jobs. These two factors are typical destination variables. People that board the bus at a specific bus stop, will usually return to that bus stop later in the day. Therefore no distinction is made between boarding and alighting stop.

Table 2: Factors that influence the demand for bus transport according to the literature. The boarding stop contains factors in four categories, while the alighting stop only has one category of factors.

Boarding stop				Alighting stop
Spatial	**Population**	**Trip**	**Bus service**	**Spatial**
Density	Age	Trip purpose	Quality of service	College enrolments
Distance to bus stops	Population	Travel time	Fare	Number of jobs
Urban design	Income/car ownership		Reputation	Distance to bus stops
Diversity	Bicycle ownership			Accessibility
Accessibility	Ethnicity			

5 Data assessment

The previous chapter concluded with an overview with factors that influence the demand for bus transport according to the literature. In order to include those alternatives in the regression analysis, indicators about those factors need to be available. This chapter examines several data sources and the available data that represent the factors from literature. The first section assesses the data which is available about the population characteristics. The second section describes the data which is available about bus service characteristics and the third section shows the available spatial data. The final section presents the chipcard data which is available for this study.

5.1 Population characteristics

5.1.1 Centraal Bureau voor de Statistiek

The Dutch Statistics Agency (Centraal Bureau voor de Statistiek, CBS) has data available about various socio-economic characteristics on different spatial levels. The three spatial levels which are relevant for this study are:

- Neighbourhood
- Squares of 500m
- Squares of 100m

Since the study focuses on 2012, data from 2012 should be used. Some useful or necessary data is however not (yet) available for 2012, so for those variables the data of 2011 is used. In case data about variables is available on different spatial levels, the spatially most detailed level is used. In those cases the consistency of the data is checked (see section 6.1).

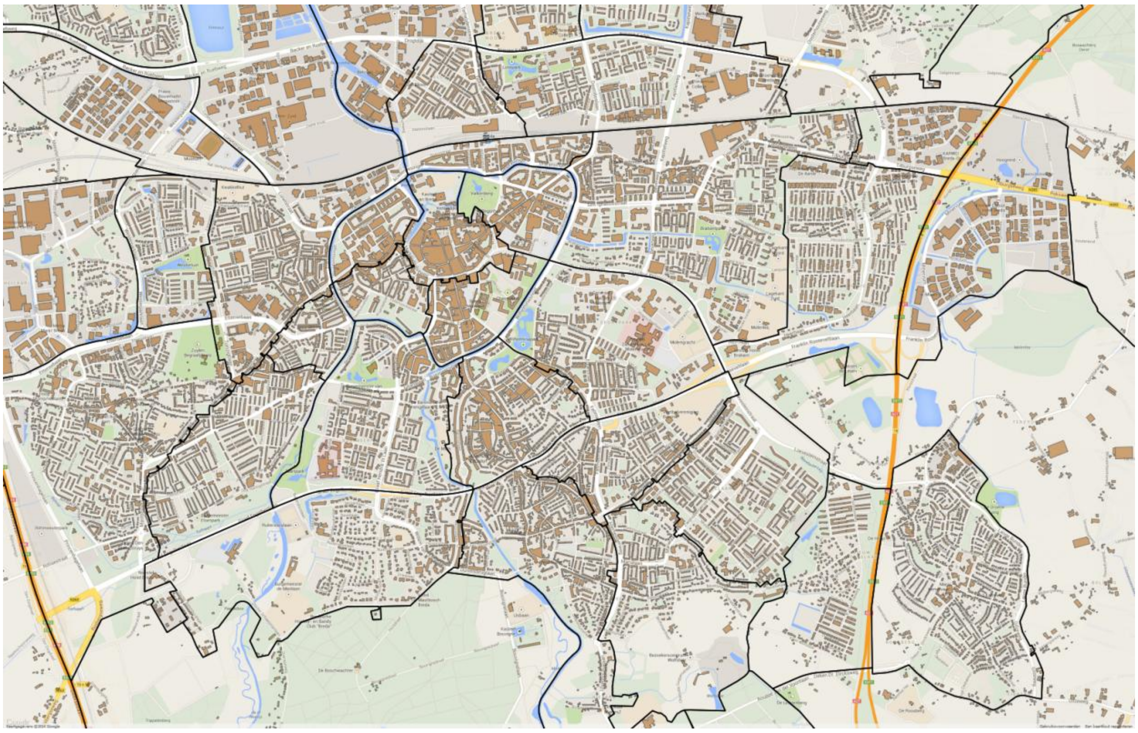


Figure 10:Neighbourhoods in Breda. © Kadaster / Centraal Bureau voor de Statistiek, 2013

Figure 10 shows some of the 56 neighbourhoods in the municipality of Breda. Since the study focuses on 2012, data of 2012 is used. The data for 2011 contains however more variables than the data of 2012, of whom a few are considered relevant for this study. The following data (2012) from CBS on neighbourhood level is considered to be relevant (CBS, 2013):

- Total number of cars
- Average number of cars per household
- Percentage of western foreigners
- Percentage of non-western foreigners

The variables in the data of 2011 that is not available yet for 2012 but is considered to be relevant are the following (CBS, 2012):

- Number of social security eligibilities (bijstandsuitkeringen)
- Average distance to school with VMBO (following the road)
- Average distance to school with havo or vwo (following the road)
- Average distance to a train station (following the road)

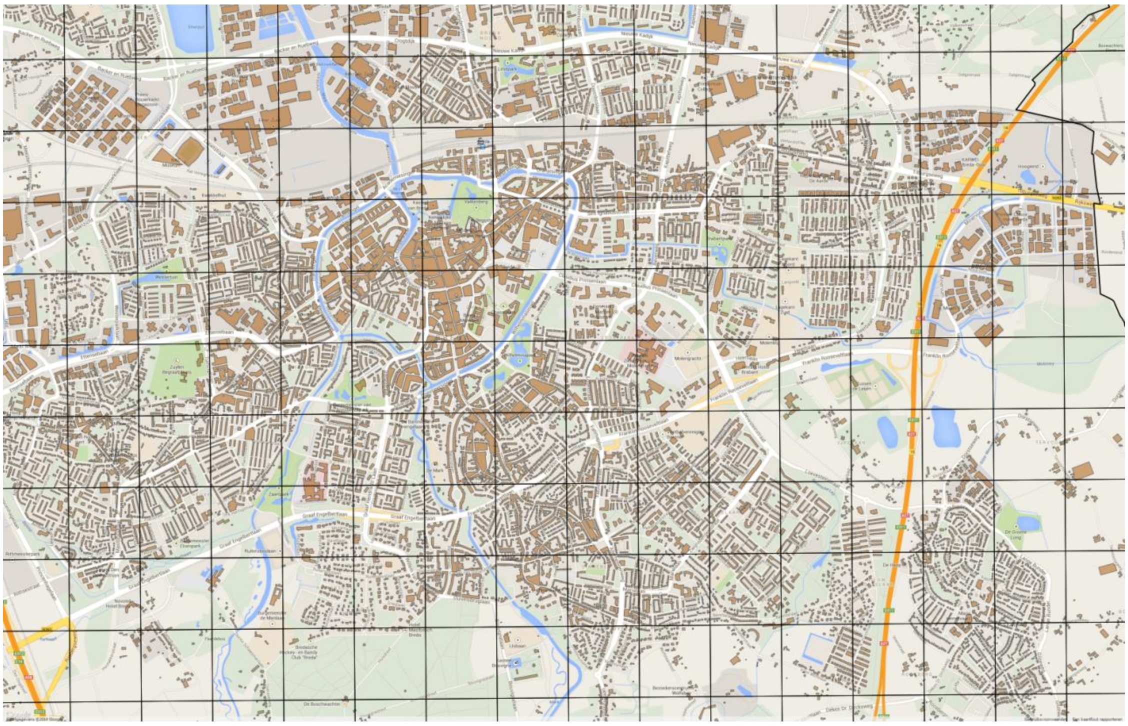


Figure 11: Municipality of Breda with squares of 500m. © Kadaster / Centraal Bureau voor de Statistiek, 2013

Figure 11 shows the municipality of Breda with squares of 500m, containing data from CBS. The data at this spatial level from the CBS that is considered to be relevant for this study are (CBS, 2013):

- OAD (omgevingsadressendichtheid, average address density)
- Percentage of low income households
- Percentage of high income households

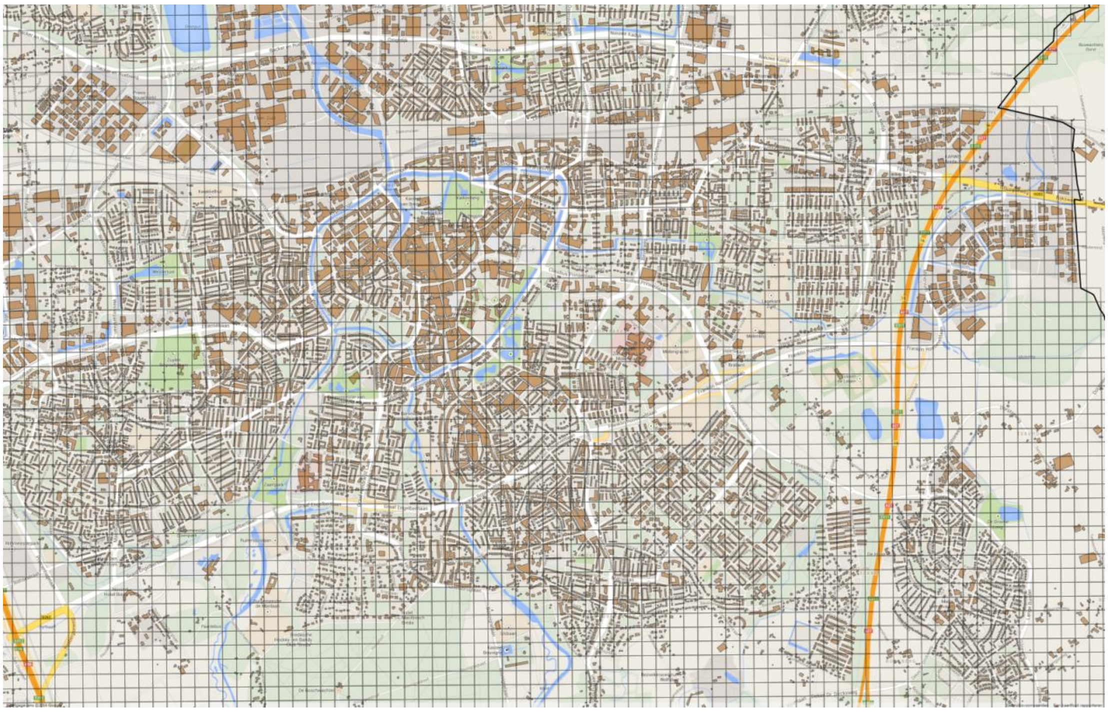


Figure 12: Municipality of Breda with squares of 100m. © Kadaster / Centraal Bureau voor de Statistiek, 2013

Figure 12 shows the municipality of Breda with squares of 100m containing data about several socio-economic characteristics. The following characteristics are selected (CBS, 2013):

- Number of people
- Number of people 0-14 years
- Number of people 15-24 years
- Number of people 25-44 years
- Number of people 45-64 years
- Number of people 65 years and older
- Percentage of western foreigners
- Percentage of non-western foreigners

5.1.2 Municipality of Breda

The municipality of Breda has information available about the number of jobs in every neighbourhood in Breda (breda.buurtmonitor.nl). The data is available for the years 2003 to 2012. Other data that is available from the municipality that might be relevant and complements the data from CBS is the average income per household on neighbourhood level. The most recent data that is available is from 2009, but according to CBS the average disposable income in Noord-Brabant in 2012 was more or less equal to 2009, namely 0.3% lower (statline.cbs.nl). It is assumed this difference is similar in Breda. The variables from the municipality of Breda that will be used are:

- Number of jobs
- Average disposable income (2009)

5.1.3 Dienst Uitvoering Onderwijs

Dienst Uitvoering Onderwijs (DUO), the Education Executive Agency of the Ministry of Education, Culture and Science publishes education related data on their website (DUO, n.d.). Data from DUO that has been selected for this study are the following:

- Number of pupils per 4-position postcode
- Number of pupils per school (secondary education)
- Number of pupils per intermediate vocational educational institute (MBO)
- Number of students per university of applied science (HBO)
- Number of students per university (WO)

5.2 Bus service characteristics

The bus service characteristics are obtained from two primary sources: Veolia Transport and OpenStreetMap. Information about the routes of the bus lines and the travel times are obtained from the *busboekjes* with the timetables and the bus stops per bus line, from the website of Veolia Transport and from the line maps. The locations of the bus stops are obtained from OpenStreetMap, which is an open source mapping project.

5.3 Spatial data

The spatial data contains the following characteristics: roads, bus stops and buildings with their function. The roads and bus stops are obtained from OpenStreetMap and the locations of those bus stops are visually checked using Google Maps and Streetview. The contour shapes of the buildings in the study area are also obtained from OpenStreetMap, although they are only used in figures. The Basisadministratie Adressen en Gebouwen (BAG) is used to identify the function of each address in the study area, as well as the floor area of the addresses (Kadaster, 2013). The BAG is made available through Open License by Kadaster and is downloaded from the Geoservice portal from the Vrije Universiteit Amsterdam (geoplaza.vu.nl).

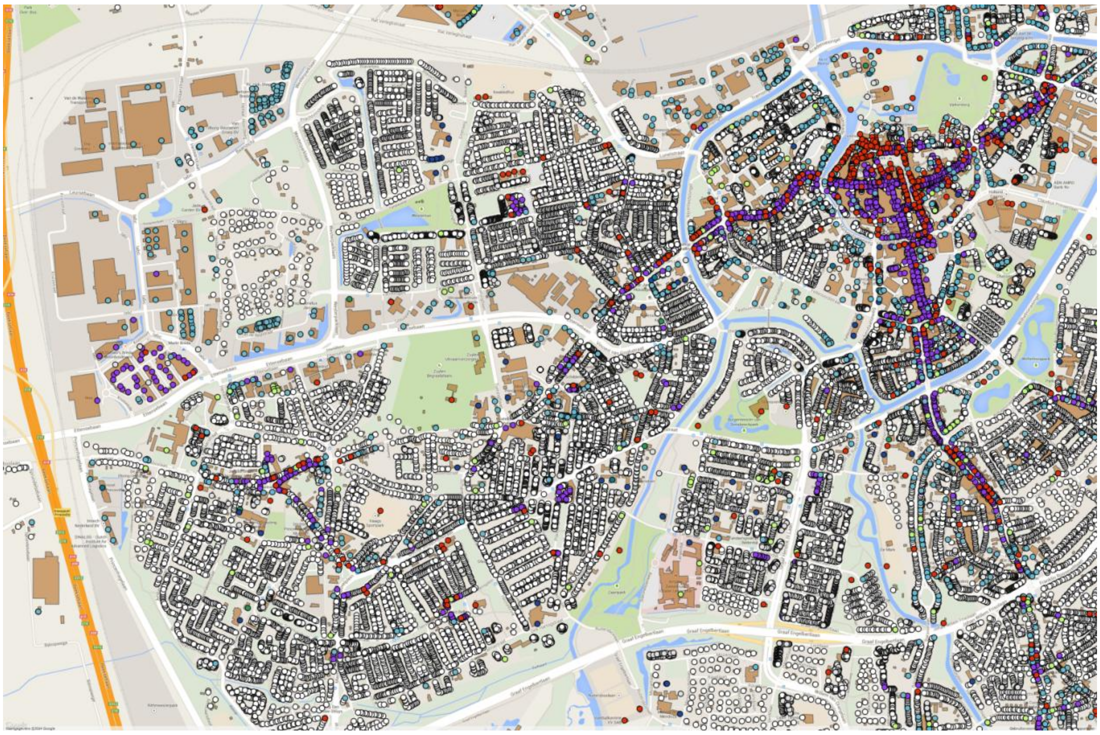


Figure 13: Map of Breda city centre and Breda West including the addresses and their function. The white circles represent addresses with the function 'woon' (residential), purple the function 'winkel' (retail) and red the function 'bijeenkomst' (social gathering). © Kadaster

5.4 Chipcard data

For this study chipcard data of 2012 is used, which is supplied by Veolia Transport. Four chipcard data sets are used. The first dataset contains the number of boarders and alighters per hour per bus stop per bus line in Breda in 2012. Both urban and regional bus lines are included. The set contains 331 excel files: 28 bus lines with a separate file for each month (some lines do not operate during July). The second dataset contains the number of passengers between each pair of bus stops per bus line in Breda that have a direct bus connection, summed over the year. This set contains 70 excel files: 28 bus lines with a separate file for the weekdays (summed), Saturdays and the Sundays. Not each bus line operates during the weekend, so not each bus line has three excel files.

Figure 14 shows a screenshot of one of the excel files in the first dataset. It contains the hourly number of boarders and alighters at bus line 1 on April 1st.

VEOLIA TRANSPORT

Veolia Datawarehouse - Occupancy Dashboard

		Breda, Argusvlinder (13789)		Breda, Aurelia (11247)		Breda, Brabantplein (11252)		Breda, Centraal Station (11010)		Breda, Centrum (11239)		Breda, Dijkplein (11255)		Breda, Draaiboom (13797)	
01-04-2012	07					1	0	2	7			1	0	1	0
	08							1	2						
	09	0	1	1	0	1	1	8	11	2	6	2	0	3	0
	10	2	1	3	0	3	0	10	13	2	15	3	0	1	0
	11	3	0	2	0	1	1	8	17	1	7				
	12	2	3	1	3	0	3	8	4	7	5	1	1	4	3
	13					0	2	5	5	4	1	0	2		
	14			0	3	0	3	18	5	7	3	1	3		
	15	2	1					4	10	9	1	2	0		
	16	0	1					4	2	3	1	1	0		
	17					0	1	8	6	1	1	2	2		
	18	2	0			0	4	8	3	1	0			0	4
	19	0	1			0	5	12	2	1	1	2	1	0	1
	20	0	1	2	0	0	1	12	5	2	0	0	2		
	21							3	0	1	0			0	1
		11	9	9	6	6	21	111	92	41	41	15	11	9	9
02-04-2012	04	1	0	1	0	2	0	0	9	1	1	1	0	3	0
	05	2	2	1	0	5	2	23	36	3	1	3	0	3	0
	06	3	0	3	1	3	6	106	15	6	0	2	1	1	1

Figure 14: Screenshot of one of the chipcard datasets. It shows the number of boarders and alighters per hour per bus stop at bus line 1. Each bus stop has two columns, the left column contains the number of boarders during the hour and the second column contains the number of alighters. © Veolia Transport

The third data set contains the number of boarders and alighters per hour per bus stop per bus line. The set contains 240 excel files: 20 bus lines with a separate file for each month. The fourth file contains the number of passengers between each pair of bus stops per bus line in Tilburg that have a direct bus connection, summed over the year. This set contains 57 excel files: three excel files per bus lines (one for the weekdays, one for Saturdays and one for Sundays).

Figure 15 shows a screenshot of one of the excel files in the third dataset. It contains the number of passengers at some bus stop pairs of line 1 in Breda on all weekdays in 2012.

VEOLIA TRANSPORT

Chipcard Finance
Travel Movements - Passengers - Crosstab

Show the number of passengers per Farestop to Farestop.

Boarding / Alighting	Breda, Argusvlinder (13789)	Breda, Aurelia (11247)	Breda, Baronielaan (11248)	Breda, Bergschot (13790)	Breda, Brabantplein (11252)	Breda, Cartier van Disselstra	Breda, Centraal Station (11010)	Breda, Centrum (11239)	Breda, Dijkplein (11255)
Breda, Argusvlinder (13789)	8	14		1	35		1,502	2,019	41
Breda, Aurelia (11247)	9	136		4	17		2,054	2,047	64
Breda, Brabantplein (11252)	62	25		99	11		14,174	2,166	116
Breda, Centraal Station (11010)	3,407	3,067	1	2,394	10,087		1,475	8,223	5,234
Breda, Centrum (11239)	3,738	2,104		279	1,171		8,028	33	2,329
Breda, Dijkplein (11255)	226	60		23	129		2,307	1,445	8
Breda, Draaiboom (13797)	20	22		3	501	2	3,621	1,169	28
Breda, Fellenoordstraat (13805)	115	37		13	175		1,878	621	114
Breda, Heerbaan (11266)	78	18		3	304		4,551	1,599	104
Breda, Ignatiusstraat (11095)	111	167		85	102		54,465	1,924	540

Figure 15: Screenshot of one of the chipcard datasets. It shows the number of passengers on line 1 on all weekdays in 2012. The columns represent the boarding stops and the rows the alighting stops. The origin-destination matrix contains bus stops that are not served by line 1, in this screenshot for example Baronielaan and Cartier van Disselstraat. These stops should be removed. © Veolia Transport

Although, or because, the datasets contain large amounts of data, some data is missing or wrong. The datasets do not contain information about transfers, so a journey from A to C with a transfer at B is registered as a trip from A to B and a trip from B to C. From the data it can not be deduced that the trips are connected with each other. Data at some bus stops during some time periods is missing and some bus stops are included twice, but with another unique number. The datasets with the boarding and alighting stops (like in figure 15) contain stops that are not served by that bus line, which need to be filtered out of the data.

5.5 Conclusion

This chapter assessed the data which is available about the factors that determine the demand for bus transport according to the literature study. For the population characteristics most data is obtained from the CBS, which has data available about for example the number of people per age category, ethnic groups and number of cars. The data is available at the spatial scales neighbourhood, squares of 500m and squares of 100m. Other population characteristics are obtained from DUO about education related factors and the municipality of Breda about income and the number of jobs.

Data about the bus service characteristics are obtained from Veolia Transport and OpenStreetMap. The spatial data is mainly obtained from OpenStreetMap (e.g. roads, location of bus stops, etc.) and Kadaster (Basisregistratie Adressen en Gebouwen). Finally, the chipcard data is supplied by Veolia Transport which contains the hourly number of boarders and alighters in Breda and Tilburg in 2012 and the number of passengers between each bus stop pair on yearly level. The data does not contain data about transfers.

6 Processing of data

The previous chapter presented the data that is used in this research. It showed that the data comes from different sources and at different spatial levels. The data needs to be fused to one layer with data. The data is processed using Quantum GIS, which is open source Geographic Information System software. Since a more detailed spatial level gives more detailed spatial information, which is beneficial for this study, the data is merged to the squares of 100m. In order to enable the data fusion, one major assumption needs to be made: it needs to be assumed that the characteristics are uniformly spread over the data object, e.g. a uniformly spread of the population within a neighbourhood. This assumption allows data fusion based on area. Figure 16 shows an example of how data about the population in a neighbourhood is converted to squares of 500m.

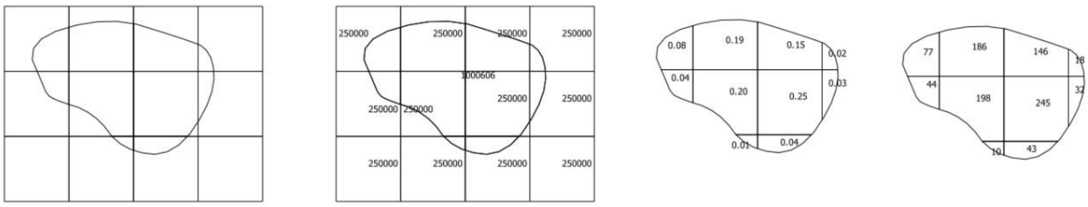


Figure 16: The distribution of a population of 1000 people over squares of 500 meter. The first figure gives the neighbourhood and the squares; the second figure gives the areas of the squares and neighbourhood; the third figure gives the proportion of the area which sums up to 1. The last figure gives the population per (part of the) square. The sum of these populations is equal to the population of the neighbourhood.

The first section checks the consistency of some variables which are available at multiple spatial levels. The second section elaborates on the creation of variables which need to be created from the available data or need additional processing.

6.1 Checking consistency data

Since some data is available on multiple spatial levels, the consistency of the data needs to be checked. The consistencies of the population and of the number or percentages of foreigners are checked in the following subsections.

6.1.1 Population

The population in the 100m squares in each 500m square is summed and compared with the population in the 500m squares. Figure 17 shows the number of occurrences of the population difference between the 500m squares and the summed population of the 100m squares. The population difference between the 500m and 100m squares is zero (rounded to five-folds) in more than half of the squares of 500m (374 of 606). The average population difference is -1 per 500m square, which is -0.34%. This difference is considered limited enough to use the 100m squares, which offer a spatially more detailed impression of the population in each catchment area in Breda.

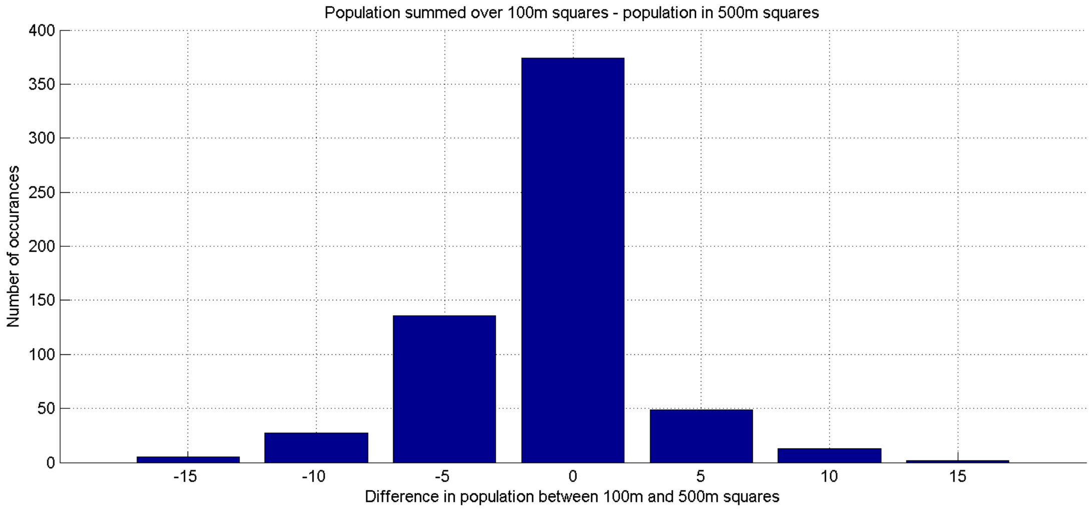


Figure 17: The number of occurrences of the differences in population between the 500m squares and the summed 100m squares. I.e. there are 136 squares of 500m with 5 more people than the summed population in the corresponding 100m squares and 27 squares of 500m with 10 people more than in the corresponding 100m squares.

6.1.2 Foreigners

The CBS data in the layers of 500m and 100m squares only offer percentages of foreigners in different categories (e.g. category 1: >67% non-western foreigners, category 2: 45-67% non-western foreigners, etc.). Only the data on neighbourhood level provides absolute numbers. Since the number of foreigners might be an important factor in the number of bus passengers, it is investigated if it is possible to use the data from 100m level. Table 3 gives an overview of the number of foreigners according to several sources. Since the data at the 500m and 100m squares are only in categories, the average of the categories are used to calculate the absolute number of foreigners (e.g. for a category of 10-20% foreigners 15% is used). For the last row the population of the 100m squares are used and the percentages of foreigners come from the neighbourhood level. It appears the combination of 100m squares and neighbourhood give a very good result, while it is spatially much more detailed than the data on neighbourhood level.

Table 3: Number of foreigners according to different sources

	Source population	Source percentages	Western foreigners	Non-western foreigners	Dutch	Total
CBS Statline	-	-	18.993	19.275	138.133	176.401
Municipality of Breda	-	-	19.080	19.315	138.158	176.553
GIS (CBS)	Neighbourhood	Neighbourhood	19.094	19.317	137.880	176.291
	500m squares	500m squares	19.048	14.022	139.548	172.618
	100m squares	100m squares	18.943	20.541	133.125	172.609
	100m squares	Neighbourhood	19.038	19.218	137.561	175.817

6.2 Creating additional variables

Although most of the data is ready to use, some relevant variables need to be created, either manually or using collected data. The following variables are created:

- Dummy variable for Central Station
- Dummy variable for higher education
- Floor area per function within catchment area
- Percentage of floor area per function (all floor area is 100%)
- Number of addresses per function within catchment area
- Floor area with shopping centre functions ‘winkel’, ‘bijeenkomst’ and ‘logies’
- Percentage of area per function
- Dummy variable for shopping centres

The creation of the dummy variables for Central Station and higher education is pretty straightforward: the dummy for Central Station has a value of 1 for the central bus station and value 0 for all other bus stops. The bus stops next to Universities and Universities of Applied Science have a value of 1 for the dummy variable for higher education, while the other stops receive a value of 0 for this variable. The BAG data is used to calculate the floor area and number of addresses per function within each catchment area. Figure 18 shows a part of Breda with the addresses and the total floor area per function per catchment area. The floor area per function within each catchment area is used to determine the percentage of floor area per function.

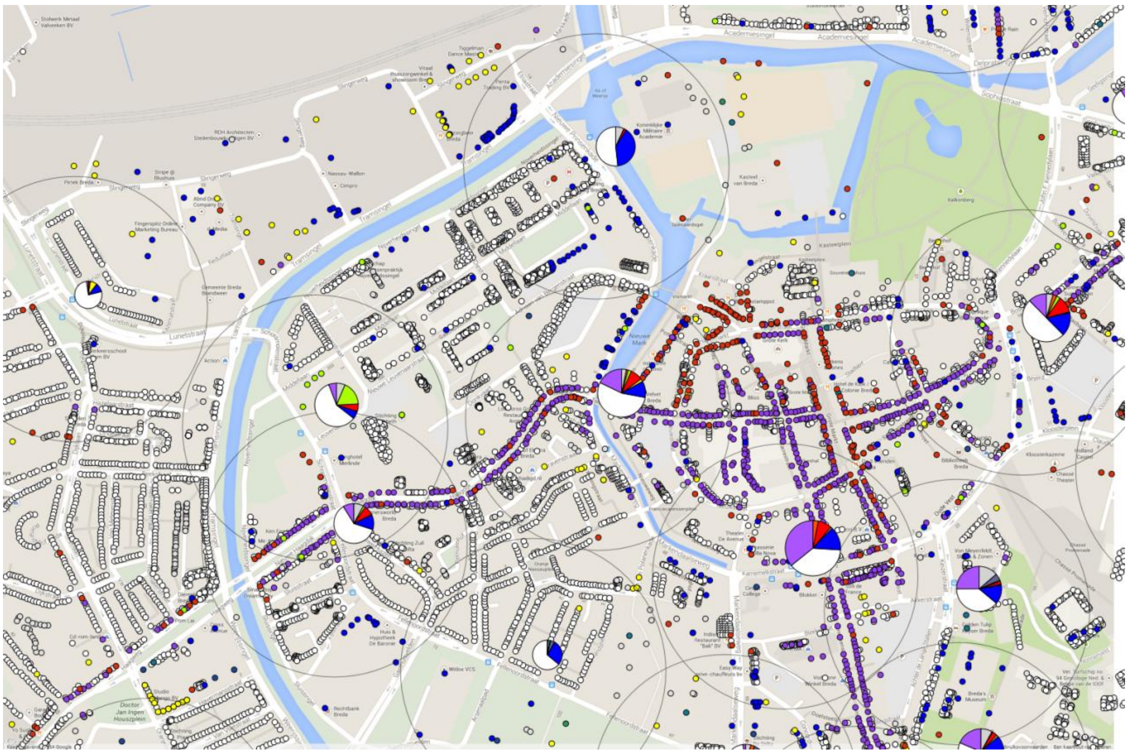


Figure 18: Map of Breda city centre and west of the city centre with the addresses in this area. The size of each pie diagram depends on the total floor area within the catchment area. © Kadaster

The dummy variable for shopping centres is created to represent the bus stops located at the shopping centres. The chipcard data shows these stops have more boarders and alighters than the other surrounding stops with almost only residential buildings. Several considerations can be made by determining proper boundaries for this variable. It needs to be determined whether only the shops within the catchment area are considered or all shops in the shopping centre. Both options have their pros and cons. If only the shops within the catchment area are selected, it is possible that shops in the same shopping centre are not considered. Yet it is conceivable that the shops are seen as one shopping centre which implies that the distances between the shops are neglected by shopping bus passengers. In this case the total floor area of the shopping centre should be considered, although this raises the question when a shopping centre is within the catchment area and when not. If only a very small part of the shopping centre is within the catchment area, should it be included or not? Since this question is not easily answered, only the shops within the catchment area are considered.

Neighbourhood shopping centres mainly attract people from a small area around the shopping centre so they do not attract many bus passengers. District shopping centres however serve an area that is large enough for people to take the bus to the shopping centre. The bus stops at those shopping centres should be included in the dummy variable. Table 4 shows the district shopping centres in Breda and Tilburg as categorized by the respective municipalities (Breda, 2002; Tilburg, 2013) with the floor area retail and catering industry in the catchment area of the nearby bus stops. According to table 04 a floor area of 6000 m^2 seems to be the most appropriate limit for the dummy variable for the shopping centres.

Table 4: District shopping centres in Breda with the floor area of retail and catering industry within the catchment area of the bus stops. In case multiple bus stops are at a similar distance from the shopping centre both bus stops are included.

Municipality	Shopping centre	Floor area retail and catering industry (m^2)
Breda	Hoge Vucht	9.826
	Heksenwiel	6.897
		6.867
	De Burcht	6.130
	Valkeniersplein	6.747
Tilburg	Heyhoef	16.340
	Westermarkt	23.688
	Wagnerplein	16.042
		15.499
	Besterdring	15.696
	Jan Heijnsstraat	10.495
	Korvelseweg	8.154
		16.965

6.3 Conclusion

This chapter elaborated on the processing of the data which is used for the study. The data is gathered at several spatial levels and merged to the squares of 100m. To enable this data fusion, it is assumed that the characteristics are uniformly spread over the data object (e.g. uniformly spread over a neighbourhood). The population in the 100m squares is consistent with the population in the 500m squares. Combining the percentage of foreigners from the neighbourhood level with the population from the squares of 100m results in numbers of foreigners which are comparable with the official numbers from CBS.

Some relevant data cannot be imported directly from the source, but need to be created. Examples of such variables are the dummy variables for the Central Station and the presence of higher education institutes. Other variables that are created are related to the floor area of addresses per function, like the floor area of offices or the number of addresses with a residential function within each catchment area.

7 Comparison of Breda and Tilburg

In the previous chapters the data which is used for the analysis and model development are assessed. Since the models are developed for Breda and validated for Tilburg, it is important to assess the main differences between Breda and Tilburg. This is necessary to understand where potential differences originate from. The differences between Breda and Tilburg are categorized in the categories spatial characteristics, population characteristics and bus service characteristics. The first section assesses the major spatial differences between Breda and Tilburg. The second section elaborates on the differences between the population characteristics. The trip bus service characteristics are assessed in section three.

7.1 Spatial characteristics

At first glance the major spatial difference between Breda and Tilburg is obvious, namely the orientation of the districts in the city. Breda is a concentric city with the city centre almost perfectly in the middle which is physically separated from the rest of the city by the Singel. The city centre of Tilburg is at the eastside of the city, so the distance between the Central Station and the neighbourhoods at the Westside is large.

As stated in the introduction, both Breda and Tilburg have higher education institutes. The Universities of Applied Science in Breda have 20,800 students at seven locations while the Universities of Applied Science in Tilburg have 13,600 students at four locations. Tilburg also has the Tilburg University with 12,952 students at one location. The largest higher education institutes in Breda are located at 1.8 km (Avans, 10,000 students) and 1.3 km (NHTV, 3,560 students) from the Central Station (Euclidean distance). These distances are larger in Tilburg, namely 2.7 km (Tilburg University, 12,952) and 2.5 km (Fontys, 11,000 students) from the Central Station. Tilburg University is however located next to a train station.

The previous chapter showed there are some differences in the district shopping centres as well (see table 04). The average floor area of retail and catering industry within the catchment area of the bus stops at the district shopping centres in Tilburg is at least twice of the average floor area in Breda.

One of the main spatial differences is that the densities in Tilburg are higher. Related with this is that the average floor area of all address functions within the catchment areas is higher in Tilburg, Only exceptions are the functions office and social gathering.

7.2 Population characteristics

Because of the higher densities in Tilburg, the average number of people per age category in the catchment area of the bus stops in Tilburg is higher at all categories. The percentages of people per age category are however similar at Breda and Tilburg. The average number of western and non-western foreigners in the catchment areas is higher in Tilburg. The percentage of western foreigners is slightly higher in Breda (11 versus 9%), while the average percentage of non-western foreigners is higher in Tilburg (14 versus 11%).

The average percentage of low income households within the catchment areas is higher in Tilburg (40 versus 36%), while the average percentage of high income households is higher in Breda (17 versus 10%).

7.3 Bus service characteristics

The difference between the bus networks of Breda and Tilburg can be assessed through the number of occurrences of each trip length in the network with a direct connection. Therefore the origin-destination matrix is used and for each bus stop pair with a direct connection the travel time is determined. Counting the number of occurrences of each trip length results in figures 19 (Breda) and 20 (Tilburg). The figures show the pattern is very similar at Breda and Tilburg. The longest trip within the municipality of Breda without transfer is 47 minutes, while the longest trip within Tilburg has a length of 65 minutes. This difference is mainly caused by the shape of the cities: the distance between the outskirts of the city and the city centre is fairly limited, while the distance between the western outskirts of Tilburg is much longer. Since there are more bus stops in Tilburg, the total number of bus stop pairs with a direct connection is larger.

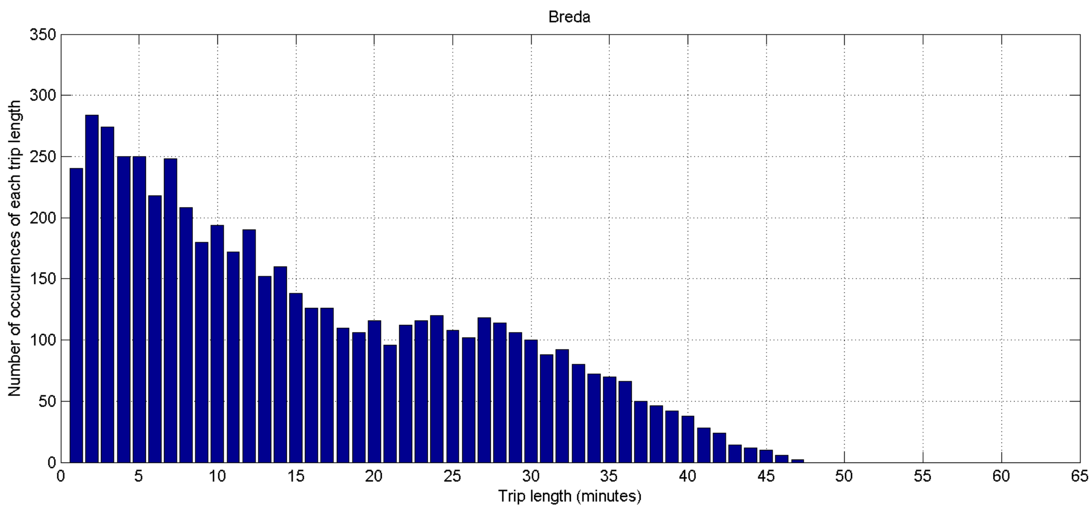


Figure 19: Number of occurrences of each trip length of the bus stop pairs in Breda with a direct connection.

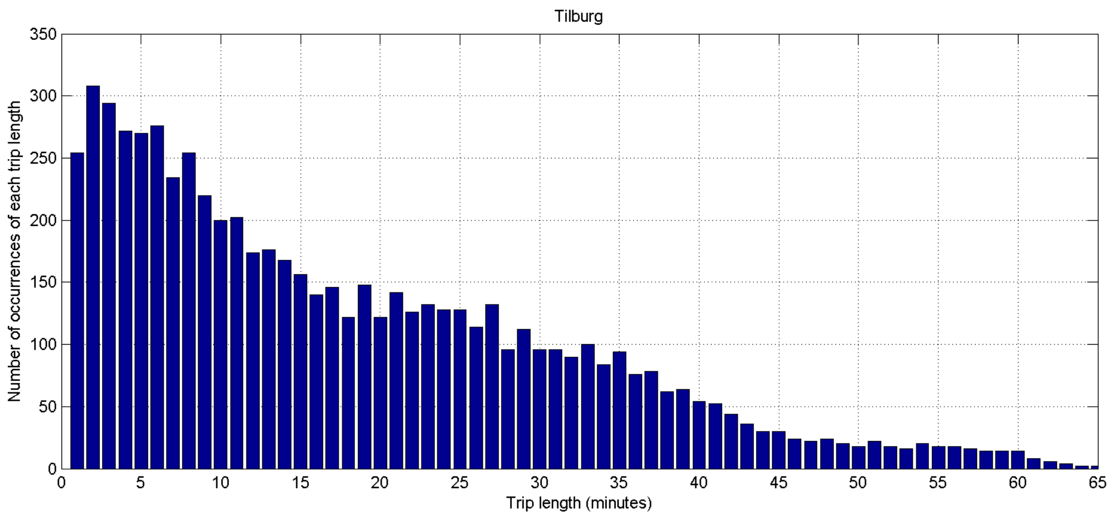


Figure 20:Number of occurrences of each trip length of the bus stop pairs in Tilburg with a direct connection.

As stated in the introduction, the bus network in both Breda and Tilburg consist of 10 urban bus lines, but the network in Breda has 18 regional bus lines and the network in Tilburg 15. The total number of bus stops is slightly higher in Tilburg, namely 191 in Tilburg and 185 in Breda.

Another method to compare the bus network of Breda and Tilburg is to assess how the number of boarders is distributed over the bus stops. It could for example be the case that in Breda only 10% of the bus stops are responsible for 80% of all boarders in the municipality, while in Tilburg 30% of the bus stops are responsible for the 80% of the boarders. Figure 21 shows however that the distribution of the total number of boarders on weekdays over the bus stops in the municipality is very similar at Breda and Tilburg. The busiest stop accounts for 40% of all boarders and the 20% of busiest bus stops account for approximately 80% of the boarders (83% in Breda, 82% in Tilburg).

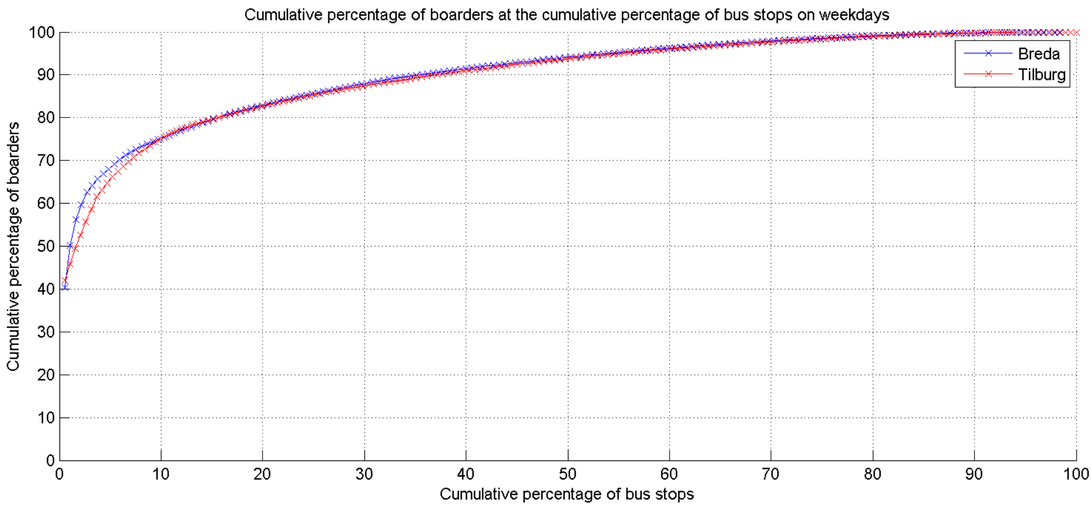


Figure 21: The cumulative percentage of boarders on weekdays at the cumulative percentage of bus stops in Breda and Tilburg. The bus stops are sorted from the largest number of boarders to the least number of observed boarders.

7.4 Conclusion

This chapter assessed the major differences between Breda and Tilburg in three categories: spatial characteristics, population characteristics and bus service characteristics. The main spatial differences are the following:

- The spatial structure (Breda concentric, Tilburg not)
- The distance between the largest education institutes and the Central Station (in Tilburg twice as long as in Breda)
- The urban density (higher densities in Tilburg)

The main population differences are related to the higher densities in Tilburg:

- Average number of people within the catchment areas of the bus stops (larger in Tilburg at all age categories). The average ratio between the age categories is however not significantly different.

- Average number of foreigners (larger in Tilburg, both western and non-western foreigners). The percentage of western foreigners is slightly higher in Breda, while the percentage non-western foreigners is higher in Tilburg.

The main bus service differences between Breda and Tilburg are the following:

- The longest trip length within the municipality without transfer (in Breda the longest trip length is 47 minutes, in Tilburg 66 minutes).
- The number of regional bus lines (the network of Breda has 18 regional bus lines, Tilburg 15).
- The number of bus stops (185 in the municipality of Breda, 191 in the municipality of Tilburg)

8 Data analysis: variation over time

In preparation of the development of the trip generation models, the variation of the number of passengers over time is explored. The first section analyses the variation of the total number of bus passengers in Breda over time. The second section shows the differences in variation over time per bus stop. The effect of events and holidays is examined in section and the effect of weather is analyzed in the fourth section.

It should be kept in mind that these analyses only consider passengers that use the chipcard to travel with the bus. People that travel with a paper ticket are not included, which might distort the analysis in this chapter. Counts of Veolia in Tilburg shows the percentage of people travelling with a paper ticket can vary a lot: from 5.5% in April to 18.9% in July. It is quite conceivable the fraction of paper tickets is higher on Saturdays and Sundays than on weekdays.

8.1 All bus stops together

Figure 22 shows the variation of the number of boarders over the year. Each bar represents the percentage of boarders in that month, with an average percentage of boarders of 8.33% (100% / 12). The figure highlights the effect of the summer period on the number of boarders. The NVS-counts were usually performed in November, which was used to determine the average number of bus passengers on workdays. If the total yearly number of bus passengers is based on November without correction factor, the total number of bus passengers was significantly overestimated. In this case, it would have resulted in an overestimation of approximately 30%. If paper tickets are however included in figure 22, the difference with the NVS counts would be smaller. Estimations of the number of boarders in the province of Fryslân support this: the number of passengers based on the NVS-counts was in 2012 30 to 40 percent higher than the number of passengers based on the chipcard data and the card sales in the bus (Fryslân, n.d.).

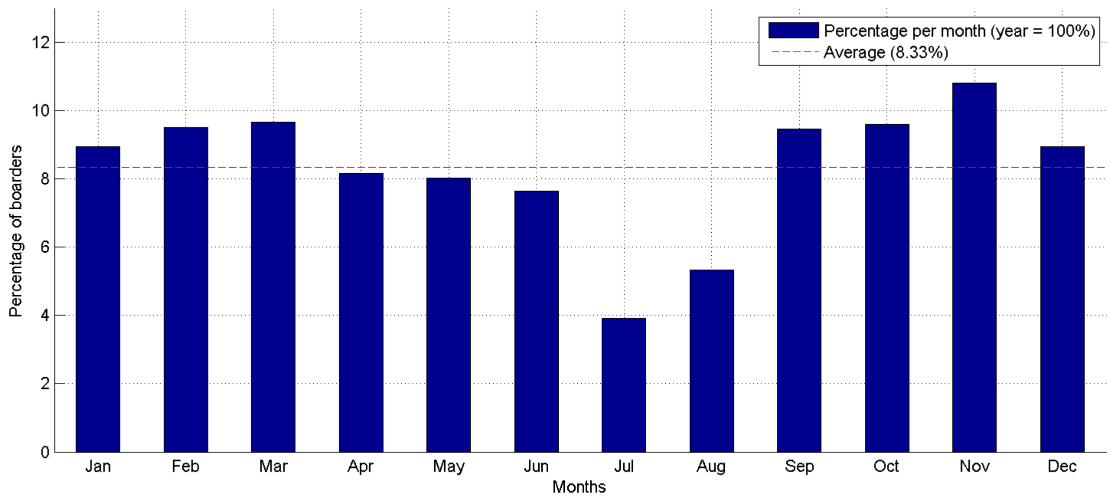


Figure 22: Variation of the number of boarders over the months of the year. The red dotted line represents the average per month

Figure 23 shows the variation of the number of boarders over the days of the week for 2012. The total number of boarders of 2012 is used to create this figure. Tuesday and Thursday appear to be slightly busier than the other weekdays, while the weekend only accounts for 10% of the number of boarders.

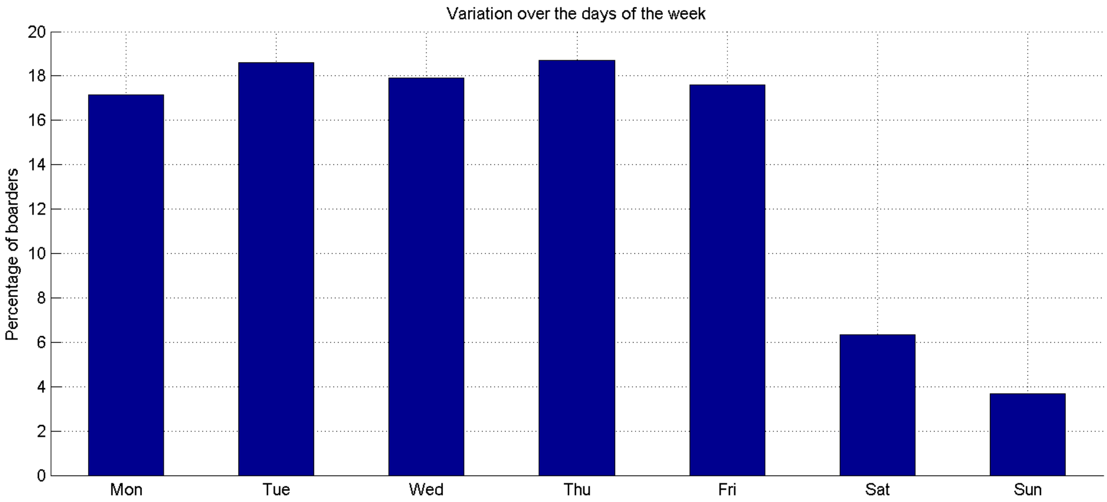


Figure 23: Variation of the number of boarders over the days of the week.

When considering the variation of the number of boarders, it appears the variation is quite similar for all weekdays (see figure 24). In the figure, each point represents the percentage of boarders during that hour, with each line summing to 100%. The morning has a peak between 8:00 and 9:00h, followed by a quieter period between 9:00 and 12:00h. From 12:00h the number of boarders slowly rises to reach the busiest hour between 16:00 and 17:00h. The Saturdays and Sundays do not have multiple peak periods, but have a constant peak between 13:00 and 18:00h. Both days follow a similar pattern, although the buses start an hour later on Sunday and the evening is busier in comparison to the rest of the day.

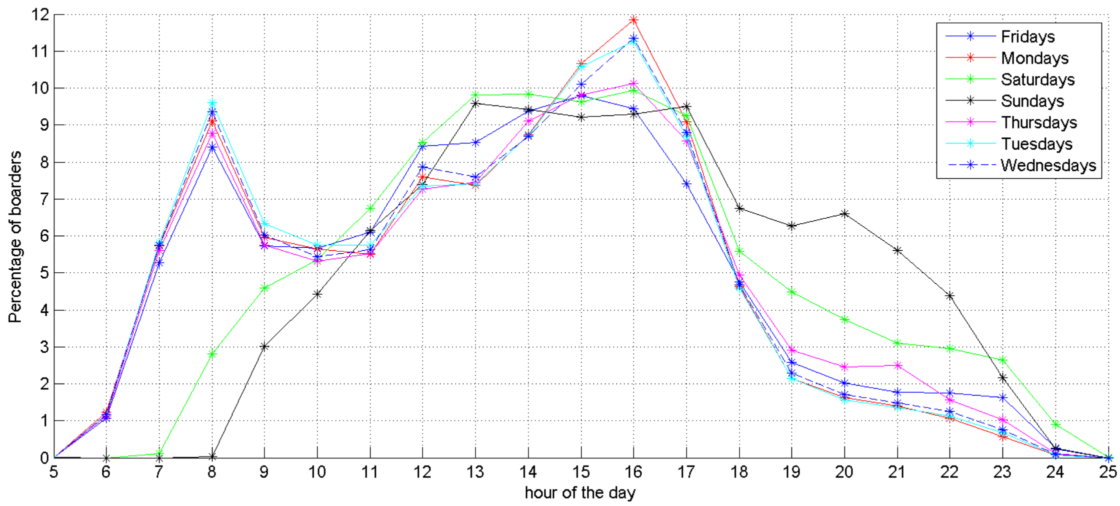


Figure 24: Variation of the number of boarders over the hours of the day per day of the week.

8.2 Bus stops separately

The previous section considered the variation of the number of boarders for all bus stops in Breda together; this section assesses the differences between the bus stops. Figure 25 shows the same information as Figure 24, but with only the boarders on the Tuesdays and all 185 bus stops in Breda separately. It highlights the differences between the bus stops. Some bus stops have their busiest hour during 7:00 and 8:00h, while others have their busiest hour between 10:00 and 11:00h or 16:00 and 17:00h.

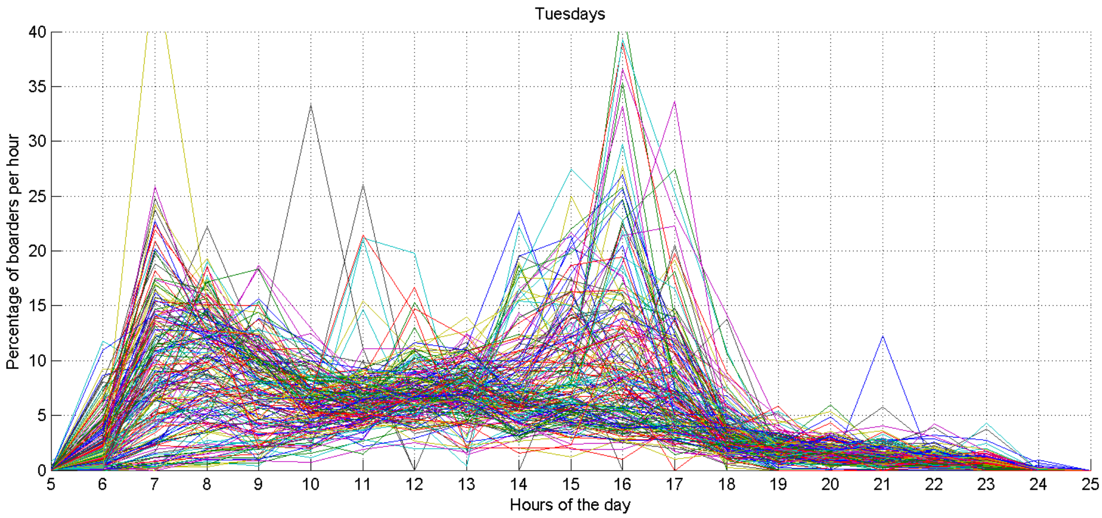


Figure 25: Variation of the number of boarders over the hours of the day at the 185 bus stops in Breda on Tuesdays in 2012. Each line adds to 100%.

The differences in variation over the hours of the day can be partially explained by their location. The following figures show a selection of the bus stops from Figure 25. Figure 26 shows bus stops with a peak between 7:00 and 9:00h with a gradually declining number of boarders during the day. Those bus stops are located in residential areas with people leaving their house in the morning to go to their work or school, while few people board the bus the rest of the day.

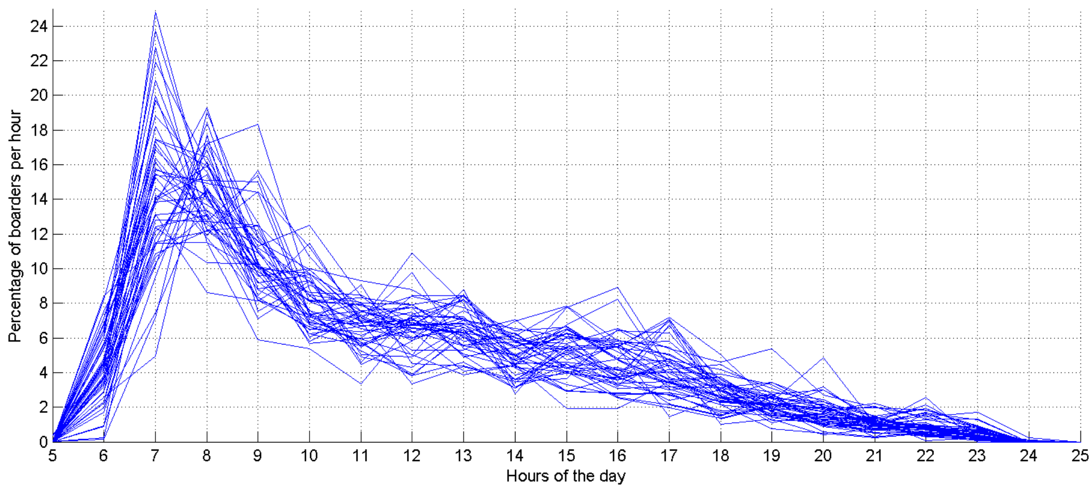


Figure 26: Variation of the number of boarders on Tuesdays at bus stops with a peak during the morning peak, with a gradually declining number of boarders during the day.

In contrary to figure 26, figure 27 shows bus stops with a peak between 16:00 and 17:00h. Those are bus stops where little people board during the morning and the first half of the afternoon, while most people board the bus during the evening peak. Those bus stops are located in or near industrial or business areas, where people board the bus to go home at the end of their working day.

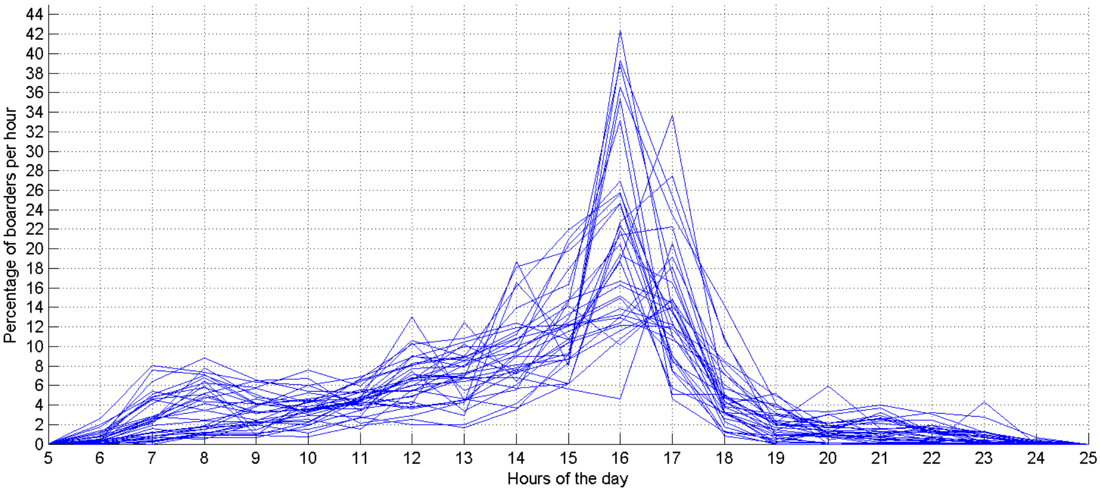


Figure 27: Variation of the number of boarders on Tuesdays at bus stops with a peak at the end of the afternoon

While the previous two figures showed bus stops with a clear peak during the morning peak hour or the evening peak hour, Figure 28 shows bus stops that do not have a clear peak. The number of boarders at those bus stops are more or less constant from 7:00h in the morning to 18:00h. These stops are typically located at the city centre, shopping centres and nursing homes.

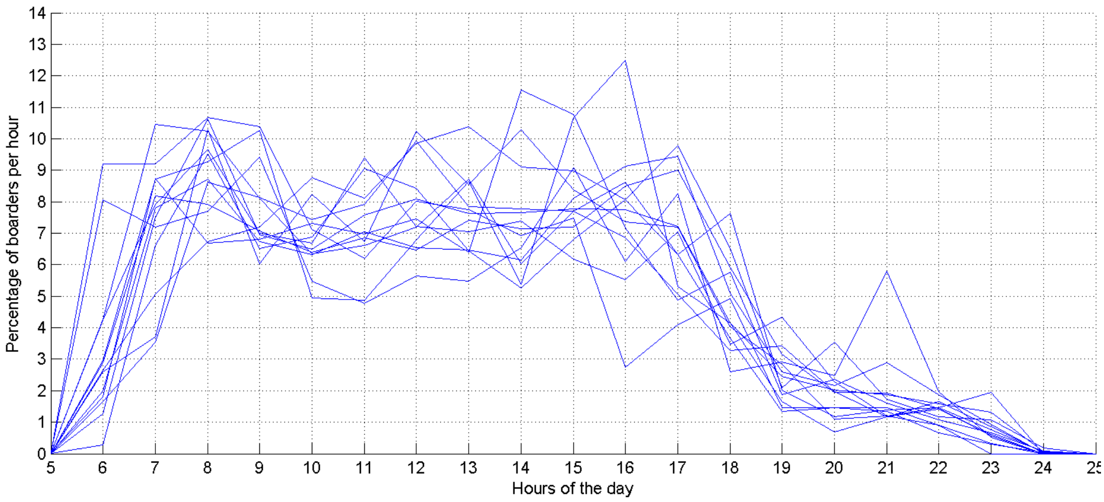


Figure 28: Variation of the number of boarders on Tuesdays at bus stops without peak periods.

8.3 Effect holidays and events

Figure 29 shows the effect of the holidays on the number of boarders in Breda. Each bar represents the number of boarders during the week, with the blue bars representing the regular weeks, while the red bars representing the official holidays in 2012. The holidays are consecutively Christmas holiday, spring holiday, May holiday, summer holiday, autumn holiday and, again, Christmas holiday. The figure shows that the number of boarders is fairly constant in the holiday weeks.

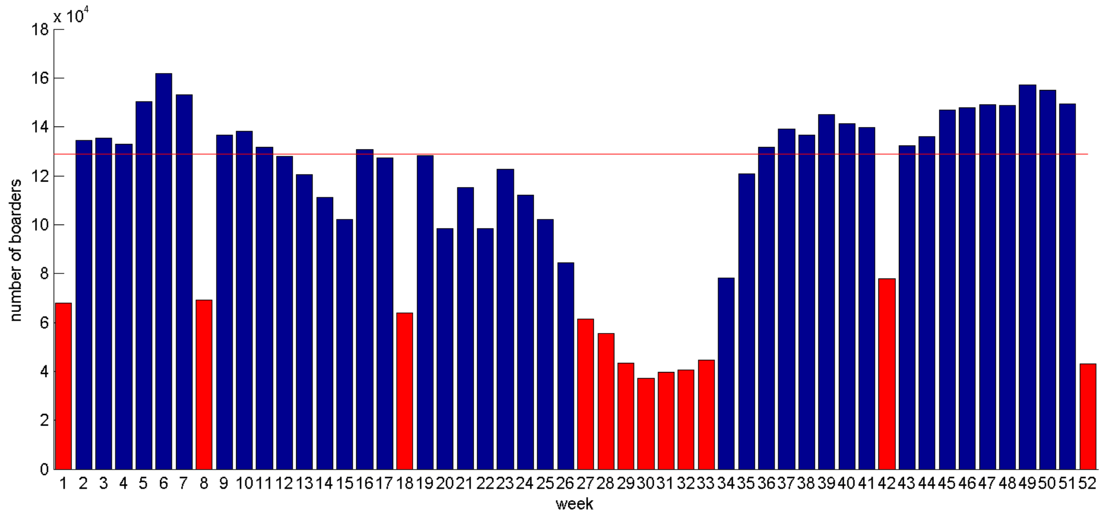


Figure 29: Number of boarders per week in 2012. The red bars represent the official holiday weeks (12 in total). The red horizontal line represents the average number of boarders during a week, excluding the holidays.

Figure 30 shows the effect of opening the shops in the city centre on Sundays (*koopzondagen*). The red lines represent the *koopzondagen* and the blue line the other Sundays. It seems that the koopzondagen are busier than the other Sundays, emphasized by the drop after 18:00h. Shoppers, and the employees of the shops, will go home after the shops close, which probably causes the drop after 18:00h. In the evening the *koopzondagen* seem to follow the regular pattern of the Sundays. Since October 2012 shops in Breda's city centre are allowed to open their doors each Sunday (BredaVandaag.nl, 2012).

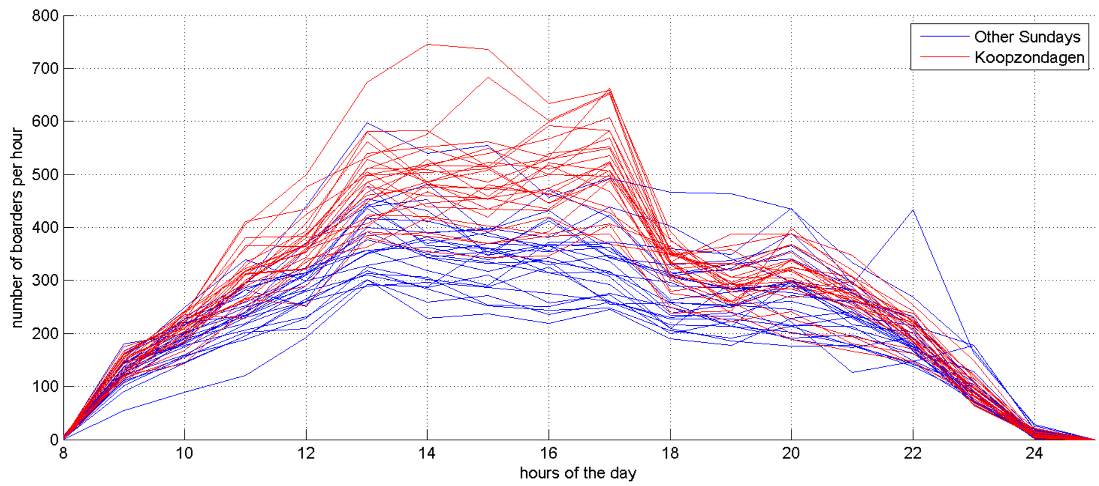


Figure 30: Variation of the number of boarders over the hours on Sundays in Breda in 2012.

The influence of events seems to be very limited. In order to assess the effect of events, the days of the week are separately analyzed. Figure 31 shows the number of boarders per hour on the Fridays of 2012. Each line in the figure represents a Friday. Apart from Carnival in the evening, hardly any influence of the events is visible from the figure. The events on all days of the week are presented in Appendix II.

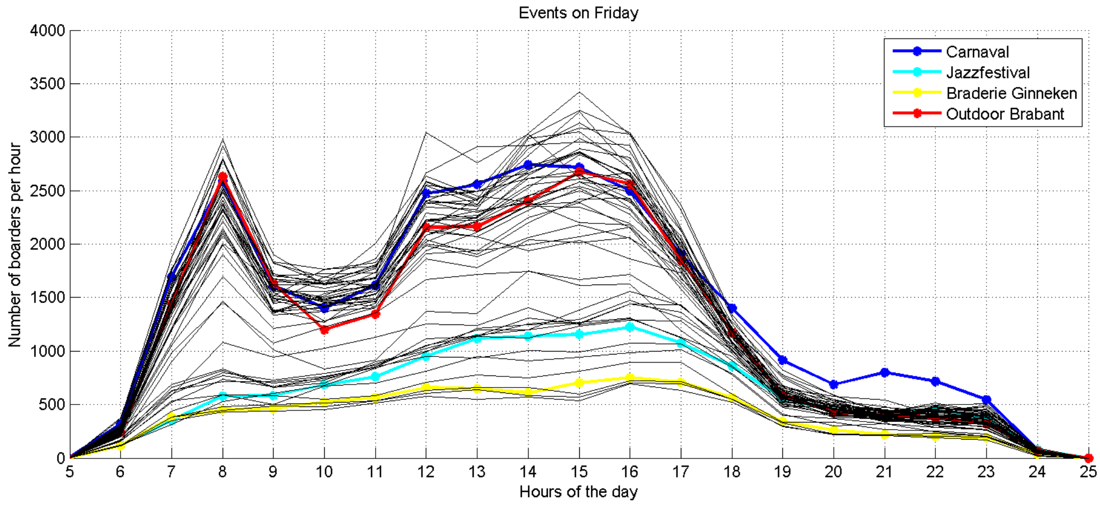


Figure 31: Variation of the number of boarders per hour on Fridays. The events in Breda on Fridays in 2012 are coloured.

8.4 Effect weather

The relation between the weather and the number of bus passengers is a rather complicated one. The bicycle is the main competing mode of the bus which is a mode that is relatively sensitive to weather circumstances. It is likely that people consider the bus to be an alternative for the bicycle with bad weather circumstances. The number of bus passengers does vary significantly over time, so the number of boarders are normalised by correcting for the effect of holiday weeks and day of the week. Therefore a correction factor is calculated for each week to correct for seasonal and holiday influences. Next to this, the daily number of boarders is corrected for the day of the week by converting them to an average day of the week (see table 23). The numbers of boarders are not corrected for the effect of holidays like Easter, Queensday and Pentecost.

Since it appears that the outliers caused by holidays like Easter, Pentecost and Christmas have a large influence on the correlation coefficients, the holidays are removed. In total ten days have been removed (New Year’s Day, Easter, Queensday, Ascension Day, Pentecost and five weekdays in the week of Christmas). Although this reduces the number of outliers, it is well possible that there are still other influences in the normalised number of boarders. The results should therefore be interpreted carefully. Analyzing the seasons separately or analyzing data of multiple years could give other results, but because of data and time constraints the analysis is limited to 2012 and all seasons together. The weather data of the weather station Gilze-Rijen is used, which is located between Breda and Tilburg (Koninklijk Nederlands Meteorologisch Instituut, 2013). The table with the range and average of the weather characteristics is included in Appendix III. Table 5 shows the correlation coefficients of the weather characteristics with the corrected daily number of boarders. The range of the variables and the average value of each weather characteristic is The significance of the average wind speed and maximum wind speed are quite low, which is not really surprising.

Really high wind speeds might make people move from the cycle to the bus, but those wind speeds are rather rare in the Netherlands.

Table 5: Correlation coefficients simple regression models of several weather characteristics with the corrected daily number of boarders

Variable	Correlation Coefficient	Significance
Average wind speed (0.1 m/s)	-.068	.199
Highest average wind speed over an hour (0.1 m/s)	.033	.535
Minimum temperature (0.1 °C)	-.173**	.001
Highest temperature (0.1 °C)	-.224**	.000
Sun hours (0.1 h)	-.199**	.000
Precipitation period (0.1h)	.157**	.003
Daily precipitation (0.1mm)	.132*	.013
Highest hourly sum of precipitation (0.1mm)	.103	.052

**: Correlation is significant at the 0.01 level (2-tailed)
*: Correlation is significant at the 0.05 level (2-tailed)

The influence of the temperature could be somewhat ambiguous. Apparently both a higher minimum and higher maximum temperature result in less bus passengers. It seems that people are more likely to take cycle or walk when the temperatures are higher. One could argue however that a lower maximum temperature (e.g. < 15 degrees Celsius) is more comfortable to cycle than a higher temperature, which would have resulted in a positive correlation. Figure 32 shows the scatter plot of the maximum temperature and the normalised number of boarders. Thomas, Jaarsma and Tutert (2009) also found the temperature to be the weather characteristics which has most influence on the number of cyclists.

The characteristic sun hours has the second highest correlation with the number of bus passengers. When the sun shines, it is more attractive to take the bicycle or walk instead of taking the bus, so the negative correlation of the sun period is quite logical. In the study of Thomas et al. sun hours was the second most important characteristic for bicycle demand, so this supports the theory that there is a strong (negative) correlation between demand for bicycling and demand for bus transport.

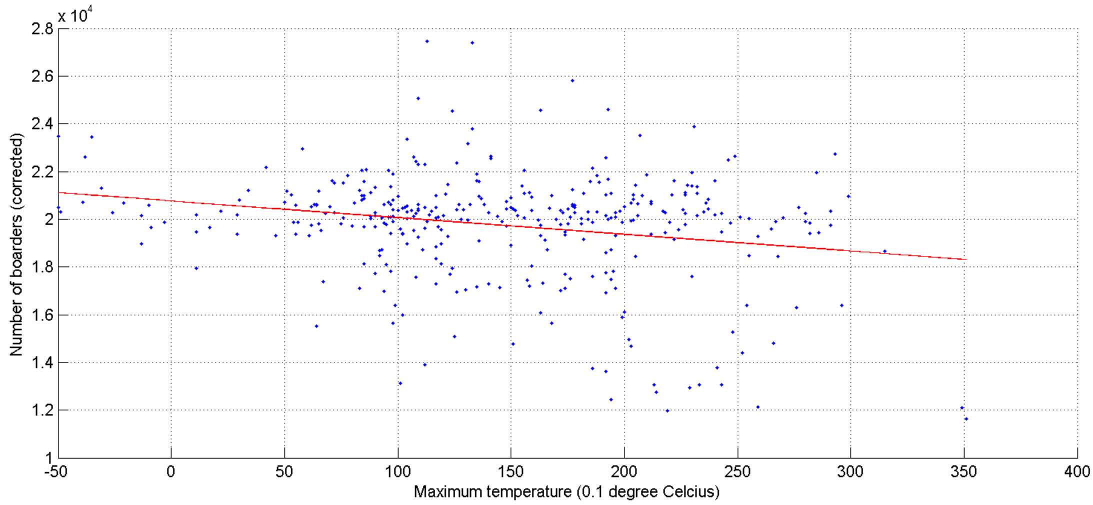


Figure 32: Scatter plot of the maximum temperature at each day in 2012 and the normalised number of boarders in Breda. The red line is the trend line.

As expected, the precipitation period and the amount of precipitation have a positive effect on the number of boarders. Again, this result is supported by the study of Thomas et al. (2009). Some of the people that usually walk or cycle will take the bus if it rains or snows. It appears that those people are somewhat more sensitive for the duration of the precipitation than for the amount of precipitation. Figure 33 shows the scatter plot of the precipitation period at each day in 2012 and the normalised number of boarders. It clearly shows why the correlation is low: the variation is quite large. Excluding the days without precipitation from the sample gives a similar correlation.

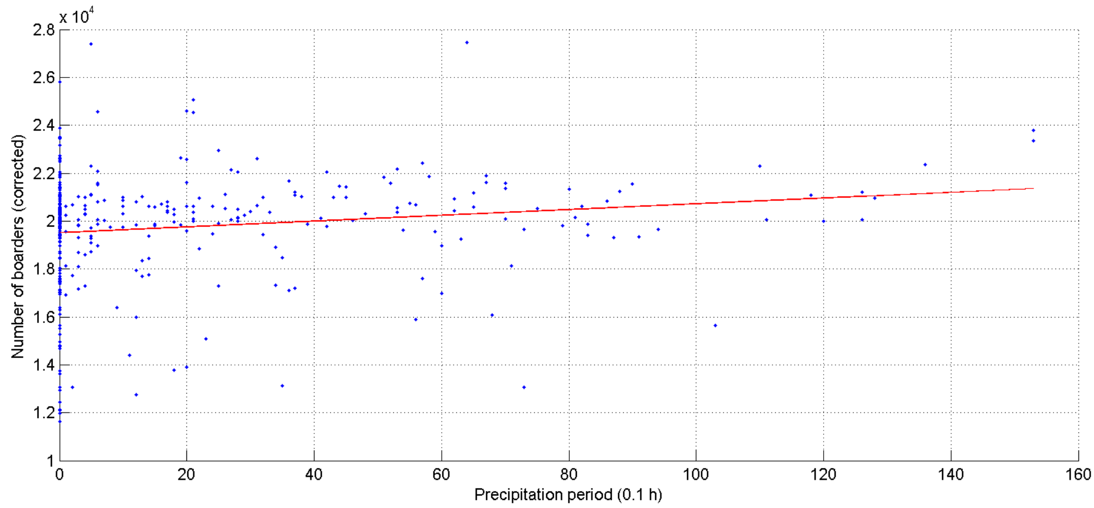


Figure 33: Scatter plot of the precipitation period at each day in 2012 at the weather station Gilze-Rijen and the normalised number of boarders in Breda. The red line is the trend line.

8.5 Conclusion

In this chapter the variation of the number of boarders over the months of the year, over the days of the week and over the hours of the day were analyzed. Also, the influence of the holidays and events were assessed. The analysis showed that the number of boarders during the summer months is significantly lower than the other months, but that the differences between the other months are rather limited. The number of bus passengers appears to be fairly constant over the weekdays, although Tuesday and Thursday are slightly busier. Approximately 90% of the bus passengers travel on weekdays, the other 10% travel in the weekend. The spread of the passengers over the day is also quite similar for all weekdays: there is a peak between 8:00h and 9:00h and between 16:00h and 17:00h. Approximately 15% of all passengers on weekdays board during the morning peak (7:00-9:00h) and 20% board during the evening peak (16:00-18:00h).

The number of passengers in the holiday periods is about half of the average number of passengers and is fairly constant over all holiday weeks. The effect of events appears to be fairly limited, although during some events the number of boarders is slightly higher for some hours. The opening of shops on Sundays seems to have a significant influence on the number of bus passengers during the opening hours of the shops, although the absolute number of passengers is still limited.

There are some weather characteristics that are slightly correlated with the number of bus passengers. The minimum and maximum temperature and the sun hours have a negative correlation while the precipitation period and daily precipitation have a positive correlation. The correlations are however so small that weather does not have a large influence on the number of bus passengers and is therefore not taken into account in the rest of the thesis.

9 Development generation models

Chapter 3 mentioned some subjects that needed to be explored in order to start with the development of the generation models. The first subject is the factors that determine the demand for bus transport. In chapter 4 a literature study was conducted to identify the key factors from literature. Chapter 5 assessed the availability of the data, which resulted in a list of variables that are used for the trip generation models. The second subject that needed to be addressed was the variation of the number of bus passengers over time. This was addressed in chapter 8. The third and fourth subjects were the number of boarders and alighters at bus stops and the size of the catchment areas. The number of boarders and alighters at the bus stops can be retrieved from the chipcard data, so only the catchment areas need to be addressed before the generation models can be developed.

The first section elaborates on the preparation for the development of the trip generation models. The second section describes the development of the trip generation model for the number of boarders on weekdays and the third section compares the included variables with other bus demand models.

9.1 Preparation regression analysis

9.1.1 Time periods

The analysis of the variation over time showed the number of boarders at the weekdays is similar, although there is little resemblance with the number of boarders on Saturday and Sunday. In addition, the main travel motives will be different on Saturdays and Sundays. The number of boarders and alighters at bus stops will probably be explained by other variables because of the other travel motives. Therefore separate trip generation models are created for these three time periods:

1. Weekdays (Monday to Friday)
2. Saturday
3. Sunday

The socio-economic variables are considered to be constant over time, so those variables are the same for all time periods. The dependent variables, boarders and alighters per time period, are different in every time period. For the weekday model, the average number of boarders on the Tuesdays in 2012 was calculated. The holidays were excluded, resulting in an average over 40 Tuesdays.

9.1.2 Input variables

The input variables were selected in chapter 5. In total 67 explanatory variables and 6 dependent variables are used. The 67 explanatory variables can be categorized in the four categories from the literature study: population characteristics, spatial characteristics, bus service characteristics and trip specific characteristics. Most variables are related to population and spatial characteristics, since most aggregated data is available and easily accessible for these categories. The six dependent variables are boarders weekday, boarders

Saturday, boarders Sunday, alighters weekday, alighters Saturday and alighters Sunday. An overview of the input variables is included in Appendix II.

9.1.3 Catchment area

In order to analyse the influence of the catchment area, four catchment areas are used for the regression analysis:

1. Catchment area of 500m
2. Catchment area of 300m
3. Catchment area of 200m
4. Voronoi diagrams

The first method assumes each person within 500m of a bus stop is a potential bus passenger for that bus stop (see figure 34). Since some people live in the catchment area of multiple bus stops, they are considered to be potential bus passengers for all those bus stops. The second method assumes the area of influence of bus stops is 300m and the third method uses a catchment area of 200m (figure 35). The final method assumes all people travel via the bus stop that is closest by. Voronoi diagrams are used to represent the catchment areas of bus stops under this assumption (figure 36). With this method the complete study area is covered by bus stops, while in methods one to three some people are not considered in the analysis at all.

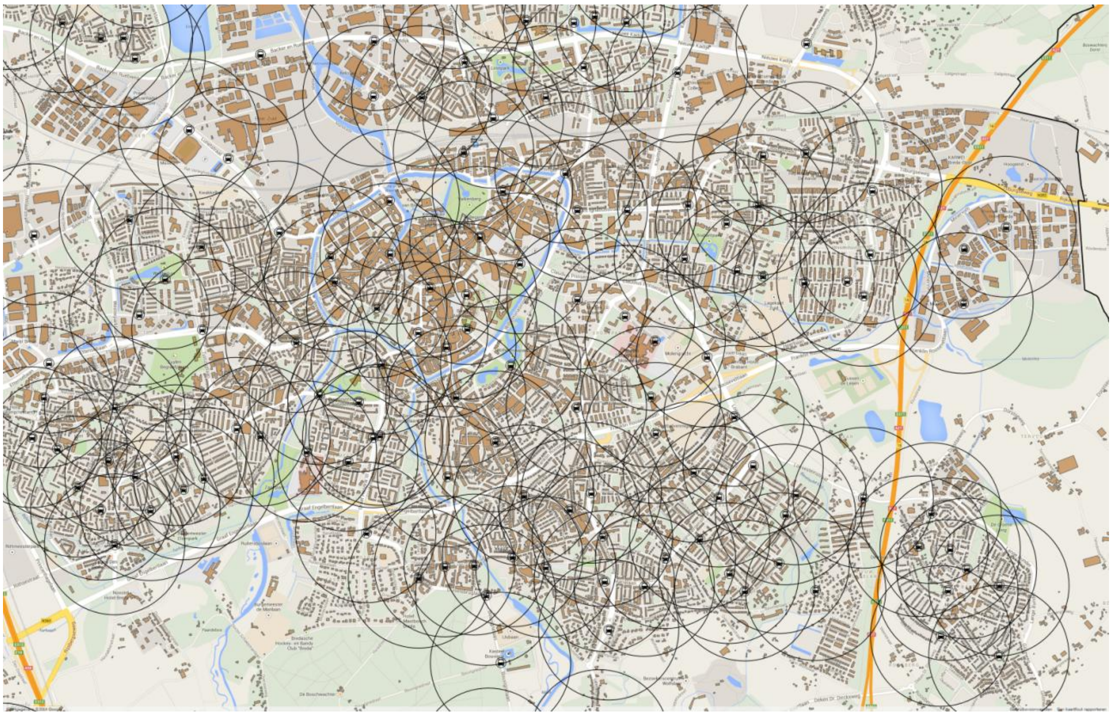


Figure 34: Bus stops with catchment areas of 500m

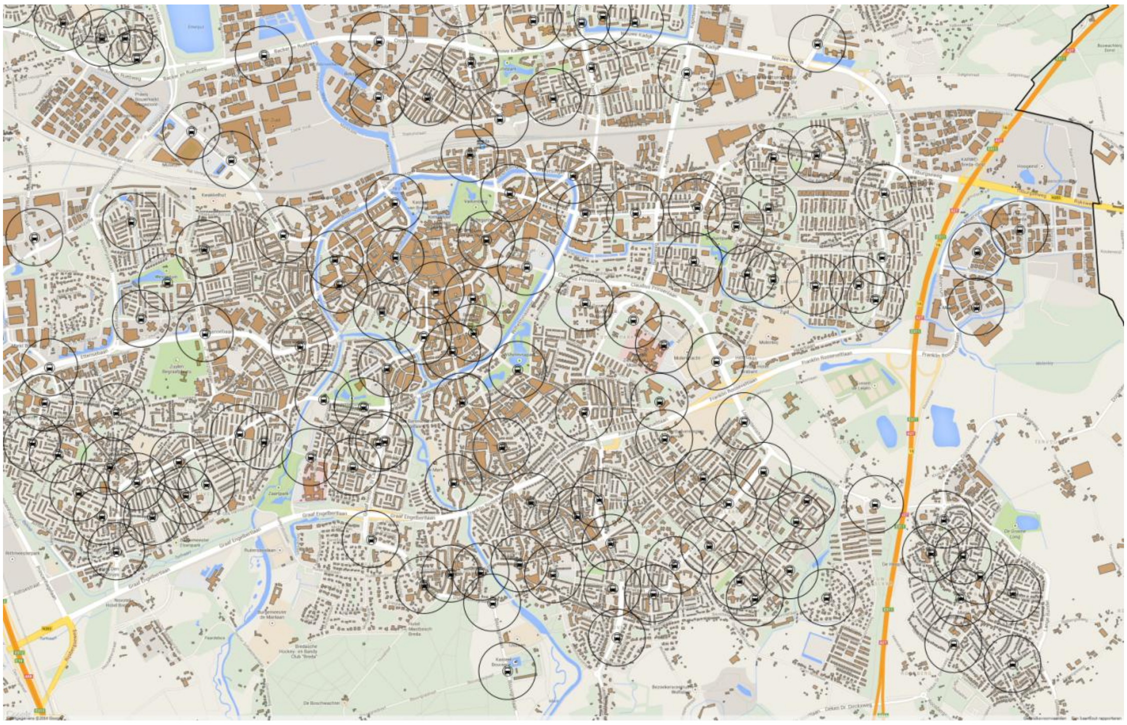


Figure 35: Bus stops with catchment areas of 200m

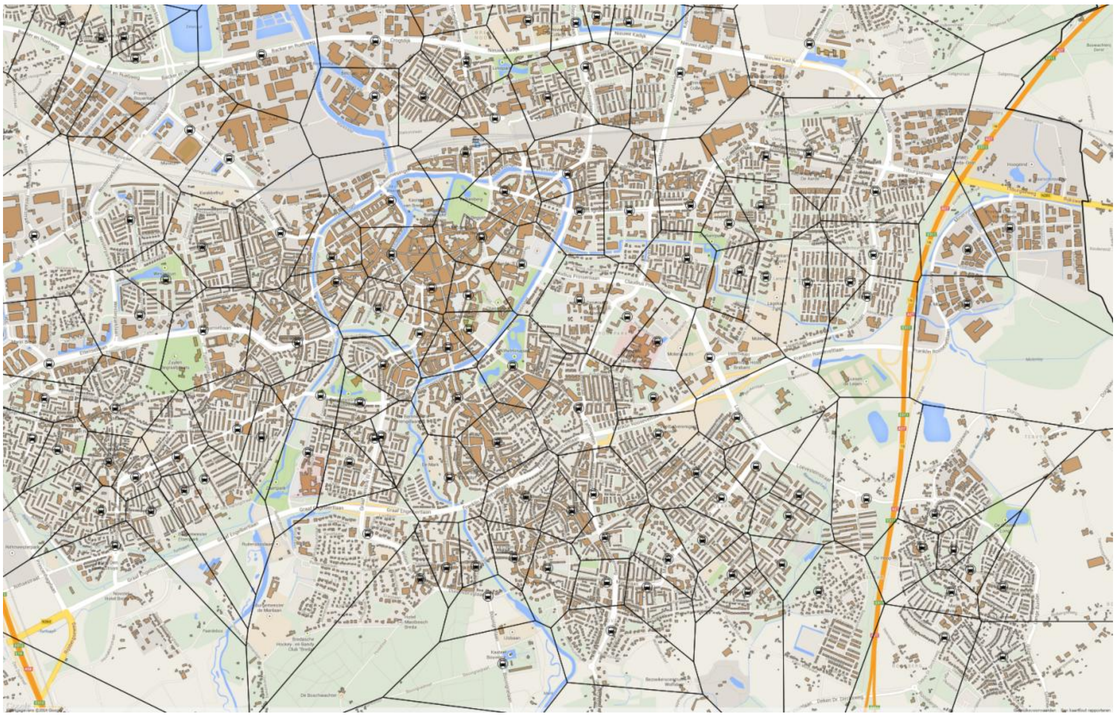


Figure 36: Bus stops with Voronoi catchment areas

The stepwise regression analysis of the number of boarders on the weekdays was performed for four different catchment areas to explore the differences. It is assumed that the catchment area with the best R Square from this stepwise regression analysis is most suitable to create the models for. Table 6 gives the R Square of the four regression models. It shows the catchment area of 500m gives the worst result. This is probably caused by the high density of bus stops; there is quite some overlap of sometimes up to seven bus stops. In case of overlapping catchment areas, the people (and other variables) in the overlapping area are counted for both bus stops since they are potential bus passengers for other bus stops. The catchment area with the best R Square is the catchment area of 200m, although the differences are very small. The Voronoi has the second largest R Square, which is almost similar to the one of 200m. The issue with Voronoi catchment areas is that it assumes all people travel via the nearest bus stop, which is a dubious assumption. If there is a bus stop with 2 buses an hour and another bus stop 300m from the other one with 12 buses an hour, most people will walk to the bus stop with 12 buses an hour, even if they need to walk 300 meter extra. Because of the small distances between the bus stops in Breda, it is very likely people consider multiple bus stops suitable to board the bus. Therefore the catchment area of 200m is considered to be more suitable than the Voronoi.

Table 6: R Square of the regression models for the weekdays (entire day)

	500m	**300m**	**200m**	**Voronoi**
R Square	0.967	0.985	0.991	0.989

9.2 Creating trip generation model boarders weekday

9.2.1 Correlation coefficients

The first stage of creating the regression models is to determine the correlation of the input variables with the dependant variable. In order to identify the variables which are important at both Breda and Tilburg, this analysis is performed for both Breda and Tilburg. Breda has 25 significant ($p<0.05$) variables and Tilburg has 14 significant variables (see table 07). The correlation coefficients of the insignificant variables are included in Appendix V. The dummy variable for the Central Station has the highest correlation coefficient for both Breda and Tilburg which is in both cases an almost perfect correlation. The second highest correlation coefficient is the number of stops that can be reached from a stop without transfer, which is quite logical since the Central Station has the highest number of boarders and all stops in the municipality can be reached from this station. The stops in the city centre are served by multiple bus lines, resulting in a high number of accessible stops without transfer and those stops have high numbers of boarders. The stops in the residential areas are only served by one or two lines and have low number of boarders. These factors together cause a high correlation coefficient.

Table 7: Correlation coefficients of the significant variables with the number of boarders on weekdays in Breda and Tilburg with a catchment area of 200m via a simple regression analysis

Variable	**Description**	**Breda (N = 185)**	**Tilburg (N = 191)**
no_stops	Number of stops that can be reached without transfer	.817**	.713**
p_014	Percentage of population of age 0-15 years	-.156*	
p_1524	Percentage of population of age 15-24 years		.212**
p_2544	Percentage of population of age 25-44 years	.145*	

Variable	Description	Breda (N = 185)	Tilburg (N = 191)
jobs	Number of jobs	.206**	.344**
P_LINH	Percentage of low income households	.152*	
SUMStud(HO)	Number of students at intermediate vocational educational institutes. Universities of Applied Science and Universities (MBO, HBO, WO)	.153*	
OAD	Average number of addresses within 100m of each address	.175*	.181*
a_station	Distance to nearest train station	-.189**	-.146*
opp_winkel	Floor area of addresses with a retail function	.171*	
no_winkel	Number of addresses with a retail function	.154*	
opp_kantoor	Floor area of addresses with an office function	.281**	.473**
no_kantoor	Number of addresses with an office function	.369**	.290**
opp_bijeenkomst	Floor area of addresses with a social gathering function	.222**	
no_bijeenkomst	Number of addresses with a social gathering function	.175*	
opp_logies	Floor area of addresses with an accommodation function	.472**	.142*
no_logies	Number of addresses with an accommodation function	.385**	.316**
opp_overig	Floor area of addresses with an other function	.188*	
% woon	Percentage of floor area with residential function	-.162*	-.198**
% kantoor	Percentage of floor area with an office function	.183*	.371**
% bijeenkomst	Percentage of floor area with a social gathering function		.165*
% logies	Percentage of floor area with an accommodation function	.386**	
% overig	Percentage of floor area with an other function	.167*	
opp_winkels_ horeca	Floor area with a retail, social gathering or accommodation function	.238**	
p_winkels_horeca	Percentage of floor area with a retail, social gathering or accommodation function	.220**	
type_CS	Dummy variable for the bus stop at the Central Station	.954**	.981**
Type_shoppingcentre	Dummy variable for the stops with opp_winkels_horeca ≥ 5000 m^2	.272**	.163*

**. Correlation is significant at the 0.01 level (2-tailed)
*. Correlation is significant at the 0.05 level (2-tailed)

Most significant variables are related to the functions of the addresses in the catchment area of the bus stops. Bus stops with a large floor area of offices, social gathering buildings or accommodations apparently have a larger number of boarders. Buildings with these three functions are often located in or around the city centre near bus stops with large numbers of boarders, while they are also located near bus stops more to the outskirts of the city, with lower number of boarders. Probably this causes these variables to be significant. Addresses with a residential function have a negative correlation with the number of boarders. The percentage of floor area with a residential function is the highest in the residential areas at the outskirts of the city, while the percentage is lower in the city centre. In the residential the number of boarders is lower than around the city centre, resulting in a negative correlation (see figure 37).

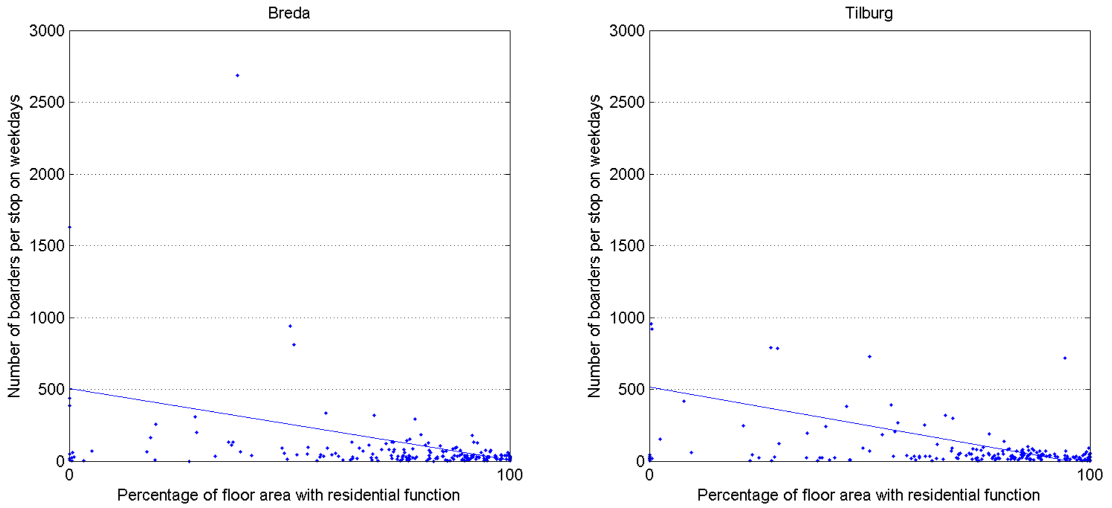


Figure 37: Scatter plots of the percentage of floor area with a residential function and the daily number of boarders at each bus stop in Breda and Tilburg. The line represents the trend line (least square line in Matlab).

The dummy variable for the presence of the Central Station has the highest correlation coefficient at both Breda and Tilburg. Figure 38 shows the scatter plots of this dummy variable and the number of boarders at each bus stop. The bus stop 'Central Station' is located at the upper right corner in both Breda and Tilburg and the other bus stops are located in the lower left corner. Because of the relatively small differences between the bus stops and the trend line the correlation coefficient is very high (> 0.95).

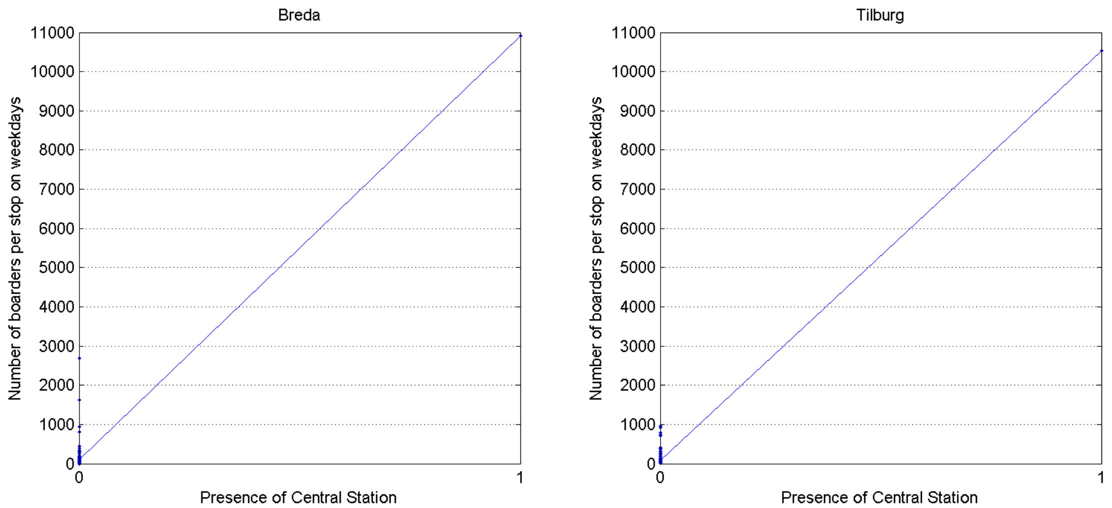


Figure 38: Scatter plots of the presence of the Central Station and the daily number of boarders at each bus stop in Breda and Tilburg. The lines represent the trend line (least square line in Matlab).

It is quite remarkable that only a few variables that are directly population-related are significant. One could expect the type of population (age, income, ethnicity, etc.) would be more important. It is remarkable that the percentage of people of 0-14 years and 25-44 years old are significant in Breda, while the 15-24 year category is significant in Tilburg. At first glance this seems to be caused by the presence of a University in Tilburg, attracting people from the category 18-24 years. The total percentage of 15-24 year old is however quite similar in Breda and Tilburg (respectively 13% and 15%). The scatter plots show why this variable is significant in Tilburg and not in Breda (figure 39): the variation in Breda is larger than in Tilburg. In Breda there is a bus stop in the upper left corner and in the lower right corner, while such outliers are absent in Tilburg.

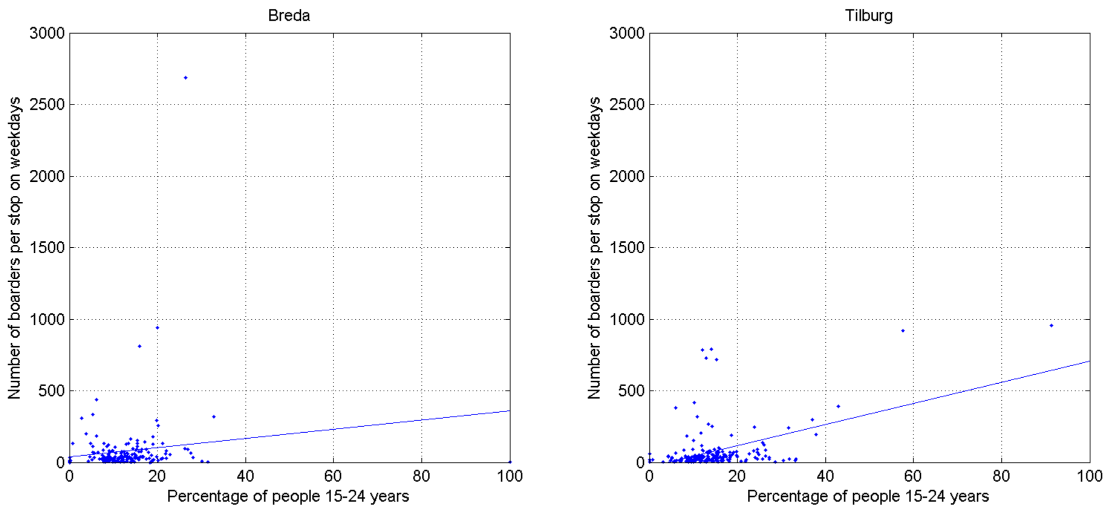


Figure 39: Scatter plots of the percentage of people of 15-24 years old and the number of boarders on weekdays at every bus stop. The bus stop 'Central Station' is not visible in the scatter plots. The lines represent the trend line (least square line in Matlab).

9.2.2 Creating regression models

From the correlation matrices of Breda and Tilburg the variables that are significant at both Breda and Tilburg are selected, together with some additional variables that are considered to be important to include in the model. This results in the following list of variables:

- No_stops
- Jobs
- OAD
- A_station
- Opp_kantoor
- No_kantoor
- Opp_logies
- No_logies
- % woon
- % kantoor
- Type_CS
- Type_centre
- P_LINH
- SUMStud(HO)
- Type_HO

Since no strongly correlated variables should be included in the trip generation model, a correlation matrix is created with the aforementioned list of variables. The variable 'jobs' is strongly correlated (≥ 0.5) with five other variables (both in Breda and Tilburg), so it is the first variable that has been removed. After removal of this variable, the variables ´no_kantoor' and '%kantoor' are strongly correlated with three variables in Breda and with two in Tilburg, so both variables are removed. The final variable that is removed is 'opp_logies' since it has a strong correlation with 'no_logies'. Some correlated variables are not removed, like 'type_CS' and 'no_stops', since both variables are considered to be important. The following variables remain:

- No_stops
- OAD
- A_station
- Opp_kantoor
- No_logies
- % woon
- Type_CS
- Type_centre
- P_LINH
- SUMStud(HO)
- Type_HO

These variables are used to create a regression model for both Breda and Tilburg. Table 8 shows the further iterative development of the model towards the final trip generation model for the number of boarders on weekdays in 2012.

Table 8: Development of the regression model for boarders on weekdays

		Breda	Tilburg
Model 1	R Square	0.97	0.991
	Standard error	149.373	76.626
Model 2	Change:	Remove OAD	
	R Square	0,97	0.991
	Standard error	149.013	76.447
Model 3	Change:	Remove no_logies	
	R Square	0,969	0.991
	Standard error	149.492	76.236
Model 4	Change:	Remove a_station	
	R Square	0.969	0.991
	Standard error	149.705	76.183
Model 5	Change:	Remove Type_centre	
	R Square	0.969	0.991
	Standard error	150.163	76.351
Model 6	Change:	Add opp_winkel_horeca	
	R Square	0.978	0.991
	Standard error	125.096	74.773
Model 7	Change:	Add OAD	
	R Square	0.979	0.991
	Standard error	123.184	74.887
Model 8	Change:	Add cars_total	
	R Square	0.979	0.991
	Standard error	122.932	74.886

The final model for the boarders on weekdays (model 8) included the following ten variables:

- No_stops
- Opp_winkels_horeca
- OAD
- Opp_kantoor
- Cars_total
- % woon
- Type_CS
- P_LINH
- SUMStud(HO)
- Type_HO

9.2.3 Compare models with stepwise regression SPSS

To check how well the developed model is doing, the results are compared with the models that SPSS creates using automatic stepwise regression. With this method SPSS tries to find the combination of variables that fits the observed data as well as possible. A model is made using the first selection of variables plus the two variables that were later added and another model is made using all variables. The results are presented in

Table 9. The table shows that the developed model (model 8) performs slightly worse for Breda, but quite good for Tilburg. The differences with the model where all variables are

uModel		Breda	Tilburg
sVariables from model 1	R Square	0.985	0.991
e+ 2 additional	Standard error	105.953	74.956
dAll variables	R Square	0.990	0.994
	Standard error	85.886	61.309

are however quite large in case of Breda: the standard error of the model of SPSS is approximately a third lower than the developed model. It seems that SPSS is able to determine quite detailed which combinations of variables gives the best results.

Table 9: Results of the automatic stepwise regression in SPSS with the first selection of variables plus opp_winkel_horeca and cars_total (17 variables in total) and with all variables

Model		Breda	Tilburg
SVariables from model 1	R Square	0.985	0.991
o+ 2 additional	Standard error	105.953	74.956
All variables	R Square	0.990	0.994
f	Standard error	85.886	61.309

Sofar the models are developed for both Breda and Tilburg to identify the variables that should be included. The final model however is only developed for Breda, while the Tilburg data is used to validate the model. Since the output of the developed model results in a negative number of boarders at some stops, these negative numbers are set to zero. Table 10 shows the R Square and standard error of the developed model of Breda with data of Breda and Tilburg. The table shows the results are actually better for Breda because of the removal of the negative number of boarders. The result of Tilburg becomes worse because of the usage of the coefficients of the Breda-model.

Table 10: Results of the model for the number of boarders on weekdays which is developed with the data of Breda.

	Breda	Tilburg
R Square	0.981	0.975
Standard error	114.917	121.846

The equation below shows the formula of the developed model for the number of boarders on weekdays. It should be kept in mind that the coefficients cannot be interpreted individually, since it is the combination of all variables that leads to the result. Apparently this combination of coefficients leads to the best fit of the observed number of boarders. It is for example more obvious if the dummy variable for the presence of a University or University of Applied Science would have had a positive influence on the number of boarders. Apparently the model overestimates the number of boarders at those stops without this variable, which is corrected by the negative coefficient for this dummy variable. The other trip generation models are included in Appendix VI.

$$Boarders_{weekday} = -128.654 + 3.721x_{no\_stops} - 0.612x_{P\_LINH} + 0.135x_{Cars\_total} + 0{,}155x_{SUMStud\ (HO)} - 0{,}044x_{OAD} + 0{,}002x_{Opp\_kantoor} + 1.074x_{\%woon} + 0.019x_{Opp\_winkel\_horeca} + 9281.374x_{Type\_CS} - 152.157x_{Type\_HO}$$

9.3 Compare included variables with literature and other models

The development of the trip generation models started with a literature study which resulted in a list of factors that influence the demand for bus transport. These factors are compared with the variables that are included in the developed model for the weekdays. The variables of the developed models are also compared with the variables included in other bus demand models. The variables included in the developed models and other models are included in Appendix VII. The developed weekday models are compared with models developed by Schmenner (1976), Klok (2010), Bechdolt & Williams (1980) and Veitch Lister Consulting (2008).

It is remarkable that only three variables are included in at least three of the four models: age, income/car ownership and number of jobs. The age variables appeared to be insignificant in the developed models, so they were not included in the developed models. A variable related to car ownership and one related to income are however both included. The number of jobs is not directly included in the models, since it was correlated with a large number of other variables. Indirectly it is however included through the variables floor area of offices and floor area with retail, social gathering or accommodation function.

Furthermore there are four variables which are included in two models: ethnicity, density, school enrolments and quality of service (frequency/headway). Ethnicity was not included in the developed models since the ethnicity-related variables had an insignificant correlation with the number of boarders. The variables density and variables related to school enrolments are included in the developed models. No variables related to quality of service were included in this study and are therewith not included in the models.

Something else which is remarkable is that only a few variables are included in the models which are related to spatial characteristics. The correlation analysis showed that especially the spatial characteristics are important factors, while the population characteristics are less important. The four models however barely include spatial variables while they include population characteristics.

9.4 Conclusion

Separate trip generation models are generated for three time periods: weekdays, Saturdays and Sundays. For each time period there is a model for boarders and a model for alighters. From the automatic stepwise regression analysis with SPSS it appeared a catchment area of 200m gives the model with the highest R Square, although the differences with other catchment areas are small. In total data of 67 explanatory variables is used. Via correlation analyses and a stepwise procedure the six trip generation models are created. The weekday models consist of ten explanatory variables and the models for Saturday and Sunday of seven variables. Bus demand models consist of very different variables; there are only a few variables which are commonly used, like age, income/car ownership and number of jobs. Especially the lack of spatial variables in these models is remarkable, since these appear to be important according to the analysis in this chapter.

10 Distribution analysis

In the previous chapter the trip generation models were developed. This chapter elaborates on the second stage of the four step model, the trip distribution stage. The first section analyzes the distribution of bus passengers in Breda according to the observed origin-destination matrix. The second section shows the calibration procedure. The calibration results of the distribution model for the weekdays are discussed in the third section and the final two sections give a sensitivity analysis and the conclusion. The calibration results of the other distribution models are included in Appendix IX.

10.1 Analysis of distribution in Breda

Figure 40 shows a map of Breda with the number of passengers on each pair of bus stops with a direct connection on the weekdays in 2012. The line thickness is scaled on the number of passengers. Unsurprisingly, most people travel to the Central Station. There are however also quite a lot of people who travel from the bus stops in the residential areas to the bus stop “Centrum” (South of the Central Station). It could be that people go there to go shopping for example, but it could also be some of those people transfer at this stop to another bus. It is possible that people board at a bus stop outside the study area, alight at for example the bus stop “Centrum” and transfer to another bus. The bus stops outside the study area are however not included, so the first part of these journeys are not included, while the second part is included. Therefore figure 40 could be a wrenched representation.

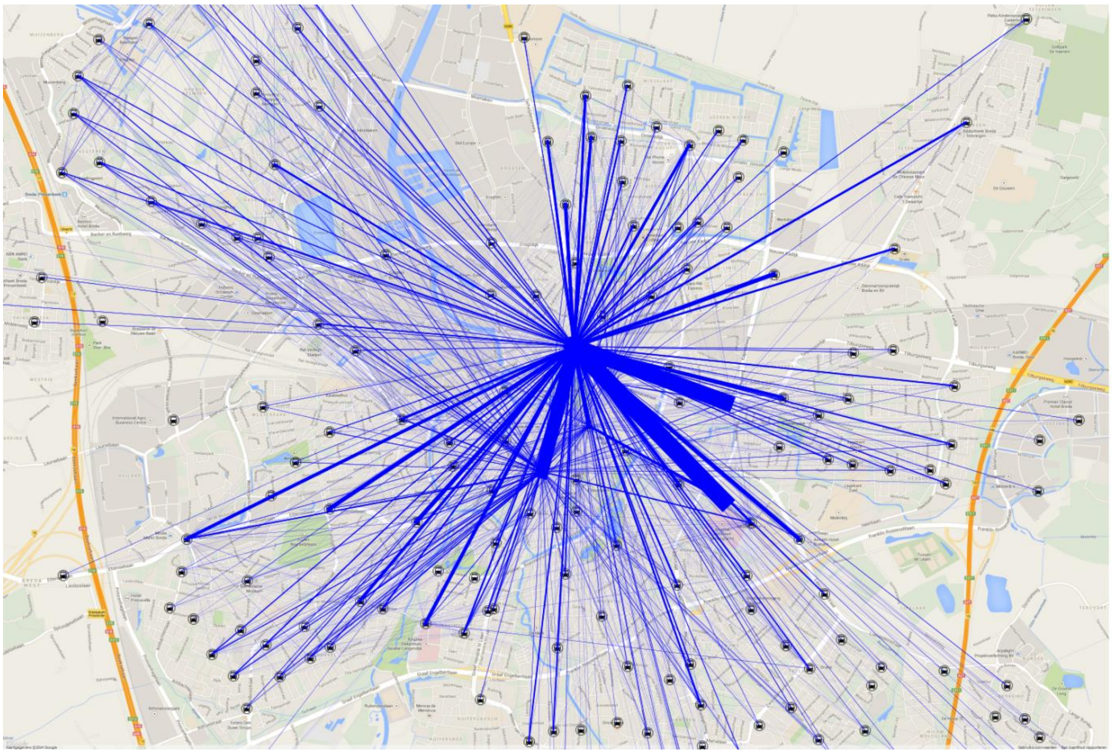


Figure 40: Map of Breda with the observed number of passengers on each pair of bus stops with a direct connection on all weekdays in 2012. Both directions (A->B and B->A) are plotted on top of each other, so only the busiest direction is visible.

The trip length distribution of Breda shows the number of trips per trip length in minutes, which is shown in Figure 41. The average trip length is 12.014 seconds with a standard deviation of 6.454. The figure shows the modal class is 12 minutes, which is mainly caused by the large number of bus passengers between the stops “Centraal Station” and “Hogeschoollaan”. The second highest peak at five minutes is mainly caused by the passengers between the stop “Centraal Station” and the stops “Centrum” and “St. Ignatiusstraat”. Figure 19 in section 7.3 shows the peaks at 5 and 12 minutes are not caused by a large occurrence of bus connections with these trip lengths, so they are caused by the number of passengers and not by the network.

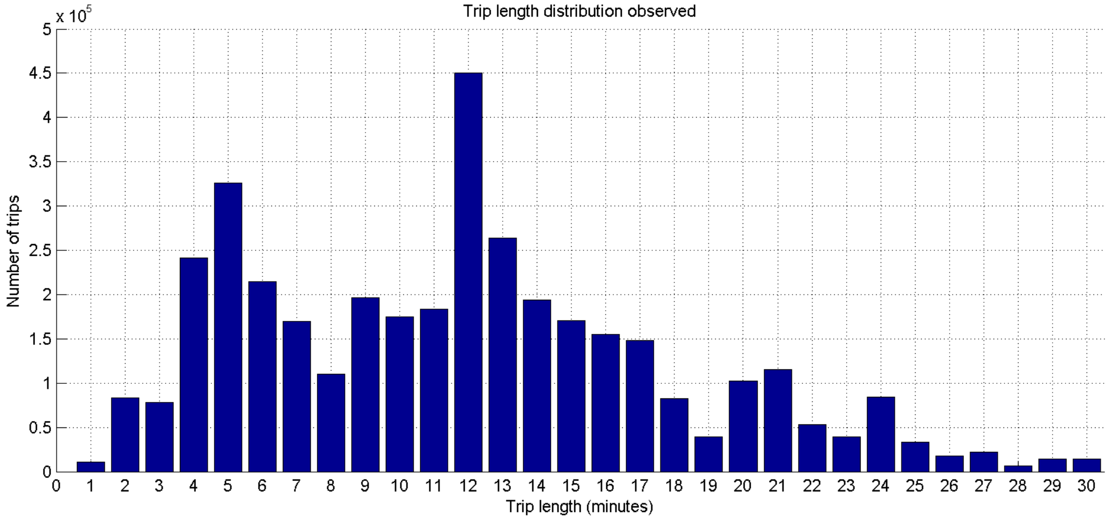


Figure 41: Observed trip length distribution of the bus trips between the analyzed bus stops in the municipality of Breda on the weekdays in 2012

10.2 Calibration procedure

The result of the generation step of the four step model is the number of boarders and alighters at every bus stop. The aim of the distribution step of the four step model is to determine the number of passengers between each bus stop pair. It is assumed the distribution takes place according to the gravity theory: the willingness to travel between two stops declines as the distance or travel time increases. Equation 10.1 gives the formula of the gravity model.

$$T_{ij} = A_i O_i B_j D_j \times f(c_{ij}) \qquad \textit{Equation 10.1}$$

With
A_i = balancing factor boarders at stop i
O_i = number of boarders at stop i *where* $\sum_j T_{ij} = O_i$
B_j = balancing factor alighters at stop j
D_j = number of alighters at stop j *where* $\sum_i T_{ij} = D_j$
$f(c_{ij})$ = distribution function

The gravity theory assumes the distribution depends on the willingness of people to travel a particular trip length. For very short trips of only a few minutes, the willingness to travel with the bus will be low, since cycling or even walking will be more attractive. At longer distances however the bus will become more attractive and for even longer trips, the car and train will become more attractive than the bus.

The distribution function is calibrated such that it meets the average trip length and standard deviation as well as possible. In order to minimize the differences between the modelled mean travel time and the observed mean travel time and the modelled standard deviation and the observed standard deviation, a goodness of fit measure is necessary. An often used goodness of fit measure is Pearson's Chi Square, which will be used here (Shrewsbury, 2012). Equation 10.2 gives its formula with the variable names used in the calibration.

$$\mathrm{gof} = \frac{(\mathrm{MMTT}-OMTT)^2}{\mathrm{OMTT}} + \frac{(\mathrm{MSTDEV}-OMSTDEV)^2}{\mathrm{OMSTDEV}} \qquad \textit{Equation 10.2}$$

Where
gof = goodness of fit
MMTT = Modelled Mean Travel Time
OMTT = Observed Mean Travel Time
MSTDEV = Modelled Standard Deviation
OSTDEV = Observed Standard Deviation

Three distribution functions are calibrated to gain the best results. The functions are the exponential distribution function (equation 10.3), the Tanner distribution function (equation 10.4) and the log-normal distribution function (equation 10.5). The first function has only one parameter (beta) and the other functions have two parameters (alpha and beta). The exponential and Tanner function are given by Ortúzar and Willumsen (2001) and the top log-normal function as used in Omnitrans of version 6.0.0 and higher (Omnitrans, 2011).

$$f(c_{ij}) = \mathrm{e}^{-\beta \mathrm{c_{ij}}} \qquad \textit{Equation 10.3}$$

$$f(c_{ij}) = c_{ij}^{\alpha} e^{-\beta c_{ij}} \qquad \textit{Equation 10.4}$$

$$f(c_{ij}) = e^{-\alpha(\log(\frac{c_{ij}}{\beta}))^2} \qquad \textit{Equation 10.5}$$

A more detailed description of the calibration procedure is included in Appendix VIII.

10.3 Results

10.3.1 Exponential distribution function

The first function that is calibrated is the exponential distribution function, since it has only one parameter and its optimized value is therefore easier to compute. Equation 10.6 shows the exponential deterrence function.

$$f(c_{ij}) = \mathrm{e}^{-\beta \mathrm{c_{ij}}} \qquad \textit{Equation 10.6}$$

The procedure of Appendix VIII was used to find the beta that gives the best fit to the mean travel time and the standard deviation. The minimum value of the goodness of fit measure was already found after seven iterations. Running the calibration algorithm with 1,000 iterations did not improve the result. To check whether the found optimum really is the optimum, the value of the goodness of fit measure was calculated for 10,000 values of beta. The beta that was found indeed appeared to be the optimum. Table 11 shows the aim values and the modelled values after calibration with the exponential deterrence function.

Table 11: Calibration values of the exponential deterrence function

$f(c_{ij}) = e^{-0.037c_{ij}}$	alpha	beta	Mean travel time	Standard deviation	Goodness of fit
Observed value/aim	-	-	12.014	6.454	0
Modelled value	-	0.037	11.562	7.090	0.080

10.3.2 Tanner distribution function

The second distribution function which is calibrated is the Tanner function, which formula is shown in equation 10.7. The calibration starts with optimizing beta with a fixed alpha of one. In the second step the optimized beta is fixed and the alpha is optimized. Then this alpha is fixed and beta is optimized. This procedure has been repeated a 100 times in order to minimize the goodness of fit measure. Table 12 shows the results of the calibration.

$$f(c_{ij}) = c_{ij}^{\alpha} e^{-\beta c_{ij}} \qquad \textit{Equation 10.7}$$

Table 12: Calibration values of the Tanner function

$f(c_{ij}) = c_{ij}^{1.045} e^{-0.119c_{ij}}$	alpha	Beta	Mean travel time	Standard deviation	Goodness of fit
Observed value/aim	-	-	12.014	6.454	0
Modelled value	1.045	0.119	12.020	6.458	5.787×10^{-6}

10.3.3 Top log-normal distribution function

The final distribution function that has been calibrated is the top log-normal distribution function, which is used in Omnitrans. Equation 10.8 shows the formula of this function. The calibration results are shown in table 13. The goodness of fit is comparable to the goodness of fit of the Tanner function, although the trip length distribution of the latter deviates somewhat less from the observed trip length distribution.

$$f(c_{ij}) = e^{-\alpha(\log(\frac{c_{ij}}{\beta}))^2} \qquad \textit{Equation 10.8}$$

Table 13: Calibration results of the top log-normal distribution model

$f(c_{ij}) = e^{-0.705(\log(c_{ij}/7.701))^2}$	alpha	Beta	Mean travel time	Standard deviation	Goodness of fit
Observed value/aim	-	-	12.014	6.454	0
Modelled value	0.705	7.701	12.014	6.454	2.853×10^{-9}

Figure 42 shows the observed trip length distribution and the trip length distribution of the three calibrated distribution models. The figure shows the exponential function overestimates the short trip lengths, while it underestimates the trip lengths from approximately 11 minutes to 17 minutes. The Tanner and top log-normal functions give a similar trip length distribution which is close the observed distribution.

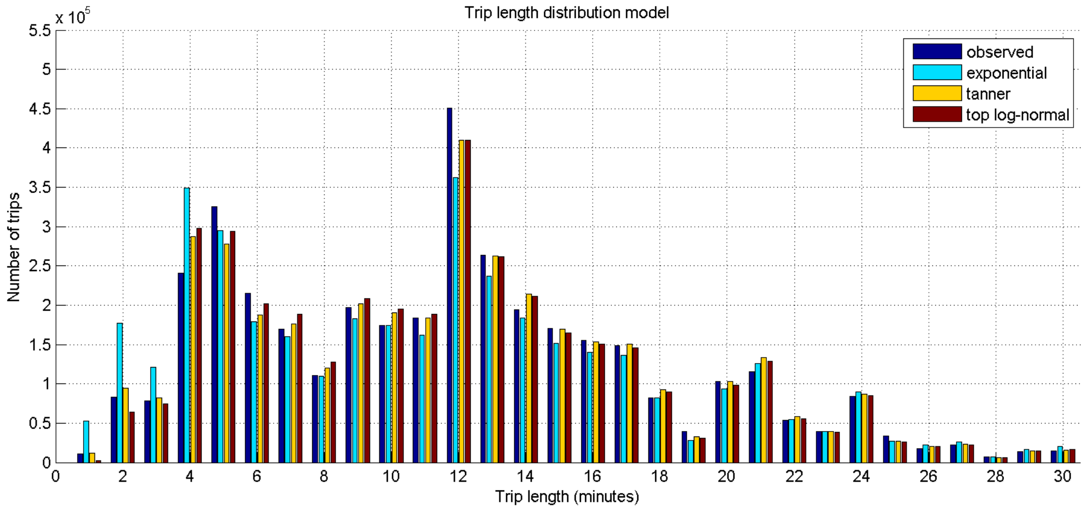


Figure 42: Observed trip length distribution and the modelled trip length distribution for the three calibrated distribution models for the weekdays

10.4 Sensitivity of models

10.4.1 Other starting values of parameters

In order to calibrate the parameters of the distribution functions, a first estimate needs to be given. Since a different first estimate might result in another optimized value of the parameters, it needs to be assessed what the influence is of other first estimates. For the beta, the first estimate was 1/OMTT and for the alpha it was '1'. The distribution functions for the weekdays are calibrated again, but with different first estimates of alpha and beta. The new first estimate for alpha is 0.1 and the new estimate for beta is the OMTT.

Calibrating the three functions again for the different first estimates gives the same results of the alpha and beta. Therefore it can be concluded that the first estimates of the parameters do not seem to influence the calibration.

10.4.2 Effect of 10% higher parameter values

To gain insight in the sensitivity of the mean travel time, standard deviation and goodness of fit, the optimized parameter values are increased with 10%. Tables 14 to 16 show the calibration results of the distribution functions for the weekdays with the effect of the increased parameter values. Table 14 shows that the effect of a change of beta on the goodness of fit is fairly limited. The goodness of fit will therefore not change much if the beta is changed.

Table 14: Calibration values of the exponential function and with a 10% higher beta

$f(c_{ij}) = \mathrm{e}^{-0.037\mathrm{c_{ij}}}$	Alpha	beta	Mean travel time	Standard deviation	Goodness of fit
Observed value/aim	-	-	12.014	6.454	0
Modelled value	-	0.037	11.562	7.090	0.080
10% higher beta	-	0.0407 (+10%)	11.457 (- 0.9%)	7.050 (+0.6%)	0.081 (+ 1.3%)

Tables 15 and 16 show the increased parameters have a small influence on the mean travel time and standard deviation, but a large influence on the goodness of fit.

Table 15: Calibration values of the Tanner function and the effect of a 10% higher alpha and a 10% higher beta.

$f(c_{ij}) = c_{ij}^{1.045} e^{-0.119 c_{ij}}$	alpha	Beta	Mean travel time	Standard deviation	Goodness of fit
Observed value/aim	-	-	12.014	6.454	0
Modelled value	1.045	0.119	12.020	6.458	5.787 x10^{-6}
10% higher alpha	1.1495 (+10%)	0.119	12.235 (+1.8%)	6.471 (+0.2%)	4.097 x10^{-3} (+7.1 x10^{4})
10% higher beta	1.045	0.1309 (+10%)	11.746 (-2.3%)	6.365 (-1.4%)	7.206 x10^{-3} (+ 1.2 x10^{5} %)

Table 16: Calibration values of the top-lognormal function and the effect of a 10% higher alpha and a 10% higher beta

$f(c_{ij}) = e^{-0.705(\log(c_{ij}/7.701))^2}$	alpha	Beta	Mean travel time	Standard deviation	Goodness of fit
Observed value/aim	-	-	12.014	6.454	0
Modelled value	0.705	7.701	12.014	6.454	2.853 x10^{-9}
10% higher alpha	0.7755 (+10%)	7.701	11.952 (- 0.5%)	6.366 (-1.4 %)	1.514 x10^{-3} (+ 5.3 x10^{7}%)
10% higher beta	0.705	8.4711 (+10%)	12.272 (+2.1%)	6.489 (+ 0.5%)	5.726x10^{-3} (+2.0 x10^{8} %)

10.5 Conclusion

Figure 43 shows the plotted lines of the three calibrated distribution functions. The effect of the additional parameter is clearly visible: the value of the Tanner and top log-normal distribution functions for short trip lengths is smaller than for slightly longer trip lengths, exactly what is observed in the data. The goodness of fit of the latter two functions both approach a minimum of zero, although the top-lognormal function gives slightly better results. When considering the trip length distributions, Tanner and the top-lognormal function perform comparable. Therefore, the top log-normal distribution function is considered to be the best performing function. Another advantage of the top log-normal function is that the top can easily be derived from the function (i.e. the top is located at 7.701, which can be read in the formula). Not surprisingly, performing a paired two-sample t-test for both the Tanner and top-lognormal functions accept the null-hypothesis that the modelled distribution and the observed distribution are equal at a significance level of 0.05%.

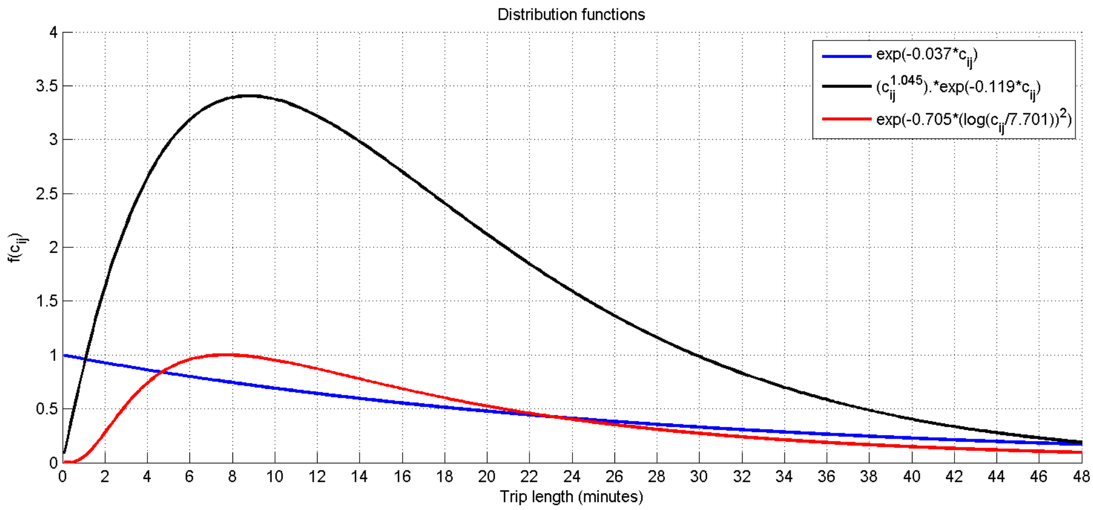


Figure 43: The three calibrated distribution functions. The top log-normal (red) and Tanner (black) functions give the best results.

11 Assignment

The distribution model that was calibrated in the previous chapter results in an Origin-Destination matrix, or, in this study, in a boarding stop-alighting stop matrix. The final stage of the four step model is to assign this matrix to the network which is formed by the bus network.

11.1 Procedure of assignment

The bus network consists of the bus stops and connections between the stops. Each of those connections is a line segment that handles passengers. Figure 44 shows a simplification of the procedure. For each pair of bus stops with a direct connection, the shortest route is selected determined to assign those passengers to. It is assumed the shortest route is the route with the least number of stops. Figure 45 shows the assignment of all bus passengers on weekdays in Breda in 2012. The thicker a line segment, the more passengers on that segment. The figure shows all major passenger flows focus on the central station. The passenger flows in the outskirts of the municipality are very small and become larger as they approach the city centre.

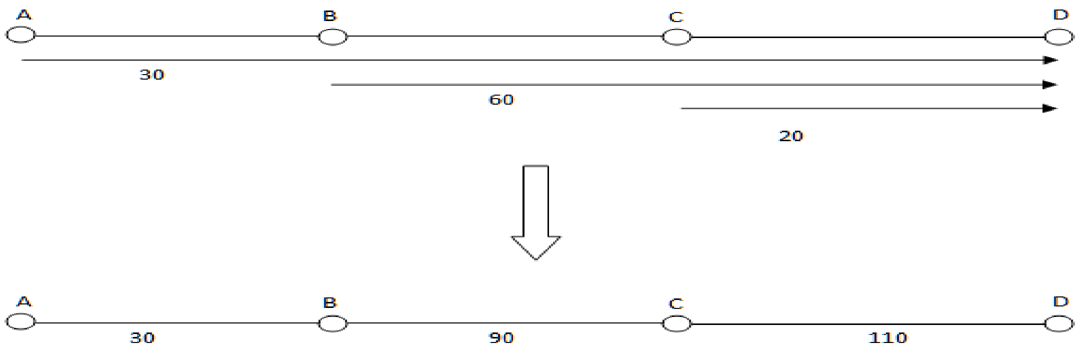


Figure 44: Procedure of assigning the origin-destination matrix to the bus network

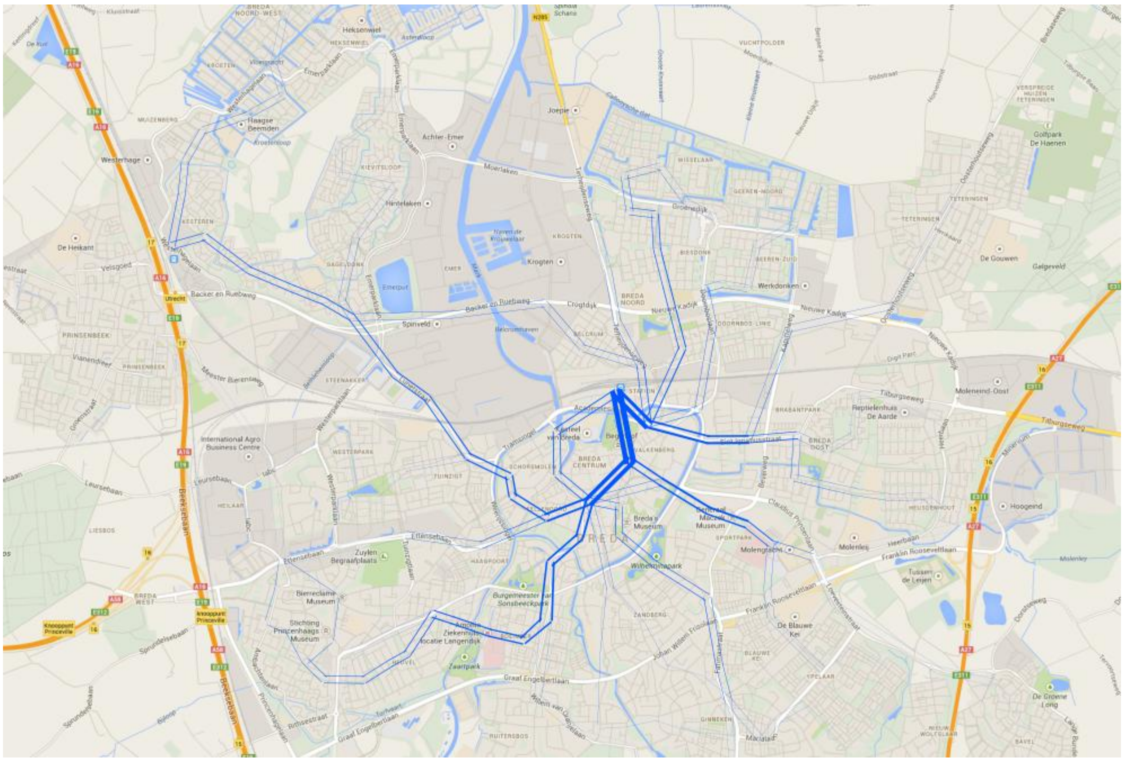


Figure 45: Assignment of the observed number of passengers on all weekdays in Breda in 2012

Figure 46 shows the assignment of the observed number of passengers per line segment (blue line). The observed number of boarders and alighters at the bus stops are used to create an origin-destination matrix using the calibrated distribution model for weekdays. In the figure this modelled number of passengers is presented with a red line. The figure shows the assignment of the modelled origin-destination matrix is almost the same as the observed origin- destination matrix, both in absolute and relative terms. The average difference between the observed and modelled number of passengers is approximately 5%. It can therefore be concluded that the distribution model fits very well.

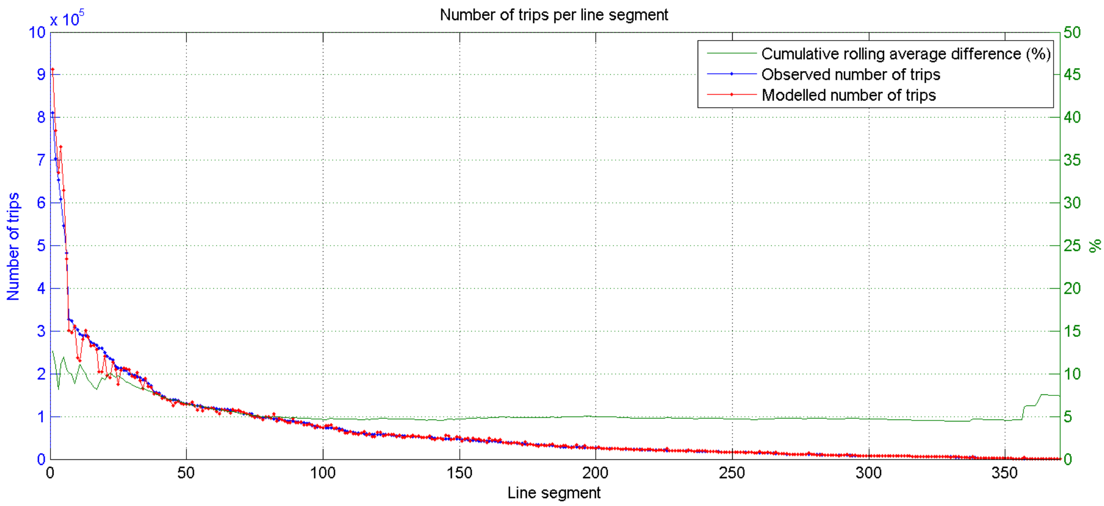


Figure 46: Number of passengers per line segment in Breda. The blue line represents the assignment of the observed origin-destination matrix while the red line represents the assignment of the origin destination matrix made with the distribution model and the observed number of boarders and alighters. The green line is the cumulative rolling average absolute difference between the observed and modelled number of passengers.

12 Validation

So far nine models have been built: six trip generation models for Breda and three trip distribution models for Breda. In order to check how well those models work, each of the nine models is validated for Tilburg. The trip generation models and the combination of the trip generation and distribution models have not yet been tested for Breda however, so these models are validated for Breda as well.

In the first section the generation models are validated for Breda and Tilburg. The second section analyses the distribution model for Tilburg. Finally, the third section validates the combination of the trip generation model and the trip distribution model. Only the models for the weekdays are validated in this chapter; for the validation of the models of Saturday and Sunday see appendices X to XII.

12.1 Validation trip generation models

The regression models are validated with two methods: performing a t-test and by summing the modelled and observed number of boarders and alighting at all bus stops in the model. The extended results are included in Appendix VIII. Equation 12.1 shows the formula of the paired t-test. The t-test is usually used to determine if two samples are significantly different or not.

$$t = \frac{\sum d}{\sqrt{\frac{n(\sum d^2)-(\sum d)^2}{n-1}}}$$

Equation 12.1

Where
d = observed - modelled number of boarders
n = number of pairs (number of bus stops)

Table 17 shows the validation results of the trip generation models. The six models are validated for Breda and Tilburg on two different manners. The table shows the weekday models, which are considered to be of most importance, perform very well with both validation methods. When the observed number of boarders at all bus stops is summed and the modelled number of boarders at all bus stops in Breda is summed, the difference is only 7%. For the number of alighters the difference is only 5% and for Tilburg the differences are only 1% and 5%.

According to the t-test both weekday models are acceptable for Breda and Tilburg, although the models of the Saturday and Sunday perform much worse. The model for the boarders on Saturday is rejected for both Breda and Tilburg, while the model for the number of alighters on Saturday is accepted for both municipalities. The models for Sunday are accepted and rejected twice.

Table 17: Difference between the observed and modelled number of boarders and alighters with data of Breda and Tilburg. Differences of 10% and smaller are coloured green.

		Breda		Tilburg	
		Difference of sum	Accepted by t-test?	Difference of sum	Accepted by t-test?
Boarders	Weekday	7%	✓	-1%	✓
	Saturday	-47%	X	-53%	X
	Sunday	28%	X	2%	✓
Alighters	Weekday	5%	✓	5%	✓
	Saturday	-26%	✓	-12%	✓
	Sunday	6%	✓	-32%	X

Figure 47 shows the map of Breda with the observed and modelled number of boarders at the bus stops on weekdays. The figure does not show all circles however. The largest two stops, 'Centraal Station' and 'Centrum' would cover multiple other bus stops. The number of observed and modelled number of boarders at the Central Station is exactly the same. The larger circles are most important, since those are the stops with the largest number of boarders. The larger bus stops east of the city centre appear to overestimate the number of boarders, while they are underestimated at the larger stops west of the city centre.

At most bus stops in Breda-Noord the number of boarders is underestimated, while they are overestimated at the smaller bus stops in the City Centre. Something remarkable happens in the northwest: the number of boarders at the bus stops west of the unbuilt land is underestimated by the model, while they are overestimated by the stops at the eastern side. The total modelled number of boarders in this district seems to be more or less equal to the observed number. The model estimates zero boarders at 48 bus stops.

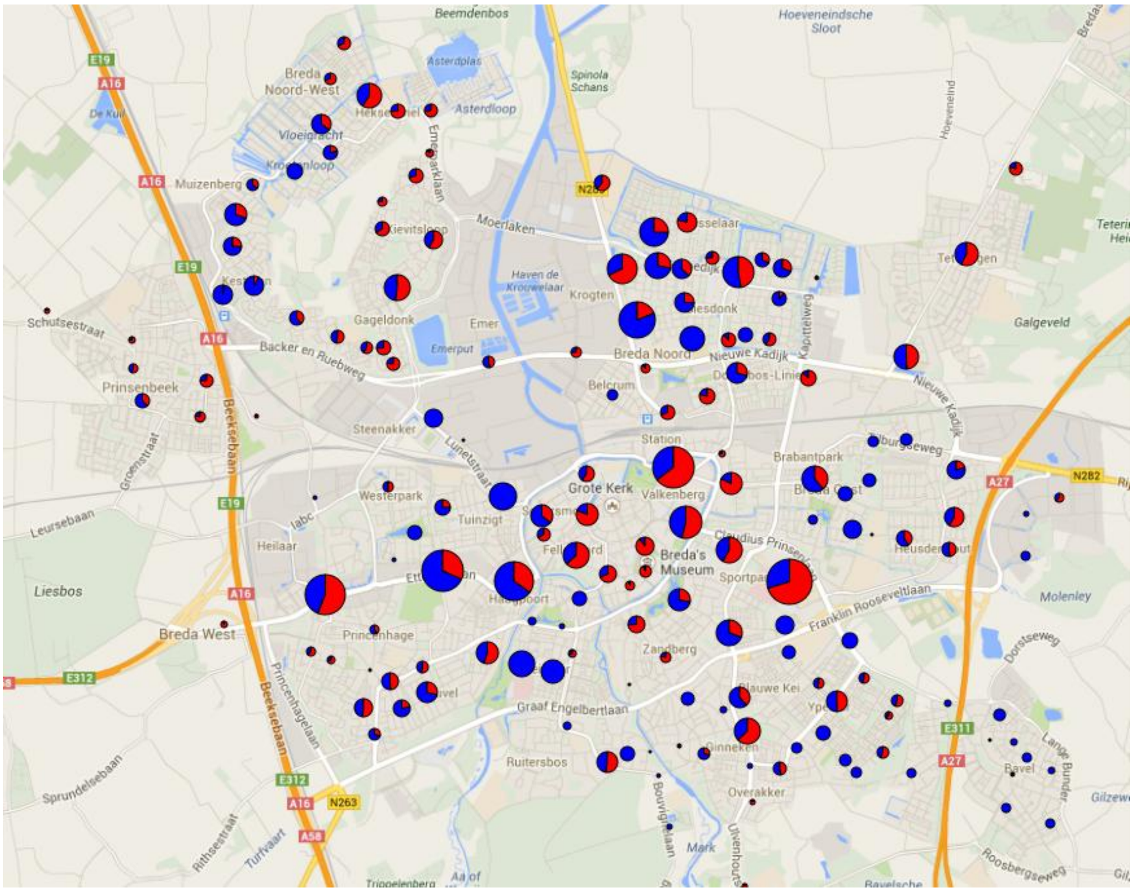


Figure 47: Map of Breda with a pie diagram for each bus stop. The area of each pie represents the observed number of boarders at the stop. The blue area represents the observed number of boarders and the red part the modelled number. The map does not show all stops, like the Central Station, since they would cover other bus stops.

Some circles are completely blue, which occurs when the model estimates a negative or zero number of boarders at the bus stop. This might occur when the combination of values of the variables accidentally result in a negative number of boarders, moreover the constant of the model is negative. The bus stops in Bavel for example are all completely blue, since the model predicts negative numbers of boarders. The bus stops have low values on the variables that increase the number of boarders (e.g. cars_total, OAD) and a high value on the variable that decreases the number of boarders (% woon).
Also some busier bus stops are allocated a negative number of boarders, like the two bus stops near the Amphia hospital, location Langendijk. It is likely that the largest part of the boarders at those bus stops is caused by the hospital. There is however no hospital-related variable included in the model.

Figure 48 shows the map of Tilburg with the observed and modelled number of boarders at the bus stops on weekdays. According to table 17 the observed and modelled number of boarders in Tilburg is approximately the same. This is however not caused by a good prediction of most of the bus stops, but by an overestimation of a part of the bus stops and an underestimation of the other bus stops. The trip generation model estimates zero boarders at 40% of the bus stops in Tilburg. At another 20% the estimated number of boarders is lower than the observed number. The other 40% of the bus stops compensate the underestimated number of boarders at the other stops.

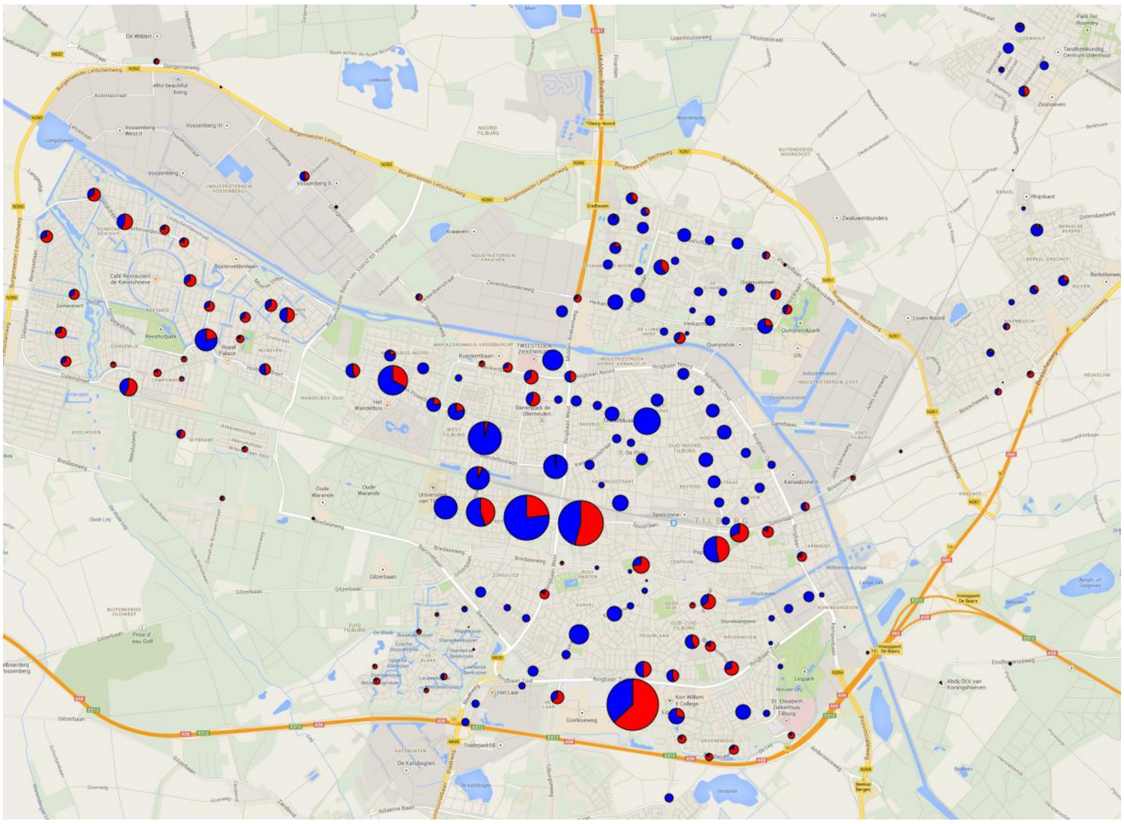


Figure 48: Map of Tilburg with a pie diagram for each bus stop. The area of each pie represents the observed number of boarders at the bus stop. The blue area represents the observed number of boarders and the red part the modelled number. The map does not show all stops, like the Central Station, since they would cover other bus stops.

12.2 Validation distribution models

The distribution models are validated via two methods, viz. via the trip length distributions and via the assignment of the origin-destination matrices. Since there are three distribution models (weekdays, Saturdays and Sundays), these three models are validated. The models of Saturday and Sunday are validated in Appendix IX.

Figure 49 shows the trip length distribution of the observed origin-destination matrix and the modelled origin-destination matrix using the observed number of boarders and alighters at the bus stops in Tilburg. The distribution seems to be quite good. The number of trips of 5 and 14 minutes are underestimated by the distribution model, while the number of trips 9 minutes is overestimated.

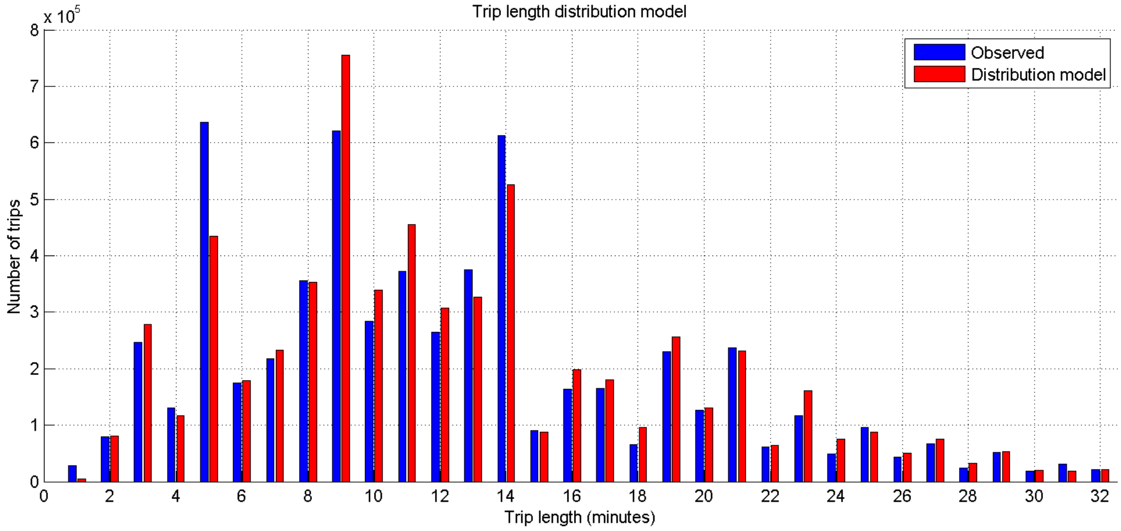


Figure 49: Trip length distribution of the observed origin-destination matrix and the modelled origin-destination matrix using observed number of boarders and alighters in Tilburg

Figure 50 shows the number of trips per line segment in Tilburg. The blue line represents the observed number of trips and the red line the modelled number of trips. For the modelled number of trips the observed number of boarders and alighters are used as input for the distribution model. It appears that the distribution model functions quite well given the fact that the urban structure of Breda and Tilburg is quite different. Both absolute and relative differences are fairly limited: the average difference between the observed and modelled number of passengers is approximately 6 to 12%.

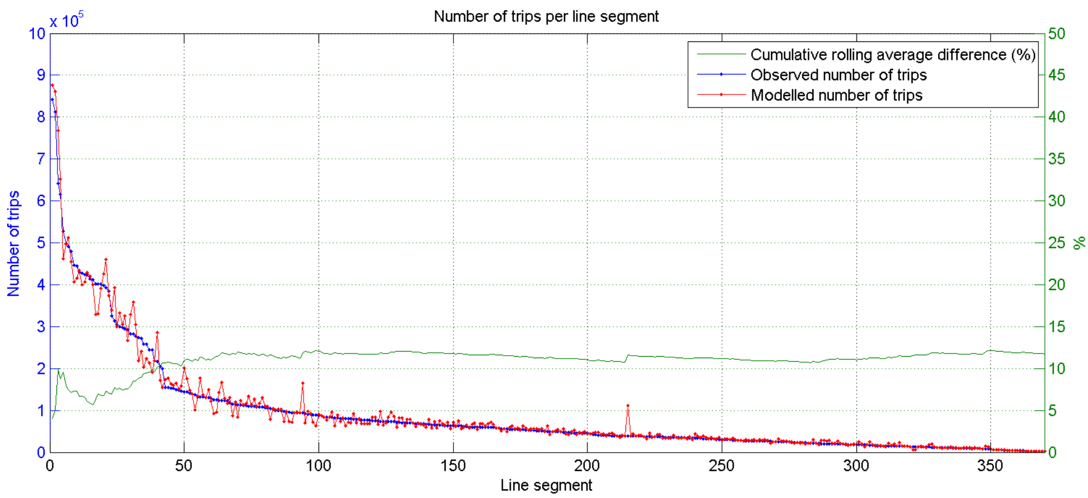


Figure 50: Number of passengers per line segment in Tilburg. The blue line represents the assignment of the origin-destination matrix as measured and the red line represents the origin-destination matrix created with the distribution model using the observed number of boarders and alighters. The green line gives the cumulative rolling average of the absolute difference.

12.2.1 Validation regression and distribution models

The previous two sections validated only the regression models or only the distribution models. This section validates both the regression models and the distribution models. The trip generation models are used to determine the number of boarders and alighters at each bus stop. The distribution models are then used to create origin-destination matrices using those numbers of boarders and alighters. The final step is to assign those passengers to the bus network. Each model is validated by the trip length distribution and the number of passengers per line segment.

The first section validates the models for Breda while the second section validates the models for Tilburg.

12.2.2 Validation Breda

Figure 51 shows the trip length distribution of Breda. The blue bars represent the observed number of trips per trip length and the red bars represent the modelled number of trips per trip length. The figure shows the models underestimate the number of trips with a trip length of 4 and 12 minutes, while it overestimates the number of trips of 6 and 13 minutes. The observed peak at 4 minutes is mainly caused by the number of trips between the bus stops 'Centrum' and 'Central Station'. The numbers of passengers that board and alight at those two stops are approximately 1,000 lower, which is exactly the underestimation of the model for this trip length.

The observed peak at 12 minutes is mainly caused by the large number of passengers between the bus stops 'Central Station' and 'Hogeschoollaan'. The total number of trips between these two stops is underestimated with 700 boarders by the model, causing the lower red bar at 12 minutes. The model overestimates the number of trips with a trip length of 13 minutes, which is mainly caused by an overestimation of the number of passengers between the Central Station and the stop 'Amphia Ziekenhuis Molengracht'.

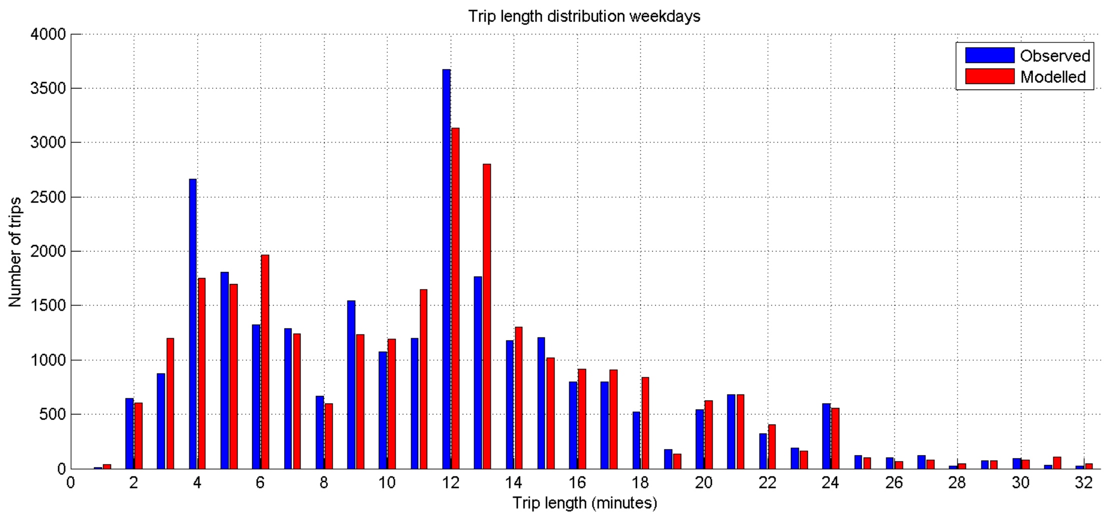


Figure 51: Observed and modelled trip length distribution

Figure 52 shows the observed and modelled number of trips at each line segment in Breda. The modelled number of trips at the busiest line segments appears to be quite similar to the observed number of trips, while the relative differences become larger at line segments with less than 1,000 trips. The figure shows that despite the relatively large differences at the individual segments, on average the modelled number of trips follows the observed number of boarders. The average difference at the line segments with at least 1,000 observed passengers is 16%.

It should be noted that the underestimation of the number of trips with a trip length of 4 minutes is visible on fewer line segments in figure 52 than the overestimation of the number of trips with a trip length of 13 minutes. The trips with a length of 4 minutes will in general use fewer line segments than the trips of 13 minutes. At first glance both figures might therefore give another image of the models.

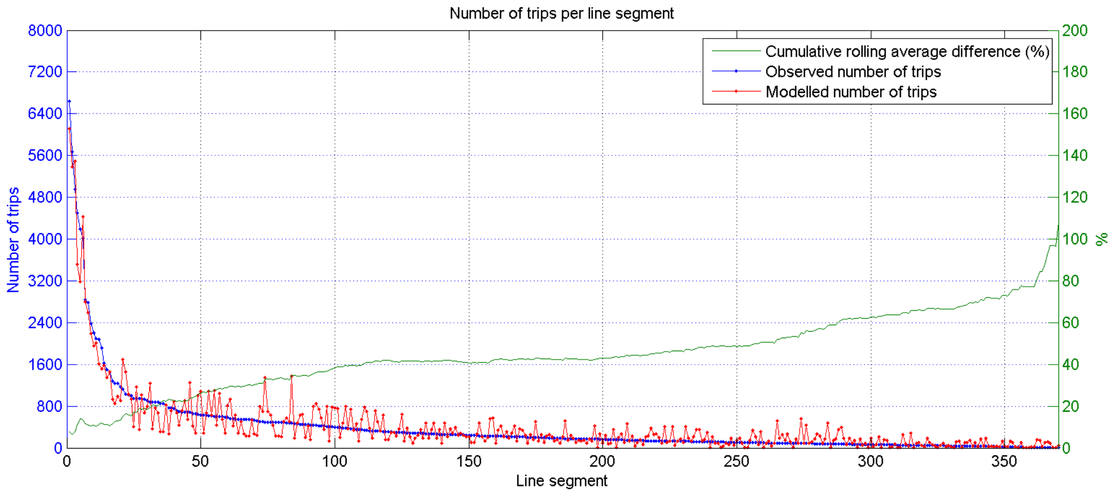


Figure 52: Observed and modelled number of passengers per line segment in Breda. The green line gives the cumulative rolling average of the absolute difference.

In order to analyze the deviation, figure 53 shows the map with the assigned number of observed and modelled trips. It appears the differences are quite limited. Once a line is overestimated at the outskirts of the city, the size of the difference usually increase as the line approaches the city centre.

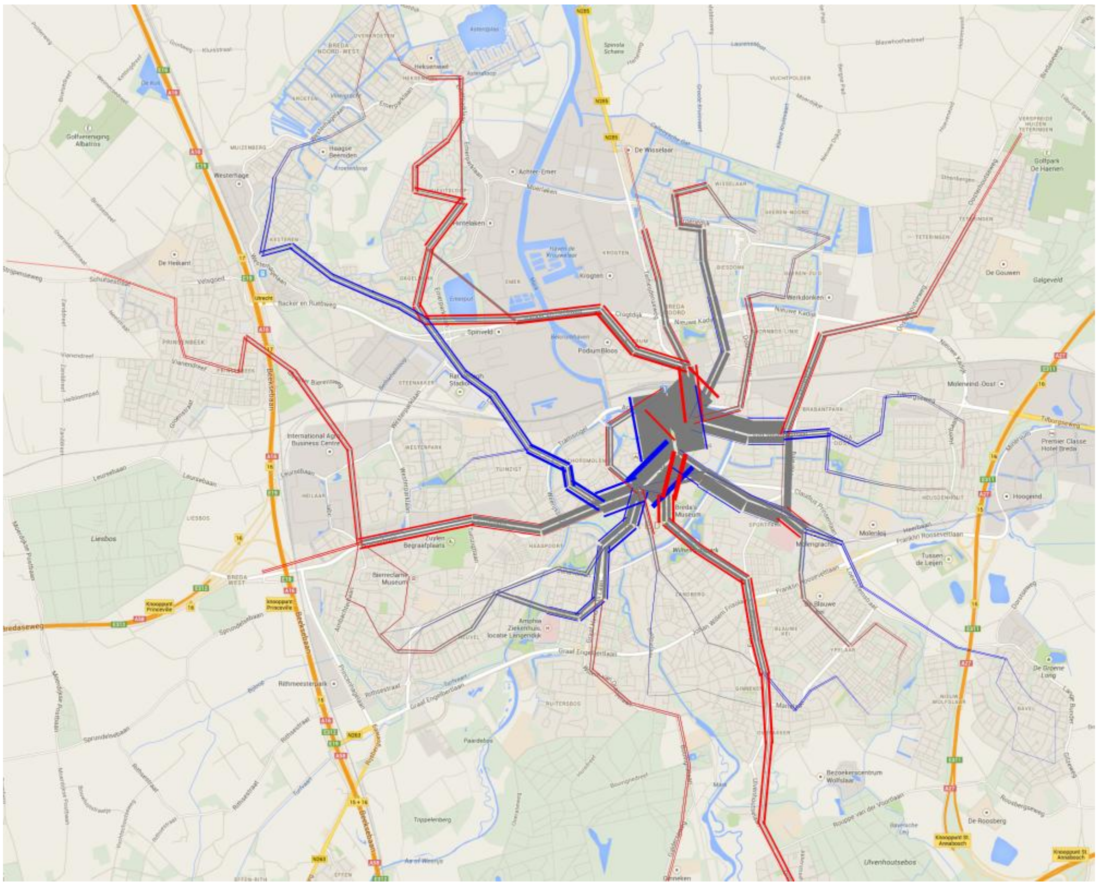


Figure 53: Observed and modelled number of passengers per line segment in Breda. Each direction at each line segment has two bars, the grey bar gives the smallest number of passengers (either observed or modelled). At line segments with a blue bar the observed number of passengers is the grey bar plus the blue bar and the modelled number the grey bar. At line segments with a red bar the modelled number of passengers is the grey bar plus the red bar, the observed number of passengers is the grey bar.

12.2.3 Validation Tilburg

Figure 54 shows the observed and modelled trip length distribution in Tilburg. The number of trips with a length of 9 minutes is underestimated by the models, while the trips of 13 and 14 minutes are overestimated.

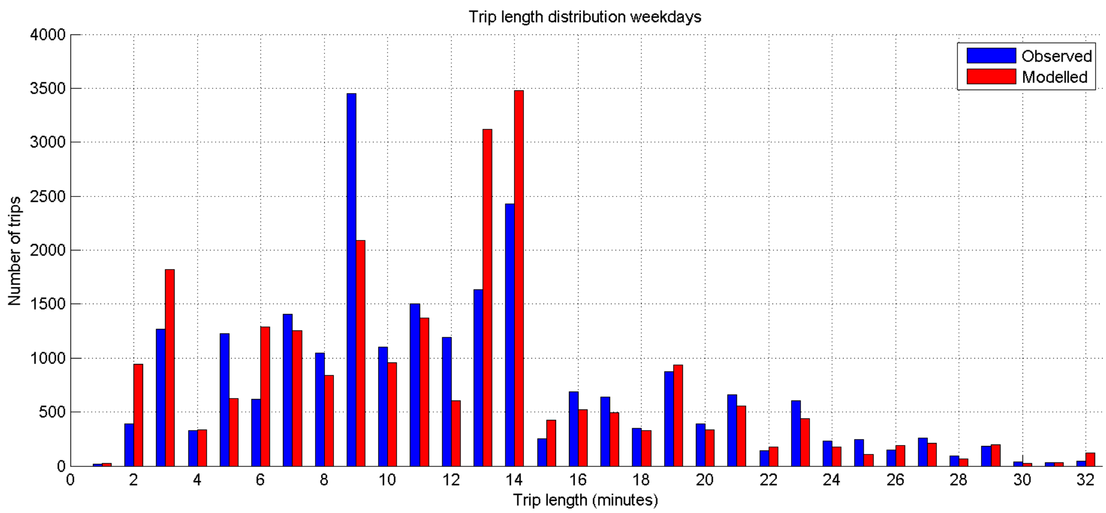


Figure 54: Observed and modelled trip length distribution in Tilburg on weekdays in 2012.

The observed and modelled number of passengers per line segment in Tilburg is shown in figure 55. The differences between the observed and modelled number of passengers are quite large. Only at a few of the busier line segments the difference between the modelled and observed number of passengers is limited. The average difference between the modelled and observed number of passengers at the segments with at least 1,000 observed passengers is 34%.

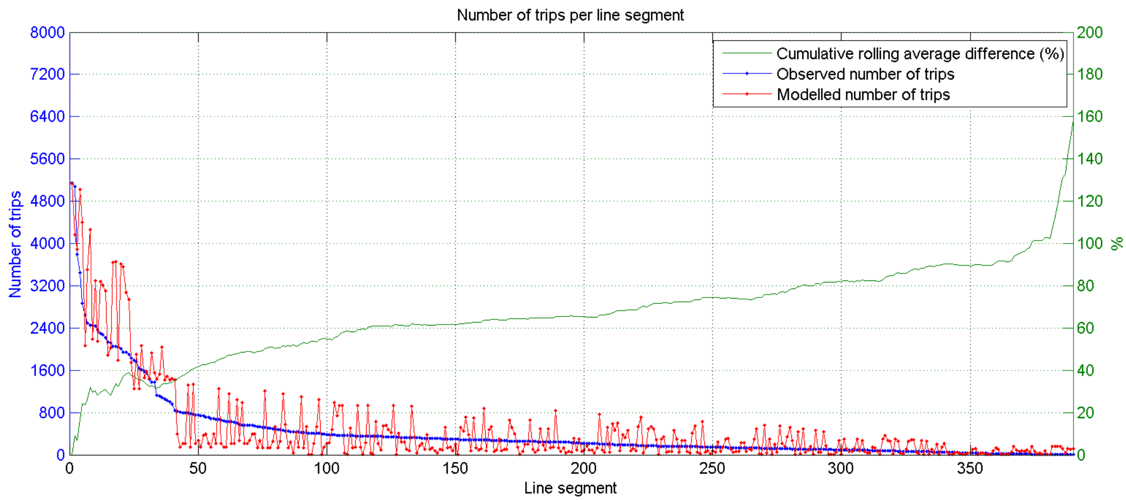


Figure 55: Observed and modelled number of passengers per line segment in Tilburg. The green line gives the cumulative rolling average of the absolute difference.

Figure 56 shows the map of Tilburg with the modelled number of passengers at the line segments. The number of passengers at the line segments south of the city centre is significantly overestimated, which is also visible in figure 55. The map in Figure 48 shows the cause of the overestimation: the number of boarders at the two stops at Stappegoor is significantly overestimated. The overestimation at these two stops affects all line segments between Stappegoor and the Central Station.

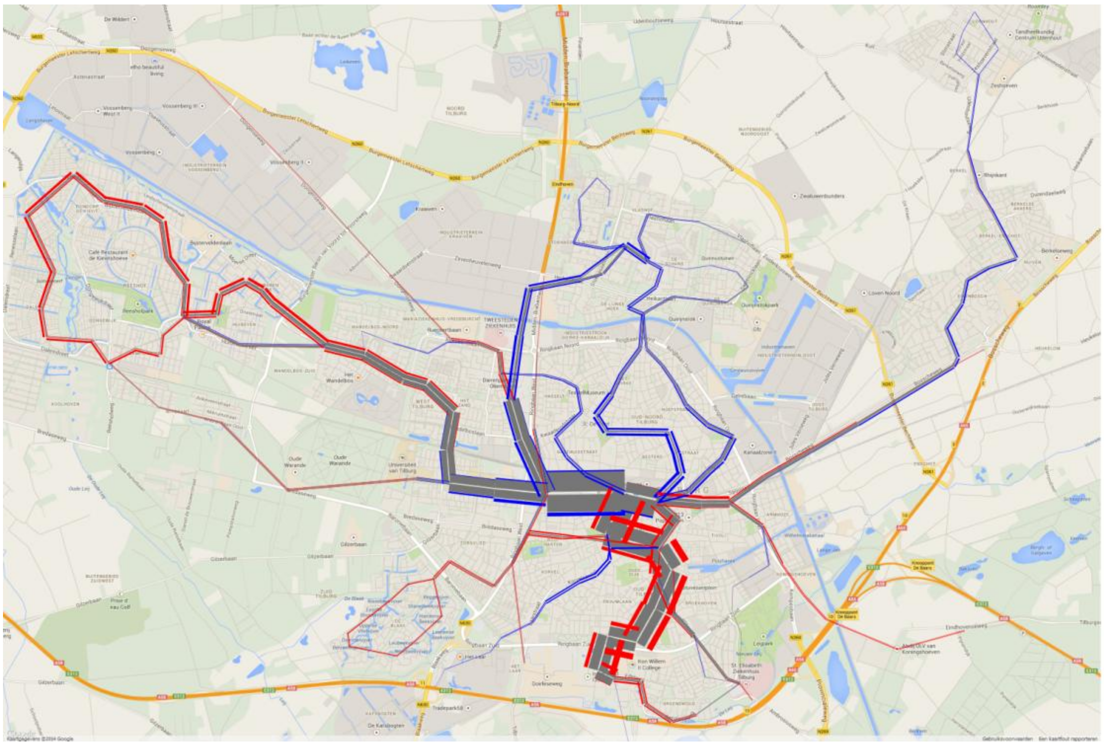


Figure 56: Observed and modelled number of passengers per line segment in Tilburg. Each direction at each line segment has two bars, the grey bar gives the smallest number of passengers (either observed or modelled). At line segments with a blue bar the observed number of passengers is the grey bar plus the blue bar and the modelled number the grey bar. At line segments with a red bar the modelled number of passengers is the grey bar plus the red bar, the observed number of passengers is the grey bar.

12.3 Conclusion

This chapter validated the developed trip generation and distribution models. The generation models for the weekdays perform quite well for Breda; although the number of boarders and alighters is different at most stops, the differences are rather limited. The total modelled number of boarders and alighters in Breda is almost equal to the observed numbers. The performance for Tilburg is worse; the differences between the modelled and observed number of boarders and alighters are larger and at more bus stops zero boarders and alighters are estimated.

The distribution models perform well at both Breda and Tilburg; they are well able to reproduce the observed trip length distributions. The models work however slightly better for Breda than for Tilburg.

The combination of the generation and distribution models gives a good result for Breda. The occupancy at some bus lines is overestimated and at other underestimated, but the differences between the observed and modelled number of passengers is limited. The differences in Tilburg are larger than in Breda which is mainly caused by the generation models. At the line segments with at least 1,000 observed passengers the average difference between the observed and modelled number of passengers is 16% in Breda and 34% in Tilburg.

13 Sensitivity analysis

The trip generation models were generated using the sample of 185 bus stops in Breda. It is possible the developed model is very sensitive for the selection of the bus stops; i.e. when a random subsample is selected which results in a model with very different regression coefficients. The sensitivity of several steps of the model development therefore needs to be assessed. The first step that needs to be assessed is the significant variables that were used to create the model. The samples of Breda and Tilburg are combined to identify which variables would have been selected for the generation model if the combined sample had been used. The second subject that needs to be assessed is the sensitivity of the regression coefficients. Random subsamples are selected from the 185 bus stops in Breda and regression models are estimated using the selected bus stops.

13.1 Sensitivity significant correlations

Table 18 shows the significant variables in Breda and Tilburg (same as in table 07) and the significant variables in the combined sample of bus stops. Almost each variable which is significant at either Breda or Tilburg is significant at the combined sample. The only three exceptions are the variables ‘opp_overig’, ‘% overig’ and ‘opp_winkels_horeca’. On the other hand there are five variables which are significant in the combined sample and not in Breda and Tilburg: ‘I_015’, ‘p_4564’, ‘a_havovwo’, ‘Type_woon_80’ and ‘Type_HO’.

Table 18: Significant variables in Breda, Tilburg and in the combined sample

Variable	Description	Breda (N = 185)	Tilburg (N = 191)	Breda & Tilburg (N = 376)
no_stops	Number of stops that can be reached without transfer	.817**	.713**	.763**
I_014	Number of people of age 0-15 years			-.126*
p_014	Percentage of population of age 0-15 years	-.156*		-.126*
p_1524	Percentage of population of age 15-24 years		.212**	.144*
p_2544	Percentage of population of age 25-44 years	.145*		.123**
P_4564	Percentage of population of age 45-64 years			-.115*
jobs	Number of jobs	.206**	.344**	.272**
P_LINH	Percentage of low income households	,152*		.127*
SUMStud(HO)	Number of students at intermediate vocational educational institutes, Universities of Applied Science and Universities (MBO,HBO,WO)	.153*		.134**
OAD	Average number of addresses within 100m of each address	.175*	.181*	.170**
A_havovwo	Average distance to nearest school with HAVO and/or VWO			-.123*
a_station	Distance to nearest train station	-.189**	-.146*	-.165**
opp_winkel	Floor area of addresses with a retail function	.171*		.110*
no_winkel	Number of addresses with a retail function	.154*		.115*

Variable	Description	Breda (N = 185)	Tilburg (N = 191)	Breda & Tilburg (N = 376)
opp_kantoor	Floor area of addresses with an office function	.281**	.473**	.378**
no_kantoor	Number of addresses with an office function	.369**	.290**	.333**
opp_bijeenkomst	Floor area of addresses with a social gathering function	.222**		.164**
no_bijeenkomst	Number of addresses with a social gathering function	.175*		.134**
opp_logies	Floor area of addresses with an accommodation function	.472**	.142*	.339**
no_logies	Number of addresses with an accommodation function	.385**	.316**	.351**
opp_overig	Floor area of addresses with an other function	.188*		
% woon	Percentage of floor area with residential function	-.162*	-.198**	-.179**
% kantoor	Percentage of floor area with an office function	.183*	.371**	.274**
% bijeenkomst	Percentage of floor area with a social gathering function		.165*	.145**
% logies	Percentage of floor area with an accommodation function	.386**		.157**
% overig	Percentage of floor area with an other function	.167*		
opp_winkels_horeca	Floor area with a retail, social gathering or accommodation function	.238**		.170**
p_winkels_horeca	Percentage of floor area with a retail, social gathering or accommodation function	.220**		
type_CS	Dummy variable for the bus stop at the Central Station	.954**	.981**	.967**
Type_woon_80	Dummy for bus stops with at least 80% floor area with residential function			-.124*
Type_HO	Dummy variable for the presence of a higher education institute (HBO and WO)			.121*
Type_shoppingcentre	Dummy variable for the stops with opp_winkels_horeca ≥ 5000 m^2	.272**	.163*	.206**

**. Correlation is significant at the 0.01 level (2-tailed)
*. Correlation is significant at the 0.05 level (2-tailed)

13.2 Sensitivity regression coefficients

From the sample of 185 bus stops in Breda, subsamples with the size of approximately a third of the bus stops are selected on a random basis. For each of the subsamples a regression model is estimated, using the variables which are included in the weekday generation model. These models are used to determine the sensitivity of the regression coefficients of the generation model. Of each of the variables the average regression coefficient as well as the standard deviation is obtained. Since the Central Station has a very large influence on the generation models, from the 10,000 models, the models of all samples which include the Central Station are separately analyzed as well.

Table 19 shows the results of the sensitivity analysis of the regression coefficients. It shows that the variation of the regression coefficients of some variables is rather high, while the variation is more limited at other variables. The standard deviation of the variable '% woon' is for example rather low in comparison to the standard deviation of the variable 'no_stops'. Especially the dummy variable for the presence of a higher education institute (Type_HO) is very sensitive for the selection of the bus stops. This is probably caused by the small number of bus stops with a higher education institute, while the number of boarders at those stops is quite high. The large number of boarders at those bus stops results in a high regression coefficient for this dummy variable. Considering the average value and the standard deviation, the negative value of the variable in the developed model becomes less odd.

The average value and the standard deviation of the regression coefficients of the subsamples with the Central Station are very similar to the sample with all 10,000 subsamples. The presence of the Central Station in the sample does not seem to have a large influence on the variables, except for the dummy variable for the presence of the Central Station. With all samples the average was about 3,200 with a standard deviation which is almost 50% higher than the average. With the samples with the Central Station the average is approximately 9,600 with a standard deviation of only 669.

Table 19: The regression coefficients of the developed generation model for the boarders on weekdays and the average and standard deviation of the regression coefficients of the models generated with random subsamples of the bus stops in Breda.

Variable	Developed model	10,000 subsamples		Subsamples with Central Station (3,346 subsamples)	
		Average	Standard deviation	Average	Standard deviation
Constant	-128.654	-75.078	83.586	-75.656	83.364
No_stops	3.721	2.834	1.634	2.811	1.621
P_LINH	-0.612	0.023	1.036	0.042	1.014
Cars_total	0.135	0.110	0.098	0.110	0.096
SUMStud (HO)	0.155	0.100	0.443	0.111	0.349
OAD	-0.044	-0.024	0.027	-0.024	0.027
Opp_kantoor	0.002	0.001	0.005	0.001	0.005
% woon	1.074	0.373	0.885	0.387	0.881
Opp_winkel_horeca	0.019	0.009	0.012	0.009	0.012
Type_CS	9281.374	3229.958	4573.096	9658.008	669.337
Type_HO	-152.157	32.108	1461.632	2.146	1166.398

13.3 Conclusion

Two sensitivity analyses were performed: the influence of combining the bus stops in Breda and Tilburg on the significant variables and the influence of randomly selected subsamples on the regression coefficients of the trip generation model for the number of boarders on weekdays. If the bus stops in Breda and Tilburg are used separately to determine the significant variables, this results in 25 significant variables in Breda and 14 in Tilburg. Combining the bus stops in Breda and Tilburg in one sample results in 29 significant variables. It can therefore be concluded that the sample of bus stops has a large influence on the significance of the explanatory variables with the daily number of boarders.

The regression coefficients of the models are also sensitive for the sample which is used to create the regression models. If a random sample of approximately a third of the bus stops in Breda is selected 10,000 times, the regression coefficients of most variables have a large standard deviation in comparison to the average regression coefficient. Especially the dummy variables for the presence of the Central Station and the presence of higher education institutes have a large standard deviation. The regression coefficients are sensitive for the sample of bus stops which is used to create the generation models for.

14 Conclusion

The study was set out to explore the factors that determine the demand for bus transport and to use those factors to develop a model that is able to forecast the demand for bus transport. This chapter answers the research questions and the main research question. The first section summarizes the results of the thesis by answering the sub research questions. The second section gives the main conclusion of this thesis.

14.1 Answering sub research questions

1. *How can bus demand be modelled according to the theory and what are the requirements of the model for this study?*

There are two methods which are suitable to estimate the number of bus passengers in a bus network, which are the traditional four step model and a direct demand model. The four step model consists of the stages trip generation, trip distribution, modal split and trip assignment, with a separate model for each stage. Direct demand models combine the stages generation, distribution and modal split and determine the number of bus passengers on bus stop pairs in one model. Since the four step model consists of separate models, it is considered to be more flexible and therewith more appropriate.

The trip generation stage can be modelled through regression models. In order to develop the regression models the following subjects were addressed:

1. Factors that determine the demand for bus transport
2. Variation of the number of bus passengers over time
3. The number of boarders and alighters at each bus stop
4. The catchment area of bus stops

The trip distribution stage can be modelled with distribution models. The distribution models need to be calibrated to approach the observed distribution as well as possible. Various distribution functions exist that can be calibrated, but three functions are chosen to calibrate and select the one that is most suitable.

The modal split stage is not considered, since the study only regards bus transport. For the trip assignment stage just the shortest route is considered (measured by number of bus stops).

2. *Which factors determine the demand for bus transport according to the literature?*

In the literature various variables are identified which can be categorized in the categories population characteristics, spatial characteristics, bus service characteristics and trip specific characteristics. The variables are presented in figure 57.

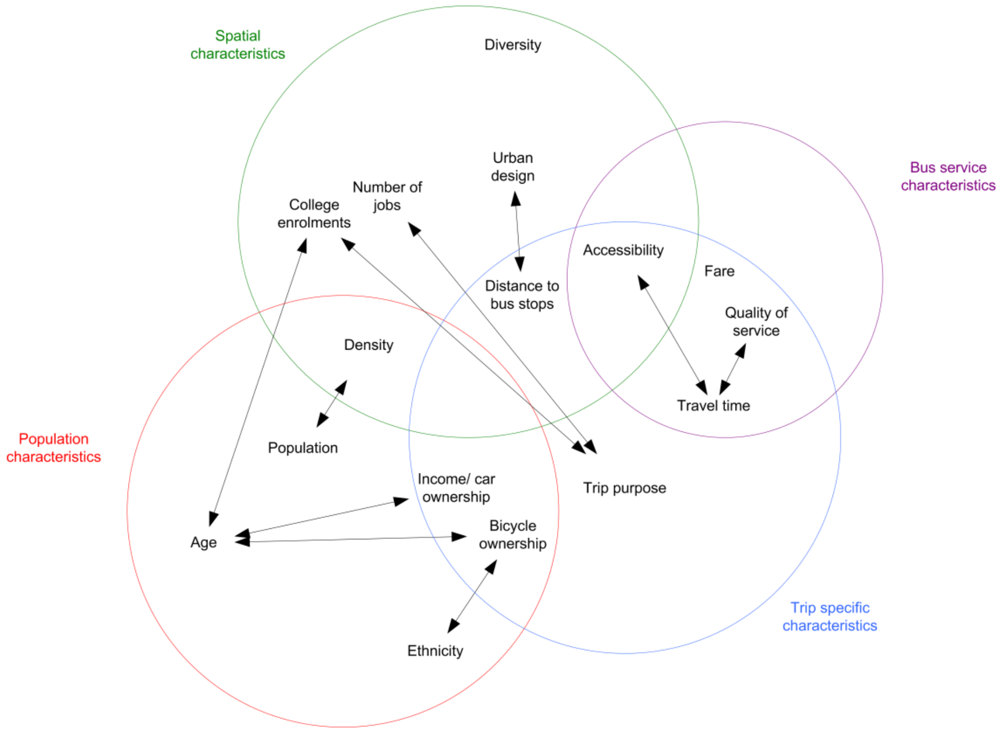


Figure 57: factors that determine the demand for bus transport according to literature

3. *Which indicators can represent the factors that were found in the second research question and which data is available about these indicators?*

Data from the CBS, DUO, Veolia Transport, municipality of Breda, OpenStreetMap and Kadaster is used as indicators for the demand influencing factors found in literature. Appendix II contains the variables that represent those factors. The population characteristics are mainly represented by the number and percentage of the population per age category, number and percentage of (non-)western foreigners, income variables and car related variables. The spatial characteristics are represented by variables related to the floor area and number of addresses with different functions and number of pupils/students at educational institutions. Bus service characteristics are represented by an accessibility measure, namely the number of stops that can be reached from a bus stop without transfer. No variables are included that can solely be categorized in the trip specific characteristics, but they all overlap with the other categories.

4. *How does the number of bus passengers vary over time?*

The percentage of boarders during each month is approximately 7.7-10.8% of the yearly number of boarders, except for July and August with only 3.9 and 5.3%. Of all bus passengers, 90% travel on weekdays with an equal distribution over the weekdays (Monday to Friday).

Two third of the other 10% travel on Saturdays and the remaining 3.7% travels on Sundays. Holidays have a quite large impact on the number of bus passengers; during the holiday weeks in 2012 the weekly number of boarders is about half of the average number of boarders. Although there is a morning peak on weekdays between 8:00 and 9:00h, during a period of four hours in the afternoon the hourly number of passengers is equal to or higher than during the morning peak.

It seems that there are some weather characteristics that influence the number of boarders. The temperature and sun hours seems to have a negative correlation with the number of boarders, while the correlation of precipitation is positive.

5. *How do the developed trip generation models look like?*

Table 20 shows the variables which are included in the six trip generation models.

Table 20: Variables of the six trip generation models

Variable	Boarders			Alighters		
	weekday	Saturday	Sunday	Weekday	Saturday	Sunday
Number of stops without transfer	✓	✓	✓	✓	✓	✓
Dummy for Central station	✓	✓	✓	✓	✓	✓
Percentage low income households	✓	✓	✓	✓	✓	✓
Floor area of offices	✓	✓	✓	✓	✓	✓
Floor area retail, social gathering and accommodation	✓	✓	✓	✓	✓	✓
Address density	✓	✓	✓	✓	✓	✓
Total number of cars	✓	✓	✓	✓	✓	✓
% of floor area with residential function	✓			✓		
Number of students at higher education institutes	✓			✓		
Dummy for University (of Applied Science)	✓			✓		
Dummy for shopping centre (>5000 m^2 floor area retail, social gathering and accommodation)		✓	✓		✓	✓

The following formula shows the trip generation model to determine the number of boarders on a regular weekday:

$$Boarders_{weekday} = -128.654 + 3.721x_{no\_stops} - 0.612x_{P\_LINH} + 0.135x_{Cars\_total} + 0.155x_{SUMStud\ (HO)} - 0.044x_{OAD} + 0.002x_{Opp\_kantoor} + 1.074x_{\%woon} + 0.019x_{Opp\_winkel\_horeca} + 9281.374x_{Type\_CS} - 152.157x_{Type\_HO}$$

6. *How do the trip distribution models look like?*

Three distribution functions were calibrated: the exponential, Tanner and top-lognormal function. The costs are represented by the travel time in minutes between the bus stops. The top-lognormal function appeared to comply slightly better than the Tanner and much better than the exponential function. Three distribution models were created using the top-lognormal function: one for weekdays (equation 14.1), one for Saturdays (equation 14.2) and one for Saturdays (equation 14.3), although the models for Saturday and Sunday are very much the same.

$$f(c_{ij}) = e^{-0.705(\log(c_{ij}/7.701))^2} \qquad \textit{Equation 14.1}$$

$$f(c_{ij}) = e^{-0.827(\log(c_{ij}/8.662))^2} \qquad \textit{Equation 14.2}$$

$$f(c_{ij}) = e^{-0.841(\log(c_{ij}/9.031))^2} \qquad \textit{Equation 14.3}$$

7. *How valid are the developed models?*

The six trip generation models are validated for both Breda and Tilburg. Table 21 shows the difference between the modelled and observed number of boarders or alighters, summed over all bus stops in the municipality. The table shows the models for the weekdays are all accepted by the t-test and also the differences are very limited (max. 7%). The models for the Saturday and Sunday perform ambiguous; although the differences of the sums are quite large for the Saturday-models, the t-test accepts the models in two cases.

The good results are mainly caused by an overestimation at some stops which is compensated by an underestimation at other stops. On average, or in total, the number of boarders and alighters becomes similar to the observed ones.

Table 21: Difference between the observed and modelled number of boarders and alighters, summed over all bus stops in the municipality.

		Breda		Tilburg	
		Difference of sum	Accepted by t-test?	Difference of sum	Accepted by t-test?
Boarders	Weekday	7%	✓	-1%	✓
	Saturday	-47%	X	-53%	X
	Sunday	28%	X	2%	✓
Alighters	Weekday	5%	✓	5%	✓
	Saturday	-26%	✓	-12%	✓
	Sunday	6%	✓	-32%	X

Three different distribution functions (exponential, Tanner and top log-normal) are calibrated for three time periods (weekdays, Saturdays and Sundays). The distribution function that is best able to reproduce the observed trip length distributions in Breda is the top log-normal function. It has two parameters which are calibrated. It almost reproduces the trip length distributions perfectly. Applying the distribution functions on Tilburg gives a result which is slightly worse than for Breda, but matches the observed trip length distribution very well.

The combination of the trip generation and distribution models gives a good result for Breda. The average difference between the observed and modelled number of passengers at the line segments with more than 1,000 passengers is only 16% and increases to 40% at the busiest half of the line segments. The performance of the combined models for Tilburg is worse. The average difference between the observed and modelled number of passengers at the line segments with more than 1,000 passengers is 34% and increases to 66% at the busiest half of the line segments.

The developed models are quite sensitive for the sample of bus stops which is used to develop the models. Combining the bus stops of Breda and Tilburg results in 29 significant variables, while the sample of Breda gives 25 significant variables and the sample of Tilburg 14. Also the regression coefficients are sensitive for the subsample of bus stops which are selected. For most variables the standard deviation is larger than the average value, which indicates that the variation is large. The dummy variables for the presence of the Central Station and the presence of higher education institutes are most sensitive for the sample of bus stops.

14.2 Main conclusion

The main research question of the study is:

Which factors determine the demand for bus transport and how can these factors be implemented in an explanatory model that is able to predict the number of boarders and alighters at a bus stop and used to predict the spatial distribution of those passengers?

It appears that the most important factors that determine the demand for bus transport are the spatial factors, like the presence of the central train station, offices, shops and higher education institutes. In other words: facilities that attract large numbers of people. In bus planning it is common practice to focus first on those locations and then on the other locations. The results of this study support this approach. Also the accessibility of bus stops is important, which was measured in this study by the number of stops that can be reached from a stop without transferring. Other factors are the address density, percentage of low income households and total number of cars.

These factors are implemented in six trip generation models which are developed for Breda (boarders: weekdays, Saturdays and Sundays and alighters: weekdays, Saturdays and Sundays). The weekday models are very well able to predict the number of boarders and alighters at most of the busier bus stops in Breda, but the performance of the weekday models is much worse when applied to Tilburg. Apparently a bus demand model developed for one area cannot simply be applied in another area, since very specific combinations of circumstances that have a large influence on the model. Also the sample of bus stops which is used has a large influence on the resulting model. The performance of the models for Saturdays and Sundays is worse than the weekday models at both Breda and Tilburg.

The distribution functions (weekdays, Saturdays and Sundays) which are calibrated for Breda are very well able to reproduce the observed trip length distribution in Breda. Applying the distribution models on Tilburg gives a very good result.

The combined models (trip generation model, distribution models and assignment) perform quite well for Breda. The number of boarders and alighters which are overestimated at one bus stop is compensated by the next bus stop, resulting in an occupancy which is quite similar to the observed occupancy. Although the generation models perform worse for Tilburg, the deviations are mitigated by the distribution and assignment.

14.3 Recommendations for further research

Repeat study with larger study area

The study area of this research was restricted to the municipalities of Breda and Tilburg. It appeared that very specific combinations of circumstances have a large influence on the trip generation models. More research is necessary to identify the variables which factors in general determine the demand for bus transport. This study mainly examined the factors that determine the demand for bus transport in Breda and Tilburg rather than bus transport in general. When the bus stops in more municipalities are included in the analysis it might be possible to determine why a variable is correlated with the number of bus passengers in one municipality and not in the other. If more cities are included in the study, the dummy variable for the presence of the Central Station could be replaced with a variable which depends on the number of train passengers at the station.

Repeat study with multiple years

This study only considered the year 2012 and studied therewith a snapshot in time. The development of the number of bus passengers over the years in combination with spatial and socio-economic changes might give valuable information about the relations between spatial characteristics and the number of bus passengers.

Use more detailed chipcard data

The chipcard data contains more detailed information than used in this study. For example the bus lines were not analyzed separately and for the distribution models data aggregated over the year was used. The data for the separate bus lines is however available and the chipcard data also contains the travel product that people use to pay for their bus trip. Also the number of passengers between each bus stop pair is known on daily level. Using this detailed data offers more insight in the travel movements and could contribute to solve shortcomings of the developed models.

Examine the effect of other explanatory variables

Although a large number of explanatory variables were considered in this study, there are some variables that might be relevant which are not included. Some of these variables are for example the directness of the bus route, possibly in relation to competing modes (e.g. bicycle routes) or the travel time rate between bus and car or bicycle. Also the frequency of bus services and the accessibility of the bus stops themselves could be important to include.

More detailed analysis of catchment area

The catchment areas of the bus stops are represented by circles with a radius of 200m. This radius was chosen since it gave the best fitted regression model in Breda. It is not known to which extent this is the optimal distance for other areas than Breda or that in other cities other catchment areas would give the best fit. Apart from this, the catchment area of bus stops is probably better represented by distance- or travel time isochrones rather than circles (Landex, Hansen & Andersen, 2006).

15 Literature

Bakker, P., Zwaneveld, P. J., Berveling, J., Korteweg, J. A., & Visser, S. (2009). *Het belang van openbaar vervoer: de maatschappelijke effecten op een rij*. Centraal Planbureau.

Balcombe, R., Mackett, R., Paulley, N., Preston, J., Shires, J., Titheridge, H., ... & White, P. (2004). The demand for public transport: a practical guide. Retrieved from http://www.demandforpublictransport.co.uk/TRL593.pdf

Bechdolt Jr, B. V., & Williams, M. (1980). Demand for bus transit in US urbanized areas. *Regional Science Perspective*, 3-14.

Breda (2002). *Detailhandelsnota Gemeente Breda*. Retrieved from http://www.breda.nl/system/files/Detailhandelsnota_April021.pdf

BredaVandaag.nl (2012). *Elke zondag koopzondag in Breda*. Retrieved from http://www.bredavandaag.nl/nieuws/economie/2012-10-01/elke-zondag-koopzondag-breda

Campoli, J. & Maclean, A. S. (2002). *Visualizing density: lower density catalog images, 0.5-10.5 units per acre* (Lincoln Institute of Land Policy Working Paper).

Centraal Bureau voor de Statistiek (2013). *Toelichting Wijk- en Buurtkaart 2012* [dataset and codebook]. Retrieved from http://www.cbs.nl/nl-NL/menu/themas/dossiers/nederland-regionaal/publicaties/geografische-data/archief/2013/default.htm

Centraal Bureau voor de Statistiek (2013). *Statistische gegevens per vierkant* [datasets and codebook]. Retrieved from http://www.cbs.nl/nl-NL/menu/themas/dossiers/nederland-regionaal/publicaties/geografische-data/archief/2013/default.htm

Centrum Vernieuwing Openbaar Vervoer (2004). Introductie chipkaart en concessieverlening in het stads- en streekvervoer. Rotterdam, the Netherlands: Centrum Vernieuwing Openbaar Vervoer: Author

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

Dienst Uitvoering Onderwijs (n.d.). Open databestanden. Retrieved from http://duo.nl/organisatie/open_onderwijsdata/databestanden/default.asp

Directoraat Generaal van het Verkeer (1981). NVS handleiding. Retrieved from publicaties.minienm.nl/download-bijlage/9477/nvs-handleiding.pdf

Domencich, T. A., & McFadden, D. (1975). *Urban Travel Demand: A Behavioral Analysis*. North-Holland Publishing Company, Amsterdam.

Eboli, L., & Mazzulla, G. (2011). A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. *Transport Policy*, *18*(1), 172-181.

Ewing, R., Meakins, G., Bjarnson, G. & Hilton, H. (2011). Transportation and land use. In A. L. Dannenberg et al. (Eds.), *Making healthy places: designing and building for health, well-being, and sustainability.* Island Press.

Fietsersbond (2011*). Fietsen in cijfers*. Retrieved from http://media.fietsersbond.nl.s3.amazonaws.com/fietsenincijfers/fietsen_in_cijfers_web.pdf

Fryslân (n.d.). Iepenbier ferfier in Fryslân: Sifers twadde healjier en jier 2013. Retrieved from www.fryslan.nl/2767/halfjaar-en-jaarcijfers/files/%5B92%5Djaar-%20en%20halfjaarcijfers%202013.pdf

Goudappel Coffeng (n.d.). De waaier van Brogt. Retrieved from http://issuu.com/gcoffeng/docs/waaier-van-brogt

Grotenhuis, J. W., Wiegmans, B. W., & Rietveld, P. (2007). The desired quality of integrated multimodal travel information in public transport: Customer needs for time and effort savings. *Transport Policy*, *14*(1), 27-38.

Guo, Z. (2003). Assessment of the transfer penalty for transit trips in downtown Boston (Master's thesis). Retrieved from http://dspace.mit.edu/bitstream/handle/1721.1/66798/52971233.pdf

Harms, L. (2006). Anders onderweg. De mobiliteit van allochtonen en autochtonen vergeleken. The Hague: Sociaal en Cultureel Planbureau.

Kadaster (2013). Basisregistraties adressen en gebouwen (BAG) 2013 [datasets]. Retrieved from http://geoplaza.vu.nl/cdm/ref/collection/gpz/id/258

Kennisplatform Verkeer en Vervoer (2013). Regionaal openbaar vervoer per 1 januari 2014. Retrieved from http://www.crow.nl/publicaties/poster-regionaal-openbaar-vervoer-per-1-januar-%281%29

Klok, E. J. (2010). *Determinants of the demand for bus transport: a model to determine the number of bus boardings in a neighbourhood* (Master's thesis). Retrieved from http://essay.utwente.nl/59703/1/MA_thesis_E_Klok.pdf

Koninklijk Nederlands Meteorologisch Instituut (2013). *Daggegevens van het weer in Nederland* [datafile]. Retrieved from http://www.knmi.nl/klimatologie/daggegevens/selectie.cgi

Landex, A., Hansen, S., & Andersen, J. L. E. (2006). Examination of catchment areas for public transport. *Traffic days*.

Lee, R., & Cervero, R. (2007). Research basis for proposed criteria of the TOD Housing Program. The effect of housing near transit stations on vehicle trip rates and transit trip generation. A review of available evidence. California: University of California, Institute of Urban and Regional Development.

Limtanakool, N., Dijst, M., & Schwanen, T. (2006). The influence of socioeconomic characteristics, land use and travel time considerations on mode choice for medium-and longer-distance trips. *Journal of Transport Geography*, *14*(5), 327-341.

Mensonides, F. (2008). Bus station Breda [picture]. Retrieved from http://www.fransmensonides.nl/volans.htm

Omnitrans (2011). *Transport modelling in OmniTRANS* [lecture pack]. Retrieved from http://www.omnitrans-international.com/en/products/training/lecturer-pack

Omroep Brabant (2007). *Veolia koopt 200 nieuwe bussen*. Retrieved from http://www.omroepbrabant.nl/?news/80194672/Veolia+koopt+200+nieuwe+bussen.aspx

Ortúzar, J. de D. & Willumsen, L. G. (2001). *Modelling transport*. Chicester, England: Wiley.

Paulley, N., Balcombe, R., Mackett, R., Titheridge, H., Preston, J., Wardman, M., ... & White, P. (2006). The demand for public transport: The effects of fares, quality of service, income and car ownership. *Transport Policy*, *13*(4), 295-306.

Pommer, E., Van Kempen, H. & Eggink, E. (2008). De staat van de publieke dienst. The Hague, the Netherlands: Sociaal en Cultureel Planbureau. Retrieved from http://www.scp.nl/Publicaties/Alle_publicaties/Publicaties_2008/De_staat_van_de_publieke_dienst

Reizigersoverleg Brabant (n.d.). *Bussen*. Retrieved from http://www.reizigersoverlegbrabant.nl/brabants-openbaar-vervoer/toegankelijkheid/bussen

Rojo, M., Gonzalo-Orden, H., & Ibeas, Á. (2012). Relationship between service quality and demand for inter-urban buses. *Transportation Research Part A: Policy and Practice*.

Schmenner, R. W. (1976). The demand for urban bus transit: A route-by-route analysis. *Journal of Transport Economics and Policy*, 68-86.

Shrewsbury, J. S. (2012). *Calibration of trip distribution by generalised linear models* (Doctoral dissertation), University of Canterbury, New Zealand. Retrieved from http://ir.canterbury.ac.nz/bitstream/10092/7685/5/Thesis_fulltext.pdf

Steg, L., & Kalfs, N. (2000). Altijd weer die auto. Sociaal- en gedragswetenschappelijk onderzoek en het verkeers- en vervoersbeleid. Den Haag, Sociaal en Cultureel Planbureau.

Taylor, B. D., Miller, D., Iseki, H., & Fink, C. (2009). Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. *Transportation Research Part A: Policy and Practice*, *43*(1), 60-77.

Tertoolen, G. & Van Uum, J. (2004). Pimp my bus! Het customizen van eenheidsworst tot massclusivity. Retrieved from http://www.crow.nl/documents/kpvv-kennisdocumenten/artikel-pimp-my-bus-het-customizen-van-eenheidswor.aspx

Thomas, T., Jaarsma, R., & Tutert, B. (2009, January). Temporal variations of bicycle demand in the Netherlands: The influence of weather on cycling. In *Transportation Research Board 88th Annual Meeting 11-15 January 2009; Washington DC*.

Tilburg (2013). Visie detailhandel gemeente Tilburg. Retrieved from https://drive.google.com/?urp=http://tilburgers.nl/weerstand-ondernemers-detailh&authuser=0#folders/0B1_OjbLnF1KDejlJV3VlX1JTQVE

Van der Blij, F., Veger, J. & Slebos, C. (2010). HOV op loopafstand: het invloedsgebied van HOV-haltes. Retrieved from www.cvs-congres.nl/cvspdfdocs/cvs10_043.pdf

Van der Zwan, A. (2011). De invoering van de OV chipkaart in Nederland. Retrieved from http://www.mobiliteitnederland.nl/pdf/Rapport_De_invoering_van_de_OV_chipkaart_in_Nederland_.pdf

Veitch Lister Consulting (2008). Zenith model establishment and validation report. Retrieved from http://www.transport.vic.gov.au/__data/assets/pdf_file/0006/34269/EWLNA-VeitchListerConsulting-Zenith_Model_Establishment_and_Validation_Report.pdf

Veolia-Transport (2013). Lijnennetkaart Breda [map]. Retrieved from https://www.veolia-transport.nl/brabant/_meta/documenten/dienstregeling-2013/Veolia_Breda_febr_13_v3_los.pdf

Veolia Transport (2013). Lijnennetkaart Tilburg [map]. Retrieved from https://www.veolia-transport.nl/brabant/_meta/documenten/dienstregeling-2013/Veolia_Tilburg_12-13_v3i.pdf

Wardman, M. (2001). A review of British evidence on time and service quality valuations. *Transportation Research Part E: Logistics and Transportation Review*, *37*(2), 107-128.

16 List of appendices

Appendix I Detailed data description

This appendix contains tables with the detailed data description of the data sources used from the Centraal Bureau voor de Statistiek as discussed in subsection 5.1.1. Table 22 contains the data at neighbourhood level for 2012, table 23 contains this data for 2011, table 24 describes the data available in squares of 500m and table 25 describes the data available at squares of 100m. The special codes are -99 999 (missing data), -99 998 (zero), -99 997 (confidential) and 0 (smaller than half the chosen unit).

Table 22: Data description of useable data from CBS at neighbourhood level for 2012

Attribute name	Description	Range in Breda	Range in Tilburg	Rounding	Minimum value (else special code)
BU_CODE	**Code of neighbourhood**	-	-	-	-
BU_NAAM	Name of the neighbourhood	-	-	-	-
POSTCODE	Most common 4-position postcode of neighbourhood	4811-4847	5011-5071	-	-
DEK_PERC	Cover percentage neighbourhood	1-5	1-6	-	-
OAD	Omgevingsadressendichtheid, average number of addresses within 1000m of a house	38-3701	49-4597	-	-
AANT_INW	Number of people	35-10665	20-12085	To five-folds	
P_00_14_JR	Percentage of people 14 years or younger	3-32%	1-34%	To integer	50 people
P_15_24_JR	Percentage of people 15 to 25 years old	6-31%	1-26%	To integer	50 people
P_25_44_JR	Percentage of people 25 to 45 years old	14-55%	4-46%	To integer	50 people
P_45_64_JR	Percentage of people 45 to 65 years old	13-43%	13-44%	To integer	50 people
P_65_EO_JR	Percentage of people 65 years or older	2-38%	2-82%	To integer	50 people
BEV_DICHTH	Population density	12-12356	11-10528	To integer	10 people
AANT_HH	Number of households	15-5425	10-7255	To five-folds	
P_WEST_AL	Percentage of western foreigners	3-18%	2-12%	To integer	50 people
P_N_W_AL	Percentage of non-western foreigners	0-38%	0-51%	To integer	50 people
AUTO_TOT	Number of cars (vehicles for max. 8 passengers)	50-7670	20-7705	To five-folds	-
AUTO_HH	Average number of cars per household	0,5-2,1	0.5-2.0	To 1 decimal	50 households and max. 2.5 cars per household

Table 23:Data description of useable data from CBS at neighbourhood level for 2011

Attribute name	Description	Range in Breda	Range in Tilburg	Rounding	Minimum value (else special code)
BU_CODE	**Code of neighbourhood**	-	-	-	-
WWB_UITTOT	Number of social security eligibilities at March 31st	0-340	0-540	Tens	50 households
AF_ONDVMB	Average distance (km) to school with VMBO (following the road)	0,4-8,5	0.2-4.8	Tenths	10 people
AF_ONDHV	Average distance (km) to school with HAVO or VWO (following the road)	0,4-7,3	0.5-5.8	Tenths	10 people
AF_TREINST	Average distance to nearest train station (following the road)	0,4-10,4	0.6-6.8	Tenths	10 people

Table 24:Data description of useable data from CBS in squares of 500m for 2012

Attribute name	Description	Range in Breda	Range in Tilburg	Rounding	Minimum value (else special code)
C28992R500	**Coordinate of lower left corner of square**	-		-	-
INW2012	Number of people	0-2855	0-2870	To five-folds	-
INW2013	Number of people	0-2910	0-2880	To five-folds	-
I2012_014	Number of people 0-14 years	0-505	0-600	To five-folds	-
I2012_1524	Number of people 15-24 years	0-510	0-785	To five-folds	-
I2012_2544	Number of people 25-44 years	0-1465	0-1180	To five-folds	-
I2012_4564	Number of people 45-64 years	0-600	0-795	To five-folds	-
I2012_65PL	Number of people 65 years or older	0-605	0-705	To five-folds	-
P_AUTO2012	Percentage of 'natives'	Category 1 (>90%) – category 4 (40-60%)	Category 1 (>90%) – category 5 (<40%)	-	10 persons
P_WAL2012	Percentage of western foreigners	Category 0 (none) - category 5 (<8%)	Category 0 (none) - category 5 (<8%)	-	10 persons
P_NWAL2012	Percentage of non-western foreigners	Category 0 (none) – category 5 (<10%)	Category 0 (none) – category 5 (<10%)	-	10 persons
P_LINH2010	Percentage of low income households	5-72%	8-72%	-	100 households and 5 low income households
P_HINH2010	Percentage of high income households	2-75%	3-60%	-	100 households and 5 high income house holds
OAD2012	Average number of addresses within 1000m of an address (OmgevingsAdressenDichtheid)	2-3828	2-4764	-	-

Table 25:Data description of useable data from CBS in squares of 100m for 2012

Attribute name	Description	Range in Breda	Range in Tilburg	Rounding	Minimum value (else special code)
C28992R100	**Coordinate of lower left corner of square**	-	-	-	-
INW2012	Number of people	0-320	0-370	To five-folds	-
INW2013	Number of people	0-320	0-	To five-folds	-
I2012_014	Number of people 0-14 years	0-60	0-95	To five-folds	-
I2012_1524	Number of people 15-24 years	0-95	0-215	To five-folds	-
I2012_2544	Number of people 25-44 years	0-185	0-145	To five-folds	-
I2012_4564	Number of people 45-64 years	0-90	0-125	To five-folds	-
I2012_65PL	Number of people 65 years or older	0-160	0-370	To five-folds	-
P_AUTO2012	Percentage of 'natives'	Cat.1 (>90%) – cat.5 (<40%)	Cat. 1 (>90%) – cat.5 (<40%)	-	10 persons
P_WAL2012	Percentage of western foreigners	Cat. 0 (none) –cat. 5 (<8%)	Cat. 0 (none) –cat.5 (<8%)	-	10 persons
P_NWAL2012	Percentage of non-western foreigners	Cat. 0 (none) – cat. 5 (<10%)	Cat. 0 (none) – cat.5 (<10%)	-	10 persons

Appendix II Effect of events on the number of bus passengers

Figures 58 and 59 show the impact of large events on the number of boarders per day of the week as discussed in section 8.3.

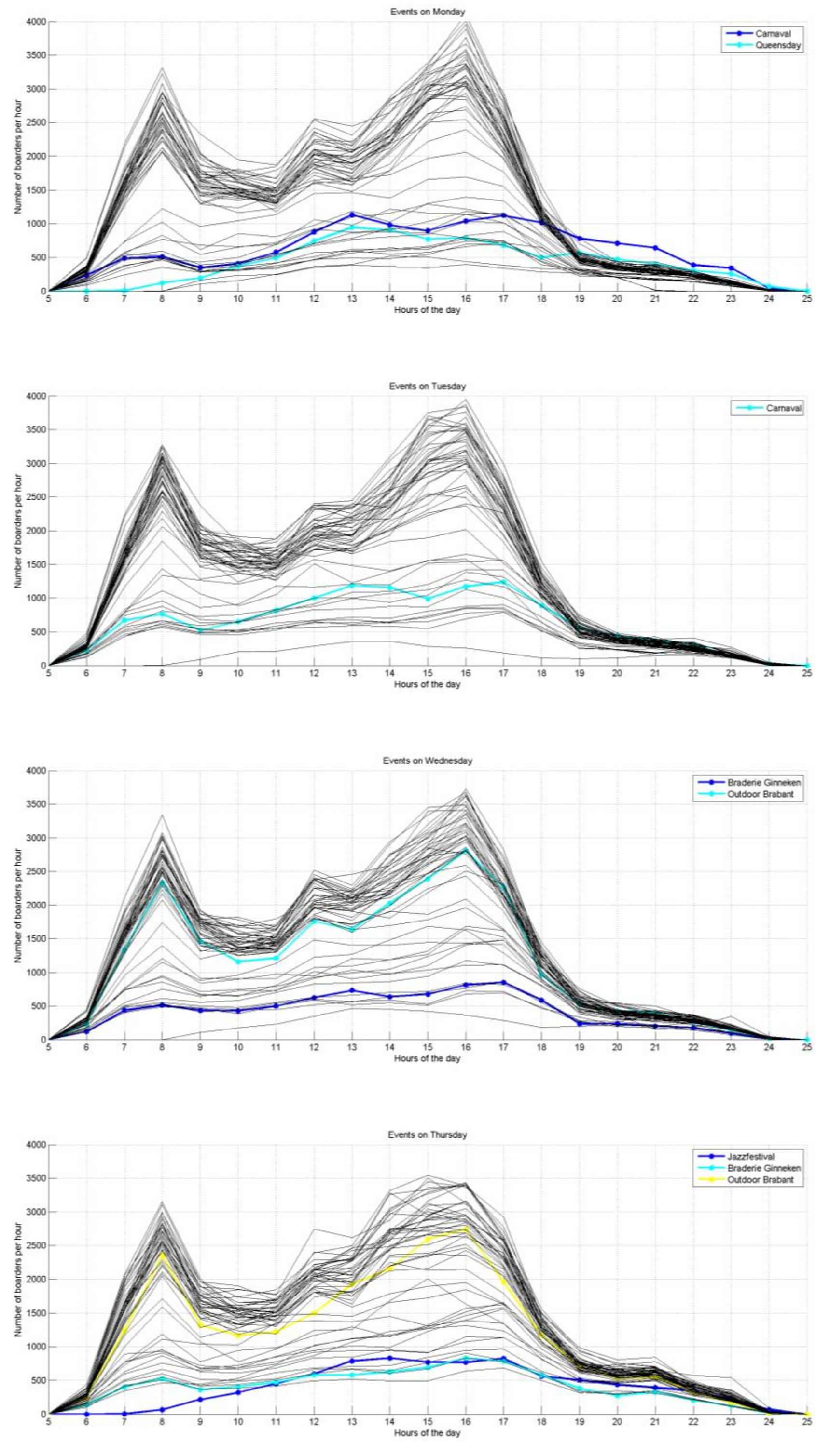


Figure 58: Influence of events on the number of bus passengers on Monday to Thursday. The top graph contains all Mondays in 2012, the Mondays with a large event are coloured. Same applies for the other graphs.

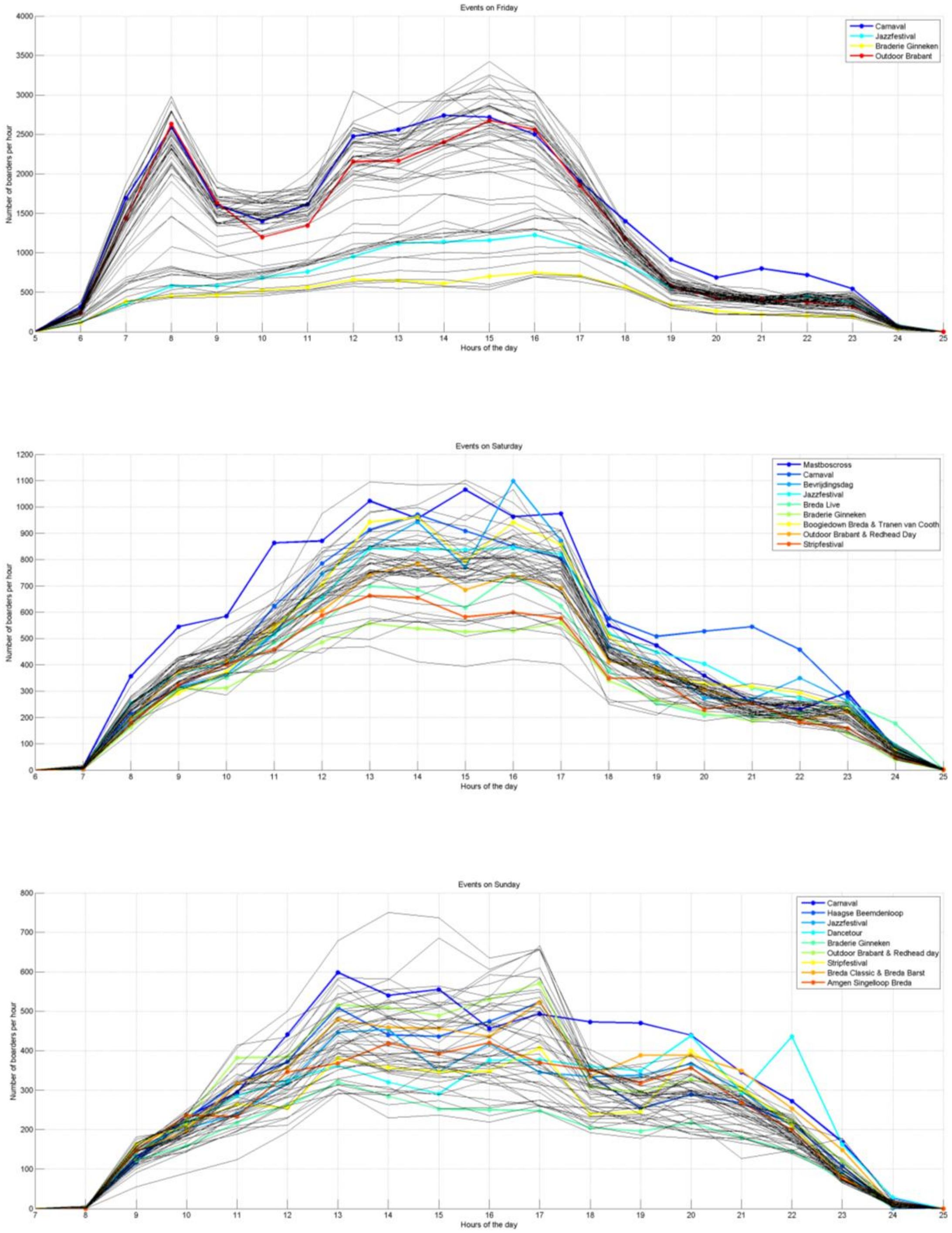


Figure 59: Influence of events on the number of boarders in Breda on Fridays, Saturdays and Sundays in 2012. The top figure contains all Fridays in 2012 and the Fridays with a large event are coloured. The second figure shows all Saturdays in 2012 and the third figure contains all Sundays.

Appendix III Weather characteristics

Table 26 shows the weather characteristics which are analyzed in section 8.4. The table shows the range and average of each variable at the weather station Gilze-Rijen in 2012.

Table 26: Minimum, maximum and average value of the weather characteristics at the weather station Gilze-Rijen in 2012. Koninklijk Nederlands Meteorologisch Instituut (KNMI), 2012.

Variable	Average	Range
Average wind speed (0.1 m/s)	35	10 – 105
Highest average wind speed over an hour (0.1 m/s)	56	20 - 140
Minimum temperature (0.1 °C)	59	-174 - 185
Highest temperature (0.1 °C)	146	-50 – 351
Sun hours (0.1 h)	47	0 – 152
Precipitation period (0.1h)	20	0 - 153
Daily precipitation (0.1mm)	23	-1 – 23
Highest hourly sum of precipitation (0.1mm)	10	-1 - 177

Appendix IV Input variables regression analysis

Table 27 gives all input variables for the regression analysis in SPSS. The variables boarders_weekday to alighters_sun are the dependant variables and the variables ‘name’ and ‘stop_number’ only function as identifying variables (unique keys).

Table 27: Input variables for the regression analysis

Variable name	Description
Name	Name of the bus stop
Boarders_weekday	Average number of boarders on the Tuesdays in 2012, excluding holidays
Boarders_sat	Average number of boarders on the Saturdays in 2012, excluding holidays
Boarders_sun	Average number of boarders on the Sundays in 2012, excluding holidays
Alighters_weekday	Average number of alighters on the Tuesdays in 2012, excluding holidays
Alighters_sat	Average number of alighters on the Saturdays in 2012, excluding holidays
Alighters_sun	Average number of alighters on the Sundays in 2012, excluding holidays
Stop_number	Unique number of the bus stop
No_stops	Number of stops that can be reached from the bus stop with transfer
INW2012	Number of people within catchment area
I2012_014	Number of people of 0 to 14 years
I2012_1524	Number of people of 15 to 24 years
I2012_2544	Number of people of 25 to 44 years
I2012_4564	Number of people of 45 to 64 years
I2012_65PL	Number of people of 65 years and older
P_LINH2010	Percentage of low income households (2010)
P-HINH2010	Percentage of high income households (2010)
OAD2012	Omgevingsadressendichtheid
SUMStud	Number of students at intermediate vocational educational institutes, Universities of Applied Science and Universities (MBO,HBO,WO)
Sum_ll_vo	Number of pupils at secondary schools
W_all	Number of western foreigners (allochtonen)
NW_all	Number of non-western foreigners
Cars_hh	Average number of cars
ll_vo_liv	Number of pupils, living location
Cars_total	Total number of cars
Jobs	Number of jobs
WWB	Number of social eligibilities (bijstandsuitkeringen)
A_vmbo	Average distance to a secondary school with VMBO education
A_havovwo	Average distance to a secondary school with havo or vwo
A_station	Average distance to a train station
Gem_ink	Average standardized income per household
P_014	Percentage of people of 0 to 14 years
P_1524	Percentage of people of 15 to 24 years
P_2544	Percentage of people of 25 to 44 years

Variable name	Description
P_4564	Percentage of people of 45 to 64 years
P_65pl	Percentage of people of 65 years and older
P_w_all	Percentage of western foreigners (allochtonen)
P_nw_all	Percentage of non-western foreigners (allochtonen)
Opp_woon	Floor area of addresses with a residential function
No_woon	Number of addresses with a residential function
Opp_winkel	Floor area of addresses with a retail function
No_winkel	Number of addresses with a retail function
Opp_kantoor	Floor area of addresses with a office function
No_kantoor	Number of addresses with a office function
Opp_industrie	Floor area of addresses with an industry function
No_industrie	Number of addresses with an industry function
Opp_bijeenkomst	Floor area of addresses with a social gathering function
No_bijeenkomst	Number of addresses with a social gathering function
Opp_gezondheid	Floor area of addresses with a healthcare function
No_gezondheid	Number of addresses with a healthcare function
Opp_logies	Floor area of addresses with an accommodation function
No_logies	Number of addresses with an accommodation function
Opp_onderwijs	Floor area of addresses with an education function
No_onderwijs	Number of addresses with an education function
Opp_sport	Floor area of addresses with a sport function
No_sport	Number of addresses with a sport function
Opp_overig	Floor area of addresses with an other function
No_overig	Number of addresses with an other function
%woon	Percentage of floor area with residential function
%winkel	Percentage of floor area with a retail function
%kantoor	Percentage of floor area with an office function
%industrie	Percentage of floor area with an industry function
%bijeenkomst	Percentage of floor area with a social gathering function
%gezondheid	Percentage of floor area with a healthcare function
%logies	Percentage of floor area with an accommodation function
%onderwijs	Percentage of floor area with an education function
%sport	Percentage of floor area with a sport function
%overig	Percentage of floor area with an other function
Opp_winkels_horeca	Floor area with a retail, social gathering or accommodation function
%winkels_horeca	Percentage of floor area with a retail, social gathering or accommodation function
Type_CS	Dummy variable for the bus stop at the Central Station
Type_woon_90	Dummy variable for the stops with %woon ≥90
Type_woon_80	Dummy variable for the stops with %woon ≥80
Type_HO	Dummy variable for the stops near Universities and Universities of Applied Science
Type_shoppingcentre	Dummy variable for the stops with opp_winkels_horeca ≥ 5000 m^2

Appendix V Insignificant variables correlation analysis

Table 28 shows the input variables which have an insignificant correlation with the daily number of boarders in Breda and Tilburg. An overview of all variables is given in Appendix IV and the significant variables are given in section 9.2.1.

Table 28: Insignificant variables of the correlation between the explanatory variables and the daily number of boarders

Variable name	Correlation coefficient	Significance	Correlation coefficient	Significance
INW2012	-.036	.624	-.066	.364
I2012_014	-.132	-.132	-.124	-.124
I2012_1524	.032	.073	-.012	.089
I2012_2544	.055	.032	-.043	-.012
I2012_4564	-.084	.666	-.105	.872
I2012_65PL	-.068	.055	-.038	-.043
P_LINH2010			.110	.131
P-HINH2010	-.092	-.092	-.039	-.039
SUMStud			.125	.086
Sum_ll_vo	.023	.761	.039	.594
W_all	.016	.829	.072	.321
NW_all	-.018	.804	-.011	.875
Cars_hh	-.082	.268	-.092	.203
ll_vo_liv	-.013	.859	-.050	.490
Cars_total	.052	.484	.061	.404
WWB	.074	.317	.036	.624
A_vmbo	-.034	.645	-.121	.095
A_havovwo	-.128	.083	-.117	.106
Gem_ink	-.027	.718		
P_014			-.110	.137
P_1524	.063	.395		
P_2544			.050	.499
P_4564	-.095	.198	-.130	.078
P_65pl	-.056	.448	-.021	.776
P_w_all	.088	.232	-.055	.453
P_nw_all	.011	.880	-.054	.460
Opp_woon	-.002	.979	-.098	.178
No_woon	.016	.829	-.067	.354
Opp_winkel			.043	.553
No_winkel			.054	.458
Opp_industrie	-.029	.694	-.037	.615
No_industrie	-.009	.898	-.048	.508
Opp_bijeenkomst			.133	.067
No_bijeenkomst			.099	.172

Variable name	Correlation coefficient	Significance	Correlation coefficient	Significance
Opp_gezondheid	.014	.847	.008	.914
No_gezondheid	-.024	.744	-.030	.677
Opp_onderwijs	.081	.274	.080	.274
No_onderwijs	.012	.873	.040	.585
Opp_sport	-.016	.832	.003	.972
No_sport	-.029	.698	.018	.803
Opp_overig	*		-.007	.928
No_overig	-.051	.492	-.035	.629
%winkel	.086	.246	.022	.765
%industrie	-.039	.598	-.048	.513
%bijeenkomst	.127	.084		
%gezondheid	-.002	.974	.002	.974
%logies			.031	.672
%onderwijs	.050	.495	.075	.305
%sport	-.021	.773	-.002	.982
%overig			-.028	.700
Opp_winkels_horeca			.095	.193
%winkels_horeca			.093	.203
Type_woon_90	-.087	.240	-.068	.351
Type_woon_80	-.130	.078	-.118	.104
Type_HO	.107	.145	.125	.086

Appendix VI Creating regression models

Section 9.2 elaborated on the creation of the trip generation model for the number of boarders at bus stops on weekdays. The models boarders on Saturday, boarders on Sunday, alighters on weekdays, alighters on Saturdays and alighters on Sundays are created in the same way. All models start with a negative constant, so when the positive variables are not large enough, a bus stop might end with a negative (rounded to zero) boarders or alighters. Because of the characteristics of regression analyses, the separate coefficients are meaningless.

The following formulas are the six trip generation models:

$$Boarders_{weekday} = -128.654 + 3.721x_{no_{stops}} - 0.612x_{P_{LINH}} + 0.135x_{Cars_{total}} + 0.155x_{SUMStud\ (HO)} - 0.044x_{OAD} + 0.002x_{Opp_{kantoor}} + 1.074x_{\%woon} + 0.019x_{Opp_{winkel_{horeca}}} + 9281.374x_{Type_{CS}} - 152.157x_{Type_{HO}}$$

$$Boarders_{saturday} = -6.928 + 0.338x_{no_{stops}} + 0.166x_{P_{LINH}} + 0.027x_{Cars_{total}} - 0.005\,x_{OAD} + 0.002x_{Opp_{winkel_{horeca}}} + 2246.167x_{Type_{CS}} - 11.025\,x_{Type_{shoppingcentre}}$$

$$Boarders_{sunday} = -17.156 + 1.126x_{no_{stops}} - 0.315x_{P_{LINH}} + 0.019x_{Cars_{total}} - 0.004\,x_{OAD} + 0.001x_{opp\_kantoor} + 0.006x_{Opp\_winkel\_horeca} + 1470.861x_{Type_{CS}} - 54.279\,x_{Type_{shoppingcentre}}$$

$$\mathrm{Alighters_{weekday}} = -127.775 + 3.979x_{no_{stops}} - 0.643x_{P_{LINH}} + 0.096x_{Cars_{total}} + 0.126x_{SUMStud\ (HO)} - 0.029\,x_{OAD} + 0.002x_{Opp_{kantoor}} + 0.886x_{\%woon} + 0.015x_{Opp_{winkel_{horeca}}} + 8127.618x_{Type_{CS}} - 141.667\,x_{Type_{HO}}$$

$$Alighters_{saturday} = -19.999 + 1.153x_{no_{stops}} - 0.340x_{P_{LINH}} - 0.001x_{Cars_{total}} + 0.005\,x_{OAD} + 0.001x_{Opp\_\mathrm{winkel\_horeca}} + 1493.062x_{Type_{CS}} - 25.330\,x_{Type_{shoppingcentre}}$$

$$Alighters_{sunday} = -9.152 + 0.425x_{no_{stops}} + 186x_{P_{LINH}} + 0.021x_{Cars_{total}} - 0.006\,x_{OAD} + 0.003x_{Opp\_winkel\_horeca} + 1766.794x_{Type_{CS}} - 24.131x_{Type_{shoppingcentre}}$$

Appendix VII Comparing model variables with other models

The variables which are included in the weekday models are compared with the factors that influence the demand for the bus transport which were found in the literature study and with other bus demand models (see table 29). The similarities and differences are analyzed in section 9.3.

Table 29: Factors that influence the demand for bus transport which were found in the literature study, the variables included in the developed weekday models and the variables included in other bus demand models

Literature	Weekday models	Schmenner	Klok	Bechdolt & Williams	Zenith
Age	*	Third quartile age of population	Percentage of people aged 15-24		Number of dependants aged 0-17, 18-64, 65+
Population	*				Residents in a household Number of blue and white collar workers Zonal population Number of households
Ethnicity	*	Percentage of population which is black		Percentage of the population of an urbanized area that is nonwhite	
Income/car ownership	Total number of cars Percentage of low income households	 Median family income		Percentage of households in an urbanized area with one or more automobiles Percentage of households in an urbanized area with incomes between $5,000 and $8,000	Level of household car ownership
Bicycle ownership	**				
Density	Average address density		Address density	Population per square mile of land	
Diversity	Percentage of floor area with residential function				
Urban design	**				
Distance to bus stops	***				

Literature	Weekday models	Schmenner	Klok	Bechdolt & Williams	Zenith
Number of jobs	Floor area of offices	Large scale employment	Number of jobs per inhabitant		Employment in 12 industry categories
College enrolments	Number of students at intermediate vocational educational institutes, Universities of Applied Science and Universities (MBO,HBO,WO)	School enrolments			Pre and primary school enrolments
	Dummy variable for the stops near Universities and Universities of Applied Science				Secondary school enrolments
					Equivalent full time tertiary school enrolments
Accessibility	Number of stops that can be reached from the bus stop with transfer				
Trip purpose	Floor area with a retail, social gathering or accommodation function	Shopping centres			
Travel time	**				
Fare	**			Average cost of trip by automobile minus the average bus fare	
Quality of service	**		Frequency	Average peak hour headway	
			Punctuality		
Reputation	**				
	Dummy variable for the bus stop at the Central Station				
				Percentage of the employed labour force in manufacturing	

*: Not significant (p > 0.05)
**: No data collected
***: Indirectly included via catchment areas

Appendix VIII Calibration procedure distribution functions

Hyman's bi-proportional calibration procedure was followed to calibrate the distribution functions (Ortúzar & Willumsen, 2001). Figure 60 shows a conceptual model of the calibration of the exponential function. First the travel time matrix is created with the travel time between the bus stops with a direct connection. In case of different travel times between bus stop pairs (e.g. because the bus stops are served by lines with different routes), the shortest travel time is selected. The number of boarders and alighters at each bus stop is collected. The origin-destination matrix is created using the travel times and the distribution function with the first estimation of beta. Equation A is used for the first estimation of beta.

$$\beta_0 = \frac{1}{OMTT} \qquad \text{Equation A}$$

A Furness procedure is followed until the total number of modelled boarders and alighters are equal to the observed numbers. Next, the goodness of fit is calculated (see section 10.2).

For the second iteration (m = 1) the new beta is calculated using equation B.

$$\beta_1 = \beta_0 \frac{MMTT}{OMTT} \qquad \text{Equation B}$$

The procedure with creating the origin-destination matrix is repeated with the new estimation of beta. For the third iteration line search is used, with the formula of equation C.

$$\beta_m = \frac{gof_1\beta_0 - gof_0\beta_0}{gof_1 - gof_0} \qquad \text{Equation C}$$

The 4th to 100th iteration either line search (equation C) or parabolic search is used (equations D, E and F).

$$a_m = \frac{(gof_{m-3} - gof_{m-2})(\beta_{m-3} - \beta_{m-1}) - (gof_{m-3} - gof_{m-1})(\beta_{m-3} - \beta_{m-2})}{(\beta - \beta_{m-2}^2)(\beta_{m-3} - \beta_{m-1}) - (\beta_{m-3}^2 - \beta_{m-1}^2)(\beta_{m-3} - \beta_{m-2})} \qquad \text{Equation D}$$

$$b_m = \frac{(gof_{m-3} - gof_{m-2}) - a(\beta_{m-3}^2 - \beta_{m-2}^2)}{(\beta_{m-3} - \beta_{m-2})} \qquad \text{Equation E}$$

$$\beta_m = -{}^{b_m}/_{2a_m} \qquad \text{Equation F}$$

When the 100 iterations are finished, the beta with the lowest goodness of fit-value is selected. For distribution functions with two variables, the procedure is repeated with the alpha instead of the beta. This entire procedure is repeated 20 times, resulting in an optimized alpha and beta.

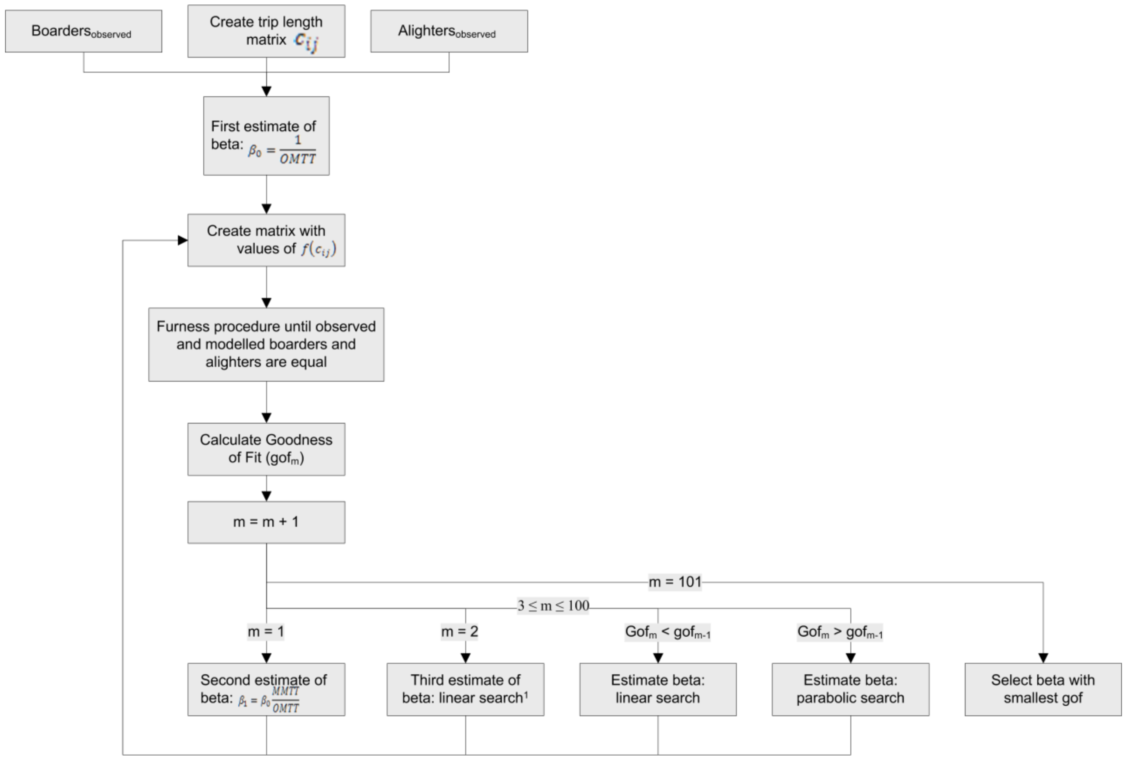


Figure 60: Calibration procedure of the distribution models

Appendix IX Create distribution models for weekend

Saturdays

For the Saturday the procedure is exactly the same as for the weekdays, but with the observed data from the Saturdays instead of the weekdays. Figure 61 shows the observed trip length distribution of the Saturday. Although the trip numbers are significantly lower, there are less outliers.

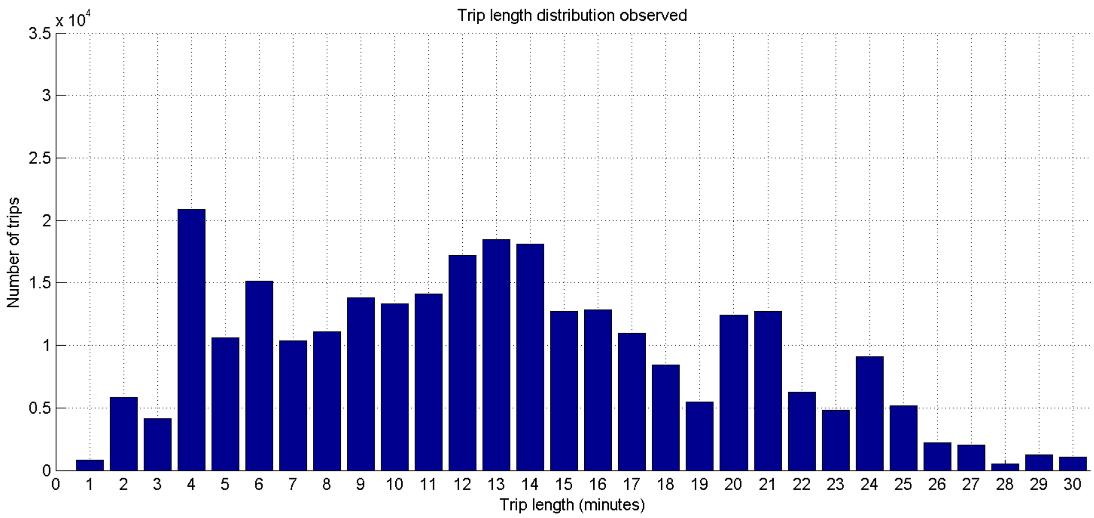


Figure 61: Observed trip length distribution on the Saturdays in 2012

Tables 30, 31 and 32 give the calibrated exponential, Tanner and top log-normal distribution functions.

Table 30: Calibration values of the exponential distribution function for the Saturdays

$f(c_{ij}) = \mathrm{e}^{-0.030\mathrm{c}_{\mathrm{ij}}}$	alpha	Beta	Mean travel time	Standard deviation	Goodness of fit
Observed value/aim	-	-	13.393	7.067	0
Modelled value	-	0.030	12.734	8.215	0.219

Table 31: Calibration values of the Tanner distribution function for the Saturdays

$f(c_{ij}) = c_{ij}^{1.303} e^{-0.130 c_{ij}}$	alpha	Beta	Mean travel time	Standard deviation	Goodness of fit
Observed value/aim	-	-	13.393	7.067	0
Modelled value	1.303	0.130	13.381	7.081	3.729 $\times 10^{-5}$

Table 32: Calibration values of the top log-normal distribution function for the Saturdays

$f(c_{ij}) = e^{-0.827(\log(c_{ij}/8.662))^2}$	alpha	Beta	Mean travel time	Standard deviation	Goodness of fit
Observed value/aim	-	-	13.393	7.067	0
Modelled value	0.827	8.862	13.393	7.067	1.575 $\times 10^{-8}$

Figure 62 shows the observed trip length distribution together with the trip length distributions of the three calibrated distribution functions. The exponential function overestimates the very short trip lengths and underestimates the longer trip lengths. The Tanner and top log-normal functions however seem to fit the observed trip distribution quite well.

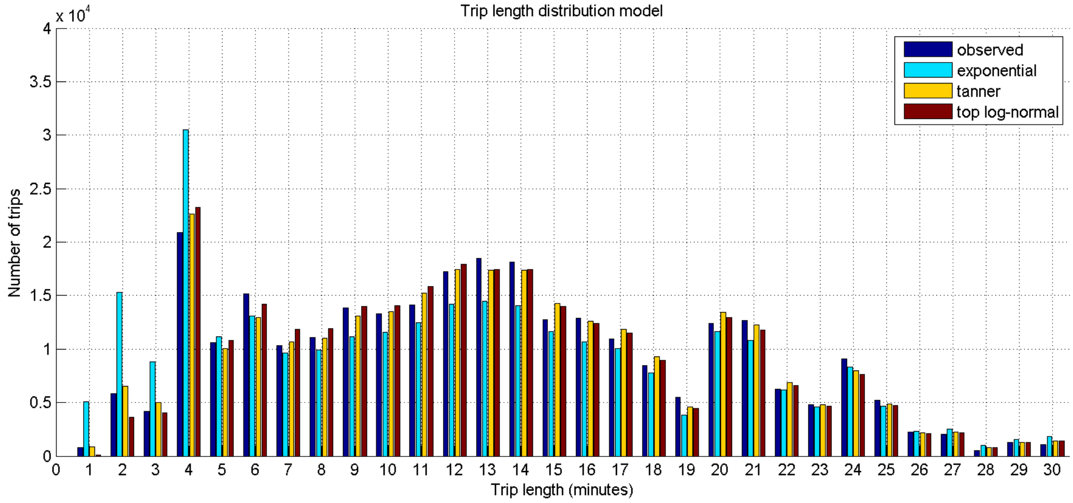


Figure 62: Trip length distribution of the observed data and the three calibrated distribution functions

Sundays

For the Sunday the procedure is exactly the same as for the Saturdays, but with the observed data from the Sundays instead of the Saturdays. Figure 61 shows the observed trip length distribution of the Sunday. The trip numbers on Sunday are even lower than on the Saturdays.

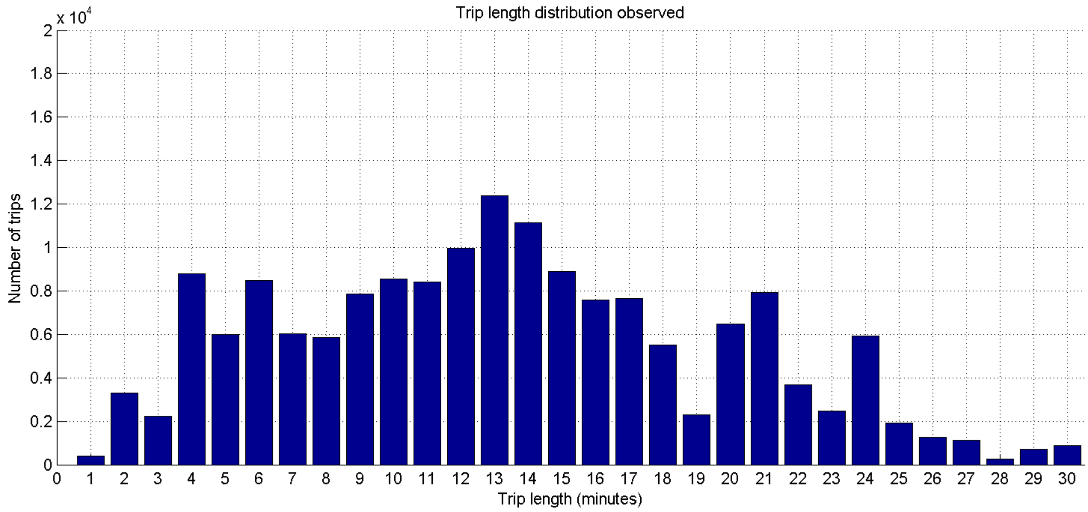


Figure 63: Observed trip length distribution of the Sundays in Breda in 2012

Tables 33, 34 and 35 give the calibrated exponential, Tanner and top log-normal distribution functions.

Table 33: Calibration values of the exponential distribution function for the Sundays

$f(c_{ij}) = e^{-0.037c_{ij}}$	alpha	Beta	Mean travel time	Standard deviation	Goodness of fit
Observed value/aim	-	-	13.576	6.811	0
Modelled value	-	0.030	12.910	7.871	0.198

Table 34: Calibration values of the Tanner distribution function for the Sundays

$f(c_{ij}) = c_{ij}^{1.370}e^{-0.132c_{ij}}$	alpha	Beta	Mean travel time	Standard deviation	Goodness of fit
Observed value/aim	-	-	13.576	6.811	0
Modelled value	1.370	0.132	13.568	6.819	1.492×10^{-5}

Table 35: Calibration values of the top log-normal distribution function for the Sundays

$f(c_{ij}) = e^{-0.841(\log(c_{ij}/9.031))^2}$	alpha	Beta	Mean travel time	Standard deviation	Goodness of fit
Observed value/aim	-	-	13.576	6.811	0
Modelled value	0.841	9.031	13.576	6.811	2.826×10^{-9}

Figure 64 shows the observed trip length distribution on Sundays and the trip length distribution of the three calibrated distribution functions. Again, the exponential function overestimates the short trip lengths, while the Tanner and top log-normal functions are quite similar to the observed distribution.

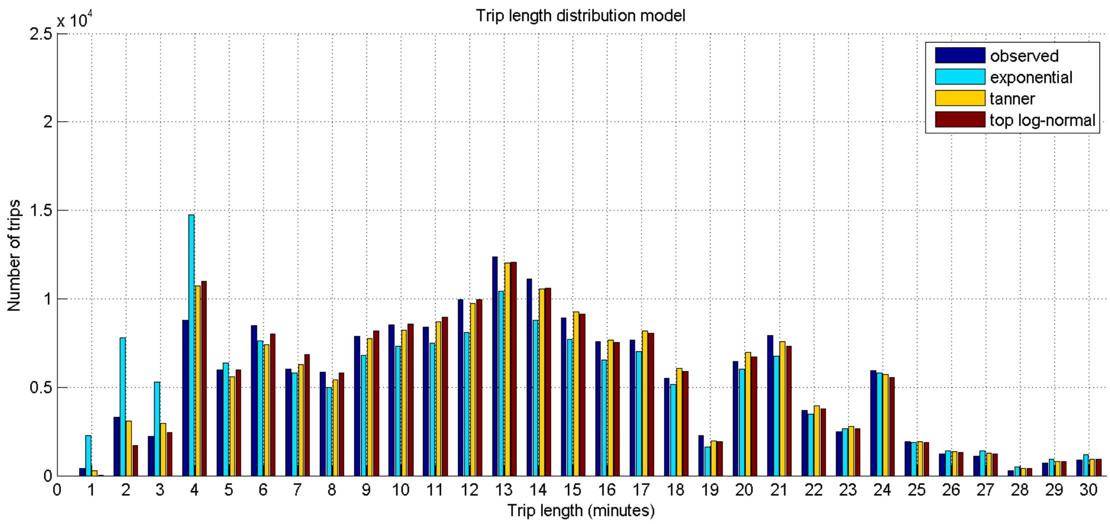


Figure 64: Trip length distribution of the observed data and the three calibrated distribution functions

Appendix X Validation regression models

Breda

Table 36 gives the total number of boarders and alighters at the bus stops in Breda according to the chipcard data and according to the models. On weekdays, the models overestimate the total number of boarders and alighters in Breda with 5-7%, which is an acceptable difference. The relative differences on Saturday and Sunday are larger, but because of the low numbers of passengers this is also acceptable.

Table 36: Observed and modelled number of boarders and alighters in Breda

Breda		Sum observed	Sum model	Difference
	Weekday	27.121	28.890	7%
Boarders	Saturday	8.278	4.362	-47%
	Sunday	4.648	5944	28%
Alighters	Weekday	26.228	27.643	5%
	Saturday	8.110	6.027	-26%
	Sunday	4.501	4.226	-6%

Tilburg

Table 37 shows the total number of boarders and alighters at the bus stops in Tilburg according to the chipcard data and according to the models. On weekdays the differences between the observed and modelled number of boarders and alighters is 1-5%. Just like for Breda, the model for the boarders on Saturday has the worst performance.

Table 37: Observed and modelled number of boarders and alighters in Tilburg

Tilburg		Sum observed	Sum model	Difference
	Weekday	25.082	24.752	-1%
Boarders	Saturday	8.463	4.008	-53%
	Sunday	5.219	5.314	2%
Alighters	Weekday	24.160	25.370	5%
	Saturday	8.136	7.129	-12%
	Sunday	5.119	3.459	-32%

Appendix XI Validation distribution models

Saturday

Figure 65 shows the observed and modelled trip length distribution in Tilburg. The distribution model that has been developed for Breda is used and the observed number of boarders and alighters in Tilburg are used as input. The figure shows the number of trips with a trip length of 9 minutes is overestimated by the model, while the number of trips at the other trip lengths is approximately equal. Since the observed and modelled number of boarders and alighters are equal at each stop, the sum of both should be equal. There is however a difference of 3,000 trips (<1%) between the models. There are some stop pairs with a travel time less than a minute (rounded to zero), which are not included in the trip length distribution, which might be an explanation.

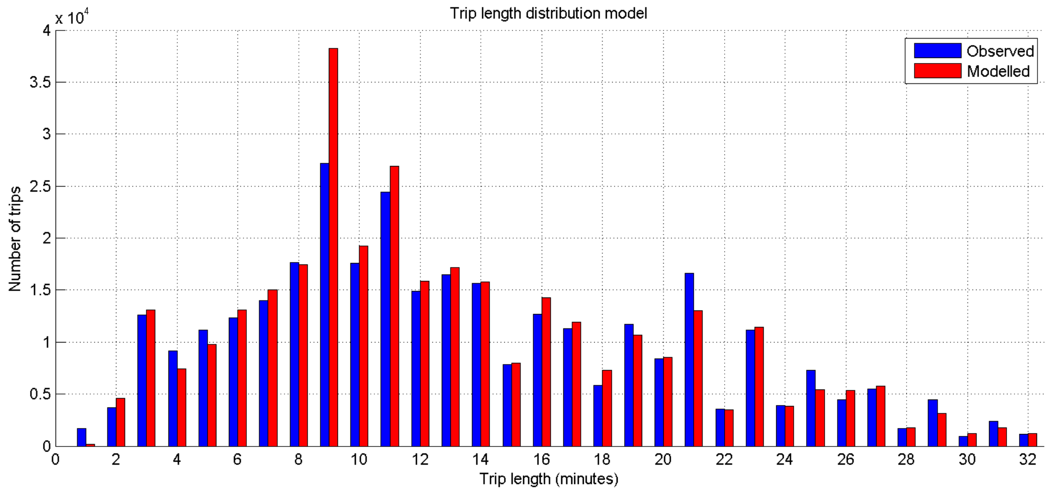


Figure 65: Observed and modelled trip length distribution in Tilburg on Saturdays.

Figure 66 shows the number of passengers on the line segments in Tilburg on Saturdays. Somehow the number of passengers at the busiest line segments is underestimated by the model. A closer look at those lines shows it are the line segments west of the city centre. Since the modelled number of boarders and alighters at each bus stop is equal to the observed number, the differences are remarkable. A cause is however not found.

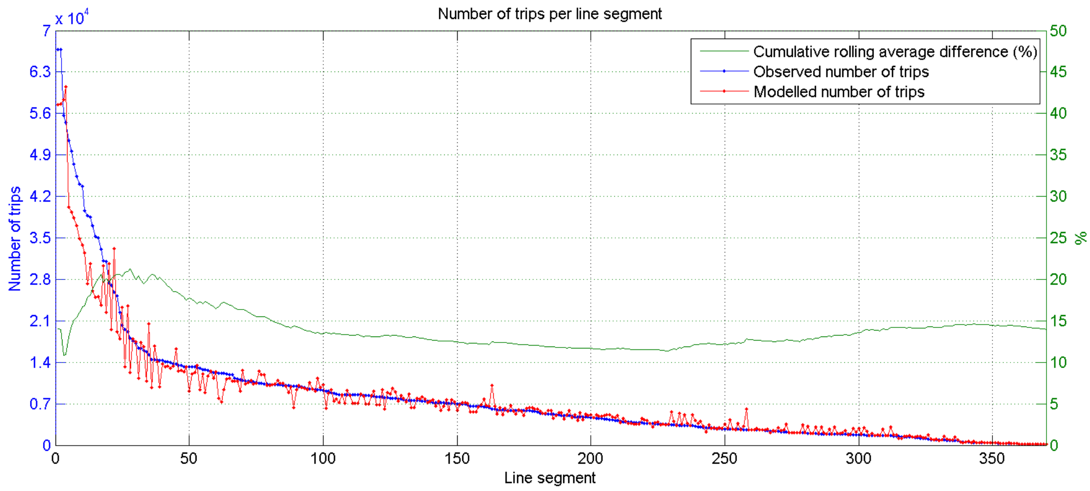


Figure 66: Observed and modelled number of passengers at each line segment in Tilburg on Saturdays. The green line gives the cumulative rolling average of the absolute difference.

Sunday

Figure 67 shows the observed and modelled trip length distribution on Sundays in Tilburg. It is very much the same as the model for Saturday, but with a lower number of passengers. Since the distribution functions for Saturday and Sundays are pretty much the same, no significantly different distribution occurs.

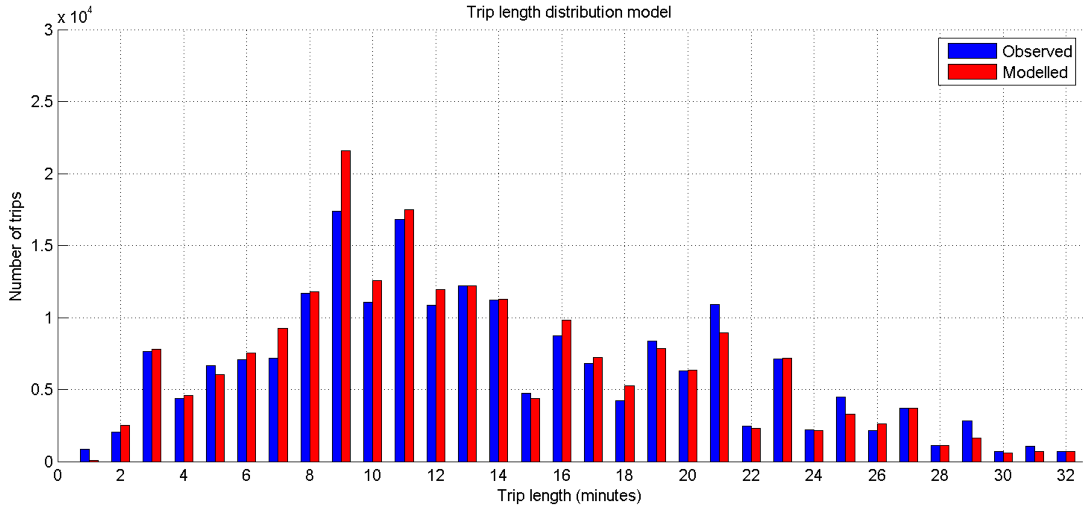


Figure 67: Observed and modelled trip length distribution in Tilburg on Sundays.

Figure 68 shows the number of passengers at the line segments. Again, the results are very similar to the Saturdays, although the number of passengers is lower.

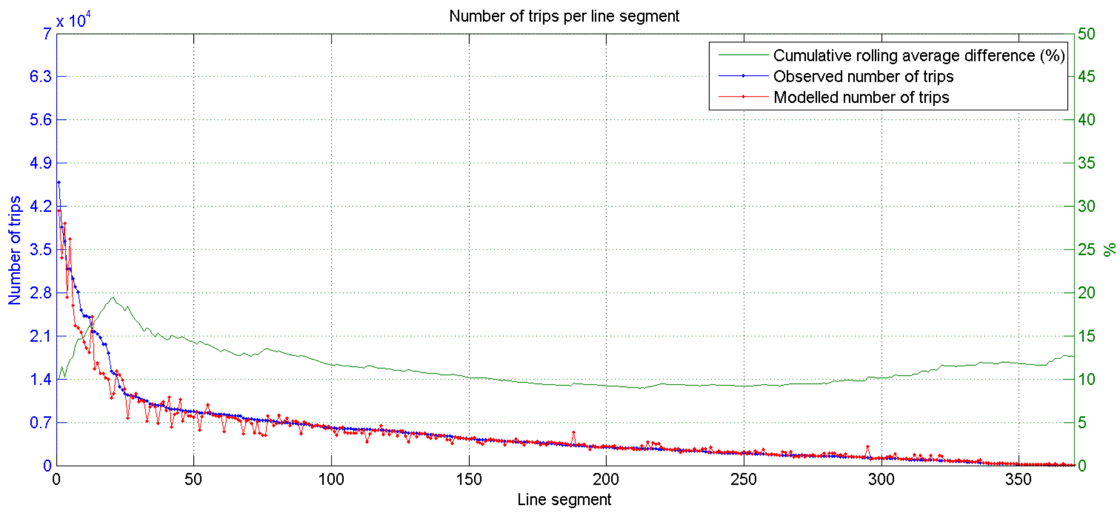


Figure 68: Observed and modelled number of passengers at each line segment in Tilburg on Sundays. The green line gives the cumulative rolling average of the absolute difference.

Appendix XII Validation regression and distribution models

Breda, Saturday

Figure 69 shows the trip length distribution of Breda on Saturdays. The observed and modelled number of passengers is quite similar on all trip lengths but one. The large number of passengers at a trip length of four minutes is caused by the number of passengers between the stop “Centrum” and “Centraal Station” (944 in both directions). The stop ‘Centrum’ is the bus stop which is closest to the shopping area in the city centre, so the large number of passengers is most likely caused by travellers with recreational purposes. The model underestimates the number of boarders and alighters at this bus stop, causing the largest part of the deviation.

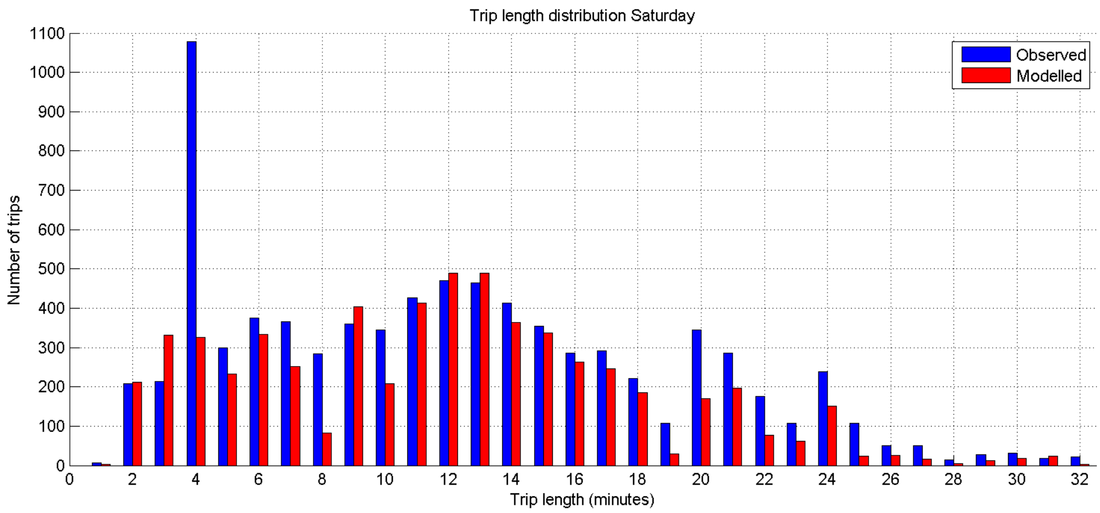


Figure 69: Observed and modelled trip length distribution in Breda on Saturdays.

Figure 70 shows the number of passengers at each line segment in Breda. Appendix X showed that the total number of passengers on Saturdays is significantly underestimated, which is also visible in this figure. The differences between the observed and modelled number of passengers at each line segment is however quite constant.

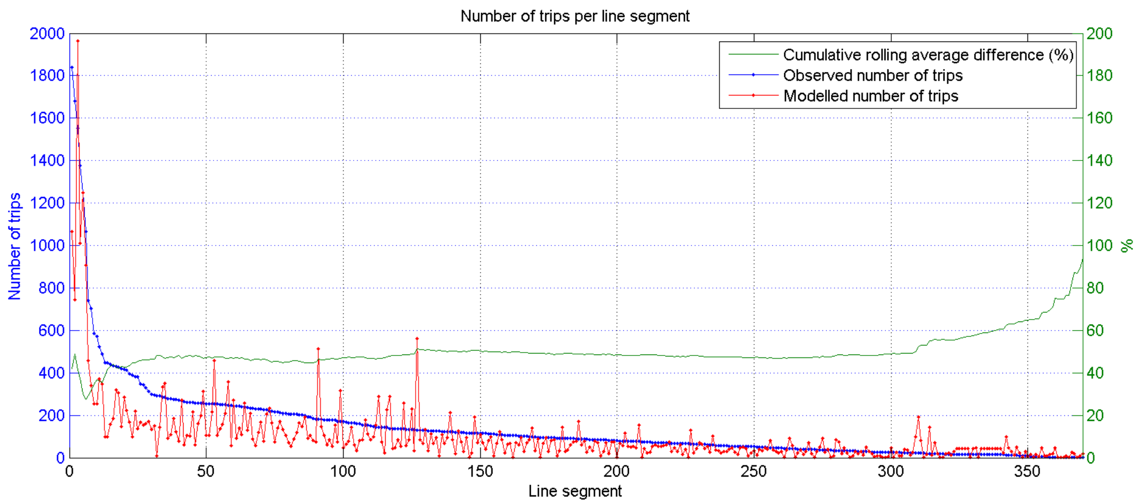


Figure 70: Observed and modelled number of passengers at each line segment in Breda on Saturdays. The green line gives the cumulative rolling average of the absolute difference.

Breda, Sunday

Figure 71 shows the observed and modelled trip length distribution on Sundays. The trip length distribution is quite similar to the Saturdays, although the difference between the number of trips on with a trip length of four minutes and the other trip lengths is smaller. The peak at four minutes is probably caused by the Sundays the shops are opened (at the *koopzondagen*), which was examined in section Effect holidays and events (section 8.3).

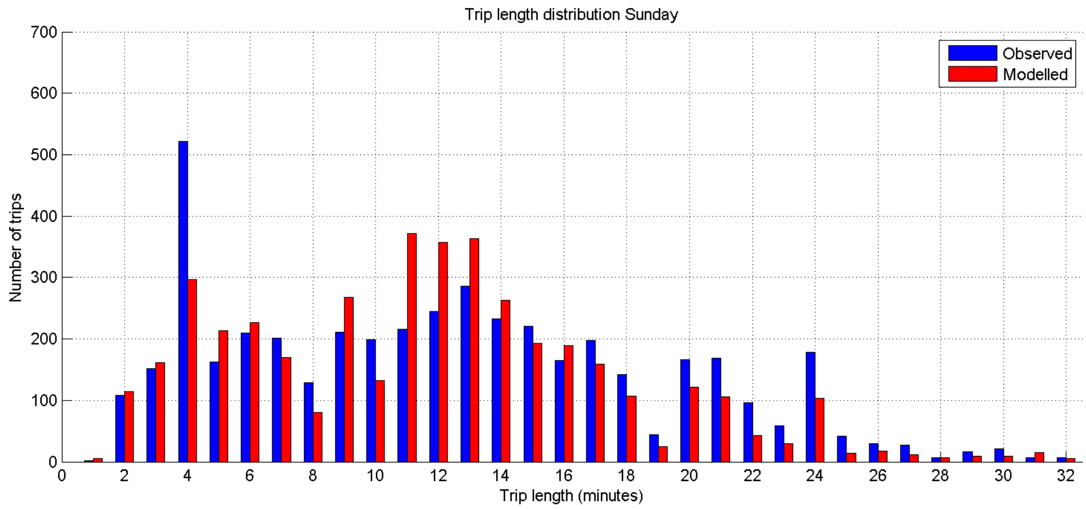


Figure 71: Observed and modelled trip length distribution in Breda on Sundays.

Figure 72 shows the number of passengers at each line segment in Breda. At some line segments the number of passengers is slightly overestimated, although the number of passengers is underestimated at more line segments. The differences are however quite limited.

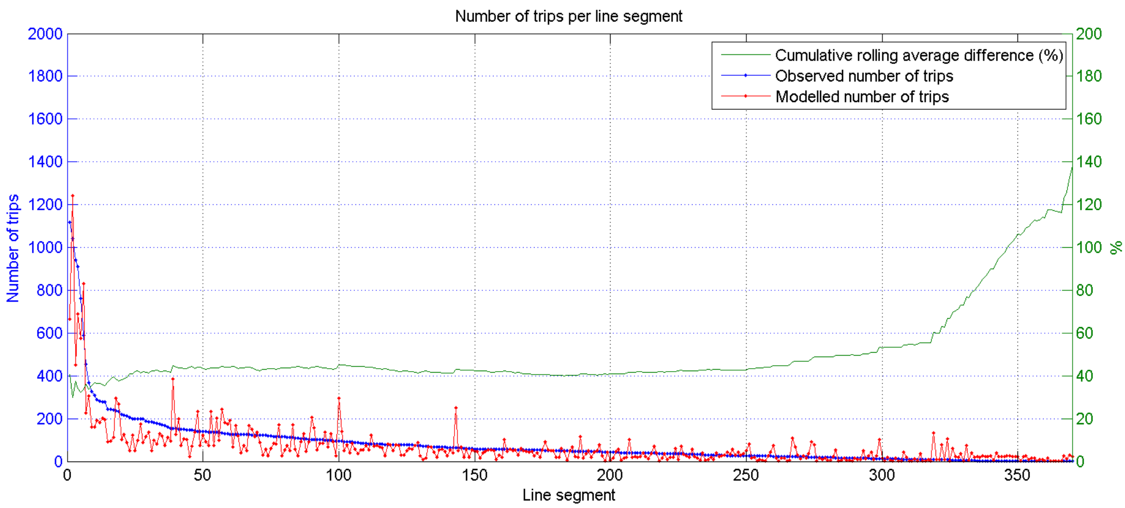


Figure 72: Observed and modelled number of passengers at each line segment in Breda on Sundays

Tilburg, Saturday

Figure 73 shows the trip length distribution of Tilburg on Saturdays. The peak at a trip length of 9 minutes is mainly caused by the number of passengers between the bus stops “Stadhuisplein” and Tilburg Central Station. The bus stop “Stadhuisplein” is the bus stop which is nearest to the shops in the city centre. The number of boarders and alighters at this bus stop is significantly underestimated by the trip generation model. Also the number of trips with a length of 11 minutes is underestimated.

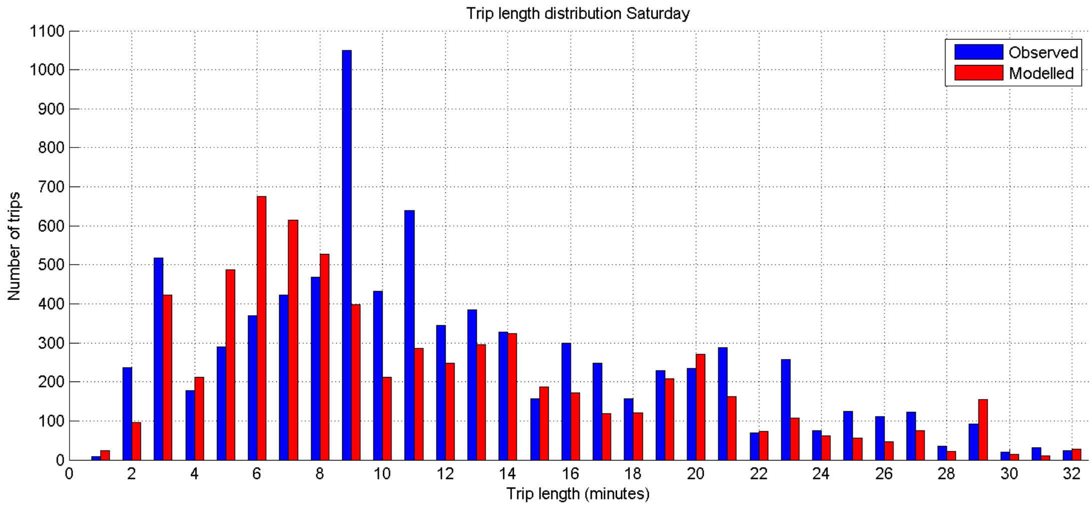


Figure 73: Observed and modelled trip length distribution in Tilburg on Saturdays.

Figure 74 shows the observed and modelled number of passengers at each line segment in Tilburg. At most line segments the number of passengers is underestimated, although there is a small overestimation at some line segments.

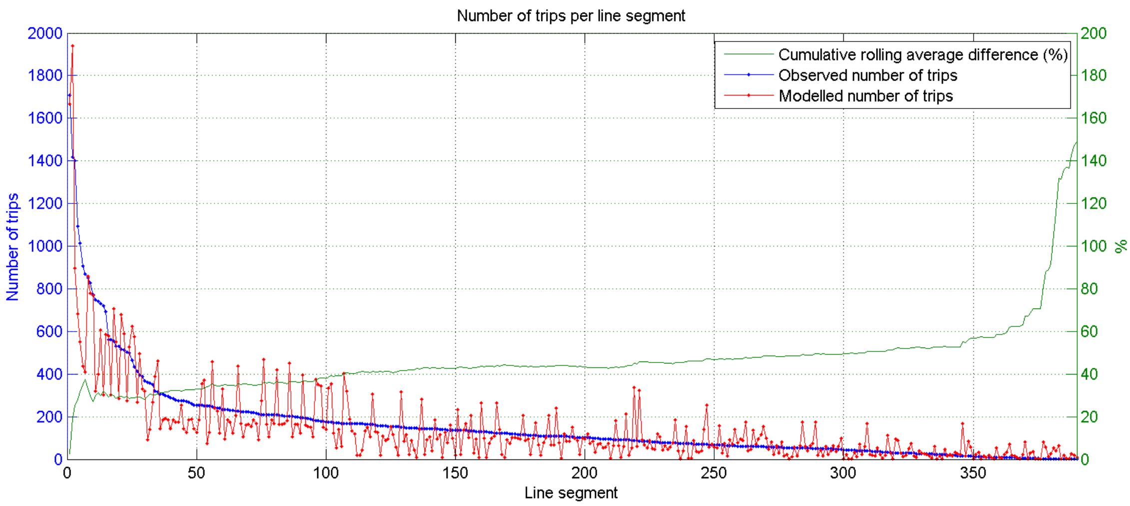


Figure 74: Observed and modelled number of passengers at each line segment in Tilburg on Saturdays

Tilburg, Sunday

Figure 75 shows the observed and modelled number of trips per trip length. The distribution is similar to the distribution of the Saturdays, although the number of passengers is lower. Also the difference between the observed and modelled number of passengers at a trip length of 9 minutes is smaller.

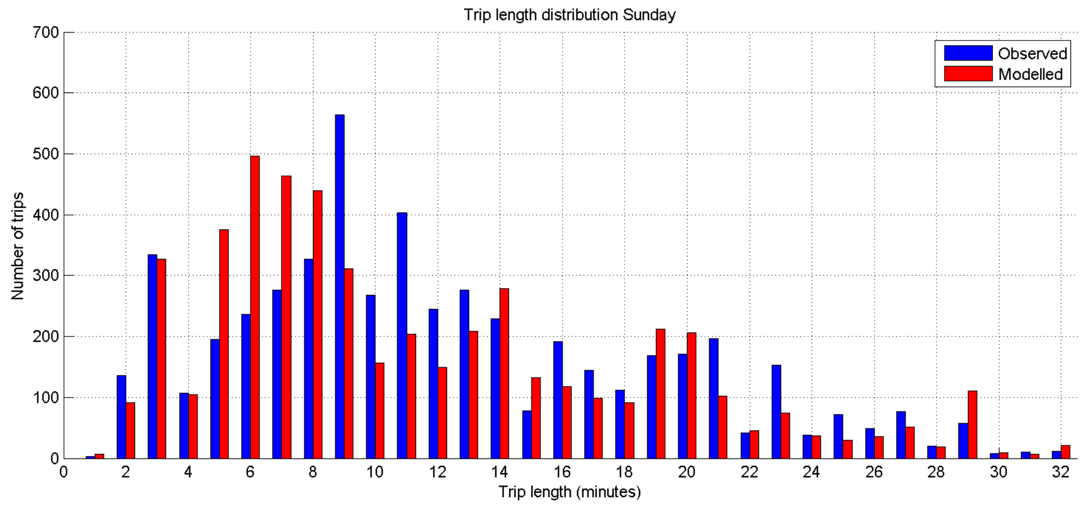


Figure 75: Observed and modelled trip length distribution in Tilburg on Sundays.

Figure 76 shows the observed and modelled number of passengers at each line segment in Tilburg. At most line segments the number of passengers is underestimated, while there are some line segments with an overestimation.

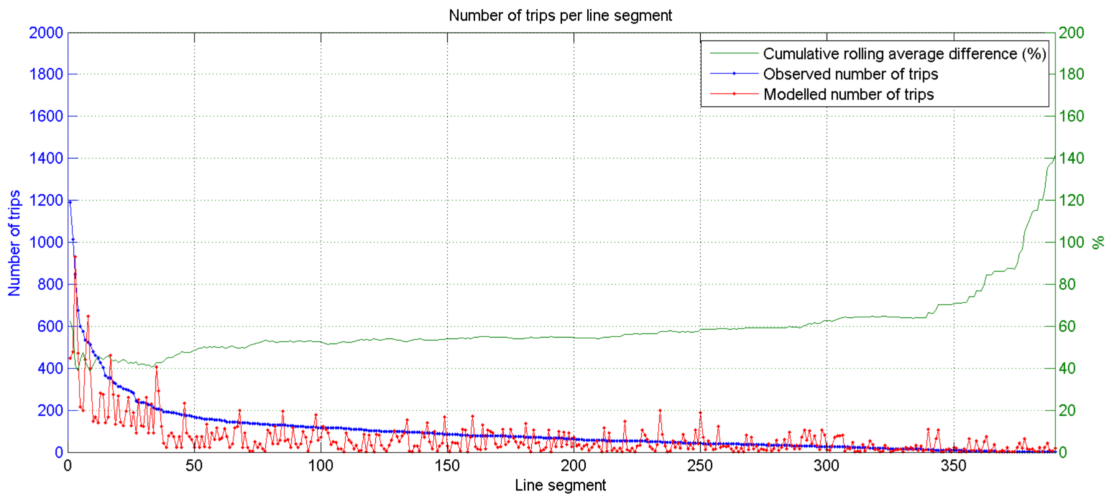


Figure 76: Observed and modelled number of passengers at each line segment in Tilburg on Sundays