STATISTICAL EXPLANATORY AND PREDICTION MODELS FOR FREE-FLOATING CARSHARING SYSTEMS

JOHANNES MÜLLER

Vollständiger Abdruck der von der Fakultät für Bauingenieurwesen und Umweltwissenschaften der Universität der Bundeswehr München zur Erlangung des akademischen Grades eines *Doktors der Ingenieurwissenschaften (Dr.-Ing.)* genehmigten Dissertation.

Gutachter

1. Univ.-Prof. Dr.-Ing. Klaus Bogenberger, Universität der Bundeswehr München
2. Professor Dr. ir. Gonçalo Correia, Delft University of Technology

Die Dissertation wurde am 01. März 2016 bei der Universität der Bundeswehr München eingereicht und durch die Fakultät für Bauingenieurwesen und Umweltwissenschaften am 02. August 2016 angenommen. Die mündliche Prüfung fand am 02. August 2016 statt.

ABSTRACT

Free-floating carsharing systems have become increasingly popular over the last years. The new mobility offer was launched without performing comprehensive target group analysis and establishing a well-structured fleet management.
This dissertation reflects and optimizes the learning-by-doing-process by giving answers to the questions which external influences have an impact on the booking demand and how booking forecasts can be performed with time series analysis models.
The basis of this work is booking data from a free-floating carsharing operator for the period of November 2011 to December 2014. At the beginning, the data is analyzed in detail on temporal and spatial level. The external influences are next to the weather land-use data, the citizens' election behavior and the local parking situation. The impacts of parking management zones and the weather turn out to be rather negligible but the other data allow to draw conclusions about the carsharing user through regression models. The typical characteristics of the customers resemble those which are found out by numerous methodologies in literature: financially well off and open for new, sustainable technologies.
The time series analysis performed better by modeling with exponential smoothing using a Holt-Winters-Filter than with ARIMA. For model calibration it is sufficient to use booking data from a period of three months.

ZUSAMMENFASSUNG

Seit einigen Jahren erfreuen sich free-floating Carsharingsysteme zunehmender Beliebtheit. Dieses neue Mobilitätsangebot wurde von verschiedenen Anbietern auf den Markt gebracht ohne tiefgründige Zielgruppenanalysen durchzuführen und ohne ein Flottenmanagement einzusetzen.
Diese Dissertation möchte diesen durch die Praxis etablierten Prozess reflektieren und optimieren, indem Antwort darauf gegeben wird, welche externen Einflüsse Auswirkungen auf die Buchungsnachfrage haben und wie mit Hilfe von Zeitreihen eine Buchungsvorhersage getroffen werden kann.
Dazu stehen Buchungsdaten eines free-floating Carsharinganbieters über den Zeitraum von November 2011 bis Dezember 2014 zur Verfügung. Zu Beginn werden diese Daten ausführlich auf zeitlicher wie räumlicher Ebene ausgewertet. Die externen Einflüsse sind neben dem Wetter Daten des sozio-ökonomischen Panels, Wahlverhalten sowie die Parksituation vor Ort. Erweisen sich die Art der Parkraumbewirtschaftung und das Wetter als eher unbedeutsam, so können mit Hilfe der anderen Daten durch Regressionsmodelle Rückschlüsse auf die Nutzer gemacht werden. Diese entsprechen dem Bild des Kunden, das bisher auch durch andere Methodiken in der Fachliteratur gezeichnet wurde: finanziell gut situiert und offen für neue, nachhaltige Technologien.
In der Zeitreihenanalyse liefert das exponentielle Glätten mit Holt-Winters-Filter ein besseres Ergebnis als das ARIMA-Modell. Zur Modellanpassung sind ein Datenzeitraum von drei Monaten ausreichend.

EXECUTIVE SUMMARY

The launch of the first free-floating carsharing system in Ulm in 2009 was an experiment for the operator. Since the success of the new system was huge, more operators joined the market and spread their vehicles in many cities all over the world.
When operators intend to implement a carsharing system in a city, several questions araise. How many vehicles should the fleet contain? How should the operating area be selected? Who will be the main customers? How can the demand for vehicles be estimated? In this dissertation, the author aimed to find answers to some of these questions. The focus was on a good understanding of the free-floating carsharing system and its users which can lead e. g. to a better definition of the operating area.

There have been a lot of studies analyzing the users' behavior usually quantitatively through surveys. What this dissertation distinguishes from existing research was the availability of real booking data of a free-floating carsharing operator. Instead of modeling the demand by intended bookings, it was possible to evaluate the observed booking behavior of the customers. The data comprised bookings from November 2011 to December 2014 for the cities of Berlin and Munich. This data was used for three purposes.

1. Analysis of booking data for a detailed description and deep understanding of the users' behavior
2. Regression models using several data sets to find explanatory variables for the varying booking demand in city districts
3. Time-series analysis for providing a precise forecast in every district

The booking data was analyzed on temporal and spatial level. The preferred time for using a vehicle was in the evening after rush hour. Since there were more bookings on Fridays and Saturdays, especially in the night, leisure activities were assumed to be a main trip purpose. Most attractive spots for trip starts and ends were places with a mix of different utilization. The combination of residence, business and shopping is typically found in city or district centers.

A map with the position of booking origins therefore also indicated where public life of a city takes place.

To figure out which factors had the most impact on the demand, several regression models with the number of bookings as dependent variable were taken into account. The independent data were land-use and spatial socio-demographic data as well as election results to characterize the milieu. Two model approaches were applied: a linear regression model and regression models for count data. The negative binomial regression proved to have the best fit for the data. Nevertheless, all models showed coherent results: Significantly more bookings were observable in districts where people were above average open for new technologies, financially well-off and prefer business or leasing cars rather than private cars. The centrality of the location is next to the availability of parking lots and the number of companies (e. g. restaurants, cafés,. . .) an important influence factor for the demand of carsharing.
The parking situation was analyzed precisely in Munich. Short-term parking lots were found to be the parking zones with the highest rate of free-floating carsharing vehicles. Other restrictions on parking did not have any effect on the number of bookings.
The influence of weather as a temporal impact on the demand was also analyzed by categorizing the weather situation on the basis of precipitation, wind and temperature for every hour of the period of analysis into "good" and "bad". The t-test was employed to compare the number of bookings during different times of the day. While the booking frequencies did not differ significantly in the data set which contains trips of all users, there were more bookings done by heavy users in the evening hours during bad weather conditions.

The time series analysis used two approaches for modeling and forecasting. The first model was a seasonal ARIMA model based on a stochastic process. The second one was an exponential smoothing model using additionally trend and seasonal smoothing parameters. This procedure is called Holt-Winters Filtering. Data sets with a period of one year, a half year, a quarter of a year and a month were compared each with both models on the basis of the Box-Jenkins approach. Holt-Winters Filtering had the much better performance and showed the best forecasting results with the quarter data set.
As a by-product of the modeling process one obtained that the spatial demand

did not vary significantly over the day. Hot spots nearly stayed the same at every time.

The results of this work can be used in many ways. The operators can transfer the regression model to cities where they plan to launch their system. The prediction of preferred destination and starts helps to define the operating area in a smarter way. The booking forecast based on the time-series model can also support the operator in detecting vehicle imbalances which is necessary for relocations. It is also conceivable to provide the forecasts for customers in the form of availability probabilities.

ACKNOWLEGEMENTS

I would like to first say a very big thank you to my first supervisor and mentor Prof. Dr. Klaus Bogenberger for all the support and encouragement he gave me. Without his guidance and constant feedback this Ph.D. would not have been achievable.

Many thanks also to my second supervisor Professor Dr. ir. Gonçalo Correia who is a recognized expert in the field of carsharing and gave me a lot of ideas and inspirations for my work.

I gratefully acknowledge the funding received towards my Ph.D. from the Federal Ministry for Environment, Nature Conservation, Building and Nuclear Safety. I really enjoyed the work within the project WiMobil and the good collaboration with the project partners especially with the Senate Department for Urban Development and the Environment in Berlin and the German Aerospace Center.

This dissertation would not have this quality without the booking data courtesy of DriveNow. My deepest thanks goes to Nico Gabriel and his team for the excellent collaboration with our institute in the last three years.

A good work is only as good as its reviewers. I want to express my biggest thanks to Williams and Harry. These guys deserve royal recognition for supporting me in the fight against linguistic mistakes. I really appreciate their efforts. A first draft was already proofread by my flatmate Tobias whom I also like to thank.

Moreover, I greatly appreciate the support received in many ways from my colleagues at the institute. A special thanks goes to Svenja and the self-declared pope of cyclists Florian for all interesting, informative, entertaining and funny discussions during our coffee breaks in the Brandl café that made stressful working periods more bearable. Some thoughts of this thesis would not exist without their inspiration.

I am also very grateful to all administrations and offices making data easily accessible. The trend for open data hopefully keeps on growing.

Finally, I am deeply thankful for my family and friends who supported me over the last years and keep believing in me during this endeavor.

PUBLICATIONS

Some ideas and figures have appeared previously in the following publications:

1. Schmöller, S., Weikl, S., Müller, J., Bogenberger, K. (2014), Empirical Data Analysis of Free-Floating Carsharing Systems, Proceedings of the Transportation Research Board 93rd Annual Meeting.

2. Müller, J., Schmöller, S., Bogenberger, K. (2014), Empirische Datenanalyse von Free Floating Car Sharing-Systemen, Berichtsband zur heureka 2014

3. Müller, J., Schmöller, S., Bogenberger, K. (2015), Empirische Datenanalyse von Free Floating Car Sharing-Systemen, Straßenverkehrstechnik 2/2015

4. Schmöller, S., Weikl, S., Müller, J., Bogenberger, K. (2015), Empirical analysis of free-floating carsharing usage: The Munich and Berlin case, Transportation Research Part C, Vol. 56, p. 34-51

5. Bogenberger, K., Müller, J. (2015), WiMobil – Wirkung von E-Carsharing-Systemen auf die Mobilität und Umwelt in urbanen Räumen, ZfAW 4/2015

6. Müller, J., Bogenberger, K. (2015), Time Series Analysis Of Booking Data Of A Free-Floating Carsharing System In Berlin (EWGT 2015, Delft, published on ScienceDirect)

7. Müller, J., Bogenberger, K. (2015), Explanatory Variables For The Varying Demand Of Free-Floating Carsharing (hEART 2015, Copenhagen)

8. Müller, J., Schmöller, S., Giesel, F. (2015), Identifying Users and Use of (Electric) Free-Floating Carsharing in Berlin and Munich (ITSC 2015, Las Palmas de Gran Canaria)

CONTENTS

1

INTRODUCTION

1.1 PREFACE

> Uber, the world's largest taxi company, owns no vehicles. Facebook, the world's most popular media owner, creates no content. Alibaba, the most valuable retailer, has no inventory. And airbnb, the world's largest accommodation provider, owns no real estate.
>
> Tom Goodwin, senior vice president of strategy and innovation at Havas Media in [66]

And the people of tomorrow? Will they still own the cars they use?

An economy consists of those who need something and those who have something that they offer on the market. This concept has stayed the same for centuries. The only thing that has changed are the products, the needs of people and also the markets.
The internet provides a powerful mechanism to bring sellers and customers together. The companies with the biggest success at the beginning of the 21st century have been those which create new markets in the form of interfaces on the world wide web. According to Goodwin, the online platforms are the places where the value and profit is because they gather the parts of people with money and the suppliers of a service.
It is not only the markets that have changed, but also the products are geared toward the individual needs of the client. Some needs appear only temporarily and the client does not necessarily need to own the product for the whole time. The internet facilitates the opportunity to share goods and services with other people.
This dissertation is about a new system of free-floating carsharing that take advantage of the broad dissemination of mobile internet. The needs of people are in this case an individual, fast and convenient mobility. The supply comes from the particular car manufacturers or companies which lease the fleet. The carsharing operator provides the interface for the customer.
The question is how this innovative mobility supply is accepted by customers and how it can help to solve the current traffic problems.

1.2 CURRENT TRAFFIC PROBLEMS FOR CITIES

In the late 1950's and 1960's, automobiles became affordable for greater parts of the population. City governments and politicians regarded an appropriate infrastructure for cars as one of the most important prerequisites for a thriving economy.
For long-distance hauling and a good connection of the rural population, a vast system of highways and freeways was built in these decades. But also cities changed their appearance in a never known celerity. Slogans like "Free roads for free citizens" were established in this time and the utopia of a car-friendly city seemed to be a most desirable goal for every mayor. The citizens' ability to reach their destinations during the day by car became associated with their quality of life.

But there are two sides of every coin.

What amounts to an increase of comfort for the driver of a car is a decrease of quality of life for all other people on the street. This problem especially concerned urban areas where street capacities could not be adjusted in an adequate way.

Prof. Knoflacher is one of the most popular vehement critics of cars. The main points of his criticism are mentioned in the following ([83]).

- *space for parking*: The benefit of an automobile is to provide mobility. But on average it serves as a means of transport for just one hour a day. This high idle time of cars entails a high demand for parking lots. So the already scarce (public) space in cities got even scarcer. Parking lots must also be regarded under an economic aspect. An average parking lot measures 20 sqm. Compared to the price of a square meter in an average downtown area in Europe one parking lot has a value of 500 to 2000 € per month ([84], p. 31).
- *pollution*: One liter of diesel produces around 2.5 kg of CO_2 ([68]). It is estimated that traffic is responsible for around 20 % of the German carbon emissions ([137]). Other exhaust gases that are by-products of combustions in engines are fine particles like PM_{10} which can cause various lung diseases (e. g. [107]).

- *noise*: A lively city is loud. But cars increase the noise level in a way that most people feel bothered. Studies (e. g. [82]) affirmed the assumption that noise provokes stress and as a consequence many typical widespread diseases.

Knoflacher regards the city life in the medieval times as the height of city planning. The high population density promoted short ways of diurnal activities such as work. Cars did not save time but bring people to spend more time for their ways according to Knoflacher's observations. And this in turn causes further isolation for people in cities.
In comparison to pedestrians and cyclists, car drivers can travel a longer distance. Cars brought cities the chance to grow enormously. But the resulting car-friendly infrastructure made the people more and more dependent on cars. The mobility system (including the low costs for gas) and the comfort and ease of using a car are the main reasons people can hardly be moved to give up their car-ownership despite of all the aforementioned disadvantages.

1.3 SOLUTION

The mobility behavior is a fundamental part of one's life and will only be changed if the advantages for oneself are noticeable. The solution can be to restrict motorized private transport or to promote other alternatives. Restrictions for cars are always difficult to enforce. Inner city tolls, an increase in taxes for gas and private cars are usually held back since politicians fear for votes.
Generally, there are two main options to solve the problems of congested cities: Reducing the quantity and distance of trips and creating an attractive alternative to a private car. These alternatives are e. g. walking, cycling or using public transport. Next to the use of other transport modes one solution is to use the car more efficiently. The idea of carsharing is that many people use and share a comparable small number of vehicles.
The idea of carsharing arises from the fact that purchasing and maintaining a private car is expensive. Carsharing gives the opportunity to use a car for a particular trip without owning the vehicle. Usually, it is organized by companies who launch their system in mainly urban areas. The fleet vehicles are available for every member in stations with fixed parking lots.

The internet changed the carsharing market radically. The first thing that has become noticeable for the customers is the greater convenience of the system.

They could check the availability of vehicles and reserve them online. Next to it, a new kind of carsharing system appears. This new system has no fixed stations but an operating area wherein the vehicles are allowed to be parked. Thanks to GPS and mobile internet, users can easily find the position of the cars in their vicinity.
The attractiveness and popularity of these free-floating carsharing systems have increased in the last years constantly. It seems that they meet the needs of people. The younger generation does not necessarily need to own a car for being mobile. In 1995, the average age of a car buyer was 46.1 years. By 2015, it has raised to 53 years ([96]). Young adults tend to use the new sharing platforms and profit from a supply which follows their individual mobility needs. Futurologist Lars Thomsen speaks in [135] about a change in the attitude towards ownership: The usage and availability of a thing becomes more important than its possession. According to him, the present time is the "age of access".

The question then arises, how this new mobility supply can help to overcome the challenges of urban car traffic.
The research project *WiMobil* ([22]) analyzed the effects of free-floating carsharing on the mobility and environment in urban areas. The impact of the system was considered regarding the three above mentioned problems of traffic.
Noise is a locally very different occurring problem in a city which is hard to quantify without an adequate number of measuring points. Carsharing is supposed to have a too little impact on the traffic that it could change any measurable improvement.
Because of the better environmental record of free-floating carsharing vehicles, pollution can lessen. The crucial point is that customers give up their private car and involve more in environmental friendly transport modes in their personal mobility behavior.
The relinquishment of private car-ownership can also have a positive impact on parking space. Carsharing first brings more cars in the city. They need additional parking space even if they are in use more often than private cars. The positive effect of a smarter use of public space just occurs when the shared mobility let the citizens' car-ownership decrease and the non utilized parking spaces are turned into spaces with a higher parking turnover or in liveable urban spaces.

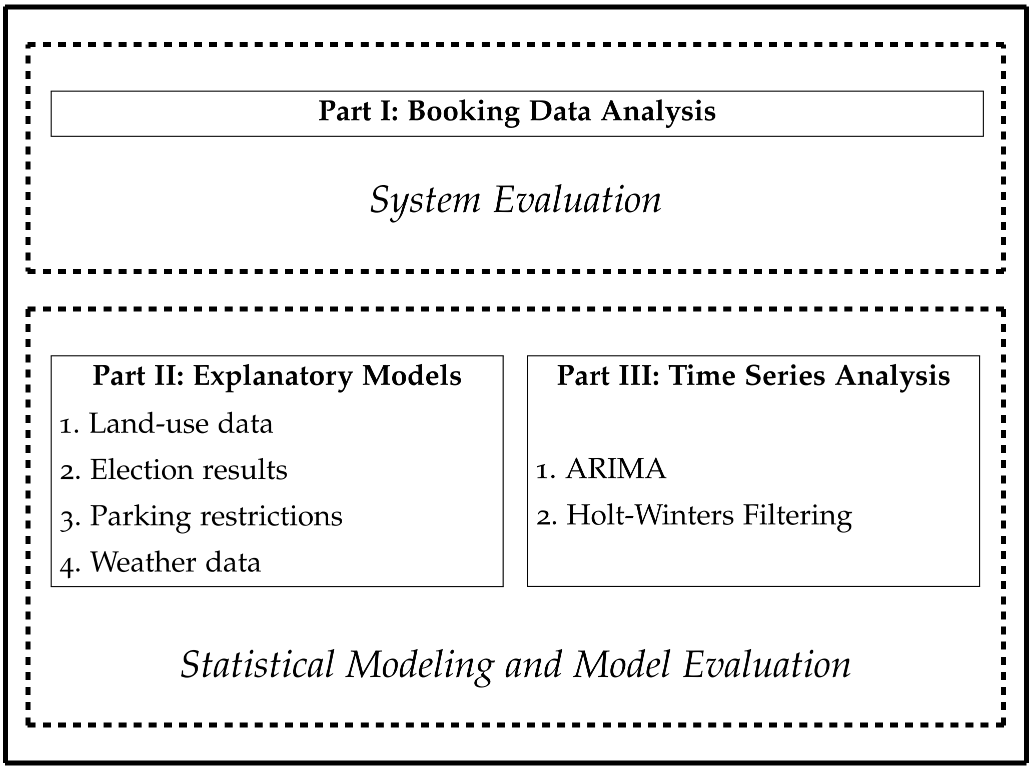


Figure 1: Outline of the dissertation

Under the assumption that free-floating carsharing has a positive effect on the traffic and parking situation, it is for the public benefit to raise the attractiveness of this mobility supply and make more people use carsharing. There are several solutions conceivable to accomplish this purpose. The solution of this dissertation is to identify the users and understand in a good way the purpose of their trips. One further goal is to support the operators by giving them tools to optimize their system. This thesis do the groundwork and focuses on two problems. The first concerns the definition of the operating area in a city. Another point of optimization is the utilization of the fleet. Relocation strategies can help to bridge possible gaps between demand and supply of cars. A precise forecast method for the demand of carsharing vehicles at a particular time and place is therefore supplied.

1.4 OUTLINE OF THE DISSERTATION

The dissertation consists of three parts. Firstly, the available FFCS booking data of Berlin and Munich are evaluated on a spatial, temporal and spatio-temporal level. A specific focus is on the temporal investigation of origins and destina-

tions of trips on a district level.
The modeling part starts with the search for variables explaining the demand of carsharing vehicles. Land-use data as well as the voting behavior of people are considered to describe typical characteristics of customers. Moreover, parking restrictions are inquired into their effect on the booking frequency in a district. A further potential temporal impact on the booking numbers that is discussed is weather.
In the last part, the FFCS booking data are considered only. For optimizing the system, a precise forecast is developed and tested. Two methods - both based on the Box-Jenkins approach - are compared. A time-series modeling with ARIMA and exponential smoothing with a Holt-Winters Filter.
A sketch of the structure of this dissertation is visualized in Fig. 1.

2

STATE OF THE ART

Although the general idea of carsharing is a few decades old the kind of carsharing this thesis is about is quite new and innovative.

The first organization that offered carsharing in a way as it is understood today is the in 1948 founded *Sefage* (Selbstfahrergemeinschaft) in Switzerland ([127]). It was a privately organized service which was deployed mainly for economic reasons. Customers were typically those people who could not afford and own a car.
The market grew slowly until the mid-1980s. 100 000 participants in the four biggest carsharing organizations were counted at this point. There were around 200 carsharing services available in the whole of Europe. The most successful were *Mobility* in Switzerland and *Stattauto* in Germany. Both are still existing station-based carsharing systems. The most successful one at the moment is *Flinkster* with 55 % share in the market and 250 000 registered customers ([39]).

2.1 THE DIFFERENT KINDS OF CARSHARING

A new idea is successful if the benefits are directly noticeable. In the case of carsharing it generally means that it must be cheaper just to use a car than to own it. And the extra time needed to reach the car must be tolerable. The idea of carsharing was consequently born in neighborhoods where people who thought about a smart and efficient use of cars.

- *neighborhood carsharing*: A car has non-negligible fix costs that make people think about sharing a vehicle with people of their social environment. One option is to share the car with a neighbor or friend and allow him to drive the car. This is obviously only practicable if there is confidence in the driver's ability. It is also popular to start a little organization where every member pays constant fees and additional fees for the usage of the common car(s). This is helpful because fix costs like taxes, insurance contributions and repair costs will be distributed fairly. The characteristic of neighborhood carsharing is the social component. It is organized privately, car owner and renter normally know each other.

It is typical that it does not take a long time until someone is making money with a good idea. This is also valid for carsharing.

- *station-based carsharing/Station cars*: Especially in the proximity of rail stations and in residential areas car sharing operators offer customers several cars for short-time rental. This is favorable because customers typically are non-car-owners and often need to get to the cars by public transport. After rental the cars have to be brought back to the same station. Customers pay in some cases an annual membership fee. If they want to use a car they have to book in advance. The costs per trip consist generally of reservation fees per hour plus a charge per driven kilometer. This primary distance-based price model has stayed typical for all station-based carsharing systems.

The two main disadvantages of this system are obvious. First, only round trips are feasible and second, the vehicle is generally not directly reachable but the renter has to make a trip from his home to the car. This last issue is a problem of the accessibility of the car. A solution - and in a way a revolution - came with the availability of mobile internet via smart phones and will be analyzed in detail later. Earlier, the first disadvantage could be compensated by allowing one-way trips.

- *one-way carsharing/multi-nodal shared-use vehicles*: This kind of carsharing system provides cars which could be returned at an arbitrary station thus allowing the customer one-way trips. The risk of such a system is the unknown distribution of cars. In the case of non-stationarity the carsharing provider can only work economically by relocating cars. Since the process of distribution is often influenced by too many factors the effectiveness of such a system is hard to predict. One famous provider is *car2go black* ([25]) operating in 8 German cities.

This classification is proposed by Barth and Shaheen in [7].

The new technology of mobile internet devices makes it possible to book and use carsharing vehicles more flexible. This was first realized by the automobile industry. Since their car sales are continuously easing or at most stagnating they tried to find a way to make an attractive offer for the changed requirements of

the European (and also American) market. This meant a change of the business strategy. They begin to see themselves more as a provider of mobility solutions than a seller of cars.
Daimler and the rental car company Europcar were the first who started a carsharing joint venture in 2009 called *car2go*. Ulm, a city with around 120 000 residents in the south of Germany, was the first test market for the new form of carsharing, which is called *free-floating carsharing* and will be abbreviated in this thesis by FFCS.

- *free-floating carsharing*: The cars do not park at fixed stations but at public parking lots in the city. The cars are usually booked only some minutes in advance. Start and end of a trip must be in the operating area of the carsharing provider. This contains in most cases the city center and outskirts with an adequate population density.

In this work the focus will rest on this kind of carsharing. There are several other kinds of "carsharing" like peer-to-peer carsharing or carpooling. However, they are not discussed in this dissertation.

2.2 THE CITIES OF THE ANALYSIS: BERLIN AND MUNICH

The basis of this thesis is booking data of a carsharing provider operating inter alia in Berlin and Munich. Before detailed information about the data will be given, the reader will get a glimpse of the demography and traffic-oriented statistics of the two cities.

Berlin is a city with a unique history. The division of the city until 1990 that had lasted for more than 40 years still has an impact today. While the western part of the city flourished during the economic boom in the 1950's and 1960's, the east German part of the city was rebuilt under the principles of town construction of the German Democratic Republic with prefabricated slab-constructions ([49], p.210/211). Today, the gaps between eastern and western parts of the city have reduced. But the differences are still visible in some points. Fig. 2 showing the level of motorization in private households indicates e. g. that the border between West and East is in a way still noticeable. Additionally, Mitte and Prenzlauer Berg developed in the last years to the most attractive, trendiest districts. Berlin did not only become the capital of Germany but also the home for many alternative, creative artists, lateral thinkers and start-up

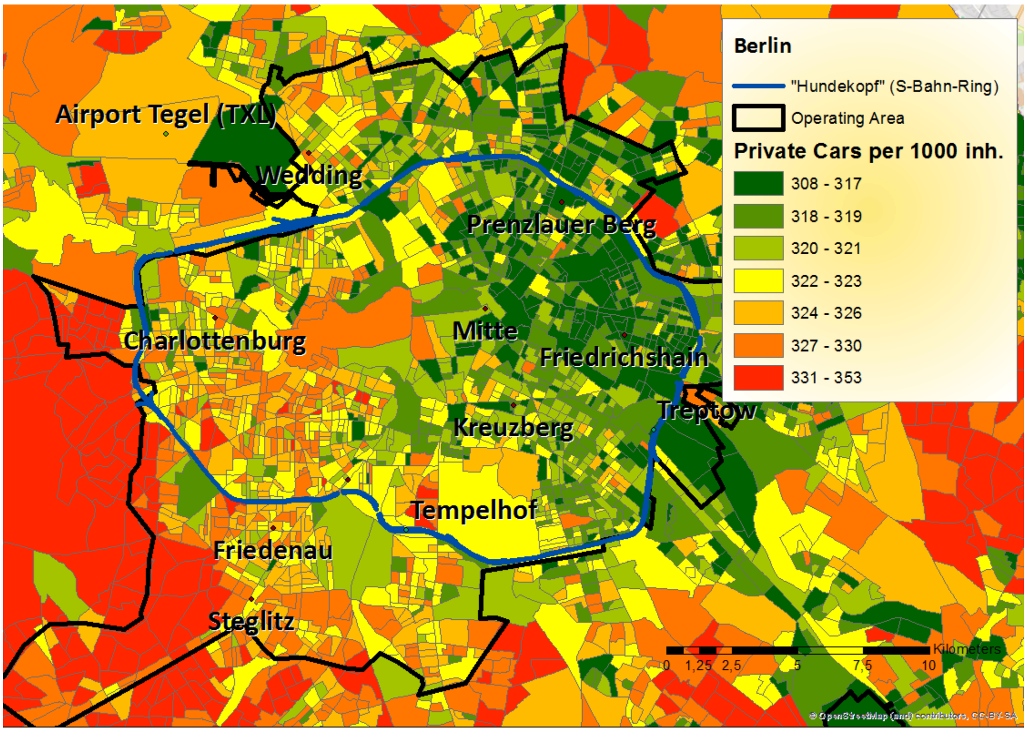


Figure 2: Map of the private car density in the districts of Berlin. The airport Schönefeld (SXF) is in the south of the city.

companies. For many Berliners, traditional values like possession have lost its importance.

Munich in the south of Germany is the capital of the state of Bavaria and the metropolitan area has become one of the most prosperous economic regions in Europe in the last decades. Eight companies appearing in the Forbes Global 2000 have their headquarters in Munich ([60]) and the city has one of the lowest unemployment rates in Germany. Many people are committed to tradition and tend to adopt a conservative attitude. This for instance is evident in the comparison of private car ownership (Table 1) between Berlin and Munich. Fig. 3 shows that even in central districts like Schwabing the rate is just average whereas new built districts (e. g. Riem) populated mostly by families have a relatively low car ownership rate.
To get a better impression the most important facts about the two cities are listed in Table 1.

So the start of the free-floating carsharing system in these two cities was an experiment on a very heterogeneous field. In the beginning it was unclear how people would accept this new kind of transport. For an operator it would have

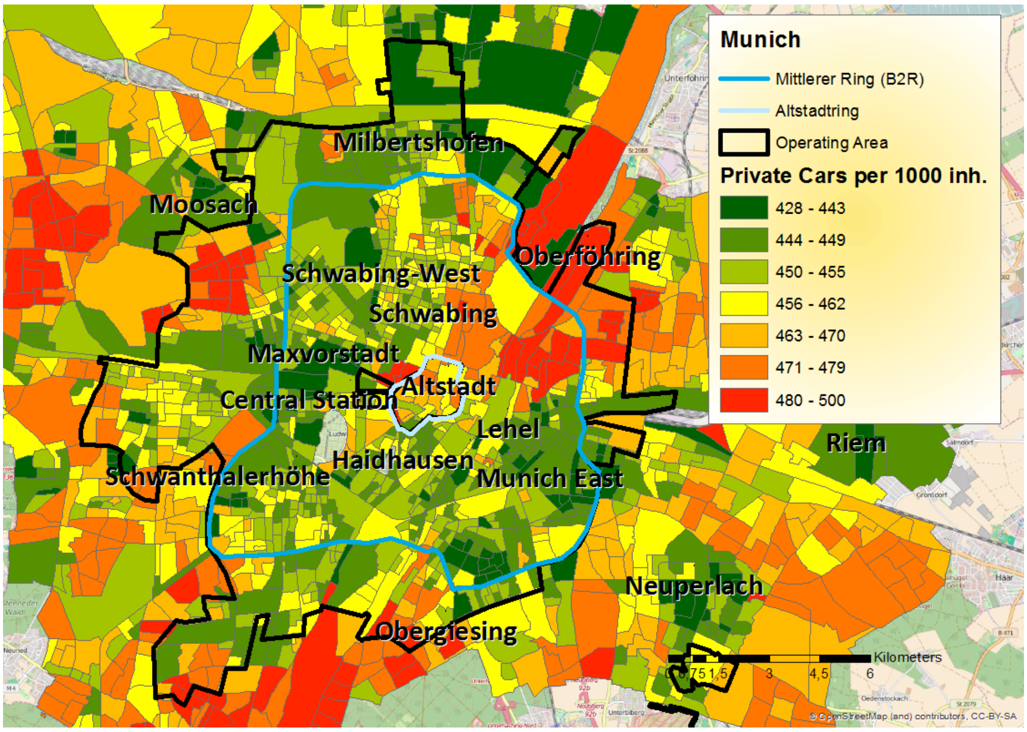


Figure 3: Map of the private car density in the districts of Munich. A satellite of the operating area is Garching 20 km in the north of the city and the airport 40 km far away in north-western direction.

	Berlin	Munich
population (city) (2014) ([1], [10])	3.37 m	1.43 m
population density (per sqkm) ([2],[10])	3 887	4 531
area (sqkm) ([3],[92])	892	311
purchasing power (index) ([64])	91.6	135.6
unemployment rate (October 2015) ([20])	10.2 %	4.4 %
registered cars (private) per 1000 (2014) ([89])	342	493
registered cars in total (2014) ([129],[131])	1 149 520	664 645
admitted driver's licenses for cars (2013) ([88])	31 610	17 165
subway (suburban train) stations ([13]([40]),[9]([41]))	173 (166)	100 (150)
modal split (ways) (public transport/car) ([125],[72])	27 %/30 %	21 %/33 %

Table 1: Comparison of Berlin and Munich

Berlin			
cambio	61 veh., 25 stations	since 2008 [12]	stationary
car2go	1200 veh.	since Apr. 2012 [24]	flexible
DriveNow	1040 veh.	since Sep. 2011 [44]	flexible
Flinkster	200 veh., 65 stations	since Nov. 2001 [38]	stationary
multicity	250 veh.	since Mar. 2011 [33]	flexible
Stadtmobil	1800 veh., 130 stations	since 2007 [128]	stationary
Munich			
car2go	500 veh.	since Jun. 2013 [24]	flexible
DriveNow	500 veh.	since Apr. 2011 [44]	flexible
Flinkster	110 veh., 42 stations	since 2003 [37]	stationary
Stattauto	450 veh., 118 stations	since 1992 [132]	stationary

Table 2: Carsharing operators in Berlin and Munich (state of Dec. 2015)

been helpful to know what the main indicators for a successful FFCS system are. The operating area could have been concentrated to promising districts which in result would have raised the profit of the carsharing company.

Air pollution is in both cities an issue. While Munich's measuring stations quantify the average PM_{10} pollution with 16-27 µg, it is between 22 and 32 µg in Berlin. On up to 48 days in 2014, the limit of 50 µg was exceeded in Berlin. In Munich, this limit for fine particles was surpassed on 8 to 17 days ([136]). Parking pressure is very high in both cities but because of the higher population density, a more severe problem exists in Munich. As one can see, the environmental problems of the motorized individual traffic are noticeable in both cities.

Smart solutions for these traffic problems are therefore much needed. Carsharing can help to reduce the number of private cars. A high number of carsharing vehicles from different providers is available in both cities. Table 2 shows the most important providers and their fleet size. There are also some other competitors like *tamyca* and *drivy* running station-based systems with private vehicles. The operator *citeecar* providing a station-based system with private parking space unfortunately went bankrupt in Dec. 2015 ([32]).

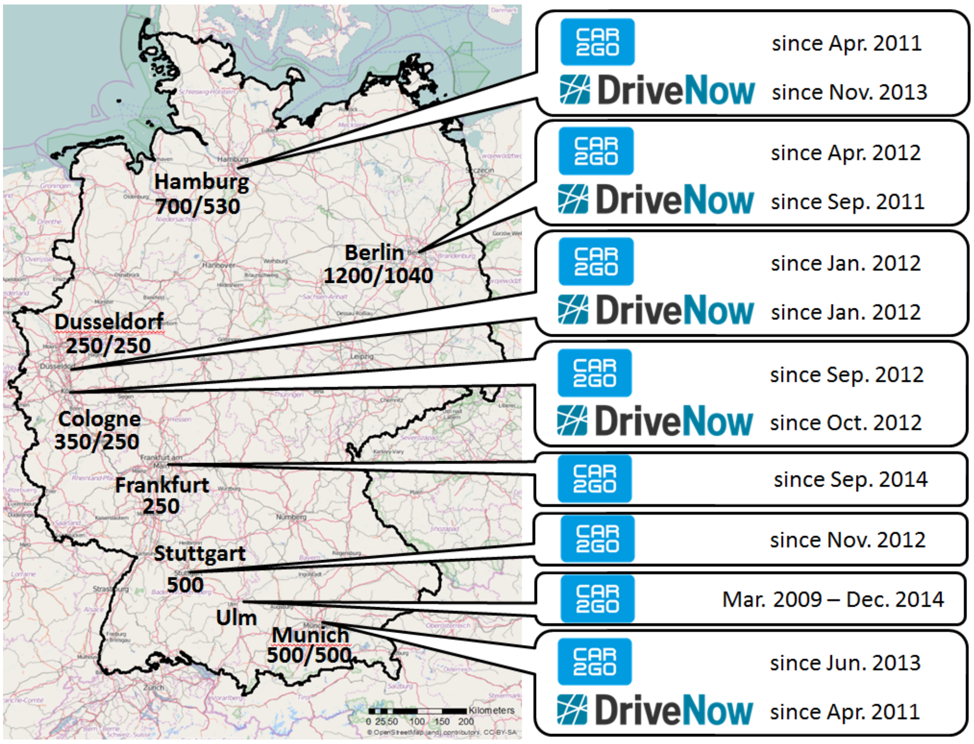


Figure 4: Operating cities of *car2go* and *DriveNow* in Germany with the actual fleet size (state of Dec. 2015).

2.3 FREE-FLOATING CARSHARING IN GERMANY

The annual report of the federal association of carsharing (*Bundesverband Carsharing*) mentioned FFCS for the first time in 2012 ([138]). But *DriveNow* had already launched its systems in June 2011 and September 2011 in Munich and Berlin, respectively. Naturally, there have been a few technical problems in the first months which reduced the quality of service. And also the heterogeneous mixture of customers – from permanent users to some who just tried it once – makes an analysis of the very first booking data not advisable. Fig. 4 shows all German cities where the market leaders *car2go* and *DriveNow* have launched their services for the last seven years. It is obvious that they concentrate their system to dense, high populated areas of the country.

With the increasing number of customers, the number of vehicles in the FFCS fleet has also grown. The provider started with 300 vehicles in Munich and 300 vehicles in Berlin. In agreement with the urban administration in Munich the fleet was not be increased for the first two years. At the beginning, the municipality was non sure about the dimension of the impact on the rare parking space in the city center. The FFCS operating area in Munich therefore contains several restrictions. The city center (Altstadt) and also numerous streets in Schwabing where only residents are allowed to park are prohibited for FFCS parking. In districts with a high parking pressure the administration demands fees for parking. Residents pay 30 € per year and are allowed to buy a parking permit which is valid in only one quarter. The carsharing operator instead has the right to park its vehicles everywhere and must in return purchase a permit for every quarter (1800 € p.a. and vehicle, Effective November 2015, [26]). The road sections that are exempted from the operating area are the resident parking zones as shown later on in Fig. 8.

In September 2013 the city council decided to allow an increase of the FFCS fleet up to 500 vehicles under the condition that less than 300 are within the circular road B2R (Mittlerer Ring) at any particular time. Berlin's fleet counts at the moment, stand at November 2015, around 1000 vehicles. Regarding the restrictions of the business area Berlin is more relaxed. Only a few streets close to the pedestrian precinct in Mitte are not for parking.

The provider uses social media platforms and his blog ([45]) to advise the customer to special parking restrictions (e. g. during the Oktoberfest in Munich) or prize competitions. The customers on the other hand can remark proposals for a better customers service and usability of the cars.

The current situation of FFCS in Germany is very positive. In August 2014, the manager magazine *Wirtschaftswoche* reported the surpass of the break-even in the FFCS sector ([120]). *DriveNow* and *car2go* are the two main operators whereby *DriveNow* is more successful. The reason for instance is that their operating area is concentrated to high populated areas in the inner cities and therefore guarantees a high utilization of the fleet.

Vehicles of the fleet are only available for members. Instead of permanent membership fees a one-time registration fee of 30 € is required. But there are often promotions which make the registration free. A reservation of the vehicle can be done up to 15 minutes in advance for free. The additional reservation time is handled as parking time which costs 0.15 € per minute and is for free between 12am and 6am. The regular price for a trip lays between 0.31 € and 0.34 € per minute and depends on the vehicle. Since the distance has no influence on the charge, the vehicles are not assumed to be driven in the most eco-friendly way. There are also prepaid minutes-packages (e. g. 500 minutes for 0.27 €/min) as well as hourly packages with limited kilometers available ([48]). There are 3h, 6h, 9h or 24h packages which cost 29 €, 54 €, 79 € or 109 €, respectively, and

include up to 200 km for free.
Some destinations require an additional fee. Trips to and from airports are examples of such areas. If the airport Berlin Tegel (TXL) or Schönefeld (SXF) is start or destination of the ride, it causes additional costs of 4 € or 6 € more per trip ([46]). The fees for the airport in Munich (MUC) are 12 € ([47]).

2.4 LITERATURE REVIEW

Carsharing is by no means a niche topic in research. With the increasing acceptance of the system and the access of new target groups, carsharing and especially FFCS has become an important research area. This section presents current research fields and results with a focus on external impacts on carsharing and demand prediction approaches.
The growing potential of carsharing has not only been observed in the management reports of the operators but also in several studies. A project of the Federal Environment Ministry (BMUB) and the Environment Protection Agency (Umweltbundesamt) established that 4 % of the German population has already used carsharing. More than 20 % of the respondents could imagine to become carsharing customers ([14]). The market research institute *infas* predicts a high potential of this market, too ([73]). Main target groups are persons who own a driver's license and use a car at most on one day of the week and public transport at least once a week.
It is obvious that carsharing has to be considered in the context of a multi-modal transport offer. Martin and Shaheen evaluated in 2011 how people's mobility behavior has changed after becoming carsharing customers. It is pleasant to see that the number of trips with public transport did not reduce whereas rides by bike enhanced significantly ([97]).
More important is the cannibalization of motorized individual transport. The project EVA-CS (Evaluation of the new flexible Carsharing offers in Munich) and WiMobil focused on the effects of carsharing on the parking situation in public space. A crucial factor is the rate of users who relinquish private car ownership because of carsharing. The determined abolition factor is in both studies 1:3 meaning that for every FFCS vehicle three public parking lots became available again ([134],[15]). Other studies get less pessimistic results. The Bundesverband Carsharing assumes that 4-8 cars are replaced ([95]) whereas every vehicle of the station-based carsharing system in Bremen should reduce the number of private cars by 8-10 ([65]). In a study by Firnkorn and Müller in

Ulm ([56]) every fourth respondent stated to forgo a car purchase if the system would last for a long time period.
Other researchers focused on the optimization of the operating business. The distribution of vehicles is assumed not to be appropriate at some times. A relocation of fleet vehicles can change this. One of the first works for this topic is published by Kek et al. in 2005 ([79]). They created a simulation for a one-way carsharing system. By their proposed relocation strategy the provider could gain a 10 % reduction of parking lots and a cost savings of 12.8 %. In later publications ([81]) they present an optimization-trend-simulation. The simulation model tested with data from a CS operator in Singapore suggests an operation that results in a reduction of 50 % staff cost. Jorge et al. used in [76] a mixed-integer programming model to find profitable locations for a one-way carsharing system. This approach was first presented by Correia and Antunes in [34] where the practical usefulness was also proven by a case study in Lisbon, Portugal. An extension of the model working with real-time vehicle stock information is described in [35]. Weikl and Bogenberger developed a relocation model for free-floating carsharing in Munich. By zoning the operating area of the FFCS provider they first detected rough spots with an unfavorable demand-supply rate and then proposed relocations on a micro-level. The model was tested within an existing carsharing system in Munich showing positive impacts on the operator's profit ([140], [141], [142]).
There are not many FFCS booking data available so researchers often fall back to traffic simulations. The agent-based software MatSim has proven to be useful for the integration of carsharing. Axhausen and Ciari from the ETH Zurich specialized on the implementation of FFCS in this simulation tool ([6], [31], [30]).
All simulations work with assumptions about the demand of carsharing. Mendes Lopez et al. for instance worked in [101] with a stochastic demand model discretized in time and space that is mostly based on travel times in the road network.
In most other works about modeling, the demand of carsharing is based on booking data which is gotten by accessing and reading the API (application programming interface) of the FFCS operator. The interface is normally used by smartphone applications and websites to provide the current distribution of available cars in the fleet. Capturing booking data via API seems to provide an exact image of real bookings but it should be treated with caution. In a study by Brockmeyer et al. (*civity study*, [19]) booking data of FFCS operators in Berlin was collected by this method. Since they could only observe if a vehicle was

available or not they could not distinguish if it was a service or a customer trip. It is supposed that their calculated trip duration is longer than the one with the original data set. But instead of the temporal use of a vehicle of around 3-4 hours they observed a time of 62 minutes. That means, in consequence, data captured via API can be thrown into great errors. Nevertheless, they should not be regarded as completely useless. Co-author of the *civity study* Weigele recognized some errors in the methodology like the overestimation of the assumed booking time ([118]).

Other studies like [139] took these data to measure the influence of particular point of interests (POI) on the number of bookings. Their approach is the zero-inflated Poisson regression. As base grid, Wagner et al. used squares with an edge length of 100 meters. Bookings as the dependent variable of the model are aggregated per cell as well as several POIs they have taken into consideration as independent variables. The zero-inflated model design excluded those cells which does not show any booking such as parks or other parking prohibited areas. The significant variables with a positive influence on the number of bookings are e. g. bars, (take-away) restaurants, the airport and areas with inhabitants that earn less than 500 € per month. A negative correlation was however observed e. g. in regions with a high educated population. Some factors like the income and education are very peculiar regarding their tendency because customer surveys in the project WiMobil identified well educated men that are in average 33 years old as typical users ([94]). Using spatial regression models with API accessed booking data make thus a characterization of carsharing customers possible. With the aid of this knowledge it is now possible to make demand predictions.

This idea will also be adopted in this thesis but applied with real booking data. A study from De Lorimier and El-Geneidy for Montréal's station-based *communauto* already tried to explain varying booking demands. By applying a multilevel regression analysis they showed that the vehicle age, the user concentration and the vicinity of stations are important factors for a high vehicle usage.

However, for understanding and predicting the use of FFCS, it is necessary to create an encompassing picture of the customers. A classic way to characterize a typical customer and his mobility behavior are surveys which can help to find attributes of an average user. Among other studies, Cervero's characterization of station-based carsharing users from 2001 ([28]) and 2002 ([29]) are among the most famous works in the carsharing research area. In his surveys,

more than 62 % of the respondents were female, the average yearly income was about $ 50 000. The study also found out that the analyzed carsharing system was mainly used during afternoon peak times for non-work purposes. An interesting result is the kind of household the users live: One third of them lived alone and every fourth shared their home with non-related adults. Cervero called them the "non-traditional" households. Although his works focused on the US market and station-based systems the kind of variables he considered seem to be helpful to draw a picture of a carsharing user.
Morency et al. also identified in [106] gender and age as significant impact variables on the carsharing behavior. Moreover, the user behavior in the previous four months directly influences the current usage frequency. Kawgan-Kagan focused in [78] on female carsharing users and revealed that female early adopters show generally a higher bike affinity and a lower open-mindedness towards new technologies than male users.
In another study by Celsor and Millard-Ball ([27]) that is based on [102] the authors emphasize the importance of the neighborhoods. They summarized the results from other researchers in four factors: parking pressure, the ability to live without a car, high population density and a mix of use of a district. Some of the points will also be considered in this dissertation.
Kumar and Bierlaire took these and further research results and modeled on base of these influence factors potential spots for carsharing stations in Nice ([90]). A study from Prettenthaler in Graz ([116]) from 1999 shows the young age of the users, too. 85 % of the respondents were between 25 and 44 years old. Since the study is some years old, it is questionable if this distribution of age is still valid for current systems. Next to the age, the education and the environmental awareness of the customers seem to have a significant influence on the frequency of use.
Stillwater et al. analyzed in [133] moreover the dependency of public transport on carsharing. Whereas the neighborhood of a light rail station have a positive impact on the demand of carsharing, regional rail availability decreases the number of bookings. An overview table of relevant studies from 1989 to 2013 about CS target groups is written by Hinkeldein et al. in [69], p. 182-186. The listed research works are analyzed regarding their query criteria like mobility related attitudes, lifestyle, family status and leisure activities.

The first work that also analyzed FFCS systems is done by Kortum and Machemehl in 2012 ([87]). The evaluated data of *car2go* in Austin showed a

	authors	year	system	location	strategy and results
[8]	Barth, Todd	1999	OWCS	CA, USA	comparing waiting time for three scenarios (non-predictive, historical predictive, exact predictive)
[80]	Kek et al.	2005	OWCS	Honda ICVS	prediction for 3-hour intervals with Neural Networks (NN), regression, selective MA, Holt's model, best results with NN
[61]	Froehlich et al.	2009	SBBS	Barcelona	prediction for 10, 20, 30, 60, 90, 120 min with 4 techniques: last value, historic mean, historic trend, Bayerian Network (BN). Best results with BN (8 % prediction error)
[77]	Kaltenbrunner et al.	2010	SBBS	Barcelona	two methods: last known value, ARMA with FIR low-pass filter, mean absolute error for 60 min: 1.39 bikes
[16]	Bourgnat et al.	2011	SBBS	Lyon	two step prediction: 1. non-stationary amplitude (linear regression), 2. hourly fluctuations (AR(1) with exogenous variables)
[62]	Gallop et al.	2012	SBBS	Vancouver	prediction with seasonal ARIMA and weather data: $+10°C \rightarrow +16.5\,\%$ bookings, $R^2 = 0.95$
[119]	Regue, Recker	2014	SBBS	Boston	prediction with a General Boosting Machine in comparison to Neural Networks for 20, 40, 60 min. Better forecasts in every case: $+1.33\,\%, +8.7\,\%, +13.27\,\%$, mean error: 0.55, 1.42, 1.68 bikes
[114]	Parikh, Ukkusuri	2015	SBBS	Antwerpen	Markov process with penalty level, 58.93 % reduction of time intervals showing a bike supply less than the penalty level

Table 3: Selected literature about prediction models in sharing systems (OWCS = one-way carsharing, SBBS = station-based bikesharing)

high acceptance and use of the system in areas with a high population and household density. A high percentage of citizens between 20 and 39 as well as students or government workers have also a positive effect. The last factor stems from the fact that many governmental agencies entered into a contract with the operator to reduce their own vehicle fleet.
Regression models are usually based on these user information and allow by this a demand prediction. A literature review about general approaches of carsharing demand estimation is published by Jorge and Correia ([75]).

These methods are nevertheless difficult to use for an exact booking frequency forecast. The prediction approach made in this dissertation is thus time series models. An overview over some relevant works about forecast approaches in sharing systems is given in Table 3.

Part I

DATA DESCRIPTION AND SYSTEM EVALUATION

As elucidated before FFCS can be one component to solve the problems of traffic and life quality which are caused by motorized individual transport. This dissertation aims to analyze the positive effects and reveal the potential of these new systems.

The first part of the dissertation consists of a detailed analysis of the present booking data. This focuses on obtaining a deep understanding of the user behavior. The spatial and spatio-temporal analyses of bookings aim to figure out what are attractive spots (hot spots) in the city for users and how they vary over the day. The temporal evaluation is useful for understanding and presuming the purpose of the trips made with carsharing vehicles. All further studies serve a better comprehension of the use of the system.

Next to the general booking data analysis the focus is on the explanation of FFCS vehicle demand. The author wants to find out reasons for the success of the system and what are probable inhibitions. Several data sets are taken and analyzed in relation to the booking data. Explanatory variables are land-use data, election results of the national parliament (Bundestag), parking restriction zones and weather data. Detailed information about what kind of variables the data sets contain is written in chapter 3. With exception of the weather data, all quantities are on a spatial level. Regression models as well as significance tests are chosen to find meaningful variables. These are standard methods in statistics to find explanatory factors. These models help to forecast booking hot spots in other cities which have no FFCS system yet. They will not be primarily made for identifying cities which promise a high potential. The optimization lays in the better definition of the operating area which could with the help of the models focus on districts with an estimated high demand.

Moreover, these models help to characterize the user. Although there is no data about the user provided the external spatial data can give information about the customer. As Seign showed in his thesis ([124]) most of the users do live in the area where the bookings start (see Fig. 5).

Modeling and explaining carsharing demand require a big research effort. It is often of interest to find the environmental and traffic effects of FFCS systems. A simulation is an adequate method but only as good as its assumptions. When

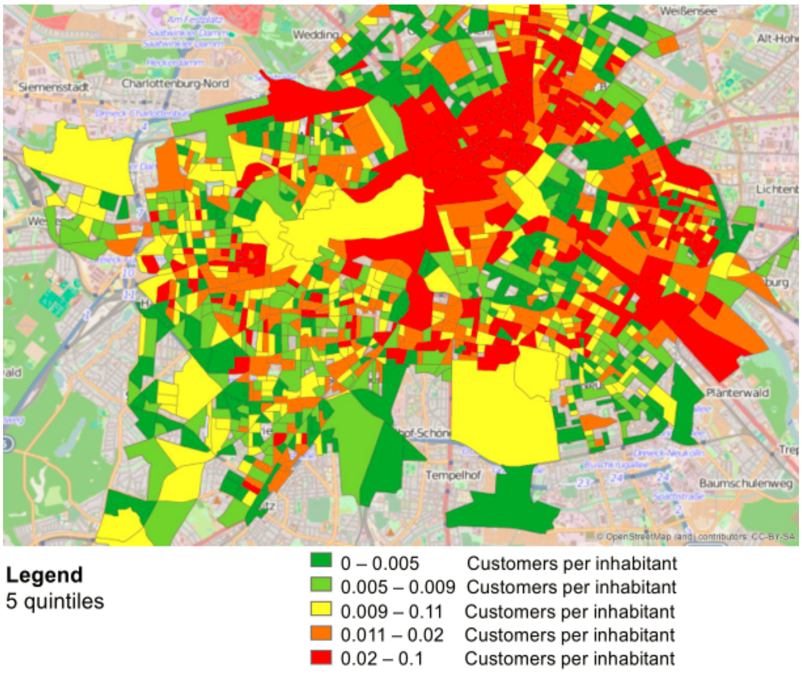


Figure 5: Customer addresses: Customer per inhabitant in Berlin, cited from [124], p.43

simulating FFCS, it is hard to predict in which city districts a FFCS system is well accepted. The analysis of real booking data is a way to remedy this lack of information.

The third and last part of the dissertation focuses on the optimization of the operating system. The system can be used in a better way if the capacity could be increased. For this purpose the operator has to know at what time how many vehicles are demanded in a particular area. Time-series analysis is the chosen method for finding a precise forecast for each district. By offering a forecast for the demand of cars and comparing it with the current supply in the districts a provider is able to estimate if a relocation is economical.

But a demand prediction is not only useful for the operator. Fliegner stated in 2002 ([57]) for classic carsharing that availability, reliability and especially comfort gain loom large for the improvement of service quality. By the provision of forecasts for the customers, the operator is able to enhance the user-friendliness. Some people may only decide to use this kind of rental service permanently if they can be sure of the reliability of the system. A prediction of the vehicle distribution can be implemented in the app of his smartphone and tell him if a carsharing vehicle is probably available at a certain time in a particular district. Next to more reliable and confident customers, new user target groups can be reached and a better popularity of FFCS systems ensured.

3

DESCRIPTION OF THE DATA

3.1 BOOKING DATA

The basis of all the following analyses is the booking data of a FFCS provider operating in Munich and Berlin. The data contain information about every trip made with one of the vehicles of the fleet. Every row in the tabular represents a trip, the columns contain the particular information about one.

An overview of the used back-end data can be found in appendix A.1.

The spatial and temporal information, available for every booking, is the most important. Since the GPS coordinates of the start and end of a trip are most precise, information in columns like `NAM1`, `JORO` and `JORI` will not be regarded. `NAM1` exists just for the provider to assign the trip to a particular city. Due to the fact that intercity trips e. g. from Berlin to Munich are not allowed every vehicle is unambiguously assigned to its home town. In May 2013, the provider changed this restriction for Cologne and Dusseldorf and allowed intercity trips between these two cities. The address of the origin of a trip is automatically created from the GPS coordinates and is shown for the customers online. It facilitates the search for an available car. But due to the text format of the field it is difficult to use and furthermore unnecessary because of the GPS coordinates. These are listed in the columns `LATO` (latitude) and `LOTO` (longitude) for the origin and in `LATI` and `LOTI` for the end of a trip. The coordinate system used in the notation is WGS 1984.
It is important to mention that only the start and end position of the vehicles are recorded. There is no tracking made by the provider but the position of each vehicle can be checked at all times if necessary. The non-availability of the exact route of the trip makes it hard to estimate whether a trip was a one way or a round-trip or to assume the purpose of the trip.
The car model is for the distinction between electric vehicles and conventional cars. This work uses bookings in general and does not focus on the different propulsion engines. The license number helps to arrange the trips by vehicle.

This is necessary for instance to calculate idle times of a vehicle.
The reservation time marks the point when the vehicle is blocked for the user whereas the start time indicates the eventual beginning of the booking. The variables `START_TIME` and `END_TIME` are the only temporal figures included in the further analyses. From the exact dates in the data it is possible to assign a trip e. g. to a day of the week or to a particular hour of the day. This is very useful for drawing accumulated booking histories.
Start and end mileage are not essential for a customer and will therefore not be part of the analysis. A greater focus will definitively be on the `DISTANCE` field which contains the difference of the two prior fields. Since billing just depends on the time in the driving (`MINF`) and parking mode (`MINH`) the driven distance is irrelevant for the customer. But it is an important indicator to understand the use of a FFCS vehicle.
The start and end zip code play a role in some analyses. Postal code areas can be used as a grid to divide the city into districts. In this thesis different background maps are used as grids. The great advantage of postal code districts is the transferability to other cities. Different district divisions based on traffic or household models cannot be applied directly to other towns without the according model.
The `PIDN` represents a customer and is used to identify the frequency of use. FFCS is a new form of transportation and it is strongly suspected that there are several customers who just tried the system once. So it seems that in order to reveal meaningful explanatory variables a focus on frequent users might be helpful.
`TRIP_TYPE` is a nominally scaled variable. A FFCS trip can be private, for business or for service. The first two are the trips the provider earns money from; service trips are made by the operator for various reasons. They can be necessary e. g. because of severe damages to a car, a lack of cleanliness the customers reported or - in the case of an EV - a low battery.
The variable `RES_TYPE` contains information about the three kinds of reservation: online, mobile or spontaneously. For the analysis of the use of carsharing, it is not essential how the booking is made. Therefore, this variable is not used in this work.
Additionally, the time between the end of a trip and the start of a new booking was calculated. These idle times of a vehicle were noted in a new column.

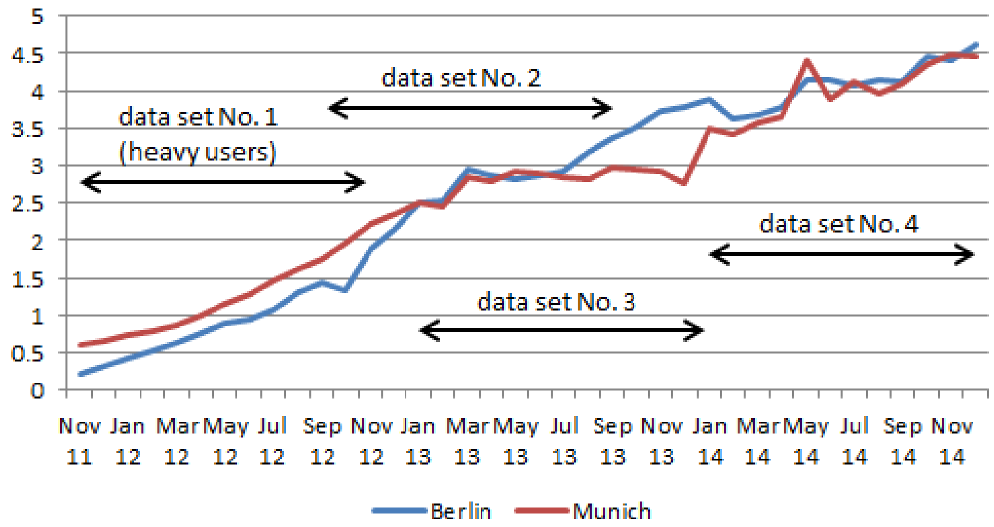


Figure 6: Number of bookings from Nov. 2011 until Dec. 2014 standardized by the total number in each city.

As usual, technical problems can appear during the booking process. For this reason, only the plausible trips are regarded for further analyses. The first stage of the data cleansing was to skip over service trips in order to only analyze and understand user behavior. The only other criteria for selecting a booking are data errors or errors occurring while recording the data. For instance, the trip is skipped if the average speed is theoretically more than 200 km/h or the booking start takes place after the end. Bookings with missing or NULL-values in one of the coordinate cells were erased, too. Thanks to the large number of bookings available, skipped data did not have an impact on the analysis.
The booking data is used from four different time periods. An overview over the data periods is shown in Fig. 6.

- data set No.1: Nov 2011 - Oct 2012 (heavy user data set)
- data set No.2: Sep 2012 - Aug 2013
- data set No.3: Jan 2013 - Dec 2013
- data set No.4: Jan 2014 - Dec 2014

This work has developed over three years and the data have been for every analysis as up-to-date as possible. This is the reason for using different annual data sets. The booking data is usually taken from the time periods of that data it is compared or modeled with.
The first data set was filtered such that only bookings of heavy users stayed

in the data set. A heavy user is defined as a customer who contributes to 80 % of the bookings done by the most frequent users. There are several reasons for considering only this special group of customers and will be explained later on. One reason is that the system was launched in both cities in 2011. It was supposed that there were a lot of customers who just tried the system for a couple of times but have no general interest in carsharing. For a characterization of users it is useful to focus on frequently customers only.

3.2 LAND-USE AND SOCIO-ECONOMIC DATA

Frequently asked questions in research about FFCS are how and where the system is used. Another focus is on the user of this carsharing system. The operator, municipalities and researchers wonder what characterizes a FFCS customer and for which purpose he uses the vehicle. To explain the use and to describe the user surveys can be conducted. But booking data offers a new chance to find relationships between origins or destinations of the trips and the use of land in these areas.

Land-use data is helpful to describe the people living in a certain area. Under the assumption that most people start their trip where they live, this census data also characterizes the drivers. By this correlation between the number of trips and the socio-economic variables of the land-use data can be found. And also without this assumption the data can generally help to understand in which districts of a city FFCS works well.

One large data set of land-use is available for regression analysis. This data was collected in 2012 by the geo infas institute, now the nexiga geomarketing company. The institute provides data for some German cities within different spatial precision. The grid of the present data is the so-called "district grid" (internal designation KGS22; KGS means Kreisgemeindeschlüssel ("county-borough-code")). The size of a district is comparable with a block in US cities with a length of 400 to 500 meters. The business area of the FFCS operator contains around 1863 districts in Berlin and 982 in Munich with a mean area of 0.18 sqkm and 0.22 sqkm, respectively. The provider describes the data as follows:

> The "district grid" (KGS22) is introduced in the official classification as a subunit. It originally comes from areal units comprised of polling districts with 400 households in average that have a maximum of homogeneity. ([109])

To get an impression of the grid the reader may take a look to the already mentioned Fig. 2 or Fig. 3 where the private car density is calculated for cells of the KGS22 grid. Next to this quantity, there are a lot of other variables available for each cell. They contain information about important factors of the population and land-use, e. g.

- demographics data: % sex, % age (categories), purchasing power, ...
- household data: % with 1,2 or 3 and more children, % single, yuppies (young urban professionals), DINKS (double income no kids), ...
- number of companies: # services, # hotels, ...
- miscellaneous: rent [per sqm], private car density, ...

The detailed description of all data is presented in Table 24 in appendix A.2.
Infas provides no information about the type of data mining they used for their data compilation. But the quality is assumed to be very high. In the technical guideline for the Act of federal geo-reference data (Technische Richtlinie Bundesgeoreferenzdatengesetz - TR BGeoRG [21], section 1.1.2.5.) from 2012 the Federal Ministry of the Interior proposed to synchronize the quality of their georeferenced address data in accordance to the geo infas data. That is also a sign for the up-to-dateness of the present geodata.
Additionally to these variables, the factors "street length" and "area size" are considered. The street length represents the number of public parking lots. Therefore only street types where parking is usually possible are regarded. The OSM streets of type "primary", "secondary", "tertiary", residential" and "living street" were selected for this purpose. The area size is also taken into the model since the district sizes differ and may need to be standardized.
The theory and application of the performed regression analysis between the land-use variables and the number of bookings is described in section 5.1 and 6.1, respectively.

3.3 RESULTS OF THE BUNDESTAG ELECTION 2013

In conversations and talks with other carsharing experts the opinion arose that the success of FFCS depends mainly on the milieu of the urban district. Socio-economic data are one instrument to characterize the environment but they do not cover all variables to describe if a district is multi-faceted or a conservative residential district.

One method to measure the milieu is to consider the voting behavior. Intuitively, trendy districts with a high rate of freelancers and generally alternatively living population tend to vote for more progressive, left wing parties whereas residential or even rural areas prefer to vote conservatively and right-wing. In other words the more urban a city district is the more difficult is an election victory for a conservative party ([110]).
Under the assumption that the voting results are an indicator for the urbanization of a district it makes sense to set the results of the election in relation to the aggregated booking data of a district. The regression analysis will thus be performed for these data in the same way as for land-use data.
Afterwards, it is necessary to interpret the results by detecting which milieus are represented by positively and negatively correlated parties. In a study from 1995 ([115]) Petersen exposed that customers of station-based carsharing systems are mainly Green party voters which show equally a high ecological awareness.

The most useful election to indicate the general voting behavior is the election for the national parliament, that is in Germany the Bundestag. The last election of the Bundestag was on September 22nd, 2013.
Germany is split into 299 constituencies (*Wahlkreis*). These are again divided into districts with one polling station each, called polling district (*Stimmbezirk* or *Wahlbezirk*). Every inhabitant of this district must elect in the respective polling station (or alternatively via postal vote).

In Fig. 7 one can see the twelve (six) constituencies and 1701 (704) polling districts of Berlin (Munich). The best chance to achieve a sound correlation relation to the FFCS booking data is to use the polling districts as grid. The constituencies are simply too imprecise.
Among others, the following parties stand for election: CDU/CSU (center-right), SPD (center-left), Die Linke (far-left), Bündnis 90/Die Grünen (in following: Die Grünen, Greens), FDP (liberals), NPD (far-right), Piraten (Pirates), AfD (right-wing populist). There are some minor political parties which have not been considered.
At the election of the Bundestag, every voting citizen has two votes: With the first one (*Erststimme*) people vote for the direct candidate of their constituency who usually comes from the CDU, CSU or SPD or in Eastern parts of Germany not uncommonly from Die Linke. The second vote (*Zweitstimme*) is the more

	Berlin	Munich	Hamburg	Cologne
CDU, CSU	28.5 %	37.8 %	34.4 %	33.0 %
SPD	24.6 %	23.9 %	37.8 %	29.8 %
Die Linke	18.5 %	4.6 %	7.5 %	8.1 %
Die Grünen	12.3 %	14.1 %	10.6 %	14.1 %
FDP	3.6 %	7.7 %	2.0 %	6.0 %
AfD	4.9 %	4.5 %	3.4 %	3.5 %
Piraten	3.6 %	2.5 %	2.5 %	2.6 %
NPD	1.5 %	0.4 %	0.6 %	0.6 %
Others	2.5 %	4.5 %	1.2 %	2.2 %

Table 4: Second vote results of the Bundestag election 2013 for Berlin, Munich, Hamburg and Cologne.

important one and decides about the ration of the parties represented in the Bundestag.

The data which was used for the regression analysis are neither absolute nor percentage values. Taking these absolute results becomes a problem when one transfers a regression model to another city because the general tendency in cities varies markedly. A solution is to regard the differences of the results in each polling district and the average of the city or the constituency. In consequence, the voting behavior in a district becomes comparable to other regions. The results of the first vote are subtracted by the average in each constituency. The reference values for the second vote are the results of the corresponding city that are listed in Table 4.

The cities of Hamburg and Cologne appear as the regression model of Berlin will be transferred to and compared with these cities. The FFCS operator thankfully provides additionally to the booking data for Berlin and Munich a data set for these two cities. The data are taken from the period of November 2013 to January 2013. Both systems were launched in 2013; the service in Hamburg just started in autumn 2013.

One entitled objection to the proposed approach may be that the voting behavior rather depends on the current election programs of the parties and the confidence with the politicians than with the place of residence. This seems evident regarding someone's individual voting behavior. But the effects are usually

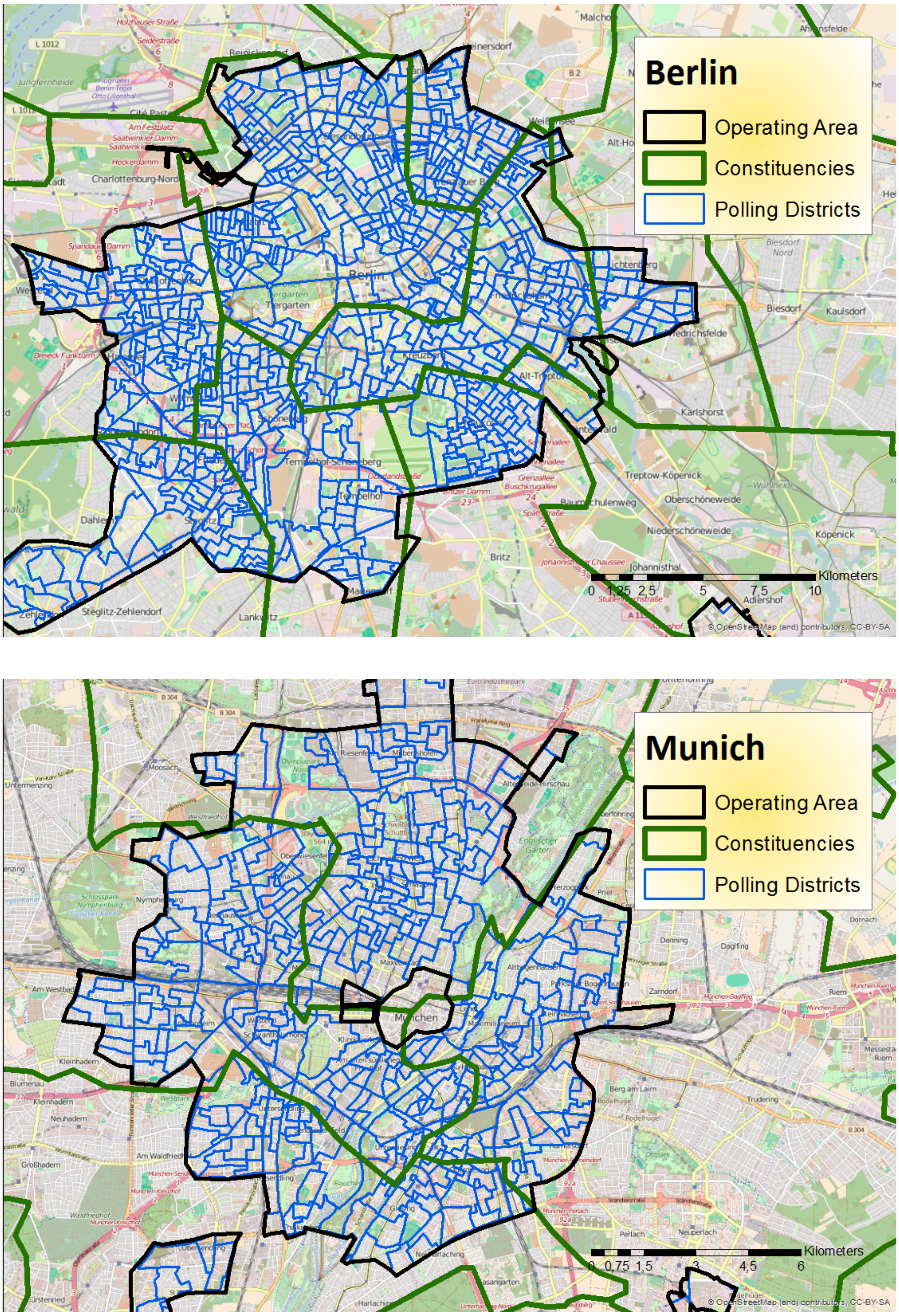


Figure 7: Constituencies and polling districts in Berlin and Munich at the Bundestagswahl 2013.

trends for the entire population. The rate of floating voters is not equal for rural and urban areas but can be at least assumed as evenly distributed in a city. Under this presumption the performed normalization of the election results is sufficient for obtaining a model which is not only transferable to other cities but also to upcoming elections.
The voting behavior is therefore more or less an indicator for the milieu of a district. Election forecasts work with that knowledge by weighting chosen district samples on base of socio-demographic data or former election results ([50]). The assumption that the rate of floating voters does not change significantly over the times is also underlined by the fact that gerrymandering is prohibited. The name stems from Massachusetts' governor Elbridge Gerry who signed in 1812 a law that changes the form of the constituencies in a way that the opposition won in only every fourth district with more than 51 % of the votes in total ([55]). He used the circumstance that specific areas generally prefer a particular party. Since his division of the state makes some constituencies look eye-catching ([54]), e. g. like a salamander, this form of election fraud is called gerrymandering.
All data for the Bundestag election – the results as well as the geo-referenced shapefiles – were kindly provided by the municipalities of the respective cities (Amt für Statistik Berlin-Brandenburg, Landeshauptstadt München, Statistikamt Nord, Stadt Köln) free of any obligations and – with the exception of Hamburg – even free of charge.

3.4 PARKING DATA

As described in the introduction public space in cities has become a more and more contested issue. The public space that parking lots need is disproportionately high compared to the number of people who use it. Therefore the municipal governments can establish parking fees in distinct areas to decrease the parking pressure. Details are determined in §45, sec. 1b it. 2a StVO (Straßenverkehrsordnung, German Road Traffic Act, see [23]).

> The road authorities give the necessary instructions [...] in connection with the marking of parking options for residents of urban districts with a considerable lack of parking space by entirely or temporary restricted reservation of parking space for the eligible groups of persons or by arrangement of exemption of the arranged parking management measures.

There are two important things noted in this act. First, the parking management is managed locally. The road authorities – i. e. in cities the municipal government – determine in which areas the parking fees have to be paid. And second, parking management is only allowed in areas with a high parking pressure. In case of a judicial review, the road authorities have to prove the "considerable lack of parking space". In conclusion, parking license areas are always areas with a supernormal parking pressure. Licensing parking areas can thus be an indicator for parking pressure but launching a system for parking fees also mostly intends to regulate the stationary traffic. The definition of parking pressure is difficult and depends on the time of the day, too. Taking restricted parking areas into account and setting them in correlation with FFCS booking data can thus not directly measure the influence of parking pressure on carsharing but an analysis can find a potential preference of carsharing users for a particular parking zone. This could in consequence be useful for municipalities to promote carsharing.
In Berlin, there are 40 parking zones with 103 210 parking lots that are currently part of the parking management ([126]). The parking fee for 15 minutes varies between 0.25 € and 0.75 €. They can be paid directly at the vending machines or cashless via mobile phone. Residents can purchase a license for the zone where the car owner is registered.
In Munich, the situation will be analyzed very precisely. After a council order from October 2005, the urban administration in Munich was forced to optimize the parking situation in the city. Next to the improvement of the living environment of the residents one goal of the order was to use the parking space more effective ([93].) As one result the parking management was introduced in most parts of the inner city. They distinguish the parking license areas generally into three zones.

- *resident parking:* Parking lots in these areas are reserved for residents only. The residents have to buy a license for their area for 30 € a year. Only one license per car is permitted. The regulation is valid on every day except on Sundays and holidays from 9 am till 11 pm.

- *mixed parking:* These zones work like resident parking zones with the difference that also visitors without a parking license are allowed. They have to pay a fee of 1 €/h (max. 6 €/day).

- *short-term parking:* Everybody has to pay a fee of 1 €/h for parking. The maximum parking time is limited to 2 h.

parking zone	details
no parking license area	
mixed parking	1,00 €/h, max. 6 €/day; free with parking license
short-term parking	until 6 pm: 1,00 €/h, max. 2 h; from 6 pm on free with a parking license
short-term parking	until 6 pm: 1,00 €/h, max. 3 h; from 6 pm on free with a parking license
short-term and resident parking	9 am-11 pm: 2,50 €/h, max. 2 h; from 7 pm on free with a parking license
short-term and mixed parking	until 6 pm: 1,00 €/h, max. 2 h; from 6 pm on mixed parking
stopping and parking prohibition and mixed parking	from 7 pm on mixed parking
from 7 pm on resident parking	
mixed and resident parking	until 6 pm mixed parking; from 6 pm on resident parking
mixed parking with parking disk	max. 2 h; free with parking license
mixed parking with parking disk	max. 4 h; free with parking license
parking zone	until 7 pm: free with parking license, max. 3 h for visitors with parking disk; from 7 pm on for free
stopping and parking prohibition	
resident parking	residents with a parking license only

Table 5: Parking license areas in Munich. Green colored zones are allowed for carsharing vehicles at all time, red ones are always prohibited. The yellow zones are the short-term parking zones, in the orange areas it is allowed to park at some time.

The concrete realization of the parking management is more complex than this distinction suggests. The 62 zones were not classified in resident, mixed or short time parking, but every street got a certain regulation that is oriented towards the three kind of parking zones. The exact progress of the determination of the parking zone areas was supported scientifically by the TU Munich and described in detail by Hanitzsch et al. in [67]. As consequence of discussions with the citizens and representatives of the local economy, not just three but 13 different kind of parking areas were built. The detailed distinction was made to respond in the best way to the interest of the residents and the on-site companies that are dependent on parking space. Table 5 shows the list of the different parking areas. Their distribution over the city is mapped in Fig. 8.

Mixed parking zones have the biggest percentage of all areas. Fortunately, FFCS vehicles are permitted to park in these streets. When ending the trip in a temporarily prohibited area the customer usually gets a message on the screen with the information to comply with the local parking restrictions. Parking

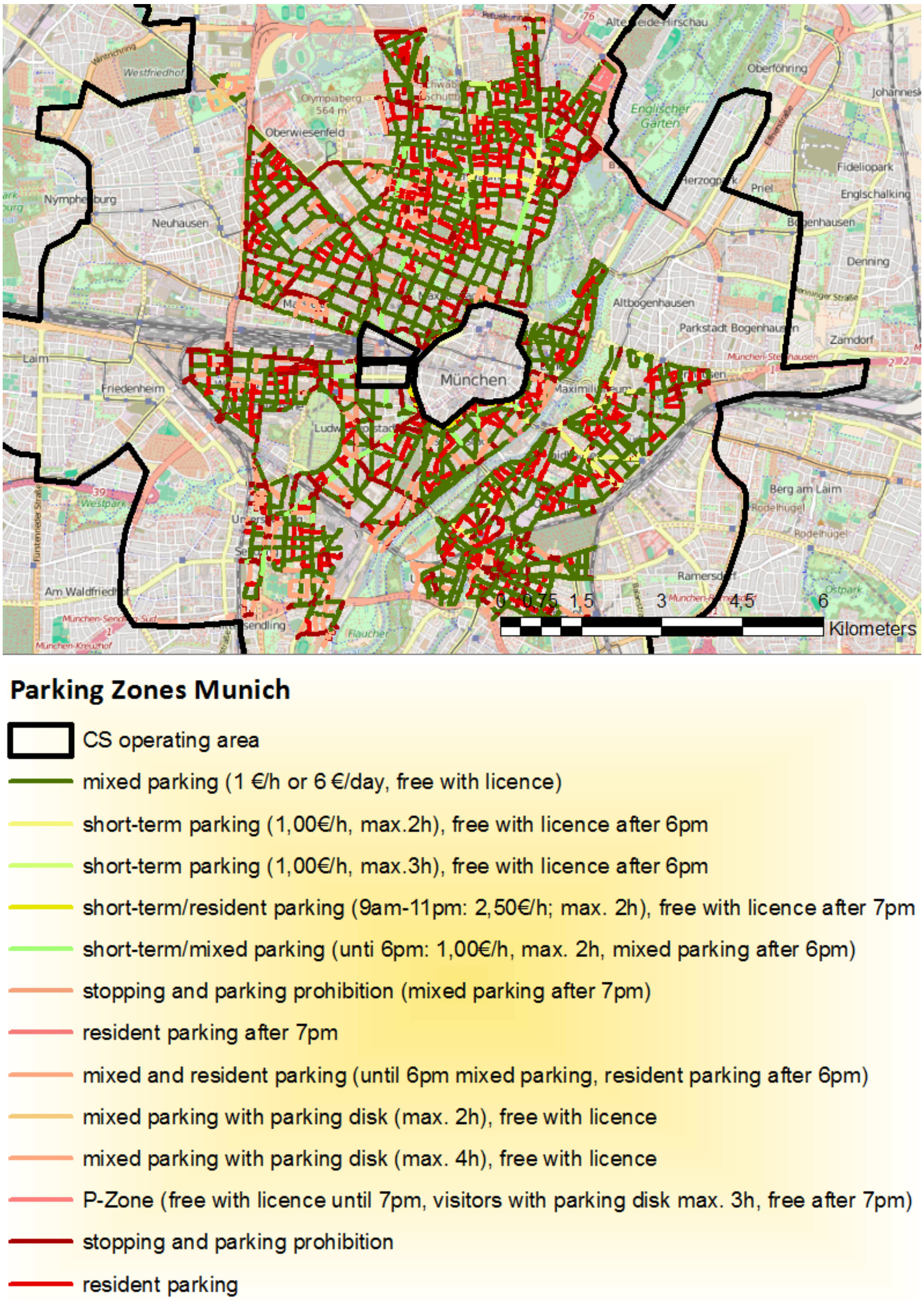


Figure 8: Parking management zones in Munich

in these areas is always a risk for the customers. If they park for instance in the afternoon in a mixed parking area that becomes in the evening hours a residential parking area he must hope for another customer to drive the car away before parking becomes prohibited. Or he expects no control of parking inspectors.
Short-term parking is simpler. If a traffic warden wants to impose a fine for the car he has to prove that the car has been parked for more than the allowed period. So he has to come again to that spot after some hours. It is first not so probable that the restrictions are checked at the time of parking. Second, it becomes more unlikely that the restrictions are checked twice to fine excessive parking durations. And third, the customer gets after some time of using FFCS a feeling for the way the system works and can expect that the car will be booked within the allowed parking time. Parking in a short-term parking zone is therefore a very low risk for the customer. If a customer gets a ticket he is the one that has to pay for it due to the terms of use. However, it is assumed that the operator handles parking fines very accommodatingly.
Parking in a permanently restricted area will be shown on the screen in the vehicles. Normally, the booking should even not be able to be finished in that area. But since the restricted zones are sometimes very small and the GPS – especially in narrow streets – does not show an exact position of the car in every case the formal restriction of parking prohibition can become useless. The official handling for parking offender is that they pay 1 € per kilometer distance from the city center additionally.

3.5 WEATHER DATA

Expert interviews with the operator of the carsharing system showed that there are more bookings during bad weather conditions. It was considered as a challenge to prove this subjective assessment in an objective way.
The presumption of the operator included that the current weather conditions play a key role for the booking. The weather data have to be available in a very precise form, at least with hourly measurements.
Fortunately, the DWD (Deutscher Wetterdienst, German Weather Service) provides for some of its measuring stations historical data about the weather conditions in the needed precision. The stations used for the present analysis are the airport Berlin-Tegel and Munich-City. Berlin-Tegel is the airport within the city (see Fig. 2). Therefore the data of both cities have a comparable precision.

Temperature	Winter (Dec, Jan, Feb)	$< -2°C$ ($< 28°F$)
	Spring (Mar, Apr, May)	$< 5°C$ ($< 41°F$)
	Summer (Jun, Jul, Aug)	$< 15°C$ ($< 59°F$)
	Autumn (Sep, Oct, Nov)	$< 5°C$ ($< 41°F$)
Precipitation		$> 0.5\,\mathrm{mm}$
Wind force		$> 3\,\mathrm{Bft.}$

Table 6: Definition of bad weather conditions

The relevant data used to describe bad weather are the temperature [in °C], precipitation [in mm] and the wind force [in $\mathrm{Bft.}$]. One option would be to take these data and find an antiproportional or proportional relationship between the number of bookings and the three variables. But it is assumed that the weather in general is more important for the choice of transport mode than the quantity of rain or the like. Therefore it is more useful to find a tolerance limit from the combination of the three variables. Exceeding this limit means that most people estimate the weather as "bad" and are probably more willing to choose a car for their trip.
The difficulty is that there does not exist a formal definition of the term "bad weather condition". Eugster considers in his diploma thesis "Einfluss des Wetters auf das Verkehrsverhalten" (Influence of the weather to travel behavior, [52]) weather data in combination with travel purposes and travel distances. His approach to define "bad weather" was to use daily mean precipitation and average temperature values. A bad weather day was in that work considered as a day with a lower mean temperature than the previous day and a precipitation during the day in addition. Good weather days are on the contrary days with a higher mean temperature than the previous day with no precipitation over the day at all. One problem of this definition is that there are days, e. g. those with precipitation and a higher mean temperature, that cannot be assigned to this characteristic. Further, the classification does not help for the current case since it is too imprecise.
As a solution, the official formulations for weather news published by the DWD ([43]) were used to find an appropriate definition of bad weather conditions.

In Table 6, the conditions of bad weather are listed. At least one of the linked conditions has to be fulfilled to speak of bad weather. In the lexicon of the DWD ([42]) the Beaufort scale is explained. A wind force of 4 Bft. and more means a

wind with more than 15 km/h. Riding a bicycle as an alternative transportation mode is then definitely not comfortable.
Two data sets are analyzed with two different methods: The first one is data set No. 1, the second data set No. 3. The first approach with data set No.1 aims to check a significance in the difference between the booking distributions of the good and bad weather conditions. The second method primarily compares the number of bookings during good and bad weather conditions for every weekday and daytime. In a second step the results are tested for significance as well.

4

GENERAL BOOKING DATA ANALYSIS

As first part of this dissertation, the available booking data set is evaluated. In contrary to fixed stations where only the different booking frequencies at the stations over the day and week are relevant for research, the flexibility of FFCS raise more questions. Where do customers park preferably in the city? At what time do they use carsharing mostly? Are there any differences in the spatial distribution of vehicles during the day?

The booking data offer a clear illustration of the use of carsharing. The data is first analyzed spatially. After discussing heat maps of Berlin and Munich the bookings are considered on the temporal level and eventually in their spatio-temporal development. In the last section, various indicators about the system are mentioned as well as a consideration of the average speed and trip duration over the day and week.

4.1 SPATIAL ANALYSIS

For both cities, data from the whole year 2014 is considered (data set No. 4). Regarding the spatial analysis there is no difference between start and end points because every end point of a trip is the spot where the next trip starts. The maps in Fig. 9 show the Kernel Density of booking starts for Berlin and Munich. The Kernel Density colors squares of a predetermined raster according to the number of bookings in a fixed surrounding. The base raster is chosen very small so that the maps show a very smooth distribution of bookings.

In Berlin, the area with the most number of bookings is around Prenzlauer Berg. It is known as a modern and hip district with a progressive lifestyle. However, it has also gone through the process of gentrification and thus has become only affordable for higher earners. There are also several other spots highly demanded. The hot spots are in the centers of the districts Mitte, Friedrichshain, Kreuzberg, Schöneberg and Charlottenburg. It is in the nature of the method that all highly demanded areas are not at the edge of the operating area. The Kernel Density considers surroundings of each spot. Non-central locations are always belted by lower demanded areas and therefore not colored as a hot spot.

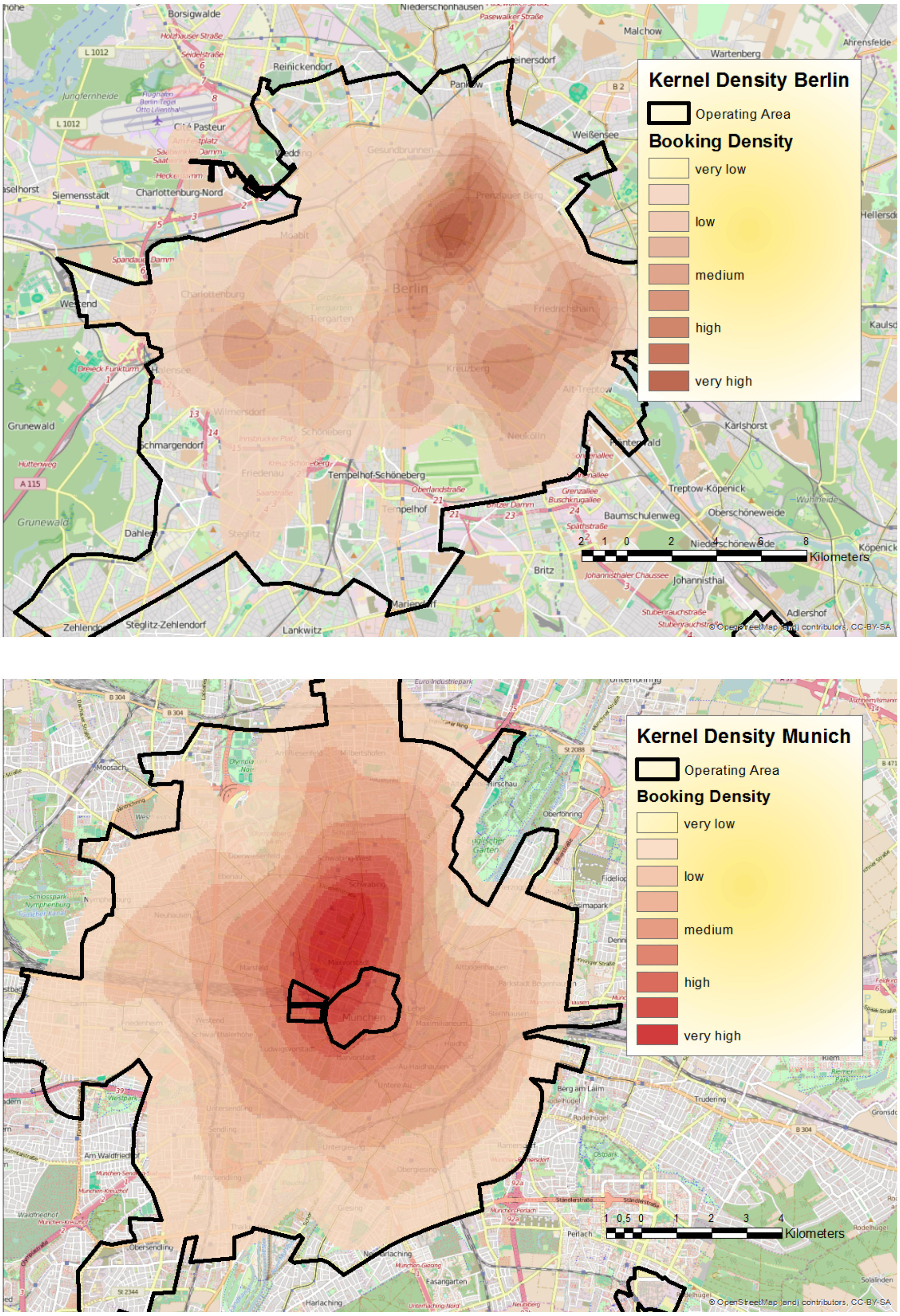


Figure 9: Maps of Berlin and Munich colored with the Kernel Density Estimation of booking frequencies

Steglitz in the south-west is an example for such a district. The later used maps in Fig. 26 and 27 for Berlin and Munich, respectively, show the district grid colored by the number of booking in quintile steps. Some districts have a high absolute number of bookings but because of the less demanded neighbor cells the Kernel Density does not show this location as a hot spot.
The situation is different in Munich. The bookings concentrate in the university district Schwabing. The density of bookings decrease in the periphery of the city. The map shows as well a high density close to the inner city that is excluded from the operating area. It can be interpreted as an indicator for a high potential demand within the Altstadtring. Munich has also in the aggregated map illustration a strong central focus. Only some parts in the east of the station Munich East have a high booking frequency which could not be seen in the map of the Kernel Density.
The heat maps do not only show the hot spots for FFCS. They also reveal the structure of a city. Whereas Munich is a mono-central metropolis, Berlin has several district centers with a prosperous city life. Berlin can hence be called a drive-in-and-out-and-through-city while Munich is a typical drive-in-and-out-city. It is interesting to see that this is also visible in the booking data.
It is moreover noticeable that there is no distinct booking concentration around public transport stations. It could be concluded from previous studies about station-based carsharing mentioned in section 2.4 that there might be a higher demand around stations for local trains. There are several reasons for the fact that this is not directly valid for FFCS systems. A possible explanation is that the search radius for the Kernel Density was chosen too large so that the fine structures of stations could not be mapped. A second more probable reason is the missing parking space in the area around the stations. There are almost no park & ride facilities in central districts available and even if they exist they have no significant higher bookings. Therefore it is not very likely that flexible carsharing is used for intermodal transport.

4.2 temporal analysis

A look at the daily profiles of bookings in Fig. 10 shows a similar shape of the function for Berlin and Munich. The explanation is therefore valid for both cities. The temporal demand over the day has on workdays a typical two-peak profile which can also be regarded for general trips. The dotted line marks the percentage trip frequencies for Germany based on the MiD 2008 study.

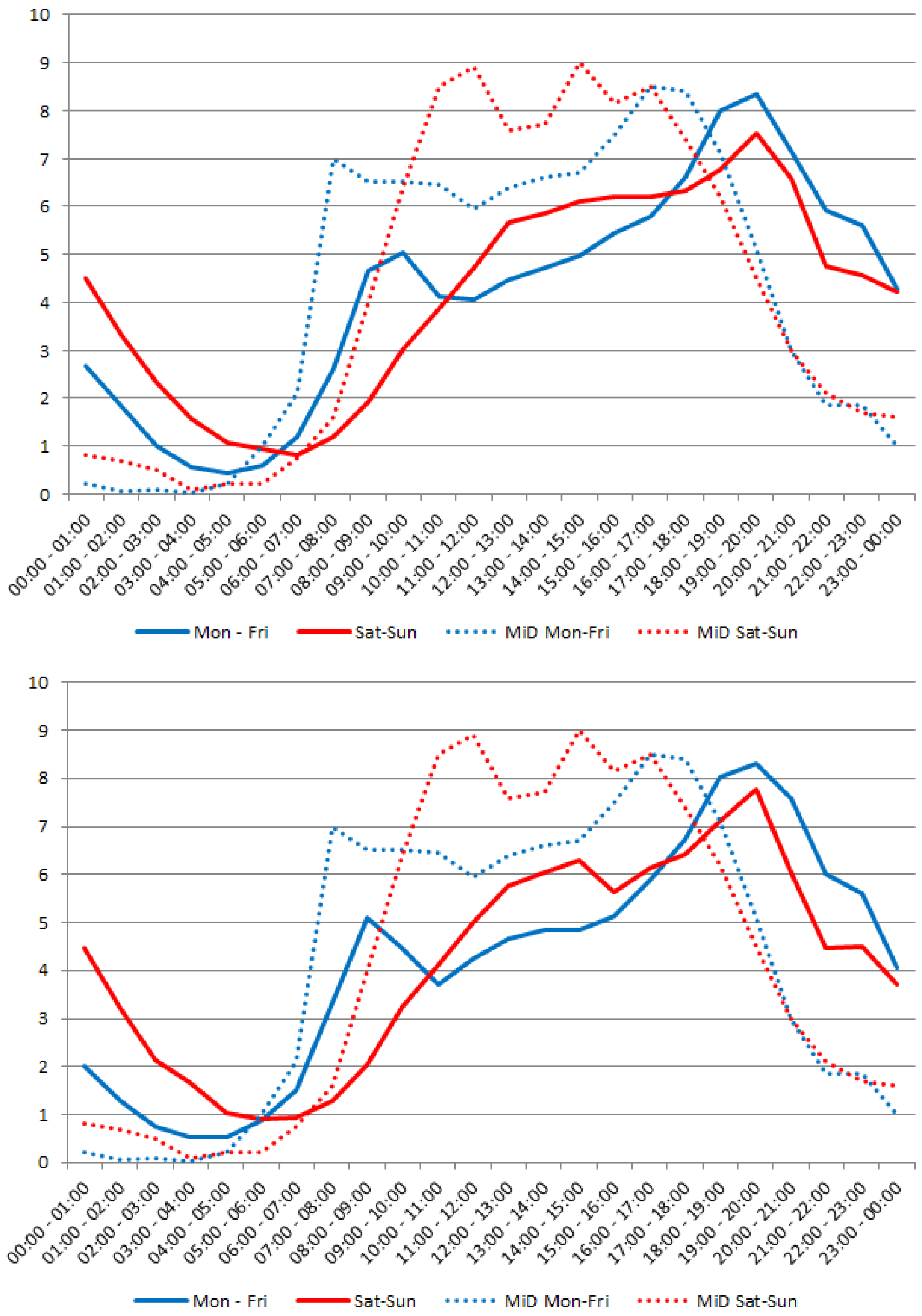


Figure 10: Percentage of trip starts in Berlin (above) and Munich (below) in comparison with the average percentage of trips starts according to MiD 2008.

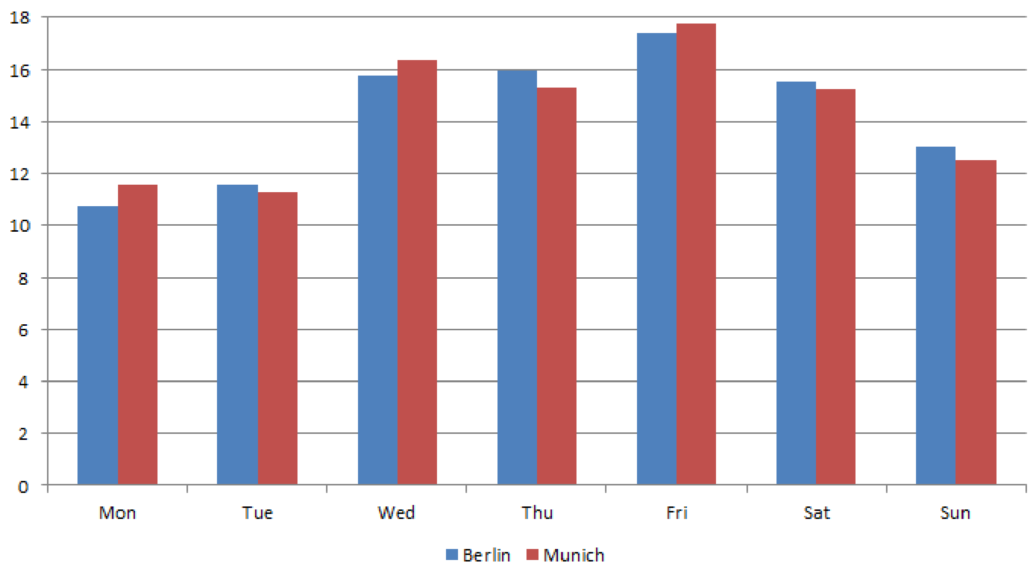


Figure 11: Bar chart showing the number of bookings (percentage) for Berlin (blue) and Munich (red) aggregated by weekdays.

However, the MiD peaks on weekdays are both a bit earlier than the local maxima of booking demand. That delay is assumed to result from two things: First, it is not very likely that FFCS is used for daily routine trips like rides to work. This would simply not be profitable for the customer. But the main use during morning hours is nevertheless business trips. That was found out by onboard questionnaires within the WiMobil research project from the DLR and BMW ([105]). Users were asked to call the purpose for their trips directly after finishing the booking. By this, the purposes for trips could be analyzed on a temporal level. It is possible that the business trips are not the routes from home to work but rides to other business meetings in the city. A second explanation is that users chose consciously to drive off-peak to save travel time and costs.

The second peak is between 5 pm and 9 pm. Trips for leisure time activities are the main purpose during that time. The use on weekends is different from workdays. It neither corresponds to the profile of the MiD. Usually, most of the trips are between 10 am and 5 pm. The booking demand for FFCS, however, concentrates on the late afternoon and evening times. This indicates again a strong use of the vehicles for non-routine trips in the spare time.

This proves to be true also in comparison to other weekdays (Fig. 11). Fridays show the highest booking frequencies. The number of bookings is on the contrary very poor at the beginning of the week. Sunday has a comparable intensity to a normal working day though.

4.3 SPATIO-TEMPORAL ANALYSIS

As it was briefly mentioned the positions of the vehicles can vary at different times. In this section, the Kernel Density maps for particular day times are discussed.
Fig. 12 and 13 illustrate the Kernel Density of all booking starts at particular times. In general, the centroids of all hot spots do not change over the day. It is just the intensity of bookings that varies. During the night when carsharing is almost not used at all there are more bookings in Friedrichshain and northern parts of Prenzlauer Berg. Munich has the most of bookings in the neighborhood around the central station. During morning rush hour the trip starts are more dispersed than on other times of the day. This is especially obvious in Munich. At noon, the focus of demand changes in Berlin also to central districts like Mitte. These are typically spots with a high shopping and job density where living is rare and expensive. This job-effect is interestingly also observed in Munich. In the northern parts at Milbertshofen where many offices of *BMW* are located the booking number raises at lunch time. In the afternoon, the spatial distribution of Munich remains the same until the end of the day with a high demand in the northern central areas close to the Altstadtring and Schwabing. This phenomenon holds also for Berlin. After the afternoon rush hour there are less bookings in central, touristy places but a higher concentration in Prenzlauer Berg and Friedrichshain.

An interesting observation was made by Weikl et al. in [142]. A conspicuous decrease of bookings appears on Monday mornings. The demand declines especially after long weekends, as e. g. Easter. It is assumed that the observed booking numbers do not represent the hypothetical demand because of a non-optimal location of the fleet. The purposes for the trips on weekends are mostly leisure time activities. Over the weekend the spatial distribution of vehicles changes to an equilibrium that matches to this purpose. On Monday morning when carsharing is again needed for business trips the spatial distribution of vehicles does not match the needs for the customers. This phenomenon is particularly visible in either residential or business districts.
The different characteristics of districts is generally observable in this spatio-

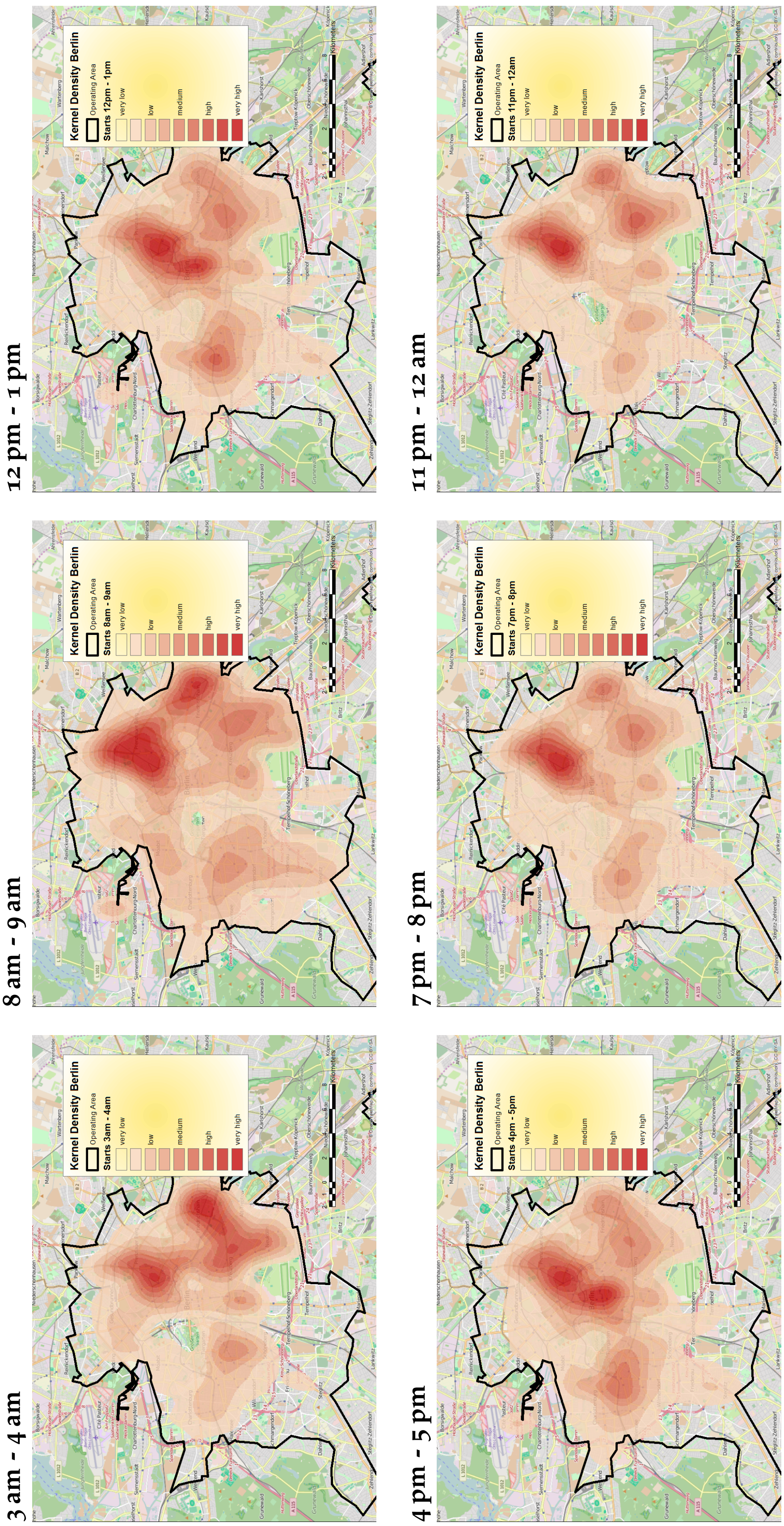


Figure 12: Maps of Berlin showing the Kernel Density for booking starts at different time slots.

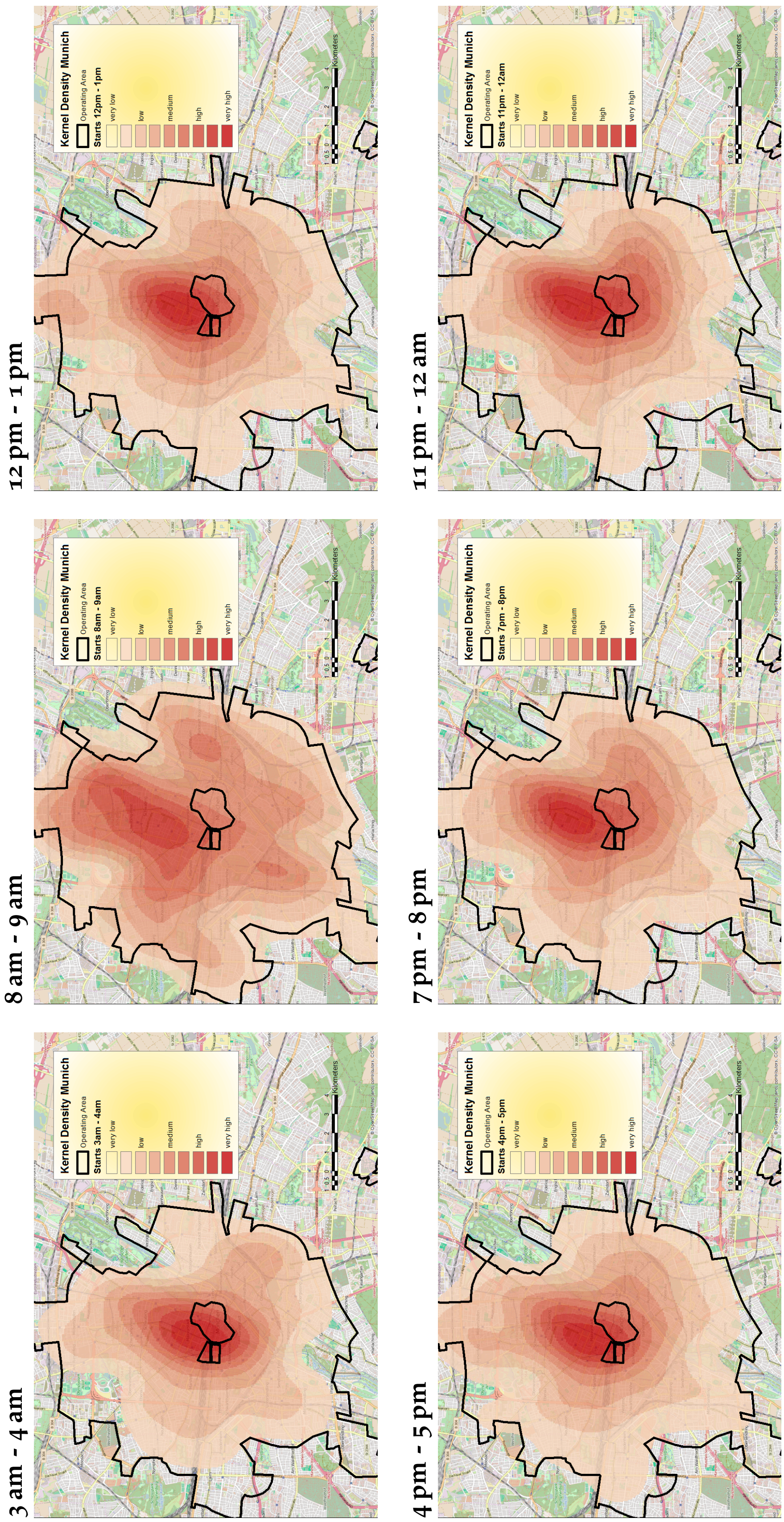


Figure 13: Maps of Munich showing the Kernel Density for booking starts at different time slots.

temporal analysis. Especially during noon it is visible which areas of a city have a high job density. The use of the system in the morning works best in residential districts and is accordingly dispersed. The focus of bookings in the evening is on spots where people can go out and eat. In Munich, job, residential and nightlife areas are located in the university district. This makes the area a hot spot for the whole time. The equivalent in Berlin is Prenzlauer Berg. Berlin, however, has some more districts that vary in their main characteristic.

4.4 KEY DATA

After describing the spatial distribution and temporal profile of FFCS bookings some benchmark values are given for a better understanding of the use of the system.

Flexible carsharing is used primarily for short-term trips. About 60 % of the trips are less than 5 km. The average distance in Berlin is 8.17 km, in Munich 13.16 km ([94]). The mean trip duration is 41.34 min in Munich, also longer than the average of 25.95 min in Berlin. This has to do with the structure of the cities which has already been explained in the first section of this chapter. Trips in Berlin go mostly from district to district whereas the trips in Munich are principally to or from the city center. Some attractive spots of the operating area are not part of the main area but far outside of the city. These satellites like Garching or the airport of Munich let the average trip duration time increase.

The around 35-40 km far away airport of Munich is an attractive target for FFCS customers. There are several reasons that make the airport to an attractive destination. The first is that a car is in general the most convenient way to travel in particular when one has luggage to carry. If a private car is available in the household the car has to be parked in the parking garage of the airport. This fee of 20 € per day ([59]) already exceeds the additional fee for airport trips of the FFCS operator (12 €). The good connection via motorway, moreover, reduces the travel time to the airport. A hired taxi costs usually more than 70 €. Also other transport alternatives score badly. The airport is not yet integrated in the railway net for regional or long-distance services. Thus a trip with the S-Bahn takes about 40 minutes to the city center. Additional time e. g. for changes to other underground lines are not even included in that time. Niehues analyzed in his master thesis ([111], p. 43-50) satellites of the operating area. The airport of Munich works in a way contrary to the regular business. Most bookings take place during the week, particularly on Mondays and Fridays. Almost all booking to the airport are in the morning between 4 am and 9 am. Trips from the airport are again mostly up after 6 pm. The enormous peak in the morning is assumed to results from a non-sufficient mobility service of public transport. The subway starts a reduced service at 5 am and runs its normal schedule from 6 am on. Flights in the early morning hours are therefore hard to reach with public transport. The cost for a ticket is 11 € and more than half as expensive as a trip with a carsharing vehicle. The price difference, however, reduces when the traffic is little like in the morning before rush hour. This makes FFCS a good supplement to public transport service.

4.5 OD ANALYSIS FOR MUNICH

As final booking data evaluation, FFCS is considered in the context of traffic. The temporal analysis has already shown that carsharing is mostly used off-peak. But there are several other open questions. Is the average speed of trips by carsharing users (due to the time-based fare) higher than of standard vehicles? Does the trip duration vary over the day and the week? Do the trip destinations change over the time in a distinct way?
The analysis is applied for Munich only. The booking data are again taken from data set No.4. For the purpose of the approach, it is necessary to eliminate those trips which are probably round trips. The following filter is applied to the ca. 750 000 trips.

- trip duration < 1 h (with exception of trips to and from cell 26 (Garching (TU Munich) in the north, not on the map) or 27 (airport, also in the north, not on the map neither)
- start and end of a trip should have at least a beeline of 800 m
- parking time: 0 min

If one of these conditions was not met, the trip was not considered for the analysis. At the end, around 550 000 bookings met the definition of the filter and served as data base for the following analysis.

At first, the duration, distance and speed of these one way trips distinguished by day of the week and time of the day of the trip start are regarded. The diagram in Fig. 14 shows the actual driven distance of the vehicle during the trip in green. It is remarkable that it stays on the same level over the whole week. There are just little peaks on workdays in the early morning. The trips from 9 pm on are in contrary a bit shorter. The longer rides result from the already mentioned airport trips that represent most of all trips from 3 am to the morning rush hour. Bookings in the evening have usually the purpose to drive home and are therefore in average shorter. The average driven distance of one way trips is 10 km shorter than the average distance for all bookings.

Due to the more or less equal driven distance the profile of trip durations (yellow) and speed (blue) average run complementary. Trips during the night are the shortest or the data contains less outliers which bias the mean distance. The average speed decreases to the morning rush hour down to 25 km/h and reaches its lowest point at 6 pm. Short time after the afternoon rush hour when most of the bookings take place the speed enhances and the trip durations start getting less.

It is hard to estimate whether customers consciously avoid carsharing during typically congested time periods. It does not hold for the morning rush hour. The reason for the shift of user peak in the early evening and the afternoon business rush hour is not assumed to be the better traffic situation but in the purpose of trips. But it can also be a soft criterion for the choice of FFCS. Evening rides are usually non-routine trips. The subjective impression of a general improvement in traffic conditions after 6 pm may let more people decide to take a car than on other times of the day.

It is interesting to see how the speed level varies in different cities. A study commissioned by the Forbes Magazine in 2008 ([117]) estimated the traffic speed in Munich to be higher, i. e. 32.35 km/h. A view to the data gives the impression that the average speed is between 25 and 30 km/h. The short trips that in most cases include time for the search for a parking space reduce this level of speed. Other cities mentioned in the study score much worse.

A more specific analysis aims to find out if users have specific destinations at particular times or if congested districts of a city have temporarily less bookings. A segmentation of the city is necessary for this purpose. A grid that provides districts of a suitable size are the official city districts of Munich (see Fig. 15). More information like the population density are not considered at this point but are provided by the statistical office Munich (Statistisches Amt München, [130]).

For all days of the week and times of the day OD matrices are created. There are four quantities of interest. The average booking frequency, the driven dis-

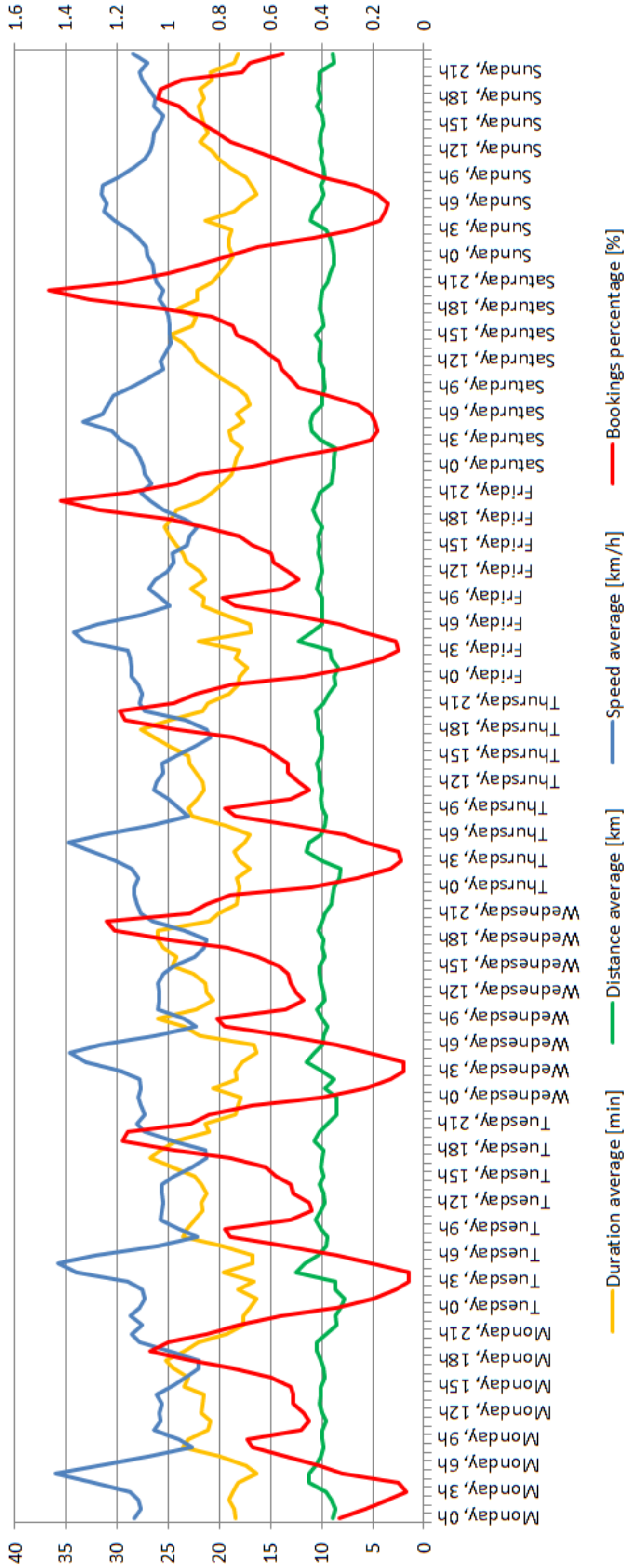


Figure 14: Average duration (orange), distance (green), speed (blue) and bookings (red, right scale) for one way trips in Munich.

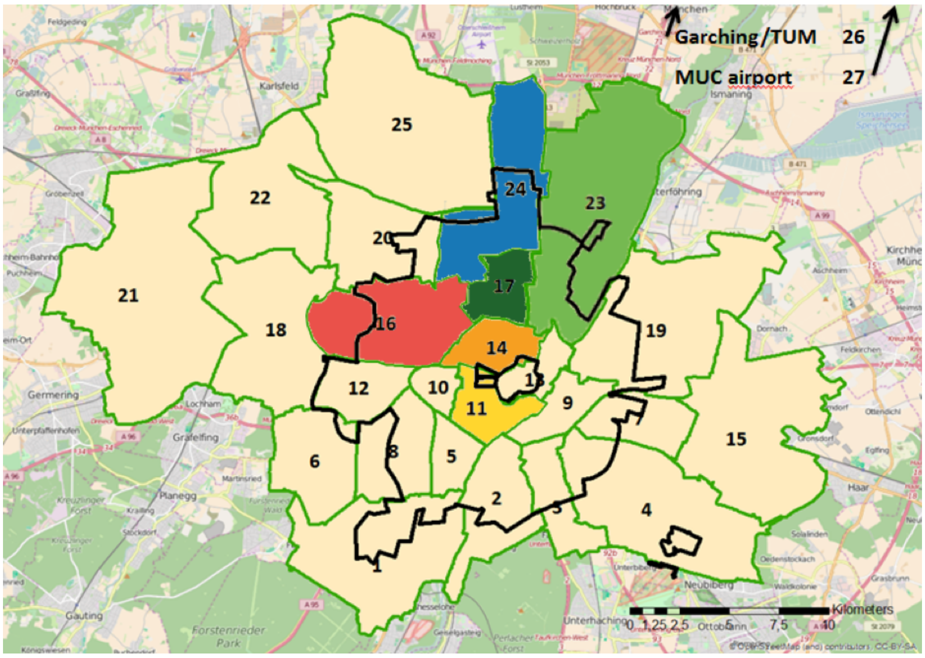


Figure 15: Map of Munich's 25 official city districts. Garching in the north and the airport are marked with 26 and 27, respectively. The colored districts are analyzed in detail in Fig. 16.

tance, the speed and trip duration. To facilitate the comparison of different time slots the matrices do not contain any numbers. The cells are instead colored by the value they contain. The color scale is for the frequencies up to 50 bookings, for the distance up to 20 km and for speed and duration the colors vary for all values between 0-60 km/h and 0-45 minutes, respectively.

The OD matrices for the other characteristics are in the appended section B.1. The ones shown in Fig. 39, 40, 41 and 42 all stem from the same time periods. An example of two OD matrices is given in Fig. 17. To find orientation in the matrices it is helpful to concentrate on some particular districts. The focus is hence on districts 11, 14, 16 and 17 the northern areas 23 and 24. One fact which plays an important part for district 14 is that the DriveNow headquarters are located in this neighborhood. The two districts in the north of the operating area (district 23 and 24) are business districts where e. g. BMW offices are located. Moreover, one eye should also be on 26, the district of the Garching satellite and 27, the airport. Trips to and from areas that are not covered by the official city district grid are assigned to district 0. The x-axis represents the destination districts while the y-axis shows the districts of origin.

First of all, the focus is on the booking frequency (Fig. 16 and Fig. 39). After mid-

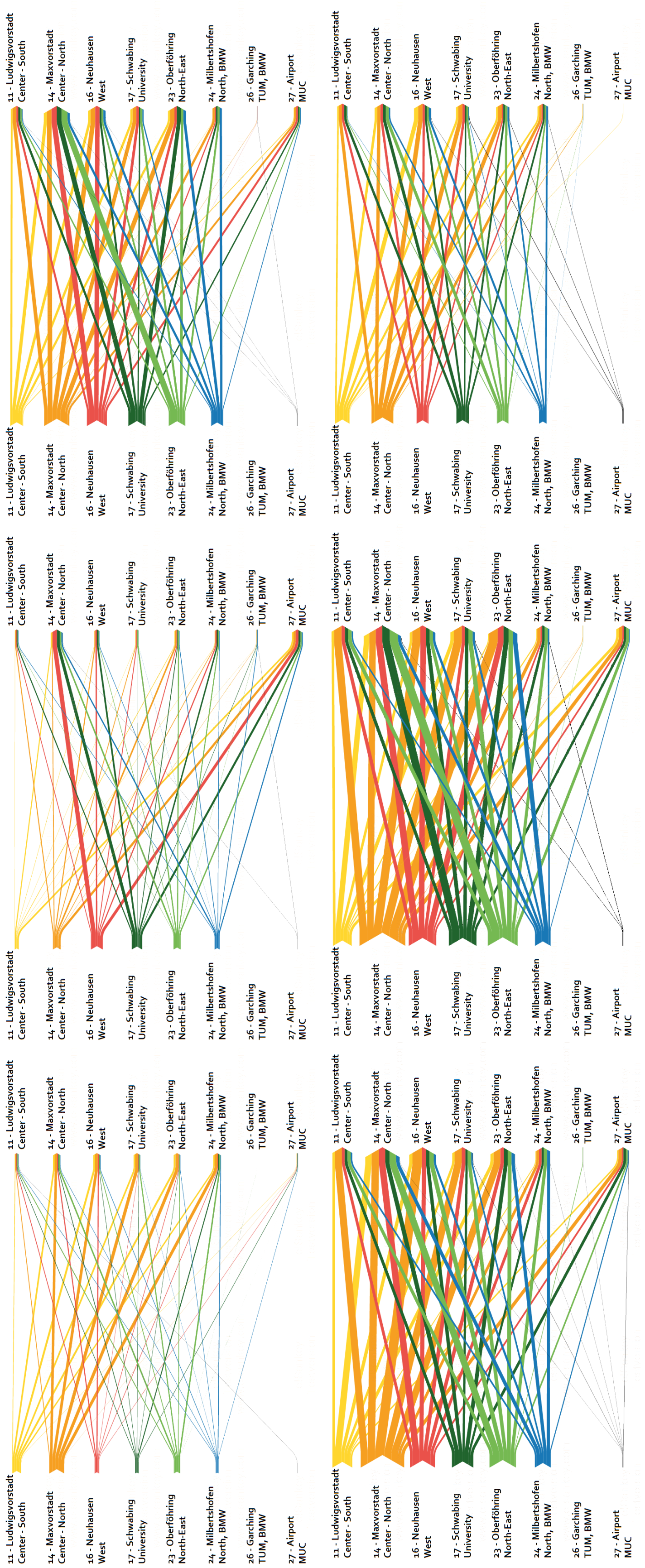
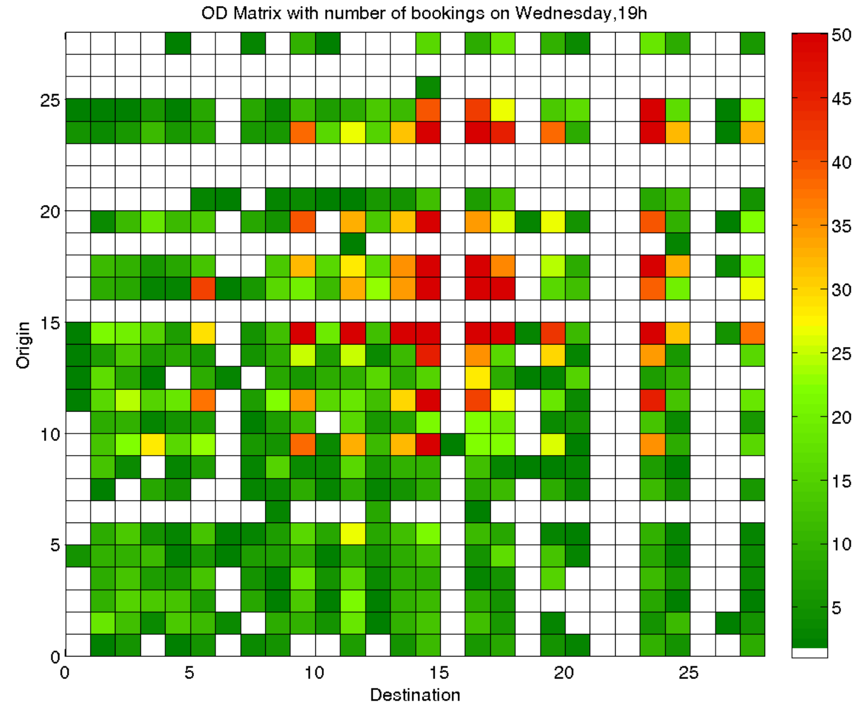


Figure 16: Flow diagrams showing the number of vehicle movements for selected districts on a typical weekday (Wednesday) in Munich at different time slots (3 am, 8 am, 12 pm, 4 pm, 7 pm, 11 pm)

number of bookings – 7 pm - 8 pm



speed – 7 pm - 8 pm

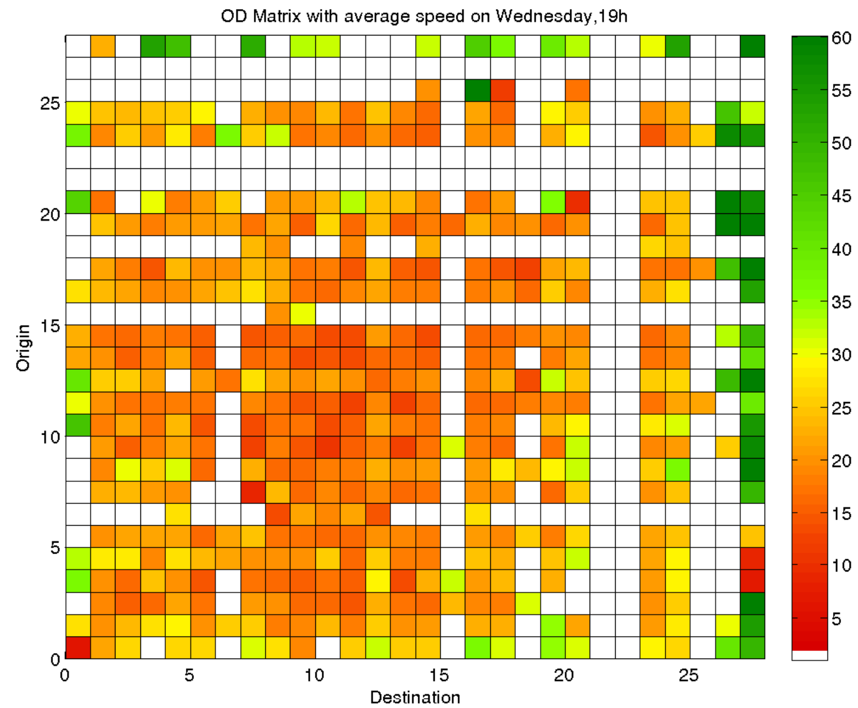


Figure 17: OD matrices showing the number of bookings and the average speed on a typical weekday (Wednesday) exemplarily at 7 pm

night, most bookings are in the city center (14, 16, 17). The time between 2 am and 5 am shows almost no booking processes. The day starts at 5 am with a very strong emphasis on trips to the airport. This is the already seen *airport-effect* in the morning that is probably caused by a weak service of public transport and non-congested roads. Airport trips with carsharing are especially attractive in the northern parts of the city, because the distance and hence the duration to the airport is small. But not only these rides are conspicious. Also trips to the northern parts of Munich (23, 24) let the number of bookings increase. It is assumed that employees of BMW working in district 24 use to drive at that time. The unusual high number of bookings is thus called *BMW-effect*. They come from all parts of the city to this district, even from the south. The usually good traffic conditions at this time have definitely an influence.
The focus of bookings goes to the city center, namely district 14, at the rush hour in the morning. The number of tendentially short trips increases. Also during noon, a lot of trips have their start and end in district 14. The BMW-effect that was already seen in the Kernel Density map becomes visible in cell 23 and 24. In the afternoon, a lot of vehicles depart from 14 to e. g. neighbor districts like 23. The evening rush hour and main FFCS booking time shows the most number of bookings in the central area of the city (14, 16, 17) as it has already been seen on the heat maps. The additional information that is given by the OD matrices is that most trips are not long journeys but end soon in one in the neighborhoods. After 10 pm, it is a similar situation as in the afternoon. Many trips are made from the center for example district 14.
The attraction of district 14 may have to do with the headquarters of DriveNow. Employees usually do not get discounts but are like BMW employees more familiar with the system than the average citizen. It is strongly surmised that many DriveNow employees are heavy users and may use the system also for routine trips during business hours or to and from their work. The booking demand in district 14 is thus supposed to be strengthened by this *DriveNow-effect*.
An attentive reader has already noticed that there are more trips to the airport than from it. This would make sense in the morning but should switch in the afternoon and evening. Trips with the airport as origin are however not so numerous even at the end of the day. An assumption is that people on the way to the airport want to save time and plan their way to not missing the flight. For the way back home there is usually no time pressure. Passengers are therefore more relaxed and accept longer travel times e. g. with the S-Bahn. This scenario is anything but desirable for the operator because vehicles concentrate unused

in this satellite and are not available for other customers.
The distance of a trip depends strongly on the origin and the destination of the journey. Due to the fact that the districts are in most cases numbered systematically, matrix entries close to the diagonal line contain the average distance of trips to neighboring districts. Over the whole day, cells close to that diagonal are mainly light green or even dark green. The distances from double-digit to single-digit cells are in contrary generally longer. Trips to Garching and the airport are normally longer than 20 km and therefore always colored red. There is no specific interpretation in the trip distances. It can just be used as an indicator for the quality of the filter for sorting one-way trips. The results show that the conditions for filtering work fine. Some outliers however also appear.
One makes a similar observation for the OD matrices of durations. Most of the short trips to neighbor districts are on average not longer than 15 or 20 minutes. Travels from one side of the city are naturally longer.
It is now interesting to see how the average speed varies over the day. The trips to destinations far outside are clearly faster because of the respective infrastructure outside the city. Journeys with a short distance and duration oscillate around an average speed of 15 to 25 km/h. These are the matrix cells at the diagonal of the matrix. The entries in the left up and right down parts are faster. Cross-city trips mostly take place on junction-free infrastructure like the Mittlerer Ring. This advantage is not noticeable during rush hour though. On every weekday the speed between 7 am and 8 am as well as 4 pm and 7 pm is, roughly speaking, the same everywhere in the city. The typical coloration of the distance and duration matrices appear usually during off-peak hours.

As conclusion of this OD analysis one obtains a better understanding of the FFCS system for the city. Heat maps are good for a first and general understanding of the system. But considering the trip destinations at different times deepens the comprehension of the user in a much better way. Demand effects caused by BMW and DriveNow employees or airport passengers are more obvious in this evaluation. Concerning the traffic in the city, it is interesting to see how the congestion on the roads during rush hour becomes noticeable in the FFCS system, too. But traffic has just a small impact on carsharing because the lion's share of trips is inner city rides that are over the whole day not faster than 25 km/h.

Part II

MODELING: EXPLANATORY VARIABLES FOR THE DEMAND OF FFCS

5 THEORY OF THE APPLIED STATISTICAL MODELS

5.1 LINEAR REGRESSION

The method of linear regression is a standard technique to quantify impacts of possible explanatory variables on a particular observation or state.
In the present case the intention is to measure the impact of the data described in section 3.2 and 3.3 on the number of FFCS booking starts.
Based on the standard declaration of variables which is in the current case adopted from [53], p. 19, the target variable $y = (y_1, \ldots, y_n)^t$ is the number of bookings. The quantity y_i, $i = 1, \ldots, n$ denotes the booking frequency per district of the respective grid. The explanatory variables $x_1, \ldots, x_k$ are generally related to y by

$$y = E(y|x_1, \ldots, x_k) + \varepsilon = f(x_1, \ldots, x_k) + \varepsilon \tag{1}$$

with the random component $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ for which $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2 \; \forall i = 1, \ldots, n$ must hold.
In a linear regression model f is a linear function. Hence (1) can be written as

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + \varepsilon. \tag{2}$$

Expressed for every district i it means

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik} + \varepsilon_i, \; i = 1, \ldots, n \tag{3}$$

$\beta_0, \ldots, \beta_k$ are called regression coefficients and determine the slope of each variable. It is also common to write $x_1, \ldots, x_k$ in matrix notation with

$$X = (x_1^t, \ldots, x_k^t) = \begin{pmatrix} x_{11} & \ldots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \ldots & x_{nk} \end{pmatrix}$$

In this dissertation, two models are considered with

- land use and social economic data (from section 3.2) (Berlin and Munich)
- results of the Bundestag election in 2013 (Berlin and Munich)

as influence factors. The reason for separating the models and not considering them as whole is the different grid the data is based on. A combined analysis of the two data sets will be described later on in section 5.2.
Six steps are proposed in ([4]) to check the model fitness.

1. Check the significance
2. Check the relationship
3. Check the redundancy
4. Check the bias
5. Check the completeness
6. Check the performance

These points and their respective tests mentioned below are also proposed by Florax and Nijkamp in [58]. The order of the stages can be changed. The variable selection e.g. can also start by checking the redundancy instead of the significance.
In the following subsections, tests and other methods for an analysis of the six stages are explained. Here, only techniques explained in ([4]) are discussed.

5.1.1 *Significance of the explanatory variables*

If an explanatory variable x_j makes a significant contribution to the dependent variable the related coefficient has to be considered. One can neglect the influence of the variable, if the slope of the regression line β_j is 0 (or ≈ 0).
A statistical standard test in this situation is the t-test. Formally, the test checks for all $j = 1, \ldots, k$ if

$$H_0\colon \beta_j = 0 \qquad H_1\colon \beta_j \neq 0$$

is valid. Let b the linear transformation of the observed data y. Then

$$b \sim \mathcal{N}(\beta, \sigma^2 (X^t X)^{-1})$$
$$\frac{b}{\sigma} \sim \mathcal{N}(\frac{\beta}{\sigma}, (X^t X)^{-1})$$

For the j-th component this yields to

$$\frac{b_j}{s} = t_j \sim t_{n-k} = \frac{1}{\sqrt{\frac{\chi^2}{n-k}}} \mathcal{N}(0, 1)$$

with s as the sample variance. Hence, the t-test with significance level α accepts H_0 if

$$|t_j| \leqslant t\left(\frac{1-\alpha}{2}, n-k\right)$$

and rejects H_0 if

$$|t_j| > t\left(\frac{1-\alpha}{2}, n-k\right)$$

The OLS toolbox offers additionally Koenker's studentized Breusch-Pagan statistic. The homonymous test checks the heteroscedasticity of the linear regression model. In a linear regression model one can regard the variance of errors σ_j^2 as

$$\sigma_j^2 = h(x_j\alpha),\ j = 1, \dots, k$$

where $h(\cdot)$ is an unspecified function possessing the first and second derivatives, $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_k) \in \mathbb{R}^{k+1}$ unrelated to the coefficient vector β and x_j an exogenous variable whereby $x_{0j} = 1$ and thus $x_j = (1, x_{1j}, \dots, x_{nj})^t$. If the model is homoscedastic, the following null and alternative hypothesis are valid:

$$H_0\colon \alpha_1, \dots, \alpha_k = 0 \qquad H_1\colon \exists j \in \{1, \dots, k\} \text{ with } \alpha_j \neq 0$$

Hence

$$\begin{aligned} H_0 \text{ is true} &\Leftrightarrow \tfrac{\partial}{\partial x_j} x_j\alpha = \alpha_0 \\ &\Leftrightarrow \sigma_j^2 = h(\alpha_0) = \sigma^2 \\ &\Leftrightarrow \text{constant variance} \\ &\Leftrightarrow \text{homoscedasticity} \end{aligned}$$

The clue of the test is that most phenomenons of heteroscedasticity can be expressed with $h(\cdot)$. The exact statistic and its distribution is described in [18] where also this derivation is taken from. If the null hypothesis is rejected one solution is to consider the robust statistics of the t-test instead of the ordinary one.

5.1.2 *Relationship between dependent and explanatory variables*

The simpleness of the linear regression model facilitates the interpretation of the coefficients. By this, a positive impact of a variable is clearly visible at the positive sign of the β_i. The greater the coefficient the bigger the influence of the exogenous variable to the target variable. If a relationship does not seem to be coherent the model should be reconceived.

5.1.3 *Redundancy of explanatory variables*

If a model includes more than two independent variables, correlation between the factors can appear. In the case of a severe (multi) collinearity at least one of the correlated variables is redundant regarding the explanation of the dependent variable.
One option to check the redundancy of an influence variable is the *v*ariance *i*nflation *f*actor (VIF). It is defined as

$$\mathrm{VIF}(\hat{\beta}_j) = \frac{1}{1 - R_j^2}$$

whereby R_j^2 is the R^2 in a linear model containing x_j as explanatory variable only. As a rule of thumb O'Brien proposed in [112] to indicate multicollinearity if $\mathrm{VIF}(\hat{\beta}_j) > 5$ or $\mathrm{VIF}(\hat{\beta}_j) > 10$. The chosen limit for the present analysis comes from [4] where variables with VIF values greater than 7.5 are seen as a problem for the model.

5.1.4 *Heterogeneity of residuals*

In an ideal model the residuals are approximately normally distributed. A biased model means e.g. a positive or negative skew in the distribution of residuals or even a bi- or multimodal one. The problem of a biased model can easily be indicated by drawing the scatterplot of the histogram of residuals for each variable. The points of the scatterplot (optionally drawn with a trend line) have in the best case a linear structure and the majority of points close to the line. Other structures such as a parabola are also visible in a skewed residual plot.
A quantitative indicator for the bias in a model is the Jarque-Bera-test (JB test). The test statistic is defined as

$$\mathrm{JB} = \frac{n}{6}\left(s^2 + \frac{(K-3)^2}{4}\right)$$

where

$$s = \frac{\mu_3}{\sigma^3} \text{ is the skewness and}$$
$$K = \frac{\mu_4}{\sigma^4} \text{ the Kurtosis of the distribution.}$$

μ_i stands for the i-th moment of the distribution.
In the present case μ_i and σ are replaced by the respective estimators $\hat{\mu}_i$ and $\hat{\sigma}$.
Bera and Jarque proved that

$$\mathrm{JB} \underset{a}{\sim} \chi^2(2)$$

The hypotheses of the tests are thus

H_0: The sample is normally distributed.

H_1: The sample is not normally distributed.

5.1.5 *Completeness of the model*

In the present case completeness would mean that almost all explanatory variables are found to describe the booking behavior of FFCS customers. An indicator that marks the incompleteness is the autocorrelation and especially the spatial autocorrelation.

Exogenous spatial autocorrelated variables can be identified by a graphic analysis. If the ordinal or interval scaled factor appears to be randomly distributed there is no autocorrelation to be supposed. But if clusters are clearly visible the variable is maybe independent from the others but definitely autocorrelated. The contrary of spatially clustered data are those which are dispersed but often appear with the same kind of different values of the variable. One can recognize them by finding the same constellation of variable values in every part of the area.

For identifying autocorrelation formally, ArcGIS offers the Spatial Autocorrelation toolbox. This works with the Moran's I statistic.

$$I_j = \frac{n \sum_{i=1}^{n} \sum_{l=1}^{n} w_{il}(x_{ij} - \bar{x}_j)(x_{lj} - \bar{x}_j)}{\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2}, j = 1, \dots, k$$

whereby w_{il} are weights with $\sum_{i,l} w_{il} = 1$. After normalization, one holds

$$Z_{I_j} = \frac{I - E[I_j]}{\sqrt{Var[I_j]}} \sim \mathcal{N}(0, 1)$$

with

$$E[I_j] = -\frac{1}{n-1}$$
$$Var[I_j] = E[I_j^2] - E[I_j]^2$$

The hypotheses are simply

H_0: Z_{I_j} is not autocorrelated.

H_1: Z_{I_j} is autocorrelated.

As alternative to Moran's I one can use Geary's C that is a similar test statistic but focused on local autocorrelation.

5.1.6 *Performance of the model*

As a final step one should check the R^2 or the adjusted R^2. The coefficient of determination is generally defined by

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{4}$$

whereby SS_{res} and SS_{tot} are the residual sum of squares and the total sum of squares, respectively. Standardizing each sum of squares with the responding degree of freedom leads to the adjusted R^2. A high value close to 1 of the adjusted R^2 is a first indication for a good explanation of the model. Nevertheless, this is not valid if other characteristics for a perfect model are not fulfilled. If all other steps are performed correctly, the R^2 value shows the proportion the explanatory variables explain the dependent variable. Otherwise the interpretation must be adapted to the results of the previous steps.

5.2 REGRESSION MODELS FOR COUNT DATA

In a second approach an advanced regression model is applied for the count of bookings with land-use data and election results as influence variables. The two data sets are based on two different grids. The KGS22 district segmentation of the city for land-use data and the polling districts need to be transferred to a base grid. The KGS22 districts are chosen as the fundamental grid. Details about how the data sets are combined are described in section 6.3.
The reason to consider different modeling methods other than the linear regression is that the scatter plots do not present a linear relation for every variable(see Fig. 19). Some resemble more of a typical Poisson distribution. A linear regression model is in such cases an indicator for a trend but the fit of the model represented by the R^2 is supposed to be quite poor. It is therefore necessary to apply other modeling methods as well. There are three regression models for count data which are applied and compared to the present data:

- the Poisson regression
- the Quasi-Poisson regression
- the negative binomial regression

All three models are part of the family of generalized linear models (GLM). The conditional density functions of the covariables $x_i = (1, x_{i1}, \dots, x_{ik})$ can hence be written in the form of a one-parameter exponential family (see [53], p. 218), i. e.

$$f(y|\theta) = \exp\left(\frac{y\theta - b(\theta)}{\phi} w + c(y, \phi, w)\right) \tag{5}$$

θ is the natural or canonical parameter

ϕ the dispersion parameter

w a weight

c an arbitrary function

To show that the Poisson regression is element of the exponential family one has to consider the density

$$\begin{aligned} f(y|\lambda) = P(Y = y) &= \frac{\lambda^y \exp(-\lambda)}{y!}, y = 0, 1, \ldots \\ \Leftrightarrow \exp(\log(f(y|\lambda))) &= \exp(y \log(\lambda) - \lambda - \log(y!)) \\ \overset{\theta := g(\lambda) := \log(\lambda)}{\Leftrightarrow} \exp(\log(f(y|\theta))) &= \exp(y\theta - \exp(\theta) - \log(y!)) \end{aligned}$$

Hence $b(\theta) = \exp(\theta)$ and $\phi = 1$. The transformation of the parameter is done by the link function g. For the Poisson regression $g(\lambda) = \log(\lambda)$ is the canonical link. With this definition of θ the density of the Poisson distribution fits the equation (5). The determination of θ is generally called link. The log-link is the natural link for Poisson regression. The fixed dispersion parameter corresponds to the fact that the expectation of the Poisson distribution equals its variance λ. One gets a problem when modeling count data which distribution does not show an increasing variance for greater mean values. A solution is to use the Quasi-Poisson distribution which uses ϕ as an additional free parameter to model overdispersion.
The third considered model is the negative binomial regression model using

$$f(y|\mu, \theta) = \frac{\Gamma(y + \theta)}{\Gamma(\theta) \cdot y!} \cdot \frac{\mu^y \cdot \theta^\theta}{(\mu + \theta)^{y+\theta}}$$

as a density for the distribution of y ([144]). μ stands for the mean, θ is the shape parameter and $\Gamma(\cdot)$ the gamma function. If the parameter θ is fixed the density is also a special case of (5). Although the dispersion $\phi = 1$ the variance is not fixed but depends on μ.

$$V(\mu) = \mu + \frac{\mu^2}{\theta} \tag{6}$$

The negative binomial distribution is member of the exponential family, too. The proof is not shown because it is unnecessary for the analysis. It turns out that the canonical link function is

$$g(\mu) = \log\left(\frac{\mu}{\mu+\theta}\right)$$

All GLM follow a structure assumption ([53]) for the conditional expectation μ_i with the linear predictor

$$\mu_i = x_i'\beta = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}$$

that is given by

$$\mu_i = h(\eta_i) = h(x_i'\beta) \text{ and } \eta_i = g(\mu_i).$$

h is a two-times differentiable response function and g the link function that is the inversion of h ($g = h^{-1}$).
The log-likelihood for every observation (y_i, x_i) is given by

$$l_i(\beta) = \log(f(y_i|\beta)) = \frac{y_i\theta_i - b(\theta_i)}{\phi} w_i$$

In cause of the assumed conditional independence of $y_i|x_i\ \forall i$ the maximum likelihood (ML) is

$$l(\beta) = \sum l_i(\beta).$$

The score function $s(\beta) = \frac{\partial l(\beta)}{\partial\beta}$ is thus given by

$$s(\beta) = \sum x_i \frac{d_i}{\sigma_i^2}(y_i - \mu_i) \text{ with } d_i = \frac{dh(\eta_i)}{d(\eta_i)}.$$

Setting the score function to 0 will result in an ML estimation for β noted as $\hat{\beta}$.
The Fisher matrix (second derivative of the ML) is defined by

$$F(\beta) = \sum x_i x_i' w_i \text{ with } w_i = \frac{d_i^2}{\sigma_i^2}$$

$\hat{\beta}$ is calculated iteratively.

The `R` output of `glm` includes a note to the significance of each β_i. The used test statistics is the Wald statistic testing generally the hypothesis

$$H_0\colon C\beta = d \quad \text{vs.} \quad H_1\colon C\beta \neq d$$

with the statistics

$$w = (C\hat{\beta} - d)' \left[CF^{-1}(\hat{\beta})C'\right]^{-1} (C\hat{\beta} - d)$$

A test for the general influence of a covariance would be equal to

$$H_0\colon \beta = 0 \quad \text{vs.} \quad H_1\colon \beta \neq 0$$

Thus the test statistic shrinks to

$$\hat{\beta}' F^{-1}(\hat{\beta})\hat{\beta} \sim \chi^2_r$$

with r rang of C. H_0 will be rejected if

$$w > \chi^2_r(1-\alpha).$$

Finally it should be mentioned that the general calculation of the ML for the Quasi-Poisson model works similar but with the Quasi-Score and Quasi-Fisher matrices instead.

$$s(\beta) \sum x_i \frac{d_i}{\phi\lambda}(y_i - \lambda_i) \text{ and}$$
$$V(\beta) = \sum x_i x_i' w_i \frac{\sigma^2_{0i}}{\sigma^2_i}$$

with σ^2_{0i} working covariances.

For comparing two models it is helpful to consider the Akaike's information criterion (AIC) of each model. The AIC measures simply said the information of the output in relation to the input. A high fit of the model does not necessarily mean a better model - it can also come from by a high number of explanatory variables. The AIC is defined by

$$AIC = -2l(\hat{\theta}) + 2p$$

whereby p is the number of exogenous variables. The less the AIC value is, the better is the information yield of the model. When it is used for one model only, it has no interpretation. But it is helpful for comparing two models.

There are some approaches for designing a measurement for the fit of the GLM. The (adj.) R^2 can just be calculated in a linear regression model. That is the reason why most coefficients of determination work with the likelihood function. McFadden's R^2 is chosen for the present analysis ([100]). It is defined by

$$R^2_{McFadden} = 1 - \frac{\log l_1}{\log l_0} \in [0, 1)$$

with l_1 likelihood of the model with explanatory variables and l_0 likelihood of the null model. A $R^2_{McFadden}$ value between 0.2 and 0.4 is already satisfying. More impact factors let the index increase. The corrected McFadden's R^2 regards this effect by penalizing a high number of exogenous variables p.

$$R^2_{McFadden_{corr}} = 1 - \frac{\log l_1 - p}{\log l_0} \in [0, 1)$$

5.3 ANALYSIS OF PARKING DATA

A regression model as it is done for socio-demographic data and the political election behavior is not useful for the parking restriction zones. Although the data is spatial they are also evaluated at different times. The first stage is to assign every booking end to a street. The nearest neighbor method applied to the booking points yields to the desired assignment (Fig. 18). It is clear that there might be some points that are allocated to the wrong parking area but errors do not have a significant effect on the results.

After this step all the number of bookings for each of the 13 kinds of parking restriction areas is known. Thanks to the temporal information of the booking they can also be distinguished by the day of the week and hour of the day. A profile of parked vehicles in the areas over the week can thus easily be provided.

One will observe a strong tendency to the mixed parking areas in the city because they build the highest percentage within the licensing area. To standardize this disparity the number of bookings are divided by the street length of each parking area. This graphical output will provide qualitative information about a general preference for a particular parking area of the FFCS customers. There will be no further analyses performed to quantify the relationship between the parking zones and booking data.

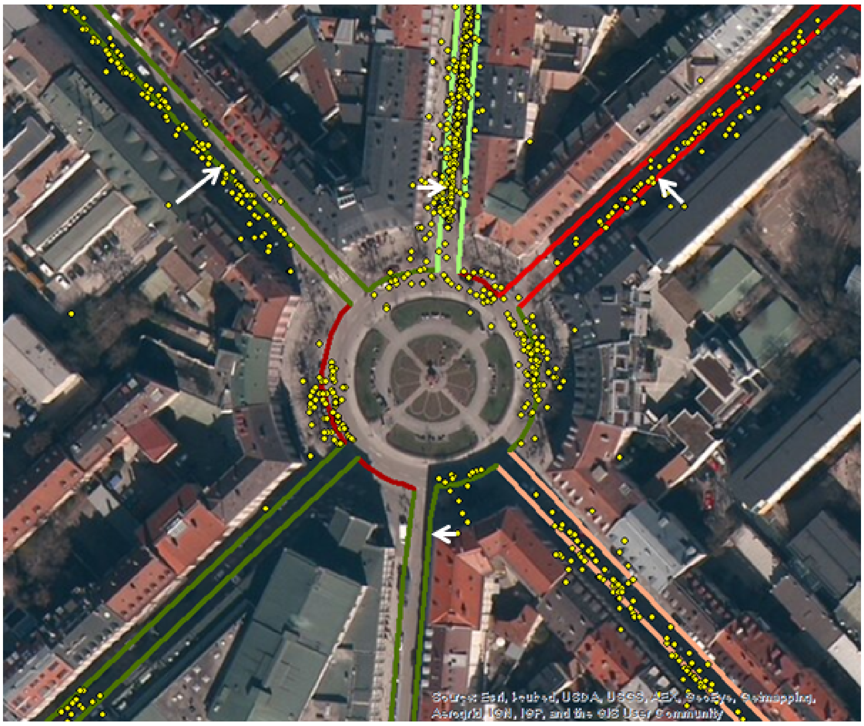


Figure 18: Booking endings around the Gärtnerplatz in Munich. Every point is assigned to its nearest part of the road.

5.4 ANALYSIS OF WEATHER DATA

The intention to analyze weather data in combination with booking data is to see if there are more bookings when the weather conditions are bad. The first method – applied to Berlin data only – just tries to prove a qualitative difference in the distribution of booking numbers while the second approach measures a quantitative difference.

For the first method data set No. 1 is modified such that only the bookings of heavy users were observed. A heavy user is defined in the same way as in section 3.1, i. e. as a customer who contributes to 80 % of all trips in a year whereby the customers were sorted by the frequency of their bookings. The idea is to measure the possible disparity of bookings during the different weather conditions for typical costumers only.

The weather conditions for all of the 8760 hours in the year were examined to establish if the weather was good or bad. The result is a data matrix consisting in the normalized number of bookings during good and bad weather conditions for every month in the columns. After that the booking frequency is noted in

an additional field in the booking table. The goal is to apply a paired t-test to the good and bad weather booking distributions for every hour.
For this reason the booking frequencies are aggregated monthly for every hour and normalized by the number of days per month with good or bad weather. The paired t-test checks if within two different samples the differences of the mean is significantly different under the assumption that the differences are normally distributed. The statistic of the test can be formulated as following:

$$T(h) = \sum_{m=1}^{12} \text{\# normalized bookings at [h] o'clock in month [m] with bad weather}$$
$$-\text{\# normalized bookings at [h] o'clock in month [m] with good weather}$$

The null and alternative hypothesis are

$$H_0\colon T(h) = 0 \qquad H_1\colon T(h) \neq 0$$

A disadvantage of this method is that the plus of bookings cannot be quantified and it cannot be distinguished between weekdays. The second method therefore analyzes the trips of data set No.3 with a different strategy. Instead of testing the booking frequencies during good and bad weather conditions regarding their differences of mean they are now compared directly.
In the beginning one assigns analogous to the first approach every hour of the data set the binary dummy variable "bad weather" depending on the current weather condition that is defined by the variables of the weather data described in Table 6. After that the frequency of bookings are – different to the first approach – distinguished in hours and weekdays. The number of bookings is normalized subsequently by the number of days of the same day of the week and the same time of the day with bad or good weather, respectively. The advantage next to the quantitative comparison is that there is no combination of day time and weekday in a year with good or bad weather condition only. By this, the problem that there were months with bad or good conditions only making the paired t-test not applicable for those months does not appear. The paired t-test is also applied to check the significance of the differences. The statistic in this case is

$$T(h) = \sum_{w=1}^{7} \text{\# normalized bookings at [h] o'clock on weekday [w] with good weather}$$
$$-\text{\# normalized bookings at [h] o'clock on weekday [w] with bad weather}$$

The switch of the order of the difference does not matter and is just important for the interpretation of the T value. The hypotheses are the same as in the first method.

6 APPLICATION OF THE STATISTICAL MODELS

6.1 LINEAR REGRESSION WITH LAND-USE DATA

The land-use data is from 2012, the booking data is taken from data set No. 3 (2013). There were also analyses done with data from data set No. 1 (2012) which led to similar results as in the evaluation described in the following.

6.1.1 *Significance of the explanatory variables*

The available land-use data consist of numerous detailed variables including e. g. age, income, number of companies etc. It is very likely that many of these variables are not independent. So a main assumption of the Linear Regression model may be violated. The Moran's I test that proves this spatial autocorrelation is done at the end of the procedure. To avoid possible autoregressive effects in the model the analysis of the land-use data starts very simple.
Instead of taking all variables into the linear model each variable is tested for its linear influence on the data. Fig. 19 shows a selection of scatterplots and residual plots for services and rents. It is clearly visible that the linear regression is not the perfect choice for an exact model for all variables. There is often a concentration of values around a point. The residual plots thus do not often show a normal distribution. Drawing the residuals in combination with the predicted value should reveal a random distribution of points. But for the number of companies, for instance, a linear trend is visible though.
An important stage in modeling the booking frequencies by land-use data is the variable selection. A first filter is the selection of those influence factors that are not significant. For the present sample size the t-statistic is approximately standard normally distributed. An absolute t-value greater than 1.96 would therefore indicate a significance for $\alpha = 0.05$. But since in a model where each variable is tested with no other factor almost every variable shows a significance. Therefore, only the very significant variables with a statistic of more than 10 are regarded in the following. The result is listed in Table 25 in appendix B.2. The sign of the t-statistic shows if the influence of the variable is

positive or negative.
Koenker's studentized Breusch-Pagan statistic testing the data for heteroscedasticity is significant for each variable mentioned in this list. It can hence be concluded that each variable is not homoscedastic when considered on its own. There are some factors with a less, but statistically still existing significance that meet the demanded uniform distribution of variance. A look to the scatter plots of this group of variables reveals that the distribution of the booking frequency for these factors is more similar to a distribution for count data and does not show a clear linear relation. The decision to take only the high significant variables into account is therefore necessary for a variable selection but unfortunately also excludes potential explanatory variables. It is already clear in this step that finding a perfect model by modeling with linear regression fails due to the heteroscedasticity of the highly significant influence factors. Nonetheless, the model serves as method to find some explanatory factors.

6.1.2 *Relationship between dependent and explanatory variables*

The next step is to take a look at the coefficients of the variables. There is generally no peculiarity visible. The booking frequency increases e. g. with the number of companies and decreases in districts with the higher rate of car ownership. The comparison of the exact coefficients is not meaningful at this time.

6.1.3 *Redundancy of explanatory variables*

For further analysis it is necessary to filter those variables that do not provide additional information to the model. Proving the redundancy of a factor is done by calculating the VIF. Instead of considering the variables separately all factors are from now on taken in one model. Variables that show a VIF value of more than 7.5 are filtered out. The exact scheme proceeds iteratively because this quantity changes with every omitted variable. The variable with the greatest VIF is first left out. After recalculating the values the filtering continues until the model consists of non-redundant and significant variables only.
For Berlin, the model reduces from 68 significant influence variables to 17 non-redundant factors. The Munich model includes just six significant and non-redundant of the originally 29 variables. They are listed in Table 7. It is obvious that variables measuring a similar quantity (e. g. number of hotels (total) and

number of hotels (small)) are strongly correlated. The updated model therefore contains in the variable group "number of companies" mostly only one variable for each company category. It is probable that the size of the company (big, medium, small) is not crucial but just the most critical representative of the category.

6.1.4 *Heterogeneity of residuals*

The output of the OLS analysis toolbox provides a histogram of standardized residuals as well as a graph of the distribution of standardized residuals in relation to predicted variable values. The figures for Berlin and Munich in Fig. 19 show a bias in the distribution. A perfect model would show a symmetric, approximately Gauss distribution of the residuals. The plot of predicted and standardized residuals should be normally distributed. As one can see this fails in both cities. Consequently, the JB test is highly significant for the chosen models.

One reason is the outliers. There are districts with a very high number of bookings. In a linear regression model the influence of outliers is over-average. They change the trend parameter more than it is necessary for the model.
The graphs also show a greater variation of residuals for higher expected values. This heteroscedasticity also reveals in a significant Koenker's studentized Breusch-Pagan test for the two cities.

6.1.5 *Completeness of the model*

The Moran's I test is performed for every variable that remains after the variable selection process. It is no big surprise that the test is significant for every variable. That means that every influencing factor builds clusters that are not likely to be random. It is a characteristic of land-use data that they are autocorrelated. Districts – regarded in that detailed way – do not differ that much between rent, number of companies or car ownership rate to their neighbor districts.

	Berlin				Munich			
	Coeff	σ	t	VIF	Coeff	σ	t	VIF
citizen data								
# citizens per sqkm	-0.001	0.00	-3.26	1.72				
affinity to leased private cars (index)	-2.88	0.46	-6.23	2.90				
frequent drivers (index)	1.99	0.54	3.69	2.84				
environmental affinity (index)	2.07	0.59	3.51	2.60				
household data								
# buildings	-1.26	0.20	-6.14	3.21				
rent	61.61	4.89	12.61	1.79				
% private cars (private)					-2.34	0.64	-3.60	1.51
number of companies								
# government agencies and administrative offices								
big					179.98	21.71	8.29	1.22
small	5.11	1.21	4.23	3.01	5.02	1.99	2.53	2.44
# banks								
medium	17.83	1.95	9.16	3.57				
# services								
big	100.36	9.50	10.56	1.74	69.61	11.28	6.17	1.42
small					4.43	0.81	5.48	2.84
# hotels								
small	9.51	1.69	5.65	2.58				
# mechanics								
medium	-13.55	3.09	-4.38	2.10				
# manufacturers					-8.26	4.18	-1.98	2.00
medium	19.78	3.77	5.25	1.94				
# other type of commerce								
medium	-40.82	10.47	-3.90	1.61				
# consulting for legal, business and investment								
medium	2.08	0.57	3.64	2.15				
miscellaneous								
# cars (total)	0.30	0.06	4.81	1.46				
street length	0.07	0.00	15.45	2.39				
purchasing power in retail per citizen	-0.04	0.00	-9.00	1.60				

Table 7: Results of the linear regression with land-use data for Berlin and Munich after omitting redundant and non-significant variables.

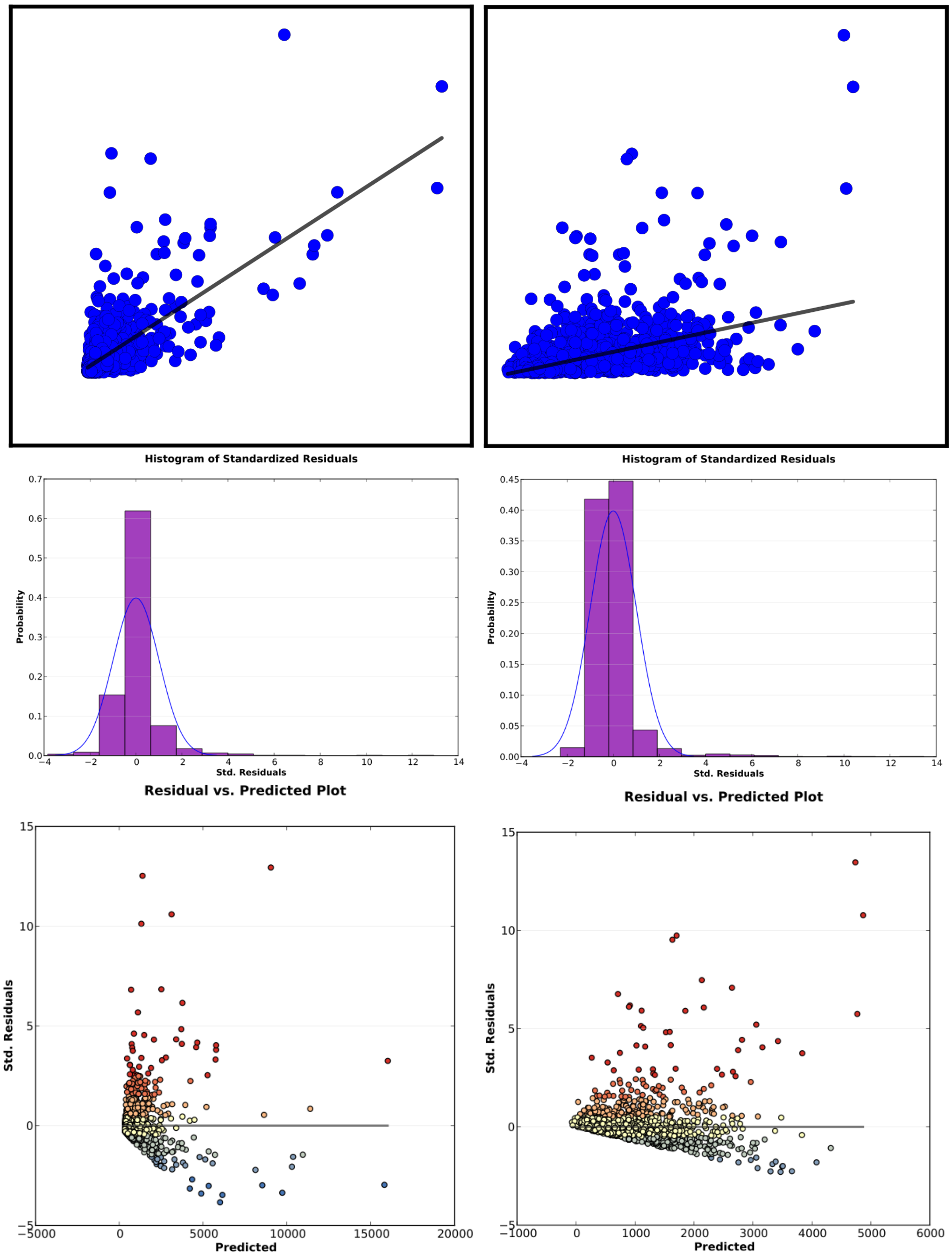


Figure 19: Scatterplot (above), distribution of residuals (middle) and a plot showing residuals in relation to their predicted values (below) for services (left) and rents (right).

6.1.6 *Performance of the model*

The previous analyses have shown that some conditions of the model are not given. The variables are spatially autocorrelated and the residuals are mostly not heteroscedastic and not normally distributed. Strictly speaking, the model is hence not useful to interpret. Berlin's adjusted R^2 reaches a value of 0.72, Munich's model has an explanation of 0.34.

6.1.7 *Interpretation*

As it has figured out the linear regression with land-use data as influence variables is not an appropriate instrument for an exact prediction. The focus of the interpretation of this model approach is therefore more on the question how the variables explain the spatially varying demand of FFCS vehicles in the city. Especially when each variable is tested separately for its linear influence the sign of coefficient can indicate a positive or negative relation to the target variable. In the final model where all significant and non-redundant variables are put together this simple interpretation must be scrutinized due to the violation of some model assumptions and especially the appearing spatial autocorrelation of the variables. The final model from Table 7 however is useful to compare the explanatory power of the two cities.
Even though the R^2 has to be treated with caution it is obvious that the land-use data is a good choice to explain the booking frequencies in Berlin. The R^2 is 0.72, twice as high as in Munich. The reason for the poor performance of the Munich model will be expounded later on.
The following interpretation thus considers the model for Berlin only. As described at the beginning of part I customers predominantly book the vehicles in that area where they live. The external factors are therefore also a characterization of the user.
A first look at the list of variables in section B.2 reveals that there are surprisingly no variables about the age of citizens in a district. First analyses from Schmöller et al. in [123] have shown a slight positive trend for citizens between 35 and 49 whereas elderly people from 65 on are a negative influence for the carsharing demand. This corresponds to the analysis from Müller et al. in [105] who found out that the age structure of a typical FFCS customer is exactly in this period of life. But it does not seem to be significant enough to pass the variable selection process.

The other variables describe different types of the socio-economy and use of space. To get an overview, they are categorized in six different categories:

1. open-mindedness
2. type of car user
3. financial situation
4. centrality
5. parking availability
6. number of companies

The first three clusters give evidence about the user, the latter ones denote spatial features of the area.

Variables like the environmental affinity index or the telecommunication type indicate the *open-mindedness* of the population. If citizens have reservations towards new technologies like smartphones and prefer fixed network telephones instead the probability will increase that there will be less demand for carsharing vehicles in these areas.

Regarding the number of FFCS bookings in a district, it also plays a part what *type of car user* lives there. Variables measuring this fact are the percentage of business and private car users as well as indices of company car drivers, for the affinity to leased private cars and for frequent drivers. All indices have a positive impact on the booking numbers. That is no wonder because they describe the best requirements for a frequent use of a rental cars. A good FFCS customer needs a general affinity to cars but also a sympathy for new and varying kind of car models which does not necessarily need to be owned. The variables moreover indicate a potential multimodal transport choice of the users. Private car-ownership impinges the number of bookings negatively. The fact that some influence variables (e. g. "affinity to leased private cars (index)" change their trend in the complete model shown in Table 7 must not be over-interpreted since the factors are not totally independent.

FFCS has time-based rates. A trip of 10 minutes costs approximately the same as a ticket for public transport. Most trips take a longer time and make this kind of transportation system only attractive for those who can afford this convenience. The *financial situation* of the citizens in a district is therefore a significant factor for the booking demand. It is apparent from variables like rent, number of dwellings in best quality and the purchasing power.

Next to the kind of people living in a district it is important to consider if the place is attractive for public life. Hypothetically, every second trip is not the way home but to a place where the user can satisfy his needs for leisure time, shopping or work. *Centrality* is therefore one big impact on the number of bookings. The linear regression reveals some variables like rent, population density, number of companies and buildings and car ownership. That are indicators for the resident and job density in an area. The company density is regarded in a separate cluster. While the rent is a clear sign for the attractiveness for an area and also for the accessibility to public transport the effect of the number of cars is quite surprising. One may assume that FFCS works better in areas with a low car ownership. To the contrary, more cars in total mean more bookings. It is important to notice that this quantity is an absolute value and not set in relation to the population. It therefore does not mean that FFCS is more successful in peripheral areas - where car-ownership is usually higher - but in districts with a greater land area and thus higher population. This reason is also supposed to be valid for the variable "numbers of houses". The sign of the population density is in a way confusing and is explained later on.
The fifth cluster represents the limiting factor of carsharing demand. It is the *parking availability*. Since FFCS customers are not allowed to finish their trip in a parking garage or on private grounds the public parking space is a crucial requirement for the system. It is taken into the model by considering the street length per district. There are only those streets selected from the road network where a parking availability at the street side is probably given. More streets lead to more bookings. It also offers a possible explanation for the negative influence of population density. A higher demand would normally let the booking demand increase but it may be that the parking situation does not provide space anymore for the demand. It can also be caused by some outliers with an extreme high population density and low public parking space. There are also some very attractive places in the city center with a high demand of FFCS bookings but a low residential density.
The last class is the *number of companies* where most of significant variables belong to. They are additionally the most informative factors and can explain up to half of the booking demand. Nearly all categories of companies appear on the list. The best explanation is given by the number of services (bars, restaurants, business services, ...), government agencies/administrative offices, banks and hotels. These are all companies that have mostly good traffic connection and a high number of walk-in customers.

As explained in the beginning, the final linear regression model does not necessarily have this fine interpretation but an interesting interpretation can be made. In the list of all significant and non-redundant factors in Table 7 there is at least one representative of every character cluster. This is a good result for the explanatory quality. Each factor is important and cannot be represented by other factors of other categories.

The linear regression model for Munich does not provide a comparable explanation as the model for Berlin. There are only 30 variables with the demanded significance, 25 are (mostly spatially correlated) number of companies. Additional factors are the company car driver density index, percentage of business and private car users, the street length and purchasing power in retail. All variables also appear in the model for Berlin. One looks in vain for variables which characterize the open-mindedness of citizens or factors which indicate the centrality of the district. Also the purchasing power does not have that impact as in Berlin.
There are several reasons which are assumed to be responsible for the low (adj.) R^2 and thus the low explanatory power of the Munich model. The purchasing power of the citizens of Munich is much higher than of the Berlin's inhabitants. The jobs in Munich are better paid and people who can afford to pay the high rents in the city center are commonly speaking those who attach importance to money and classic status symbols like an own car. That is the reason why the car-ownership rate is even in the center around the university much higher than in the suburbs of Berlin. The income is thus rather represented by the type of car user. Variables of this cluster have similar impacts as in Berlin in Table 25. That is caused by the fact that the even high percentage of private cars in the center is topped by other districts like Bogenhausen where FFCS is only moderately attractive.
The open-mindedness factors do not appear in the list because the carsharing systems in Munich seem to work independent of this attitude of people. Another point that is different from Berlin is the lively urban districts. Berlin's citizens often identify themselves rather with their so-called "Kiez" than with their city. This multi-central structure of Berlin has already been seen in the booking data analysis. Munich, in contrary, has one city center where a lot of events for leisure time take place. But exactly this central area within the "Altstadtring" is excluded from the operating area. The model thus would estimate a higher booking demand but in cause of the regulations the observed data do

not correspond to that estimation. A similar situation is given for the numerous parking zones at the street that are reserved for residents. So again FFCS is prohibited in high-potential areas. Berlin's operating area has almost no restrictions and there are hardly any pedestrian areas neither.
One other thing that has an effect on the bad explanation by land-use data in Munich: The BMW plant Munich. Many administrative offices and the headquarters of the automobile manufacturer are located in the northern parts of the operating area. Employees do not get a discount but they are more familiar with the system. Public transport in this region is acceptable but the car is the most convenient and fastest way to commute between the different office buildings.
By all these reasons, Munich's FFCS system has to be regarded with its special conditions. But it also shows the importance that the results from the linear regression model cannot be directly transferred to other cities but the special conditions have to be taken into account. Due to the questionable fit of the model the Berlin model is not applied to other cities as it will be done for other models later on.

Summarizing, the linear regression model may not be the best choice for a perfect prediction model for the booking demand but it supplies a satisfying explanation for the booking frequencies, at least in Berlin. Both social and economic data help to figure out main success factors. The characterization of the users corresponds to findings from other research studies mentioned in section 2.4. The hot spots for FFCS carsharing are well-accessible areas with a high attractiveness for jobs as well as leisure time activities.

The next section focuses on the political attitude and thus more on the social background of the users.

6.2 LINEAR REGRESSION WITH ELECTION RESULTS

The election results come from the Bundestag election in 2013. To obtain the best comparison between the data, the chosen data period for the booking data is also 2013, i. e. data set No. 3.

	Berlin		Munich			Berlin		Munich	
	t	adj.R^2	t	adj.R^2		t	adj.R^2	t	adj.R^2
CDU/CSU, 1st	-4.13	0.02	-0.76	0.00	FDP, 1st	5.52	0.04	4.21	0.04
CDU/CSU, 2nd	-5.17	0.03	-1.49	0.00	FDP, 2nd	6.95	0.06	4.03	0.04
SPD, 1st	-6.04	0.05	1.07	0.00	AfD, 1st	-11.04	0.14	-	-
SPD, 2nd	1.00	0.00	-1.02	0.00	AfD, 2nd	-9.07	0.10	-1.99	0.01
Die Linke, 1st	11.97	0.05	-2.72	0.01	Piraten, 1st	-0.32	0.00	3.24	0.01
Die Linke, 2nd	6.37	0.16	-1.18	0.00	Piraten, 2nd	3.49	0.01	2.03	0.01
Die Grünen, 1st	14.25	0.21	0.93	0.00	NPD, 1st	-10.84	0.13	-1.81	0.00
Die Grünen, 2nd	19.00	0.32	1.64	0.00	NPD, 2nd	-10.05	0.12	-0.96	0.00

Table 8: Results of the linear regression with election results standardized by the street length

6.2.1 *Significance of the explanatory variables*

The procedure of model building is analogous to the previous section 6.1. In the first stage, all variables are considered separately as it was done for land-use data. It is useful because a possible dependence between the variables is thus suppressed. A complete independence may not be given but can be hold true. Election results are usually not independent. A higher result of one party entails a lower result of other parties. And also the consideration of the differences to the average reveals that the sum of differences has to be zero. A solution is that not every polling district is analyzed. Due to the fact that no constituency is completely within the operating area (see Fig. 7) the variables can be considered independent.
Nevertheless, the significance of variables is first considered in linear models which contain only one variable. This facilitates the interpretation. Table 8 shows the t-test results and the adjusted R^2 of these separate models. The data are due to the positive impact of the street length standardized by this quantity.

The variable selection does not need to be that strict since the number of regarded influencing factors is manageable. For a first variable selection all factors with an absolute t-value of more than 1.96 are taken into consideration. That corresponds to a significance value of $\alpha = 0.05$.
The heterogeneity of the variance of the residuals is not a problem of every variable. In Berlin, the assumption holds for the results of the CDU, SPD and Piraten. The only party for which the Koenker's studentized Breusch-Pagan test

	Berlin				Munich			
	Coeff	σ	t	VIF	Coeff	σ	t	VIF
CDU, 1st	3906.45	519.20	7.52	2.99				
Die Linke, 1st	3044.20	532.84	5.71	1.78				
Die Linke, 2nd	1849.53	536.00	3.45	2.85				
Die Grünen, 2nd	8876.38	507.06	17.51	2.31				
FDP, 2nd	10210.04	1600.18	6.38	2.38	5042.32	1251.48	4.03	-

Table 9: Results of the linear regression with election results for Berlin and Munich after omitting redundant and non-significant variables.

fails in Munich is the Pirates. The results of the separate models can therefore taken seriously in most instances.

6.2.2 *Relationship between dependent and explanatory variables*

A first inspection of the influences listed in Table 8 shows no unexpected relations. Conservative or right-wing parties have a negative impact on the number of bookings while a higher result of left-wing and ecologic parties let the booking numbers increase.

6.2.3 *Redundancy of explanatory variables*

Next to the significance of the influence factors the redundancy of them need to be checked. The variables are from this point on again considered together in one model. After recursively selecting only those variables with a VIF less than 7.5 the model stated in Table 9 becomes the final solution.

The liberal party FDP is left as the only significant and non-redundant variable in the model for Munich.

6.2.4 *Heterogeneity of residuals*

The model for Berlin shows a bias in the residuals. Especially for larger predicted values the variance of residuals increases. Consequently, the JB test shows a significance. The Munich model consists of just one variable and thus has not a severe bias. In cause of the high spread of predicted values the JB test is significant, too.

6.2.5 *Completeness of the model*

It is no surprise that all variables show a significant Moran's I test with positive z-values. That means that there are clusters in the data that are not assumed to appear randomly. It is clear that election results with this kind of high spatial resolution are correlated to their adjacent polling districts. The problem is the same as for land-use data. The spatial accuracy of the data that is necessary to explain the booking data results in violation of the model assumption.

6.2.6 *Performance of the model*

The performance of the linear regression model is better than for land-use data. Some assumptions are not given either. It is the homogeneity of the residuals which does not hold for every variable, the bias of residuals and the spatial autocorrelation. But the model seems nevertheless more coherent since all influence factors have the same level and can be assumed to be independent. The adjusted R^2 for the model of Berlin is 0.46 and quite satisfying. The approach for Munich with only one significant non-redundant variable is regarded to have failed and leads to an adjusted R^2 of just 0.04.

6.2.7 *Interpretation and Evaluation*

In analogy to the interpretation of linear regression models for land-use data, the focus of interpretation is on the effect of significant variables that each has separately considered on the number of bookings. The violation of assumptions is not as severe as for land-use data. The explanatory indicator R^2 of the final model is hence more expressive than in the final model with land-use data. A conspicuity that is also apparent for land-use data is the result for Munich where the final model ends in a single-variable-model. The interpretation is thus constrained to Berlin.

A first view to the results in Table 9 is satisfying and show interpretable results. The significance and explanation of the factors are not as high as for some variables of the land-use data but it is supposed to come from less outliers in the data.

The two parties with a majority of votes are the CDU and SPD. The labor party has a large electorate in big cities whereas the center-right CDU is preferred in rural areas or the suburbs of a city. The negative impact is therefore caused by

the centrality of an area. The electorate of the SPD lives more dispersed in the city.
Interesting is the effect of smaller parties. They have a politically more polarizing position with the consequence that their voters can be characterized more precisely. The strongest significance is observed at the Green party. Nearly one third of the booking demand can be explained just by their election results of the second vote. Also the left-wing party and the liberals reveal a positive trend. Very conservative parties like the AfD or the extreme right-wing party NPD have in contrary a negative influence. The Pirates seem to have a non-homogenous spectrum of voters so that they do not have a clear impact.

To find a link to the linear regression model from the previous section the political attitude of the voters should be seen as a characteristic of the milieu of the population. The milieu comprises the first three clusters defined as result of the linear regression analysis with land-use data. Since the *type of car user* is hard to identify by the voting behavior, the focus is on the *open-mindedness* and *financial situation* cluster.
In a 2013 published study Bach and Grabka analyzed correlations between political tendencies and data of the socio-economic panel. Financially well-situated households tend to vote for CDU/CSU, FDP and the Greens while people with less income prefer SPD and Die Linke ([5]). There is no such study about the open-mindedness but voters of conservative parties typically try to keep existing values and lifestyles up. They are generally not as open to new technologies as supporters of the FDP, Die Grünen or Die Linke. The Greens are moreover typically elected by ecological-acting people ([110]).
With this background the results for Berlin are meaningful. The positive effect of the liberals can be explained with the open-mindedness of the electorate towards new technologies and individual transport. For the voters of the Greens the ecological consciousness and sustainability are additional characteristics having a positive impact on the number of bookings. The positive impact of the CDU/CSU is assumed to come from the over-average good financial situation of their voters which favors the carsharing demand. The party Die Linke stands for a fair social economic system. FFCS is supposed to be attractive in areas with a high percentage of their supporters since they do not consider private possession important.
Considering Munich, the linear regression model seems again to provide no helpful explanation. But that is no surprise; the specialties and restrictions of

the operating area are assumed to have a main impact to bias the demand. In addition to it the *open-mindedness* and *financial situation* which are figured out to be the clusters which characterizes the political attitude in the best way already had no significant effect in Munich. When omitting all non-significant variables from the model just the FDP stays. Typical green-electing districts like Haidhausen in the south-east of the center have just a slight higher demand than elsewhere.

The final model for Berlin (Table 9) is applied for the cities Munich, Hamburg and Cologne to see the transferability and usefulness of the explanatory variables. One has to keep in mind that the model assumptions are not totally given but more met than in the linear regression model with land-use data. A precise forecast for each polling district is therefore not useful. Instead, the transfer of the model aims to check the explanatory variables for their general use to identify FFCS hot spots of a city. The prediction is classified in five categories on the basis of the quintiles of the predicted booking demand. That means that 20 % of the districts with the lowest predicted demand are assigned to category 1 and so on. The same is made with the observed data. An evaluation is after that done by comparing the maps and count the percentage that is predicted correctly.
Each of the pictures in Fig. 20 and 21 show two columns with three maps. The left map shows the categories of the observed data, the figure in the middle the predicted categories. In the third map on the right the districts are colored by the difference of the categories of the predicted and observed data. A negative value which is colored green indicates an underestimation while a positive value for a cell that is then colored red shows overestimated districts. Yellow marked cells represent correctly estimated categories of estimation.

The Berlin model is first applied to Berlin itself. More than 80 % are correctly estimated or vary slightly (see Table 10). The multi-centrality is mapped in the model as well as low demanded parts of the city like Wedding in the north and Tempelhof in the south. The overestimated area in the south-west of the operating area is Friedenau-Steglitz which was just added some months before the point of observation. This explains the low observed demand which changes in the next year as it can be seen in the data evaluation of 2014 in Fig. 26.
As expected the Berlin model cannot explain the booking behavior in Munich either. The centrality is more or less correctly estimated but the model failed for northern parts of the operating area because of the BMW-effect. Areas around

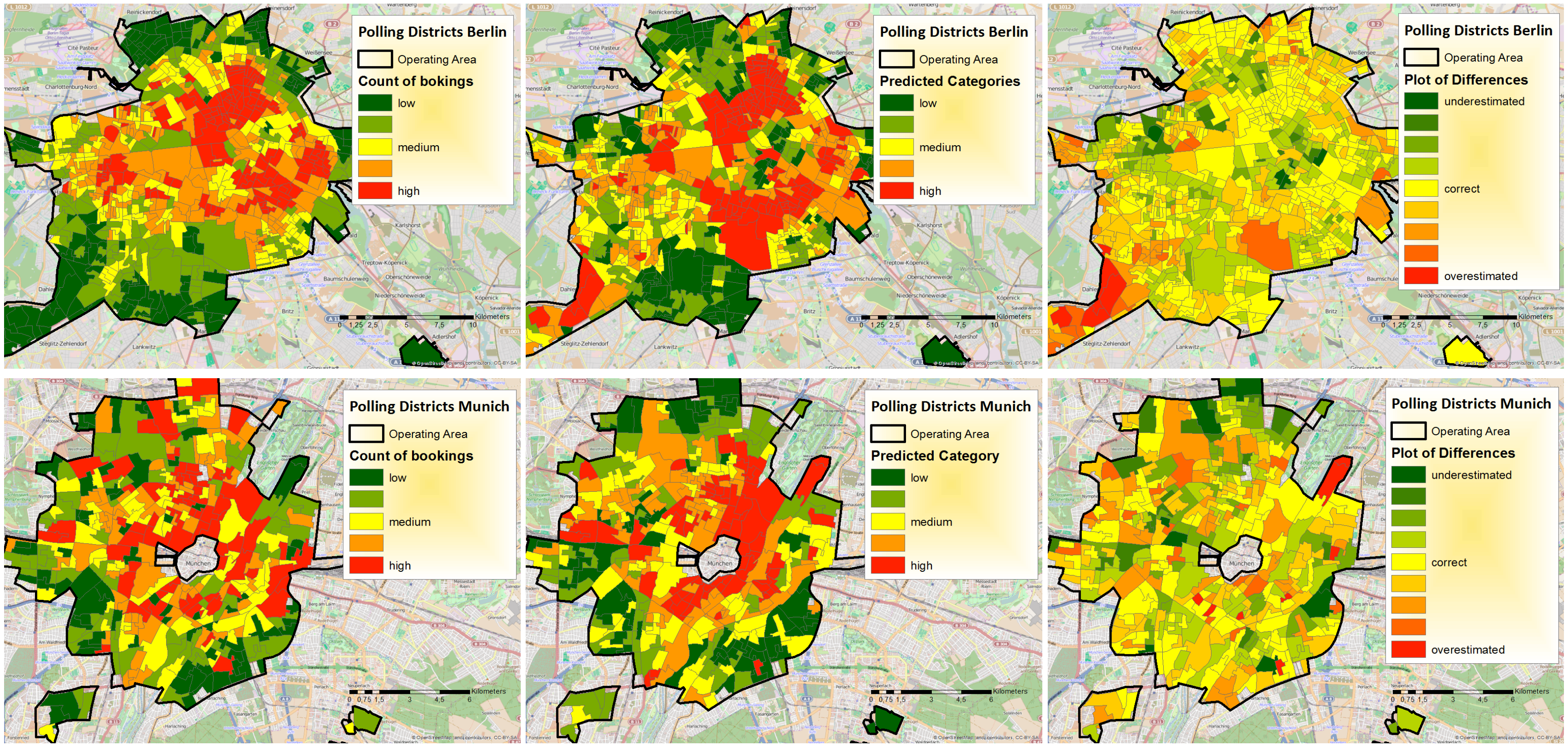


Figure 20: Maps of the cities Berlin and Munich showing the observed (left) and predicted (middle) categories for the number of bookings as well as the difference plot (right). The prediction is based on the linear regression model with election results for Berlin.

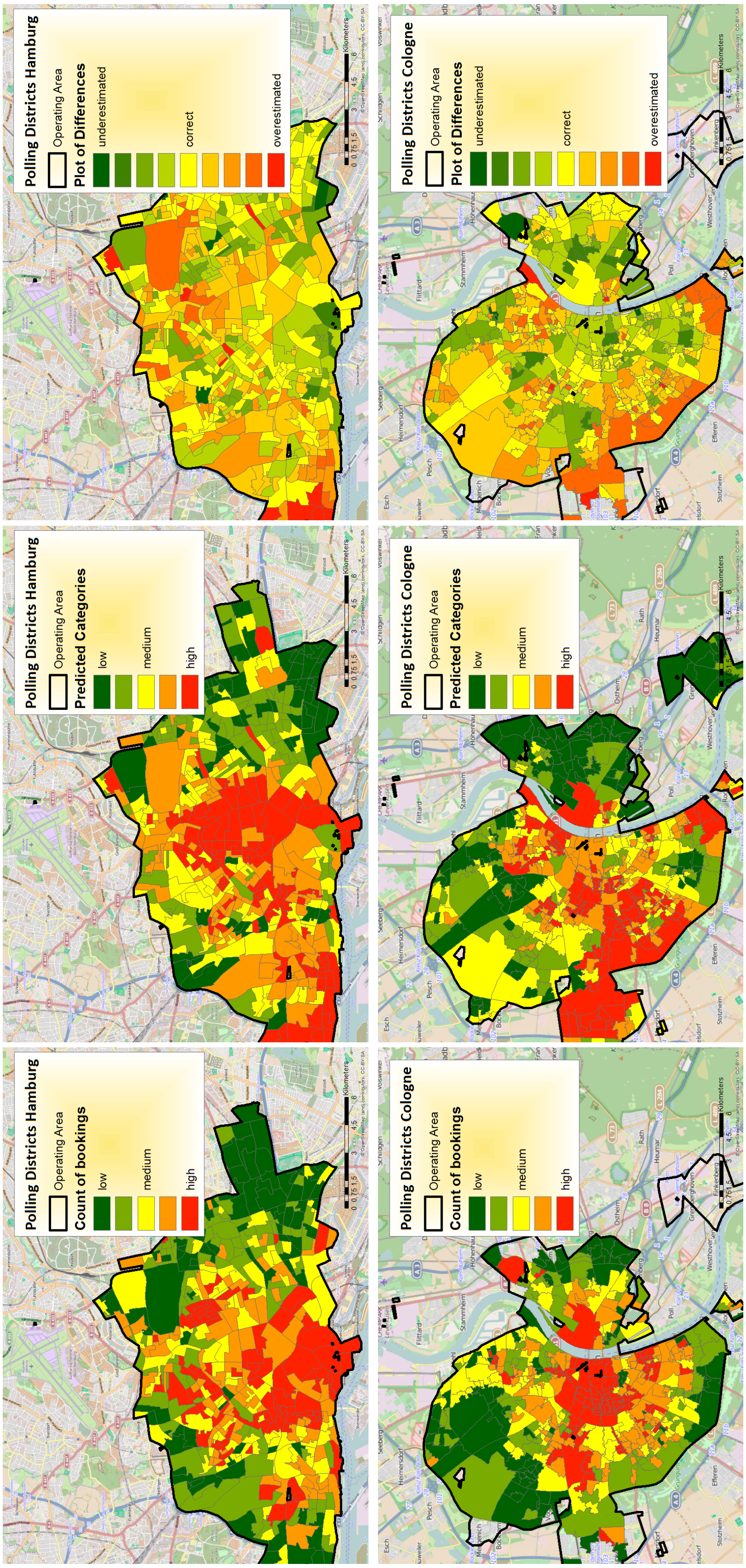


Figure 21: Maps of the cities Hamburg and Cologne showing the observed (left) and predicted (middle) categories for the number of bookings as well as the difference plot (right). The prediction is based on the linear regression model with election data for Berlin.

	-4	-3	-2	-1	0	1	2	3	4
Berlin	0.66	2.26	6.24	19.39	41.83	20.32	6.51	1.99	0.8
Munich	2.23	4.74	11.42	16.99	27.86	19.22	11.14	4.18	2.23
Hamburg	0.98	2.15	10.35	15.82	36.91	18.95	10.35	2.34	2.15
Cologne	0.99	2.97	9.5	16.83	36.73	17.43	9.11	5.35	1.09

Table 10: Quantitative evaluation of the differences between predicted and observed categories for the cities of Berlin, Munich, Hamburg and Cologne. The values are the percentage of cells with the respective difference between predicted and observed category. The prediction is based on the linear regression model with election data for Berlin.

the train station Munich East and in the west of the central station are slightly underestimated while parts of Schwabing-West (north-west of the center) and Oberföhring (north-east) have less bookings than predicted. Only 28 % of the cells are characterized in the right way and less than 2/3 are in the ± 1-buffer.
The focus of bookings in Hamburg is in the south directly next to the central station and the harbors. The Berlin model, however, displaces this spot more in the north. Also parts of the west with a high percentage of green party voters is expected to have a high FFCS demand whereas the east parts tend to be overestimated. The observed data stem from the first three months after launching the system and show therefore a lot of bookings in central districts. The percentage of 37 % correctly and 70 % slightly deviating predicted cells is satisfying but can still be improved.
The same result shows the quantitative evaluation of the Berlin model for Cologne. The comparison of the maps shows the differences of the observed and estimated hot spots clearly. The high demand for the western areas of Cologne result similar as in Hamburg from a high green electorate.

Summarizing, the linear regression model with election results as exogenous variables is a good way to represent the openness of citizens in an area. Especially electors of the left-wing, green and liberal parties have a positive impact on the booking demand. The application of the Berlin model to other cities in Germany with a similar population size yields that the aspect of centrality of an area has also to be considered to predict FFCS hot spots.

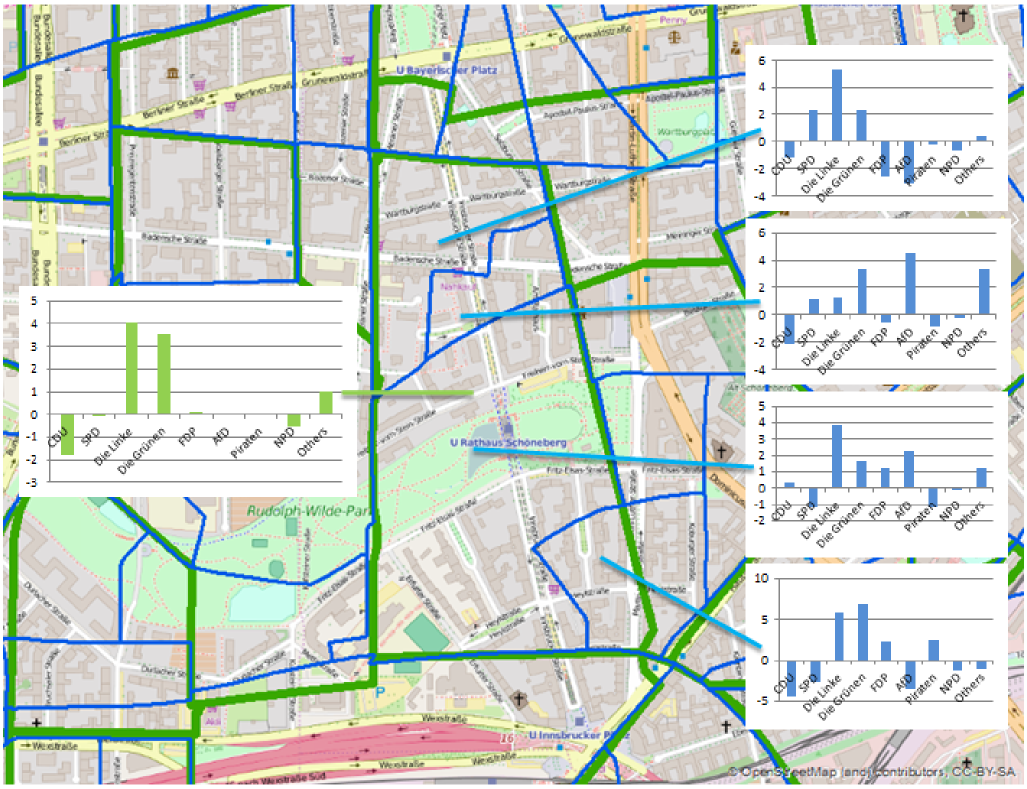


Figure 22: The election results for the KGS22 districts (green) are averaged by those polling districts (blue) intersecting the particular KGS22 district.

6.3 advanced modeling

In the previous sections, the approach was to model and predict the booking demand for FFCS systems by linear regression. As it turned out the approach cannot provide a concrete and exact forecast. It can just serve as an explanation model since some main assumptions of the model structure are violated. One problem is the target variable. The number of bookings is a count variable. Linear regression models do not necessarily exclude count data in general but it is more challenging to fulfill the assumptions of the model with this kind of data. Furthermore, the last two sections model bookings with land-use data and election results separately. For a general understanding of FFCS it is helpful to consider both data sets in one model. For this issue the data sets have to be merged spatially. ArcGIS provides a tool to add data from another set to the data of a target layer. The target layer is in the present case the KGS22 district map, the data to add comes from the polling district zone.

In Fig. 22 it is schematically drawn how the merging process proceeds. Every polling district that intersects a district from the target layer is taken into

consideration for calculating the election results. The election results for the KGS22 districts are calculated by taking the mean of all overlapping polling districts. Unfortunately, ArcGIS does not offer the opportunity to weight the districts by the area which the polling district covers from the target layer. But it is esteemed to be negligible since the results do not differ that much and the data is proven to be autocorrelated.
Next to the two data sets information on particular POIs is added to the data. Inspired by Wagner et al. in [139] the POIs are first spatially aggregated over a net of $100\mathrm{m} \times 100\mathrm{m}$ cells. Subsequently, the POI density is added to the exogenous variables. Eight categories of POI are considered. The names in brackets are the types of POIs taken from the OSM data set.

- bars (bars, cafés, fast food places, restaurants)
- late-night attractions (cinemas, nightclub)
- touristy attractions (attractions, arts centres, artworks)
- accommodations (hotel, hostel)
- ATM (ATM, banks)
- buses (bus stops)
- taxis (taxi stands)
- subway or suburban train stations (stations)

This data is supposed to be a better indicator than company counts (e. g. count of services) for the attractiveness of an area. The density is not only measured by the simple count of each POI group. Additionally the GiZ scores of the Getis-Ord-Gi*-test are considered. Background information about the test are given in [113] and [63]. By this the effect of POIs to their neighbor cells is also respected. Some further variables serving as potential explanation for the booking demand are the distance from the KGS22 district centroid to the city center and the district center as well as the size of the area of the district. A sketch of the model design is given in Fig. 23.
There is no such strict work schedule for GLM as for linear regressions models. The model building follows the general idea of this procedure proposal. First, redundant and non-significant variables are omitted from the model. After checking all model assumptions the model is examined regarding its performance. The interpretation of the results follows in the last paragraph.

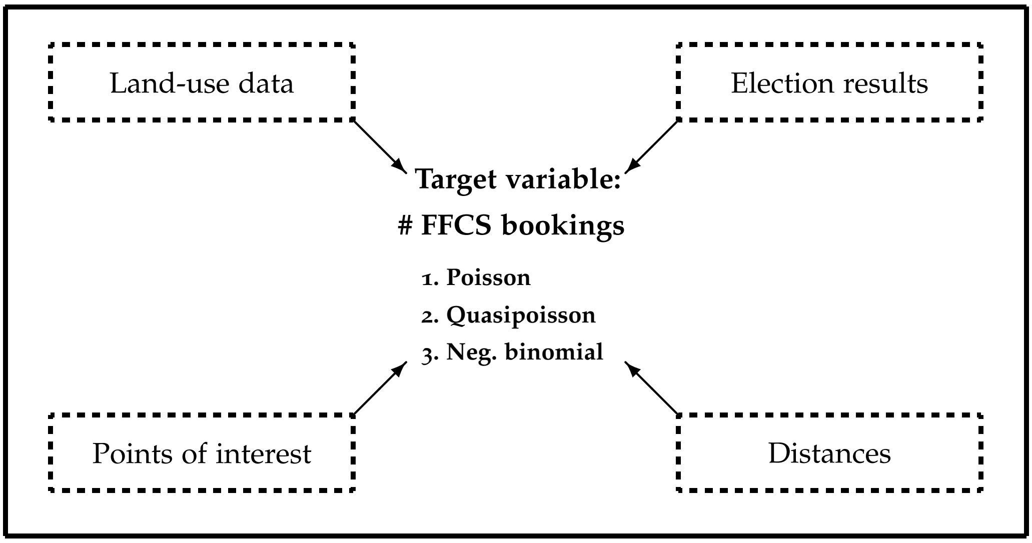


Figure 23: Scheme of the model design for the Poisson, Quasipoisson and NB model

This model approach is just performed for Berlin since Munich has turned out to be a city where the demand is hard to predict with the present exogenous variables.
The booking data is taken from 2014 (dataset No. 4) because this evaluation was performed in 2015 and the data should be as new as possible. The difference between the data sets of different years is negligible. The interpretation of the following models is thus also meaningful in comparison to the previous linear models.

6.3.1 *Variable selection*

Unlike in the approach for linear regression the variables are first analyzed regarding their redundancy. The VIF is at the beginning calculated in the saturated model. The variable with the highest VIF value greater than 7.5 is first eliminated. Analogously to the procedure for the linear regression the VIF values for every variable of the new model is recalculated and the selection continues until all redundant variables are omitted. The implementation for `R` is taken from [11].
The 108 variables left in the model are listed in Table 26 in appendix B.3. The significance of a variable depends on the underlying GLM.

6.3.2 *Performance of the GLM*

Each GLM is considered to model the booking frequencies with the recently mentioned choice of non-redundant variables as exogenous factors. The table lists the coefficient (estimate), standard error, z-value and significance level of the factor for the Poisson, Quasipoisson and negative binomial distribution. The significance levels α are 0.001 (***, 99.9 %), 0.01 (**, 99 %), 0.05 (*, 95 %) and 0.1 (., 90 %). It is remarkable that almost every variable shows a significance in the Poisson model. The z-values however vary much and reveal an extreme significance. This is a first indicator for an overdispersion that can only be modeled with the dispersion parameter ϕ in the Quasipoisson model. The summary of the Quasipoisson model affirm this presumption. A dispersion parameter of $\phi = 299.13$ is estimated. The Quasi-Poisson and Poisson model reveal the same estimation. The significance of influence factors is a point where they distinguish. A dispersion parameter close to 1 would make a differentiation needless. The high ϕ-value however is a sign that the assumption of expectation-variance-equality does not hold and the overdispersion needs to be considered.
In the following, the models are calculated again – each with its significant variables. The model with all Poisson-significant variables (Poisson model) contains 91 different factors while the model with the Quasipoisson and negative binomial significant variables (Quasipoisson and NB model) get along with 22 and 11 variables, respectively. The quality characteristic which is checked next is the homogeneity and normal distribution of the residuals. The analysis confines to a graphical comparison. Fig. 24 and 25 show the observed and predicted values for the target variable as well as the residuals of the Poisson/Quasipoisson and the NB model. This result is satisfying. There is no heterogeneity visible and the residuals are mostly normally distributed. The residuals are never less than -1 but take values up to 8. This slight positive leaning indicates that the model tends to underestimate the data. The error is more severe in the Poisson/Quasipoisson model.

Before comparing the McFadden's indices the AIC of the Poisson and NB model are compared. The AIC of the Quasipoisson model does not exist because of its pseudo-density function. The remarkable higher value of 443 511 for the Poisson-model in comparison to the AIC of 27 545 of the NB model is a strong indicator that the latter model has the better explanatory power.
Finally, McFadden's R^2 is calculated for the models containing significant non-

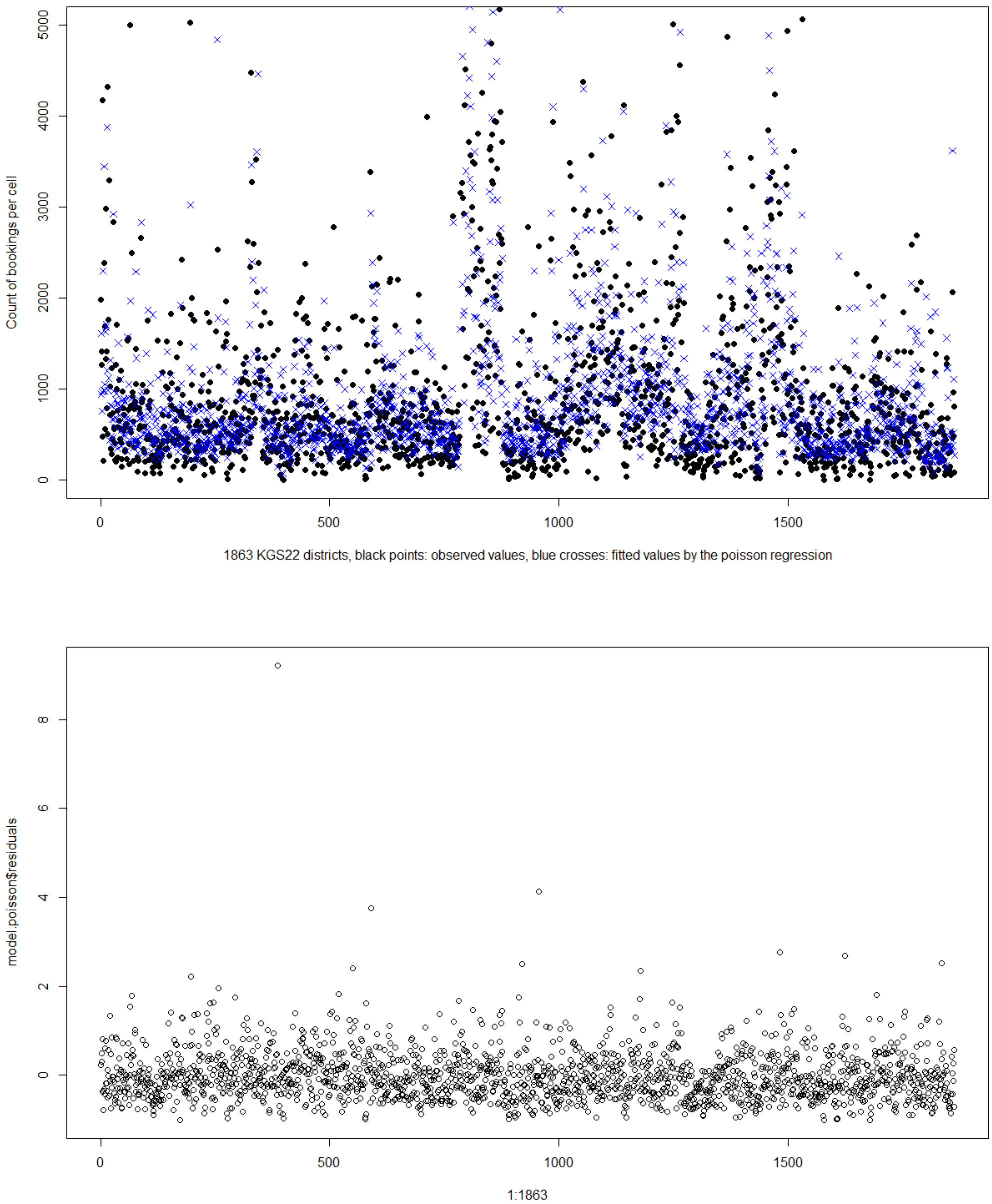


Figure 24: Plot of the fitted and observed values (above) and the residuals (below) of the Poisson/Quasipoisson model.

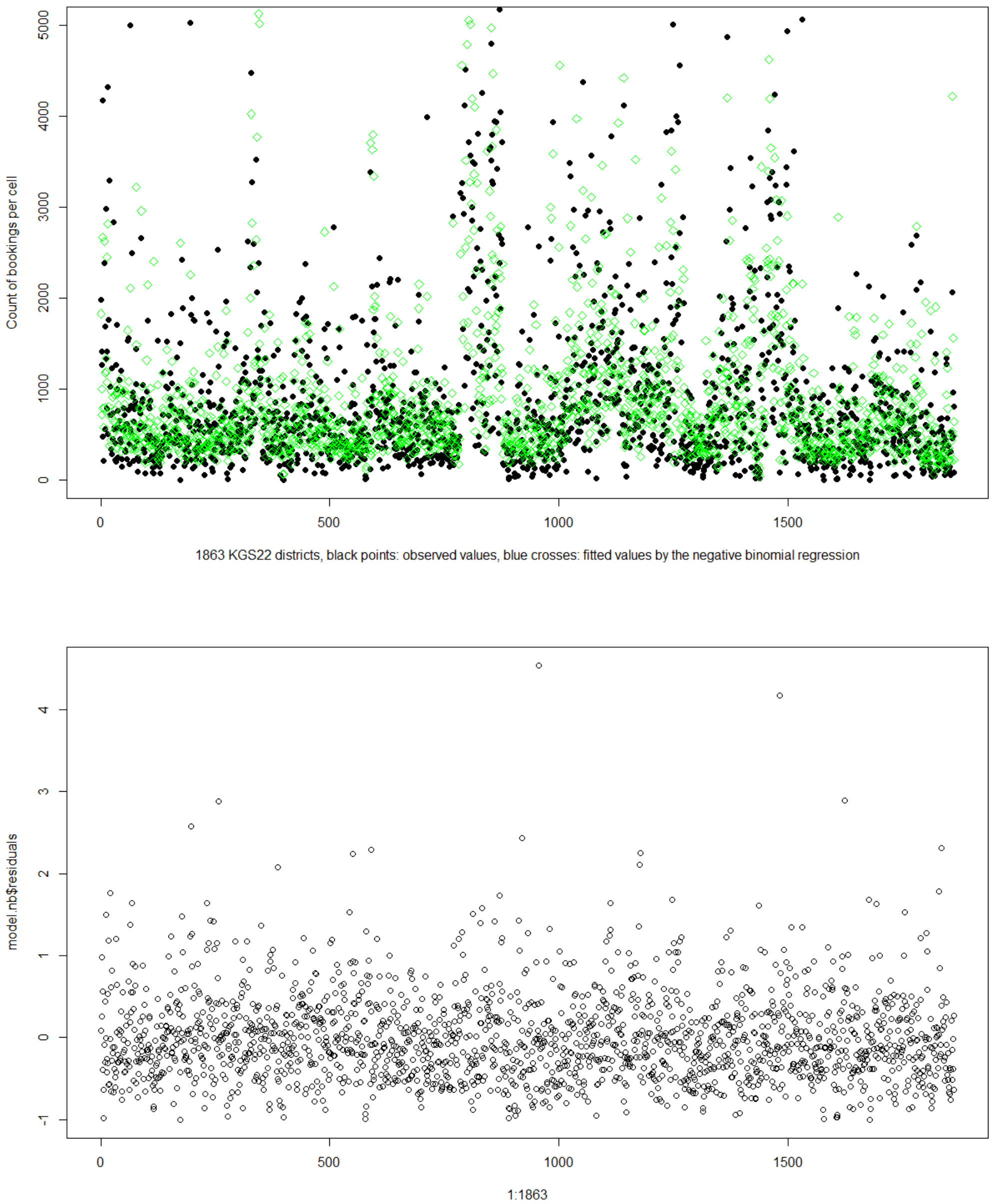


Figure 25: Plot of the fitted and observed values (above) and the residuals (below) of the negative binomial model.

redundant variables only. With the analogous reason as for the AIC this values cannot be calculated for the Quasipoisson model.

	$R^2_{McFadden}$	$R^2_{McFadden_{corr}}$
Poisson	0.790	0.790
neg. binomial	0.068	0.067

The differences between $R^2_{McFadden}$ and $R^2_{McFadden_{corr}}$ are negligible, the differences between the two models are unlikely higher. This effect is typical for all coefficients of determination. More included variables yield to a higher R^2. Hence it is important to see if the model assumptions are given. In the present case the NB model turns out to be the much more adequate model. The R^2 value is acceptable regarding the fact that even values from 0.2 up are a hint for a very good model.

6.3.3 *Interpretation and Evaluation*

The focus of interpretation is first on the non-redundant variables in Table 26 in appendix B.3. The look goes to the new density and distance variables. In the Quasipoisson and NB model, the distance to the city center and the density of bars are both significant. The count of POIs however does not play a part. Quite surprising is the fact that even in the multi-central Berlin the distance to district centers has no influence. The reason for this is the fact that not every Kiez figures out to be a hot spot and there is thus no overall trend for a concentration in sub-centers visible.
It is also interesting to see which variables do not appear in the list. It is for instance the Greens who are represented by other factors. It cannot be said which factor made it redundant but it is supposed to be variables that also have a similar central-oriented structure. Other variables however do not appear in the linear regression model. That are e. g. those which provide information about the population's age structure in a district. The percentage of small children (3-5 years old) are significant in the NB model.

In the following, only the NB model which turns out to meet the assumptions of GLM modeling better than the Quasipoisson model is discussed. There are eleven significant variables for $\alpha = 0.001$ (***, 99.9 %) in the model that includes every non-redundant factor. Regarded separately, they do not necessarily have to be significant but the sign of the coefficient can tell something

	Estimate	Std. Error	t value	Sign
citizens per sqkm	-3.654e-05	2.183e-06	-16.74	***
% citizens: 3 - 5 yrs old	0.03	0.096	0.331	
% quality of buildings: good	0.002	0.001	1.69	.
% quality of buildings: very	-0.01	0.001	-11.09	***
% households with ⩾3 persons	-0.008	0.007	-1.15	
# households with net income 900-1500€	0.004	0.001	8.064	***
density of private cars (index)	-0.005	0.001	-4.903	***
rent	0.373	0.02	22.32	***
NPD, 1st vote	-0.478	0.037	-13.01	***
distance to city center	-3.083e-05	4.911e-05	-0.628	
street length	3.455e-04	1.344e-05	25.72	***

Table 11: List of significant and non-redundant variables considered in separate negative binomial models for Berlin

about the impact. The results of the separate modeling are shown in Table 11. These significant and non-redundant variables are also evaluated together (Table 12). The exact values of the coefficients and t values differ slightly from Table 26 but always have the same sign and significance level.

The only representative of the *open-mindedness* cluster is the percentage of those who gave their first vote to the right-wing NPD. The impact is as expected negative. As variable that identifies the *type of car user* can be seen at most the car density. The *financial situation* of citizens in a district are measured by four factors: rent, percentage of buildings in good and simple quality and the number of households with a net income between 900 and 1500€. It is surprising that the latter variable has a positive influence. One explanation is that the absolute, and not the percentage values are counted. Districts with a larger area size thus could have more households and FFCS bookings but relatively seen the impact of households with low income are meaningless or even negative. The percentage of buildings showed to be significant only in the joint model. Its effect is moreover only positive in the separate model. The positive trend makes the variable part of the financial situation cluster. An interpretation for the joint model is more difficult because of interaction effects between

the variables. Influences like rent or the distance to the city center represent the *centrality* of a district. A cluster that does not appear in the NB model is the *number of companies*. The main reason is that the relation of these variables towards the number of bookings is more linear. A linear regression model is therefore a more appropriate instrument for modeling than the GLM for count data. The negative sign of the coefficient from the variable population density can be interpreted as for the linear regression models. In very dense areas the *parking situation* is too critical. The street length has therefore again a positive effect.
The age variable and the household size build an extra cluster that can best be characterized as *family* cluster. Both factors are separately considered not significant in the NB model. The percentage of 3-5 year-old children has moreover an ambivalent influence. In the separate model, their impact is positive whereas it is negative in the joint model (Table 12). Since both variables are not significant in the separate model, it makes more sense to consider their effect in combination with the other factors. The factors may represent the percentage of young families in a district. Because of the fact that the birth of a child is still a reason for most parents to buy a car and they thereby change their complete mobility behavior, this variable has a negative impact on carsharing bookings. The influence of this new family cluster is estimated to be low because its significance just appears in the joint model and can thus be interpreted as a soft additional effect.

In the following, the Quasipoisson and NB model are applied to the city of Munich and Cologne. Hamburg has to be left out due to missing land-use data for that city.
The presentation of the results is analogous to the transfer of Berlin's linear regression model with election results. The tables show the percentage of districts classified by the difference between predicted and observed categories.
The Berlin model applied for Berlin (Fig. 26, Table 13) shows as expected a very good result. The hot spots of the observed data are more central than in the linear regression model. This is caused by the different grid that is used for this analysis as well as the different data set. In the Quasipoisson model there are some districts in Mitte and the southwestern areas which are underestimated. The NB model works more than satisfying. More than 45 % are predicted correctly and over 85 % have just a deviation of ± 1.

	Estimate	Std. Error	t value	Sign
citizens per sqkm	-2.670e-05	1.837e-06	-14.536	***
% citizens: 3 - 5 yrs old	-0.28	0.06	-4.281	***
% quality of buildings: good	-0.005	8.721e-04	-5.47	***
% quality of buildings: very	-0.004	7.383e-04	-4.841	***
% households with $\geqslant$3 persons	-0.03	0.006	-5.43	***
# households with net income 900-1500€	0.004	3.603e-04	12.17	***
density of private cars (index)	-0.008	7.624e-04	-10.891	***
rent	0.21	0.01	15.231	***
NPD, 1st vote	-0.23	0.029	-8.052	***
distance to city center	-1.352e-04	7.363e-06	-18.362	***
street length	2.567e-04	1.2e-05	21.388	***

Table 12: List of significant and non-redundant variables of the (joint) negative binomial model for Berlin

The observed data for Munich (Fig. 27, Table 14) also show an increasing centrality that is caused by the same reason as in Berlin. Both models overestimate again the demand in Oberföhring whereas the BMW area in the north is once more slightly underestimated. The NB model results in a better prediction than the Quasipoisson model. Nearly 70 % are classified more or less correctly, almost 1/3 of the cells are categorized in the right way.
The city of Cologne also gets a fine prediction by both models (Fig. 28, Table 15). The trend of overestimation in polling districts with a high percentage of green electorates is eliminated. The Quasipoisson model has, as in Berlin, some underestimated districts in the city center; the NB model just fails in some northern parts of Cologne. Even if 41 % of the districts in the Quasipoisson model are predicted right (37 % NB model) the NB model has just slight deviations (± 1: 78 % (NB), 76 % (Quasipoisson)).

In all, the NB model is an excellent instrument to explain and predict hot spots of FFCS demand. The success can easily be observed by a look at the respective difference plot. And also the quantitative comparison with the Quasipoisson model and the linear regression model with election results turns out that the NB model is the most precise choice for modeling.

Berlin	-4	-3	-2	-1	0	1	2	3	4
Quasipoisson	0.16	1.74	7.63	20.37	39.98	20.53	7.73	1.74	0.11
Neg. binomial	0	0.82	5.83	21.41	45.48	18.74	6.54	1.09	0.11

Table 13: Quantitative evaluation of the differences between predicted and observed categories for the city of Berlin. The values are the percentage of cells with the respective difference between predicted and observed category. The prediction is based on the Quasipoisson and negative binomial regression model for Berlin.

Munich	-4	-3	-2	-1	0	1	2	3	4
Quasipoisson	2.41	5.73	9.46	16.7	27.16	20.22	12.98	4.53	0.8
Neg. binomial	2.11	4.02	8.85	17.3	31.29	21.23	10.97	3.72	0.5

Table 14: Quantitative evaluation of the differences between predicted and observed categories for the city of Munich. The values are the percentage of cells with the respective difference between predicted and observed category. The prediction is based on the Quasipoisson and negative binomial regression model for Berlin.

Cologne	-4	-3	-2	-1	0	1	2	3	4
Quasipoisson	0.81	1.63	8.13	16.42	40.65	19.19	9.11	3.58	0.49
Neg. binomial	0.16	1.63	7.48	20	37.24	20.81	9.43	2.6	0.65

Table 15: Quantitative evaluation of the differences between predicted and observed categories for the city of Cologne. The values are the percentage of cells with the respective difference between predicted and observed category. The prediction is based on the Quasipoisson and negative binomial regression model for Berlin.

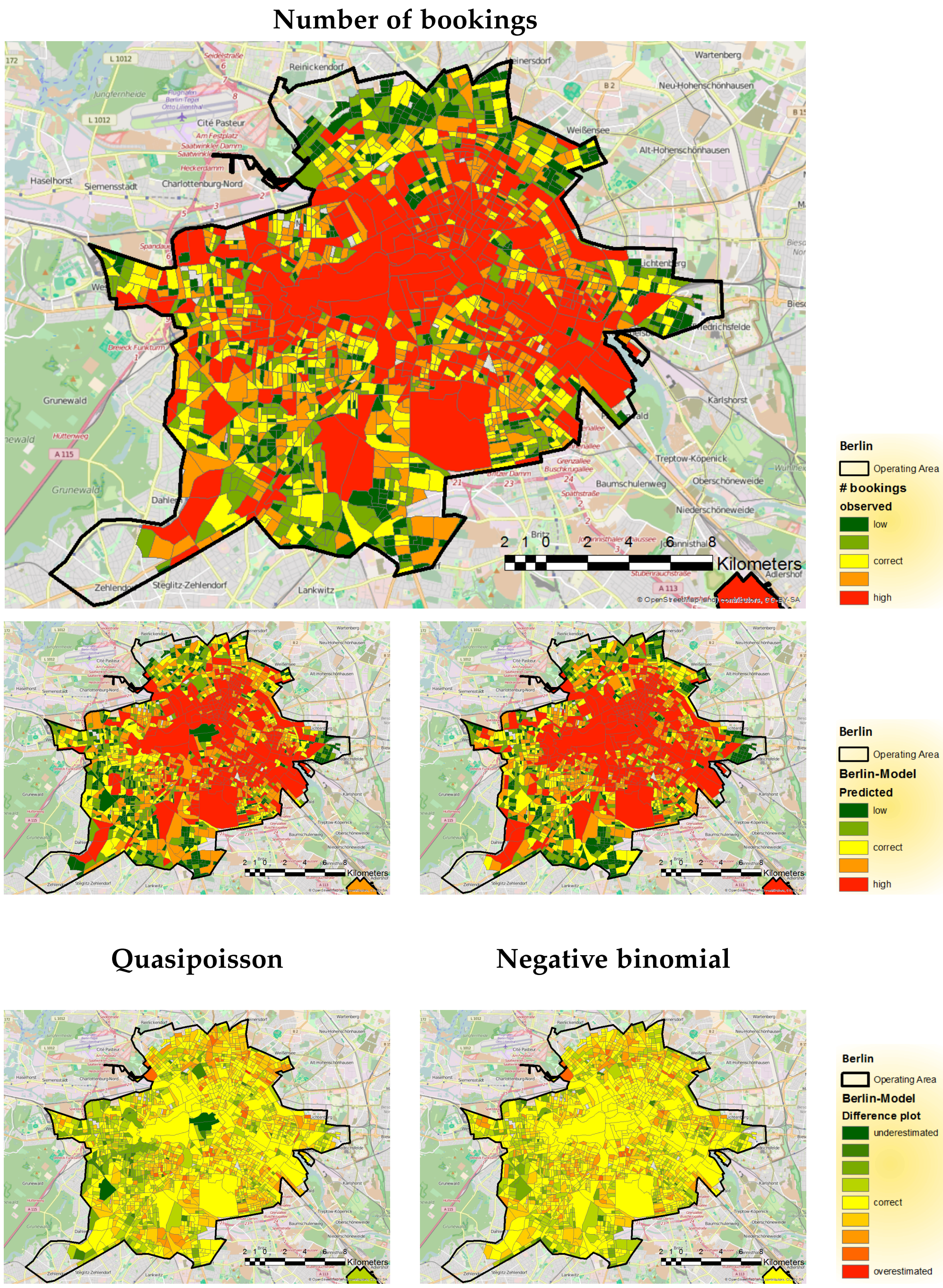


Figure 26: Maps of Berlin showing the observed (above) and predicted (middle) categories for the number of bookings as well as the difference plot (below).

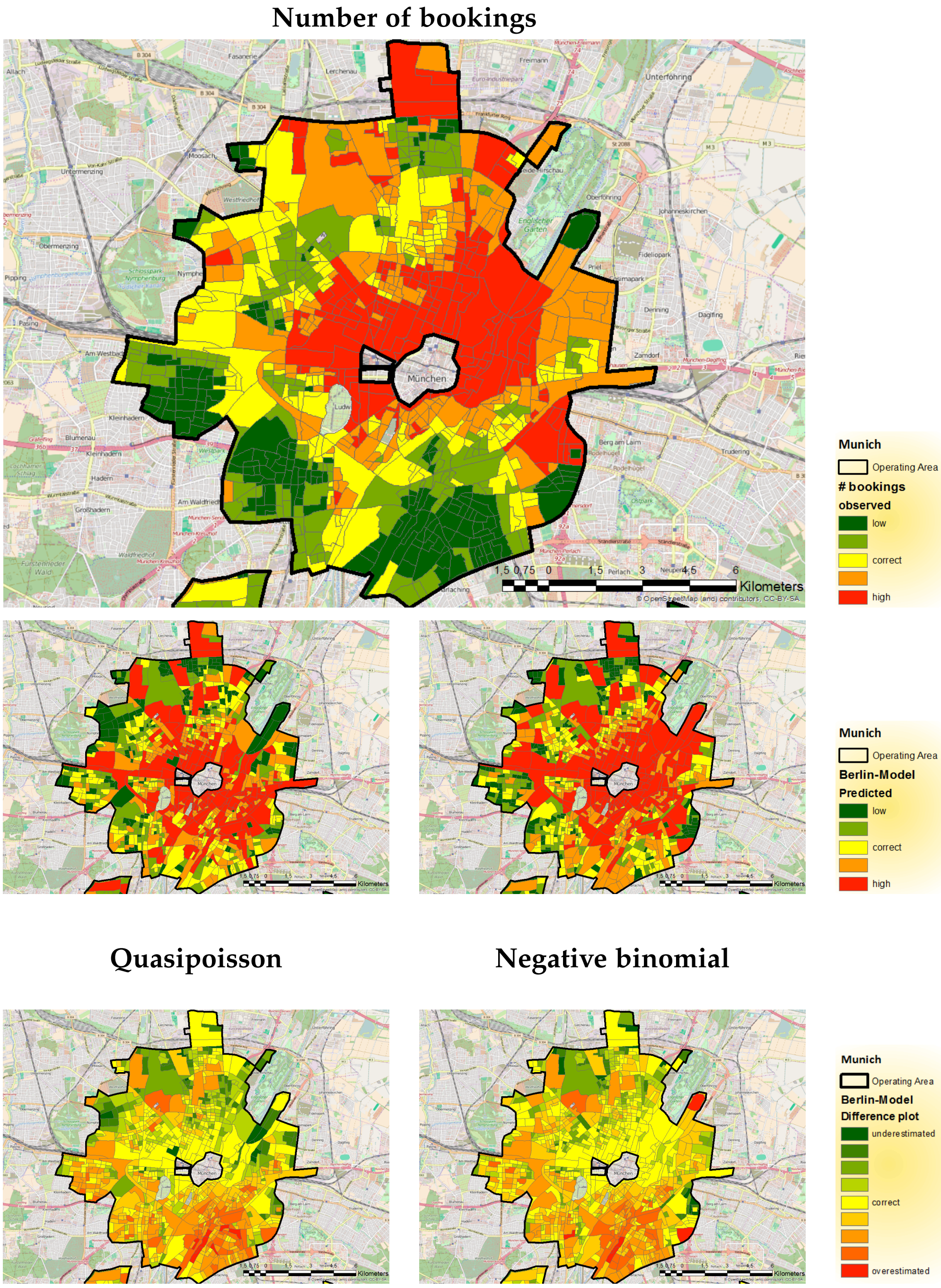


Figure 27: Maps of Munich showing the observed (above) and predicted (middle) categories for the number of bookings as well as the difference plot (below).

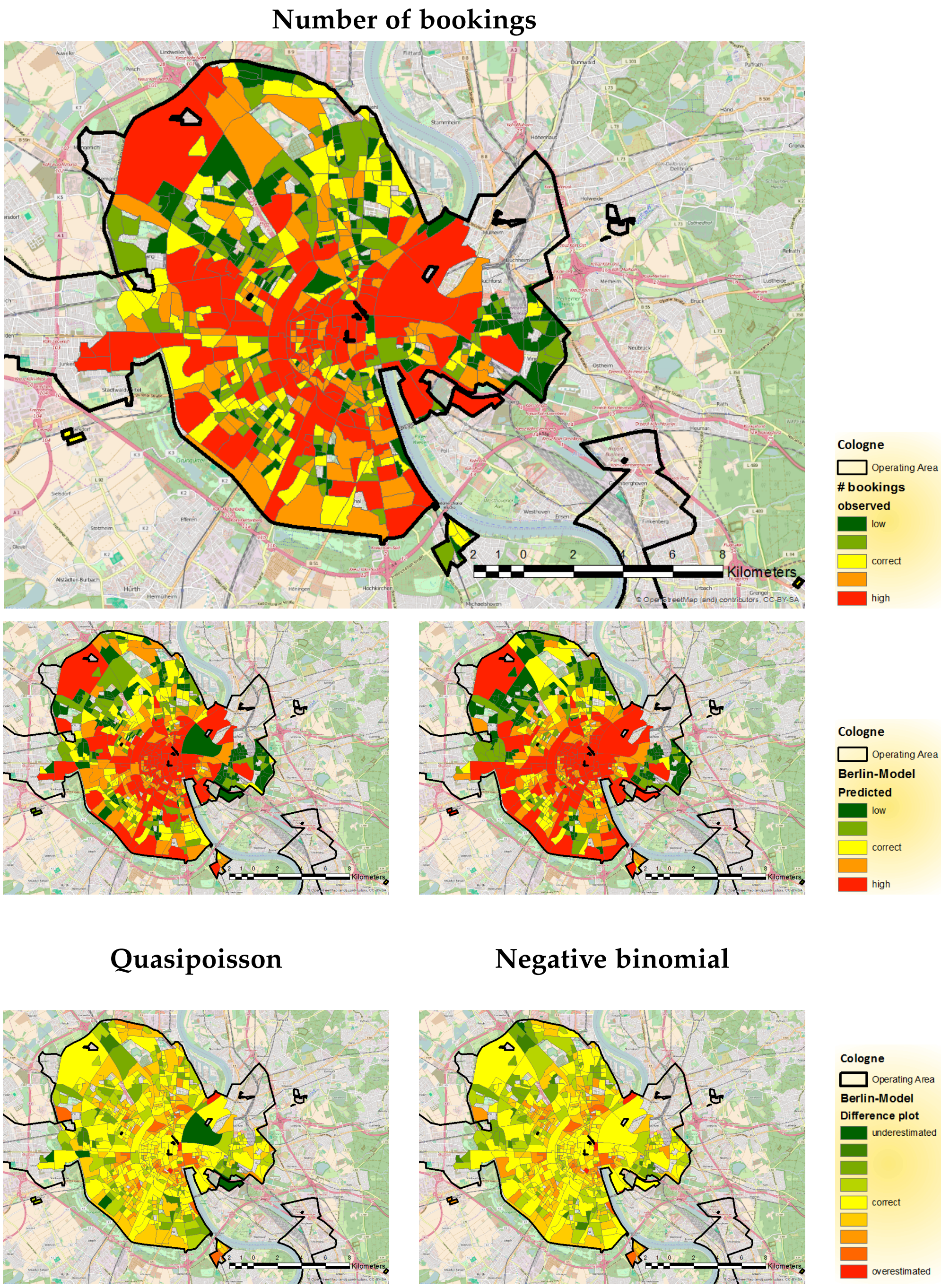


Figure 28: Maps of Cologne showing the observed (above) and predicted (middle) categories for the number of bookings as well as the difference plot (below).

6.4 ANALYSIS OF PARKING DATA

The analysis of the parking restriction areas is performed as described in section 5.3. The basis data comes from data set No. 4 (2014). The profile of standardized booking endings for every parking zone over the period of one week is shown in Fig. 29. Additionally, the trend of short-term parking zones is illustrated in Fig. 30 to emphasize their profile. The lines have not been colored as in Table 5 anymore but simply in green for permitted and red for prohibited parking.

When one considers the development of bookings over the week the typical two peaks per day during working days and the significant evening peak on every day is visible. But it is at least in the evening not specific for a special parking zone, it is conspicuous for every area. In the morning, however, there is a preference for one group of parking areas. A look Fig. 30 shows that short-term parking zones seem to be more attractive in the morning and also during the day. Especially the mixed version with resident parking during the night is very attractive in the morning.
The reason is supposed to be the high number of available parking lots in the morning that is caused by cars of residents which used these spaces during night hours and have to be driven away in the early morning. Another reason for the preference of short-term parking areas is the high fluctuation of parking vehicles in these zones. By this it becomes more likely to find a parking lot spontaneously. The cause can also have to do with the position of the short-term parking zones. They are mostly established in areas with an assumed high number of walk-in customers such as retail shops, restaurants or surgeries. These are exactly those spots that are pointed out in the previous sections as crucial for a high demand for FFCS.

It is difficult to decide what can be concluded for the municipalities. One may require to install or redesignate more parking lots of these parking areas. The chance of short-term parking is that it can be used for resident parking during night hours and thus conflicts between residents and the municipality about the use of parking space can be avoided.
Another byproduct of the analysis that is nice to have but not useful for further analyses is the rate of FFCS customers who end their trip in a prohibited parking zone. 13.8 % of the users are parking offenders within parking license areas. Since 54 % off all trips end in this area, every 13th booking in average ends at a place where parking was officially not allowed at that time. It cannot be

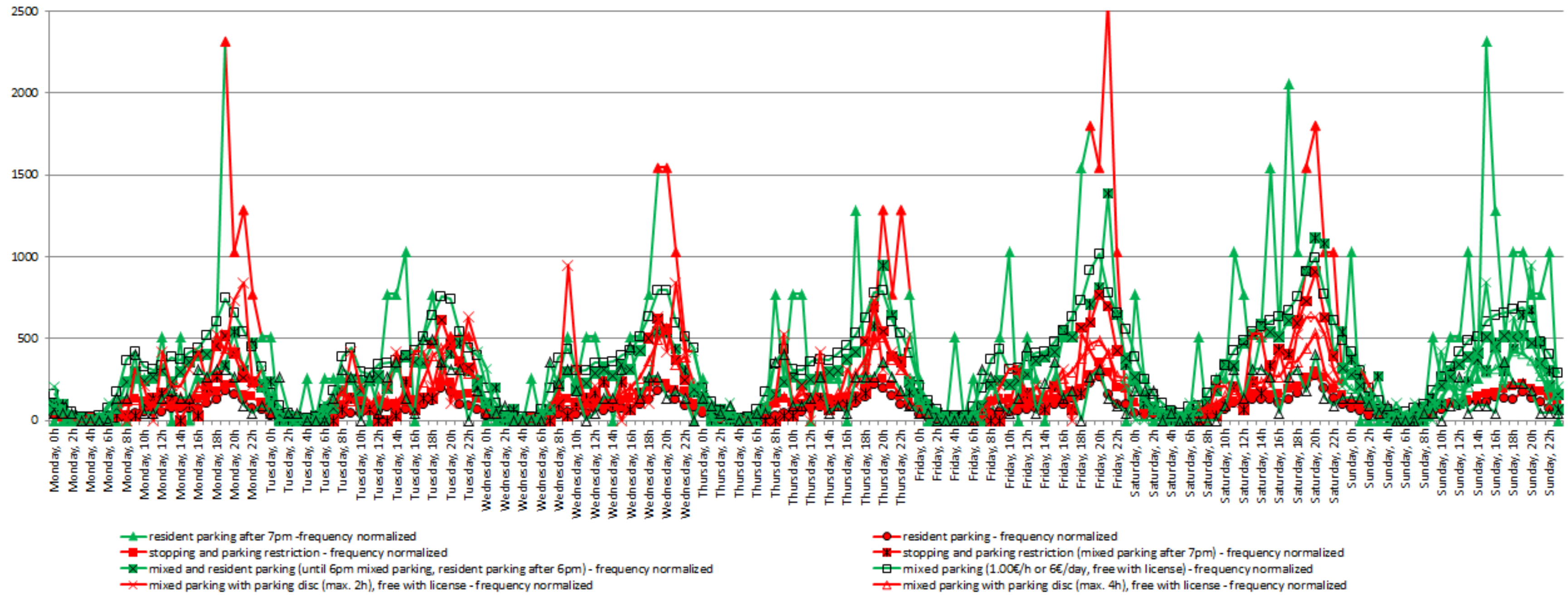


Figure 29: Non-short-term parking zones: Booking endings over the week standardized by the street length of the respective zone. A green line means that the trip ends during an allowed time in the particular parking zone. Red colored booking profiles are parking offenders.

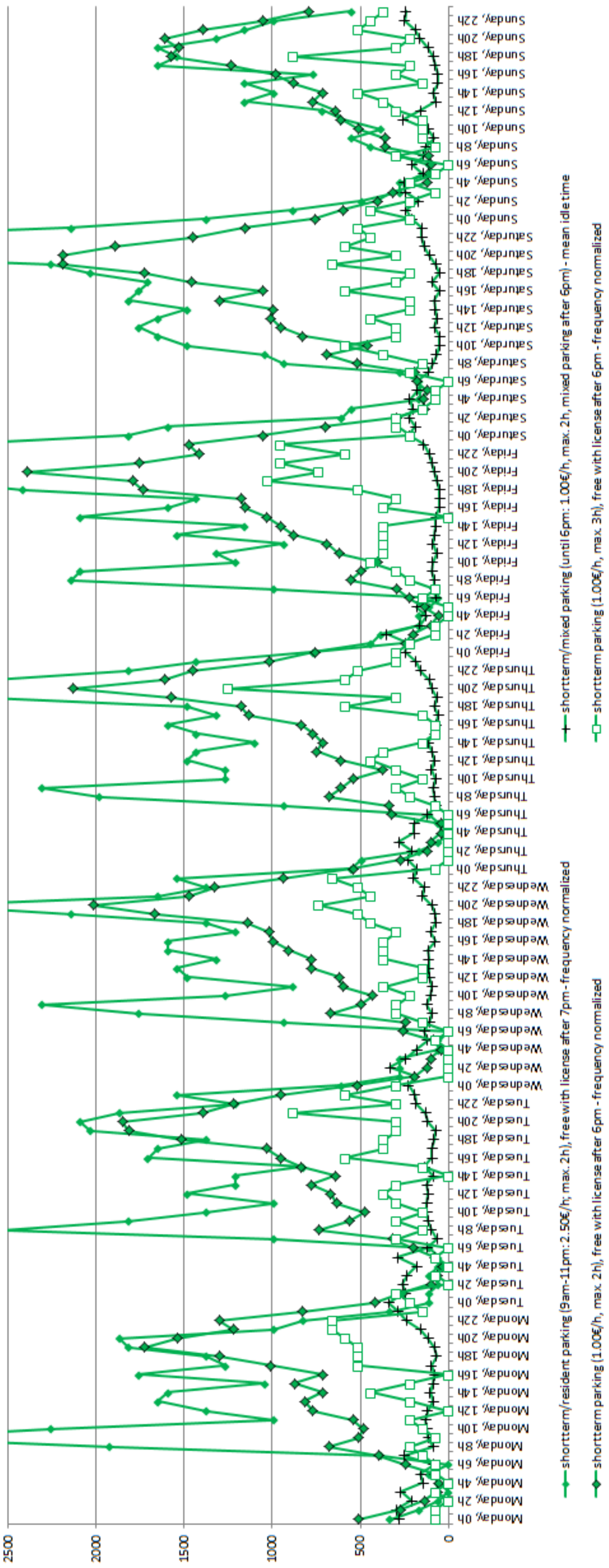


Figure 30: Short-term parking zones: Booking endings over the week standardized by the street length of the respective zone.

daytime	T	df	p-value	daytime	T	df	p-value
6am	1.189	9	0.265	3pm	1.100	8	0.303
7am	-0,777	9	0.457	4pm	1.054	9	0.319
8am	0.917	10	0.381	5pm	-0.929	8	0.380
9am	0.602	9	0.562	6pm	2.312	8	*0.050*
10am	1.449	9	0.181	7pm	2.148	9	*0.060*
11am	-1.382	8	0.204	8pm	1.135	7	0.294
12pm	0.895	7	0.400	9pm	0.555	8	0.594
1pm	1.384	9	0.200	10pm	1.053	9	0.320
2pm	0.623	7	0.553	11pm	0.557	9	0.591

Table 16: Results of the paired t-test for data set No. 1 (Berlin). The significant values are written in italic.

said if it is deliberately or by mistake done by the customers. It is probable that some customers know the rules but take the risk of a ticket instead of paying for further minutes for parking search traffic.

6.5 ANALYSIS OF WEATHER DATA

For the first approach, Berlin's booking data from data set No. 1 is used for the paired t-test. After assigning every hour of the period the binary variable "weather condition" the frequencies of booking are counted monthly and normalized by the number of good and bad weather days at that daytime, respectively. During this period of the first analysis, Berlin was having bad weather conditions for around 48 % of the year.

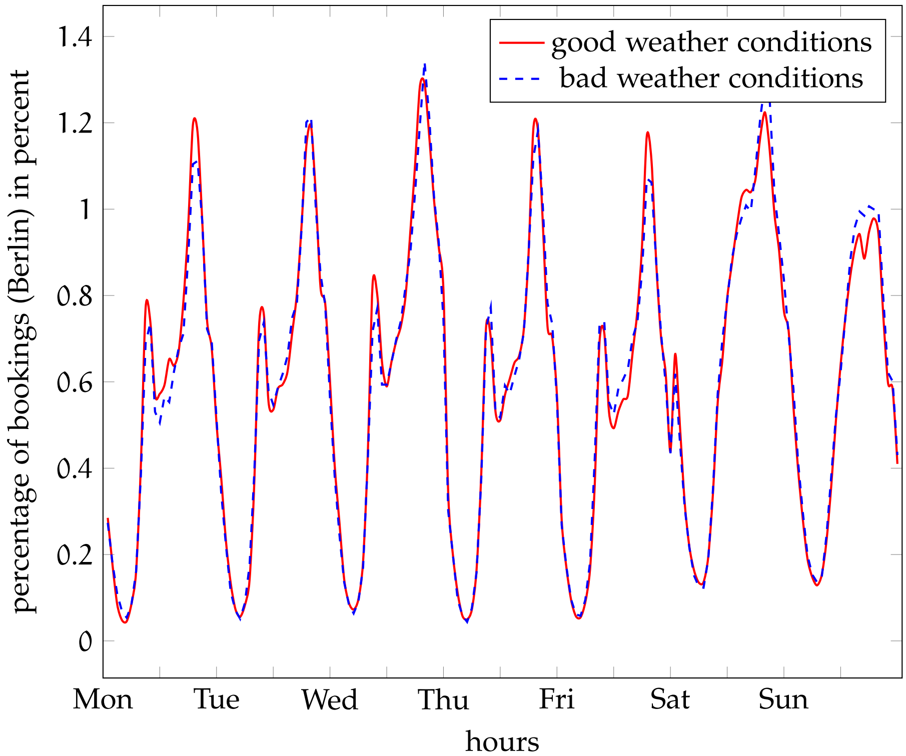
The result is noted in Table 16. The degrees of freedom depends on the number of involved months in the statistic and changes from hour to hour. If all twelve months are included in the statistic, the degree of freedom (df) will equal eleven. The reason for the reduced df is the non-appearance of both weather conditions in a month. In February 2012, for instance, it was exceptionally cold so that was no good weather condition for some day times for a whole months. In consequence this month could not be included in the statistic of the paired t-test.

One can justly criticize the method due to the low degree of freedom. Since

the test works with the assumption of a normal distribution of the T-statistics a rule of thumb is to reach a df of at least 10. The p-values should therefore be handled with caution but they are still not useless to detect a significance.
The less the p-values the higher the significance of the test. The interesting values are the ones for 6 pm and 7 pm. The p-values are in these cases less than 0.1. In combination with the positive T-value this result means that there are with a significance of around 90 % more bookings between 6 pm and 8 pm on days with bad weather conditions than on days with good weather conditions. The test does not provide information about the quantity of the plus.

The second analysis works with a similar statistic but distinguishes additionally the data between weekdays. It was moreover performed with a different data set which includes bookings of every user. During the period of analysis, 63 % and 53 % of the time were bad weather in Berlin and Munich, respectively. The graphics in Fig. 31 show the average percentage number of bookings per daytime and weekday for good (red) and bad (blue) weather conditions in Berlin and Munich.

A first glimpse on the graphics show that the differences in the number of bookings are not as conspicuous as the first analysis could have been expected. During night hours there is almost no difference in Berlin and only marginally in Munich. The peak hours in the evening do not show a consistent pattern. The differences are in both cities more visible than in the afternoon but the number of bookings is not necessary higher on days with bad weather conditions. Mondays, Thursdays and Fridays in Berlin as well as Mondays, Tuesdays, Thursdays and Sundays in Munich have on average more bookings in the evenings with good weather. This is a confusing observation because it contradicts the result of the significance test for the first data set. Nevertheless the two profiles of bookings can be regarded as nearly equal. It is therefore not necessary to give a detailed list of the quantitative differences.
Before finding an explanation, the data is first analyzed relating to the significance of the difference, too. The paired t-test is applied for every hourly pair. The days of the week are the seven samples instead of the twelve months in the first approach. The results of the test are noted in Table 17 and 18. The rule of thumb for the degrees of freedom is not complied due to the construction of the test.



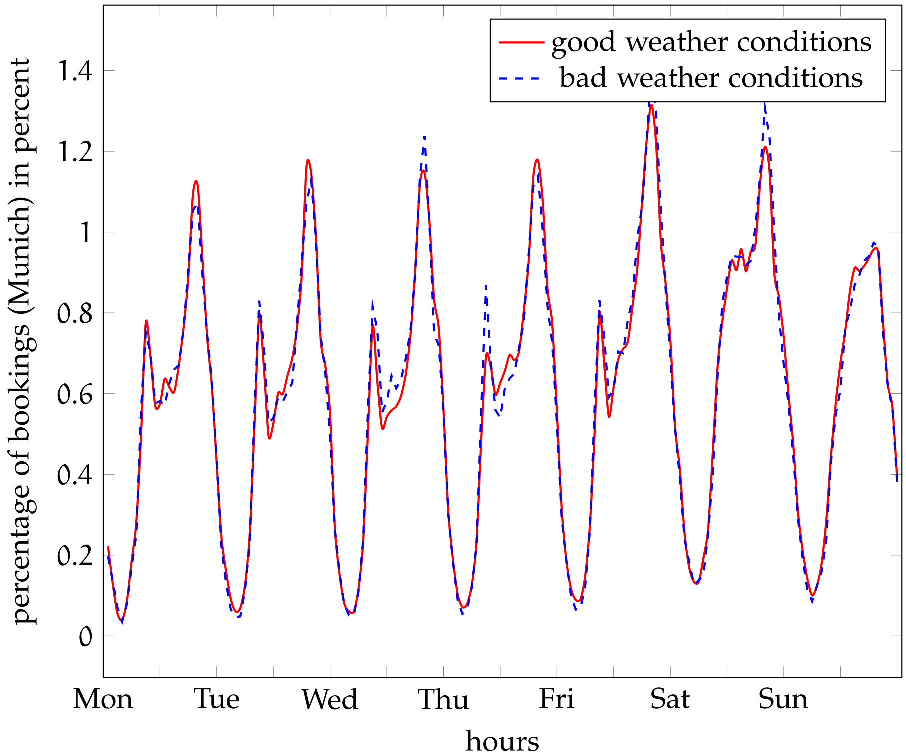


Figure 31: Profile of average percentaged booking starts per hour over the period of a year distinguished by weekdays in Berlin (above) and Munich (below). The bookings are aggregated over hours during good (red, solid) and bad (blue, dashed) weather conditions.

daytime	T	df	p-value	daytime	T	df	p-value
12am	4.651	6	*0.003*	12pm	-0.790	6	0.459
1am	0.511	6	0.628	1pm	0.347	6	0.740
2am	0.317	6	0.762	2pm	-0.516	6	0.624
3am	-1.205	6	0.274	3pm	-1.167	6	0.287
4am	0.583	6	0.581	4pm	0.932	6	0.387
5am	-0.041	6	0.968	5pm	0.978	6	0.366
6am	-0.505	6	0.632	6pm	1.329	6	0.232
7am	-0.246	6	0.814	7pm	-0.486	6	0.644
8am	1.190	6	0.279	8pm	-1.600	6	0.161
9am	-0.820	6	0.443	9pm	-2.052	6	*0.086*
10am	-0.047	6	0.964	10pm	-1.341	6	0.229
11am	0.516	6	0.624	11pm	0.203	6	0.846

Table 17: Results of the paired t-test for data set No. 2 (Berlin)

daytime	T	df	p-value	daytime	T	df	p-value
12am	4.569	6	*0.004*	12pm	0.386	6	0.713
1am	2.177	6	*0.072*	1pm	0.986	6	0.362
2am	2.526	6	*0.045*	2pm	0.367	6	0.726
3am	2.379	6	*0.055*	3pm	0.575	6	0.586
4am	5.468	6	*0.002*	4pm	0.081	6	0.938
5am	3.622	6	*0.011*	5pm	-0.652	6	0.538
6am	3.154	6	*0.020*	6pm	1.868	6	0.111
7am	-2.328	6	*0.059*	7pm	-0.733	6	0.491
8am	-0.548	6	0.604	8pm	0.458	6	0.663
9am	-1.329	6	0.232	9pm	0.036	6	0.972
10am	-0.635	6	0.549	10pm	3.382	6	*0.015*
11am	0.971	6	0.369	11pm	3.405	6	*0.014*

Table 18: Results of the paired t-test for data set No. 2 (Munich)

A 90 % significance in the differences between the number of bookings during good and bad weather can be found from 10 pm till 7 am in Munich and at 12 am and 9 pm in Berlin. The T-value is for these cases mostly positive, except for 9 am in Berlin and 7 am in Munich. This means that there are at those times significantly more bookings at bad weather conditions. During other day times there are no differences in the booking distribution or even more bookings during good weather periods. The significance for Munich during night hours can be explained by the generally low level of booking frequencies at this time so that even a little difference can be marked as significant. The significances for the rest of the hours seem to be more or less coincidentally.
Now it has to be wondered if weather really has got a measurable influence on the booking frequency. The test results show that at least in the morning and afternoon the weather conditions are negligible. But for the evening and night hours the two tests for the two different data sets show different results.

One intuitive answer is the frequency of hours with bad weather varies within the two data sets (Berlin: 48 %, 63 %, Munich: - , 53 %). That would mean that a user is more disposed to use carsharing in bad weather times. But due to the high costs of this transportation mode the FFCS customer will not use carsharing for a longer period of time. In consequence, it becomes more unlikely to consider an increasing demand during bad weather periods when these periods hold on for a long time. Schmöller et al. analyzed in [123] the FFCS booking numbers when a change of the weather took place. They compared in a contingency table the bookings during 2-5 pm and 5-8 pm with and without precipitation. The increase of 6 % on days where the weather turned bad shows that not only the current conditions but also the change of the weather is important to consider.
Another cause for these results could be the time when bad weather occurs. If hours with good weather coincided with generally attractive booking periods (such as Friday and Saturday evenings or evenings before holidays) the test would show a significance for good weather even if the current weather was not the determining factor for the user to choose a carsharing vehicle. It is a natural fact that the weather conditions are independent of the day of week. Given the additional assumption that hours with bad and good weather are uniformly distributed over the year the results of the second analysis will mean that weather cannot be considered as an important external factor for the FFCS booking frequency.

This assumption is held true so that as the only explanation for the different results of the first and second analysis the different customer groups come into consideration. The first test is done for heavy users only while the second one analyzes every booking. In consequence the results would mean that heavy users are more willing to use FFCS vehicles in bad weather periods than a customer using the system sporadically only.
This seems to be a coherent explanation. Eugster showed in [52] that for trips up to 10 km with leisure as purpose cars are significantly more used as transportation mode during bad weather conditions. So it is a matter of habit: A non-frequent customer probably has some reservations regarding the safety during the trip or is simply unaccustomed to conduct a car so that he does not use the system under difficult traffic conditions as they can appear on rainy days. A heavy user in contrast is used to the system and regards FFCS as a fixed component of his transportation mode. In the evening hours, during 6 pm and 8 pm when the trips' purpose is mostly leisure the willingness to use FFCS is raised significantly.
Eugster was not very confident with his results and proposed to consider different variables to describe the present weather situation. One of his proposals that is also thinkable for FFCS data was to look at the cloud cover and the atmosphere pressure. These could give a better realization of the subjective feeling of the current weather situation that is eventually more important for the customers' choice of transport mode than the objective condition.
One may also criticize the method for classifying the weather condition into "good" and "bad". Parts of the results are disappointing and can be caused by the fixed point when the binary dummy variable changed its value. There is the chance to put this right by introducing a probability function that outputs a likelihood of having bad weather conditions dependent on three used variables or even the two further ones mentioned above.
Another option is take the results as they are and classify the weather as an influence factor that has an impact to frequent users but not for the general use of the system. When modeling the carsharing demand this temporal factor can thus be neglected.

6.6 Conclusion

This part of the dissertation aimed to identify the user and provide a model which predicts the demand of a FFCS bookings in a city. The linear regression

models for land-use data and election results give an impression of significant influence variables. The biggest problem for modeling are the quite restrictive model assumptions.
Spatial data are usually correlated. A higher quality of dwellings for instance also means a higher rent index. But also autocorrelations appear since neighbored districts do not distinguished that much regarding the wealth or social class of the citizens. The results of the models are therefore not totally appropriate to present the exact demand. It was on these grounds chosen to estimate demand classifications. With the help of the regression models it is possible to identify variables that correlate with the number of booking frequencies. Six clusters of variables are found out to have a strong impact on the booking demand.

1. *open-mindedness*: The FFCS systems in both cities are analyzed in their first year after launching. It is clear that potential customers and early adopters are above average open for new technologies. These kind of people have generally little reserves against new things and do not mind to use different vehicles for every trip.

2. *type of car user*: It is an advantage if someone does not own a car. Those who lease their private car or use business cars are tendentiously more frequent users. As it was found out by Kopp ([85], p.14; [86]) FFCS users move more multimodally. A high percentage of classic unimodal car users entail usually a lower carsharing demand.

3. *financial situation*: A trip of 10 minutes with a carsharing vehicle is as expensive as a regular ticket with public transport. That makes FFCS only attractive for non-price-sensitive people. These are typically well-educated, high-earning users living in for their society typical districts.

4. *centrality*: Urban life takes place in central districts both in Berlin and Munich. Berlin has additional district centers with an own city culture that is very attractive for carsharing. The more variety of business, recreational and residential use is present in the districts the more carsharing bookings are observed over the day.

5. *parking availability*: The longer the street the higher is usually the number of parking lots. Street length appears to be more important than the area size of a district.

6. *number of companies*: There is a linear correlation between the number of companies and the number of carsharing bookings. Especially administrative offices and services ergo places with a high number of walk-in customers are highly frequented spots for carsharing.

In the model which combines land-use data and election results the NB model fulfills most of the conditions. This is the reason why it is the only model that is regarded for interpretation. The significant variables can be assigned to the six defined clusters. With exception of the class *number of companies* all variable groups are represented. The company count has a more linear correlation with the number of bookings. The effect of living in a *family* with small children is slightly negative and only appears to be significant in combination with all other variables.
The models fit best for Berlin. The mentioned BMW-, DriveNow- and airport-effect in Munich make the search for an appropriate model difficult. Transferring results from the regression models to other cities is hence only useful for the Berlin model. The linear regression model based on the election results as exogenous variables is already quite satisfying for Hamburg, Cologne and Munich. But due to the strong positive impact of Green voters in Berlin, districts with a high rate of sympathizers of the Greens are also predicted as high demanded carsharing areas in other cities. This reveals to an overestimation in all cities. It can thus be concluded that centrality plays a greater role for the demand than the rate of Green voters. In Berlin, these two variables are highly associated.
It was therefore a good decision to consider both land-use data and election results in a model together. The NB model provides for all cities a better demand estimation than the linear regression model. More than 85 % of the quintiles are predicted correctly or with an error of ± 1 (Munich: 70 %, Cologne: 78 %). This is a good result and throws the model in a more positive light than it is assumed by McFadden's R^2.
The analysis of booking data in relation to parking restrictions shows a slight preference of FFCS users for short-term parking zones. It is assumed that this preference is not a conscious choice of the user. There just seems to be a better availability of parking lots in these zones. Moreover, short-term parking areas are close to attractive spots in the city and thus also attractive for carsharing. But in general, the kind of parking restriction has no strong noticeable impact. Parking pressure in parking management areas seem to be perceived generally as high. Someone who decides to still use a car for his destinations in this area

does not consciously distinguish between the different parking zones.
Weather has an influence on the choice of transport mode. A carsharing vehicle becomes more attractive for frequent customers (at least one booking per month) if the weather conditions are not good (rainfall, wind, low temperatures). But generally considering, there is no remarkable difference in the number of bookings at different times. This contradicts in a sort the observations of the operator. Some explanations are thinkable.

- The increase of bookings during bad weather is more a subjective impression of the fleet manager but actually not statistically significant.

- The weather effect is just one temporal influence that is covered by more decisive weather-independent factors.

- The weather takes effect with a delay. Rain during noontime on a general good day may tempt people to use carsharing in the evening hours because their subjective feeling during the day has changed without a change of the objective criteria though.

- The distinction between good and bad weather conditions is not appropriate. One has to consider other variables like clouding cover and atmospheric pressure to represent the subjective impression of weather in a better way. Creating a probability function for the choice of a carsharing as transport mode could be an option.

This chapter helped to define the typical user of flexible carsharing systems on the one side and to figure out additional impacts on the booking demand on the other side. The satisfying transferability of the model shows that the established external influences are also valid for other comparable cities.

Part III

MODELING: TIME SERIES ANALYSIS OF FFCS BOOKINGS

7 THEORY OF TIME SERIES ANALYSIS

The third part of this dissertation focuses more on the prediction than on the explanation of bookings. Instead of forecasting the location of hot spots a precise quantification of expected bookings is provided. The two chosen approaches should work with nothing more than the booking data as input. Such a model in statistics can e. g. be a time series.
One classic time series model is the *ARIMA model* (*ARIMA*), the other is *exponential smoothing with a Holt-Winters-Filter* (*HWF*). In this chapter, the research methods are explained in theory. The performance and implementation of the time series for the booking data is written in Chapter 8. The ideas and results of these analyses are already published by Müller and Bogenberger in [103].

As already explained in the beginning of part I, the analysis of the booking data by modeling them with time series methods intends to give a precise short-time prediction that can be helpful for the provider to estimate the demand in some areas of the cities and organize relocations if necessary. To create a widely usable forecast instrument that can easily be transferred to other cities or even other sharing systems, all external impacts are neglected. Nevertheless, results from previous chapters and studies ([123],[104]) are used as support for this work.
It is apparent that there are spatial differences in the booking demand of FFCS systems. A forecast for the whole operating area in Berlin is thus not appropriate. But instead of creating a complex spatial-temporal predicting model the city is divided into adequate districts. To meet the requirements of a transferable model, a grid that is available for most cities must be chosen. The net of zip code areas works perfect for that purpose. Berlins FFCS operating area consists of 119 zip code areas, the one in Munich contains 62 different zip code areas. A particular forecast is done for every one of these districts. A short-time forecast should be used to estimate the number of bookings in the next hours. Bookings (more precisely: booking starts) are hence aggregated per hour for every zip code area.

Time series prediction according to the Box-Jenkins-Approach (BJA) that is used in this work generally consists of three stages. It is originally described and explained in detail in the book "Time Series Analysis" ([17]) by G. E. P. Box, G. M. Jenkins and G. C. Reinsel that was published in 1976. The current approach that is based on [74], p. 4, is extended by one step in this work. In Stage *A* the time series is identified regarding trends and seasonal effects. Stage *B* uses historical data on which the time series model is based on. The third one is the creation of the forecast using the particular model. In the additional Stage *D* that is officially not part of the BJA the values of the forecast are validated with historic data from the same period. Fig. 32 shows the process of the stages.

One goal of this chapter is to give an answer to practical questions like an appropriate method to model the data or a sufficient duration of historic data set to gain a good forecast.
To compare the quality of forecasts with a different period, four historical data sets are used to build the model in stage *B*:

1. *a year*: September 1st 2012 00:00 until August 31st 2013 23:00
2. *a half year*: March 1st 2013 00:00 until August 31st 2013 23:00
3. *a quarter of a year*: March 6th 2013 16:00 until June 6th 2013 15:00
4. *a month*: May 6th 2013 16:00 until June 6th 2013 15:00

The ends of the last two datasets are chosen to be in the afternoon instead of at the end of the day. In this way, the difference in the quality of the forecast for high and low demanded time periods can be compared as well. The period of the forecast always follows directly behind the period of the historical data. An overview about the design of the data sets is drawn in Fig. 33.
Next to different data sets, the applied methods are different. The first is the ARIMA model, the second is exponential smoothing, more precisely, the Holt-Winters Filtering (HWF). ARIMA models a time series as a stochastic process whereas HWF is a typical numeric procedure to simplify a time series.

The next sections are structured by the stages of the BJA. In Stage B, the two methods are described separately.

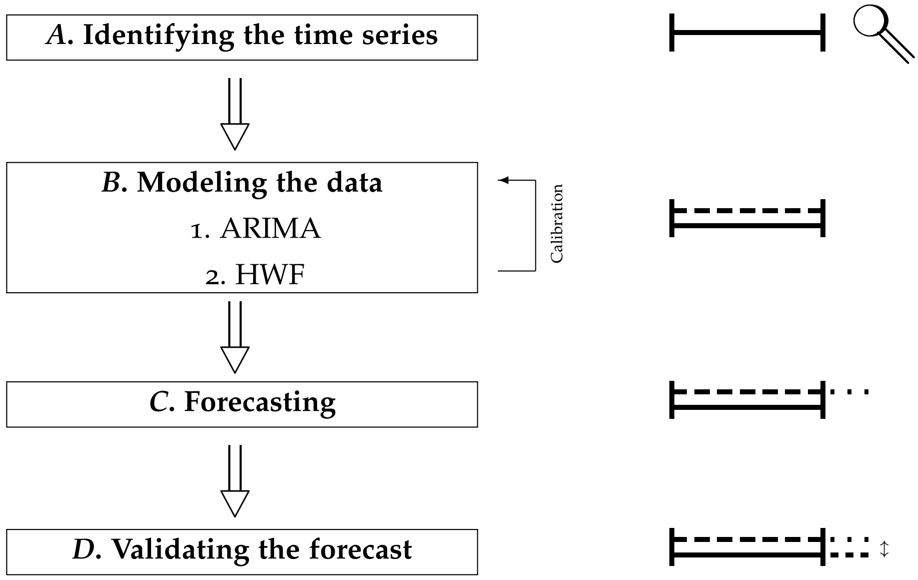


Figure 32: Scheme of the procedure of forecast

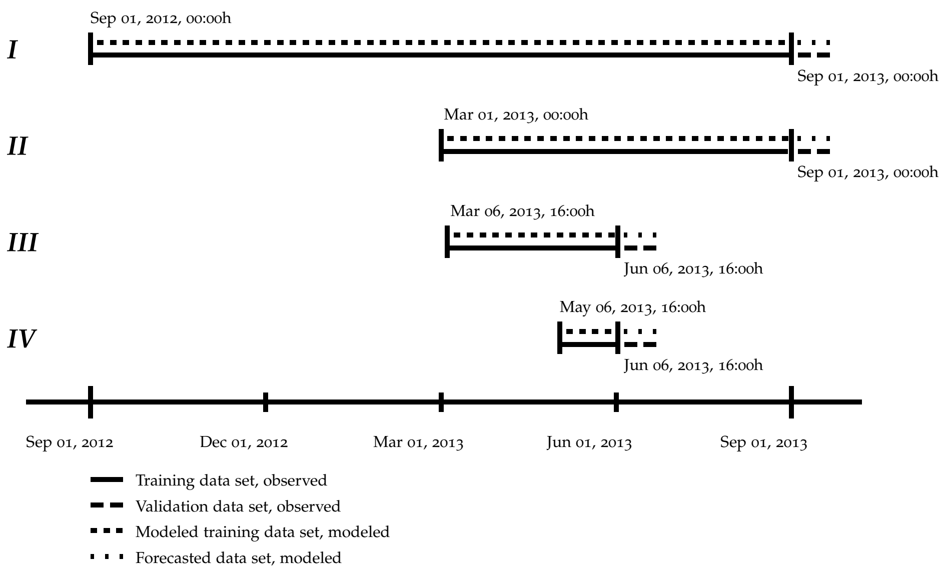


Figure 33: Overview of the different data sets used for time-series modeling

7.1 BOX-JENKINS-APPROACH: STAGE A

An essential part of modeling the data is to get an overview of the time series. For this reason Box and Jenkins proposed as first step of the analysis the identification of some key data about the time series such as the cycle and a possible non-stationarity. It is of fundamental significance to obtain an overview about the fluctuations and trends of the time series because this will be the basis for all further analyses. A choice of good identification methods are e. g.

- decomposition
- spectral analysis
- (partial) autocorrelation function
- stationary tests

In this work, the focus will lay on the (additive) decomposition, the spectral analysis and the autocorrelation function to identify trend and seasonal effects. Also other results from the general booking data analysis from chapter 4 will be used to get an impression for the time series.
This stage is essential for both methods of modeling.

The result of the decomposition analysis is a splitting of the observed time series $(x_t)_{t=1,...,T}$ in a trend, seasonal and random component noted as m_t, s_t and r_t, respectively.
`R` offers in the settings of the `decompose` function two options for the kind of splittings: the additive and the multiplicative method. As all values need to be postitive and not just non-negative as it may appear in the booking data the additive decomposition is chosen. Thus, the relation

$$x_t = m_t + s_t + r_t \;\; \forall t \in \mathcal{T} = \{1, \ldots, T\} \tag{7}$$

is valid. In the description of the function ([98]) it is explained that the trend component is determined by using the moving average (MA). This MA part is exactly the MA component of the ARIMA time series model explained in subsection 7.2.1. The equation of the moving average is noted in (18).
m_t or in other words the MA part of the time series is subtracted afterwards from x_t. In the next step, the seasonal component is calculated by averaging the trend time series for the predetermined cycle c. The graph of s_t is drawn in the output of the function for the whole time series but has because of the

calculation a recurrent appearance.
The random component r_t is determined by removing m_t and s_t from the observed data x_t.

The autocorrelation function (acf) describes the correlation at one time step of the function to previous ones. The formal definition is based on the autocovariance function v_t (see [122], p. 7)

$$v_\tau = \frac{1}{|T|} \sum_{t=1}^{T} (x_t - \bar{x})(x_{t+\tau} - \bar{x}) \tag{8}$$

with $\bar{x}$ defining the arithmetic mean of the time series x_t. The autocorrelation function a_τ is given by $a_\tau = \frac{v_\tau}{v_0}$.

$$a_\tau = \frac{\sum_{t=1}^{T} (x_t - \bar{x})(x_{t+\tau} - \bar{x})}{\sum_{t=1}^{T} (x_t - \bar{x})(x_t - \bar{x})} \tag{9}$$

The graph of a_τ is usually a bar chart that is called correlogram.

The spectral analysis is performed with the function `specest` that is part of the package `tsutil.r` programmed by R. Schlittgen. The theoretical background is also based on the definitions and explanations in this book ([121], p. 120-142).
The fundament of spectral analysis is the periodogram. A periodogram I of a time series x_t is defined as

$$I(\lambda) = T \left[\left(\frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x}) \cos(2\pi\lambda t) \right)^2 + \left(\frac{1}{T} \sum_{t=1}^{T} (x_t - \bar{x}) \sin(2\pi\lambda t) \right)^2 \right]$$

with $\lambda \in (0, 1)$. It is sufficient to consider just the field of $0 \leqslant \lambda \leqslant 0.5$. The reasons are the relationships $I(\lambda + 1) = I(\lambda)$ that is caused by the periodicity of trigonometric functions and $I(-\lambda) = I(\lambda)$ that easily follows when one looks at the square in the definition of I and remains that cos and sin are symmetric to the y-axis and $(0, 0)$, respectively.
Another way to formulate the periodogram is

$$I(\lambda) = a_0 + 2 \sum_{\tau=1}^{T-1} a_\tau \cos(2\pi\lambda\tau) \tag{10}$$

with the (empiric) acf a_0. When taking the theoretic acf γ_τ instead of the empiric acf a_τ one can prove that

$$E(I(\lambda)) \approx \gamma_0 + 2 \sum_{\tau=1}^{T-1} \gamma_\tau \cos(2\pi\lambda\tau) \to \gamma_0 + 2 \sum_{\tau=1}^{\infty} \gamma_\tau \cos(2\pi\lambda\tau) \tag{11}$$

because of $E(a_\tau) \approx \gamma_\tau$. The spectral density $f(\lambda)$ is thus defined as

$$f(\lambda) = \gamma_0 + 2 \sum_{\tau=1}^{\infty} \gamma_\tau \cos(2\pi\lambda\tau) \tag{12}$$

with respect to $\sum |\gamma_\tau| < \infty$.

The distinction between a_τ and γ_τ is just relevant in theory. The terminology "acf" will be used for the empiric as well as for the theoretic autocorrelation function. The spectrum is also a synonym for a periodogram.

One theorem is important to facilitate the interpretation of the spectrum. It is Parseval's theorem that states the equation

$$\gamma_\tau = \int_{-0.5}^{0.5} f(\lambda) \cos(2\pi\lambda\tau) d\lambda \tag{13}$$

and especially

$$\gamma_0 = \int_{-0.5}^{0.5} f(\lambda) d\lambda \tag{14}$$

The integral allows an important interpretation. "The rate of the area under $f(\lambda)$ between two frequency points and the entire area between $[0, 0.5)$ corresponds to the rate of variance which harks back to harmonics with frequencies out of this area" (cited from [121], p.129). The graph of the periodogram $I(\lambda)$ is thus showing the strength of correlation with a harmonic of the time series and the λ. Parseval's theorem shows additionally that the spectral density contains the same information as the autocovariance function in (9) with the difference that frequencies are shown instead of lags.

The function `specest` gives strictly speaking not the spectrum but the spectral estimation as output. Therefore, it is briefly mentioned what is actually calculated. An estimation of the spectral density can be

$$\hat{f}(\lambda) = \sum_{u=-q}^{q} w_u I(\lambda \cdot \frac{u}{T}) \tag{15}$$

whereby w_u are weights that can be chosen individually. The default setting is the Daniell-window

$$w_u = \frac{1}{2q+1}, \ (u = -q, \dots, 0, \dots, q)$$

The higher q is set, the more smoother the graph appears. Since $\sum w_u = 1$ the estimator is unbiased.

7.2 BOX-JENKINS-APPROACH: STAGE B

Until 1970 the time series model which was widely accepted was a deterministic model that contains a trend, a seasonal and an – usually normally distributed – random variable. Let $\hat{x}_t$, $t \in \mathcal{T}$ be the predicted value of the time series x at time t. Then the standard deterministic model can be noted as

$$\hat{x}_t = \mu_t + \omega_t + \varepsilon_t \tag{16}$$

with

μ_t trend effect

ω_t seasonal effect

ε_t random effect, mostly $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

The difficulty of this model is that it only fits perfectly for data with a fixed recurrent trend for the whole time series. But usually, the trend changes during the time. And normally, the seasonal effect neither occurs deterministically.
The two methods described below have two different approaches to skip the quite volatile trend and seasonal component. The exponential smoothing is based on the deterministic model but weights the data according to the currentness while the Box-Jenkins approach detaches from the deterministic idea and creates a model that is built with stochastic components.

7.2.1 *ARIMA*

The general idea is to see the time series not primarily as a sum or product of a trend, seasonal and random component as in (16) but to consider historical values themselves as influence for future values. This is on the one hand the data itself that is appearing in the AR (*a*uto*r*egressive) part of the time series and on the other hand the so-called MA (*m*oving *a*verage) part of the time series containing weighted historical random effects. Both the autoregressive and the moving average part are regarded as stochastic processes. That means they are a family of random variables.
For the following mathematical formulations, the notations in [17], p.8-12, served as basis. Let t be the currently observed time. An autoregressive process with parameter p is noted as $AR(p)$ and can be written as

$$AR(p) = X_t = \sum_{n=0}^{p} \alpha_n X_{t-n} + \varepsilon_t \tag{17}$$

with the random error $\varepsilon_t \sim \mathcal{N}(0, 1)$. That means that data from p time steps ago influence the current values.
The moving average process with parameter q is noted as $MA(q)$ and defined as

$$MA(q) = X_t = \sum_{n=0}^{q} \beta_n \varepsilon_{t-n} \tag{18}$$

This process is built by adding up the last q past random variables. The greater p and q are the more weight lies on historical data and the moving average. In consequence high values of p and q make the graph of the model smoother.
These two processes combined leads to the autoregressive moving average model that is abbreviated the $ARMA(p, q)$ model.

$$ARMA(p, q) = X_t = \sum_{n=0}^{p} \alpha_n X_{t-n} + \varepsilon_t + \sum_{n=0}^{q} \beta_n \varepsilon_{t-n} \tag{19}$$

The autoregressive process is the deterministic part of the model while the moving average is completely stochastic.
In most cases it is sufficient to model the data with an autoregression and a moving average process. But sometimes time series include a general non-stationarity. One solution is to consider instead of the original time series represented by the process X_t the differences $\nabla^d X_t = (1 - B)^d X_t$. B is the backward shift operator

$$BX_t = X_{t-1}, B^d X_t = X_{t-d} \tag{20}$$

so $\nabla X_t = X_t - X_{t-1}$. So one can speak of the real time series as the integrated version of the (stationary) time series. ARMA models including the parameter d additionally as variable that determines how often the time series is differentiated are called autoregressive integrated moving average models. Those are defined analogously to equation (19) with $X_t = \nabla^d X_t$ in the autoregressive term.

$$X_t = \sum_{n=0}^{p} \alpha_n \nabla^d X_{t-n} + \varepsilon_t + \sum_{n=0}^{q} \beta_n \varepsilon_{t-n} \tag{21}$$

A further extension is helpful when one observes a recurring trend. Since there is no seasonal component included in the model yet regularly deviations between real and modeled data can appear. To prevent this phenomenon the model can be expanded with AR and MA processes transferred to seasonal trends. For this it is essential to know the period the recurring fluctuations appear. Data from previous P periods will be considered as endogenous influencing factors. Analogously one includes the last Q periods to calculate the moving average or uses instead of the real times series x the D-th seasonal differentiation. Let s be the period of the seasonal component then the seasonal ARIMA model $ARIMA_s(p,d,q) \times (P,D,Q)$ is defined as

$$\begin{aligned} X_t = &\sum_{n=0}^{p} \alpha_n \nabla^d X_{t-n} + \sum_{n=0}^{q} \beta_n \varepsilon_{t-n} \\ &+ \sum_{n=0}^{P} \tilde{\alpha}_{s\cdot n} \nabla^D X_{t-s\cdot n} + \varepsilon_t + \sum_{n=0}^{Q} \tilde{\beta}_{s\cdot n} \varepsilon_{t-s\cdot n}. \end{aligned}$$

There exist also some other different notations of the ARMA processes that should be mentioned briefly (cited from [121], p. 79).

$$\begin{aligned} X_t &= \alpha_1 X_{t-1} + \ldots + \alpha_p X_{t-p} + \epsilon_t - \beta_1 \epsilon_{t-1} - \ldots - \beta_q \epsilon_{t-q} \\ X_t - \alpha_1 X_{t-1} - \ldots - \alpha_p X_{t-p} &= \epsilon_t - \beta_1 \epsilon_{t-1} - \ldots - \beta_q \epsilon_{t-q} \\ (1 - \alpha_1 B - \ldots - \alpha_p B^p) X_t &= (1 - \beta_1 B - \ldots - \beta_q B^q) \epsilon_t \\ \alpha(B) X_t &= \beta(B) \epsilon_t \end{aligned}$$

Extending this to the ARIMA model leads to

$$\alpha(B)(1-B)^d X_t = \beta(B)\epsilon_t. \tag{22}$$

The corresponding notation for the multiplicative seasonal is (cited from [122], p.138)

$$\alpha(B)(1-B)^d \phi(B^s)(1-B^s)^D X_t = \beta(B)\theta(B^s)\epsilon_t. \tag{23}$$

The whole procedure of finding the optimal parameter values for p, d, q, P, D, Q and the optimal determination of values $\alpha_1, \ldots, \alpha_p$, $\beta_1, \ldots, \beta_q$, $\phi_1, \ldots, \phi_P$, $\theta_1, \ldots, \theta_Q$ is performed by `auto.arima`. This `R` function is implemented by R. J. Hyndman and G. Athanasopoulos and follows the Hyndman-Khandakar algorithm they described in the book "Forecasting principles and practice" ([71], chapter 8.7). Here, the procedure will be briefly sketched.

1. Determing d, D: The Kwiatkowski-Philipps-Schmidt-Shin-test (KPSS-test) presented in [91] is applied for the d-differentiated and D-seasonal-differentiated time series. It tests the time series for stationarity, i. e. for a unit root. The parameters increase until the null hypothesis ("X_t is stationary") cannot be rejected.

2. p, q, P and Q are in following optimized. For parameter estimation, R uses maximum likelihood estimator (MLE) or minimizes the sum of squared errors (SSE). The two methods work with quite similar target functions and are explained e. g. in [36], section 4.5.3.0.1 and 4.5.3.0.2. With both methods the parameters α, β, ϕ and θ can be optimized. The notation of the exact log-likelihood for seasonal ARIMA models is omitted because just mentioning the long complicated formula does not bring any increment value.

Optimal p, q, P and Q are obtained by comparing the conditional Akaike's information criterion (AICc) of the different models with the optimized α, β, ϕ and θ . It is possible to choose every thinkable combination of the four parameters or to perform it stepwise. The latter option is chosen when problems with the computation time occur.
The AICc for ARIMA models is given by

$$\mathrm{AICc} = \mathrm{AIC} + \frac{2(p+q+k+1)(p+q+k+2)}{T-p-q-k-2}$$

whereby $k = 1$ if $d \neq 0$ and 0 otherwise. l is the likelihood function of each model and the AIC for ARIMA models defined by

$$\mathrm{AIC} = -2\log(l) + 2(p+q+k+1).$$

7.2.2 *HWF*

The basis of the exponential smoothing is the deterministic time series model from equation (16) without the seasonal component ω_t. Current data have more influence on the forecasted value than older ones.
For the following definitions the formulations and terms of [122], p. 44ff. served as the basis. Let $x = x_1, \ldots, x_t, t \in \mathbb{N}$ the measured data, $h > 0$ the time interval the time series is forecasted and $\hat{x}_{t,h}$ the forecasted value of x at the time of $t+h$. Then recursive definition of the (simple) exponential smoothing of x for $h = 1$ is

$$\hat{x}_{t,1} = \beta\hat{x}_{t-1,1} + (1-\beta)x_t, \text{ with } 0 < \beta < 1 \tag{24}$$

The initial value is the first observed data that has influence on the forecasted values. Mostly, it is the first value of the time series, i.e. $\hat{x}_{1,1} = x_1$. Generally, it is the mth value of the series, so $\hat{x}_{1,1} = x_m$.
β is called *smoothing parameter*. The smaller the value of β is, the more the latest

data are weighted. That means in reverse, that a value of β close to 1 makes the graph of the fitted value smoother. Finding an appropriate β is essential for a good fit of the data.
The name "exponential" comes from another way to write (24), namely as

$$\hat{x}_{t,1} = (1-\beta)\sum_{u=0}^{\infty} \beta^u x_{t-u}.$$

In 1957 and 1960 C. C. Holt and P. R. Winters published the articles "Forecasting seasonals and trends by exponentially weighted moving averages" ([70]) and "Forecasting Sales by Exponentially Weighted Moving Averages" ([143]), respectively. These two articles describe an extension of the exponential smoothing that also includes the seasonal component in the base model.
An assumption of the model is that the trend component in (16) is at least locally linear, such that it can be written as

$$\mu_t = a + bt.$$

The recursion in (24) is now applied to μ_t, b and the seasonal component ω_t. First, the seasonal effect is neglected. Then

$$\hat{\mu}_N = (1-\alpha)x_N + (\mu_{N-1} + \hat{b}_N) \text{ and} \tag{25}$$
$$\hat{b}_N = (1-\beta)\cdot(\hat{\mu}_{N-1} - \hat{\mu}_{N-2}) + \beta\hat{b}_{N-1} \text{ since } b_{t-1} = \mu_{t-1} - \mu_{t-2} \tag{26}$$

and for the fitted values in step h the following equation applies

$$\hat{x}_{N,h} = \hat{\mu}_N + h\cdot\hat{b}_N \tag{27}$$

Including the seasonal component ω_t with period s in this approach yields

$$\hat{\mu}_N = (1-\alpha)(x_N - \hat{\omega}_{N-s}) + \alpha(\hat{\mu}_{N-1} + \hat{b}_N) \tag{28}$$
$$\hat{\omega}_N = (1-\gamma)(x_N - \hat{\mu}_N) + \gamma\hat{\omega}_{N-s} \tag{29}$$

α is the *general smoothing parameter* while γ is the *seasonal smoothing parameter* and β the *trend smoothing parameter*. If β is set to 0, one would get the exponential smoothing with the seasonal component.
For the fitted values at the time h the following equations hold

$$\hat{x}_{N,h} = \hat{\mu}_N + h\hat{b}_N + \hat{\omega}_{N+h-s} \text{ with } h = 1,\dots,s \tag{30}$$
$$= \hat{\mu}_N + h\hat{b}_N + \hat{\omega}_{N+h-2s} \text{ with } h = s+1,\dots,2s \tag{31}$$

Hence there are at least s observations necessary to determine the start values. The `R` function `HoltWinters` takes the initializing values μ_0, b_0 and ω_0 from

the `decompose` function performed for the first period of the data set ([99]). That means for the present data sets that the start values come from the decomposition of the first week of the set.
The parameters α, β and γ are optimized by minimizing the squared one step prediction error. So `R` uses a standard method for optimization that is the method of Least Squared Errors (LSE). The equation uses $\hat{x}_{t,1}$ from (30) and has $m = 1$ as preset.

$$\min_{0<\alpha,\beta,\gamma<1} \sum_{t=m}^{N-1} (x_{t+1} - \hat{x}_{t,1}(\alpha, \beta, \gamma))^2$$

An important note is that `R` internally calculates with a "switch" of the parameters. Instead of α the program uses $1 - \alpha$. The same applies analogously for β and γ. In Table 21 the output values of the function are listed. To keep the notation consistently the according column headings are changed for the parameters. It is important to keep that in mind when one interprets the values.
There also exists a multiplicative Holt-Winters method. This is defined analogously to the formulas of the additive model but requires positive values for every time step. Exactly as in the case of multiplicative decomposition, this version of the method is not feasible due to time steps with no bookings.

7.3 BOX-JENKINS-APPROACH: STAGE C

7.3.1 *ARIMA*

The `R` function `forecast.Arima` makes use of `predict.Arima` which again uses the `KalmanFilter` function. The forecast procedure of the Kalman Filter is explained in detail in [108] by Morrison and Pike. The exact explanation of the algorithm would not be expedient. Anyhow, the process is summed up shortly.
The Kalman-Filter is generally a method for estimating the state vector of a linear dynamic system from noisy observations. The assumptions are

$$\begin{aligned} \tilde{X}_{t+1} &= A_t \tilde{X}_t + U_t + \eta_t \\ X_t &= H_t \tilde{X}_t + \epsilon_t \end{aligned}$$

whereby X_t is the observed matrix, $\tilde{X}_t$ the model matrix and U_t, A_t and H_t are matrices that have to be known $\forall t \in \mathcal{T}$. Also the noise covariance matrices η_t and ϵ_t need to be known for every time instant. The seasonal ARIMA model from (23) can be transferred in this notation.

The Kalman-Filter produces an estimation of the model matrix $\tilde{X}_t$ that is noted as $\hat{X}_t$ and gained by the observed data $X_t = (x_1, ..., x_t)$. The algorithm of Morrison and Pike works as a sequential estimation procedure. For every new x_t that is taken into consideration $\hat{X}_t$ has to be reestimated.

7.3.2 *HWF*

The prediction of the data set with HWF proceeds much easier than with an ARIMA model. The recursive formulation of μ_t, b_t and ω_t in (30) makes no further algorithms necessary. The values are just continued with the same definition for $t = T+1, ..., T+168$. Only for the demanded observed values x_t which do not exist for $t > T$ predicted values for x_t, i.e. $\hat{x}$ are taken.

7.4 box-jenkins-approach: stage d

This last part of the approach is not necessary for every forecast and hence not proposed in the BJA. If the identifying stage works without any problems and all advices are carried out the quality of the forecast is normally given. But since this work deals with several data sets a comparison of the different forecasts is required. The goodness of the prediction is measured by the mean absolute error between the real and fitted values of the forecasted time period. The mean over every zip code area and hour will be compared.

8 APPLICATION OF TIME SERIES MODELS

8.1 Box-Jenkins-Approach: Stage A

The first and probably most important part of the identifying stage is the determination of the cycle of the time series. All further analyses depend on this value. The decomposition, autocorrelation function and the spectrum can be an indicator if the choice of the cycle was good but they already require a fixed cycle. This is the reason why the analyses from chapter 4 of this thesis are essential. Clearly visible is the strong daily recurrent trend. But also the differences between weekends and weekdays are so strong that they could not be neglected. The decision is hence to consider one week, i. e. 168h, as an appropriate cycle for the time series. For sure, there is also a difference between summer and winter times but the data set needs to have at least two periods of the cycle. Due to the fact that the systems in Berlin and Munich just started in 2012 it makes no sense to take a cycle of one year into account.
The following evaluation of the decomposition, autocorrelation function and spectral density are only performed for the data set I since all other data set are subsets of these data.

Although the spectral analysis is the most complicated method for identification of time series and based on the acf in theory, the description starts with a closer view to the spectra. The reason is that this method can be used for clustering as it is described in the following. The data are already standardized due to a trend that was observed in the decomposition analysis which is described below.
An example of a spectrum is shown in Fig. 34. When evaluating the graphs of each spectrum estimation it is remarkable that they vary in their general profile as well as in the height of the amplitude. A higher amplitude means a higher intensity of bookings at that frequency. The height of the amplitude can be changed manually to obtain a graphically appealing result. Hence, an interpretation of the spectrum just makes sense in comparison with other spectra. There are four frequencies that are considered in detail. They are all divi-

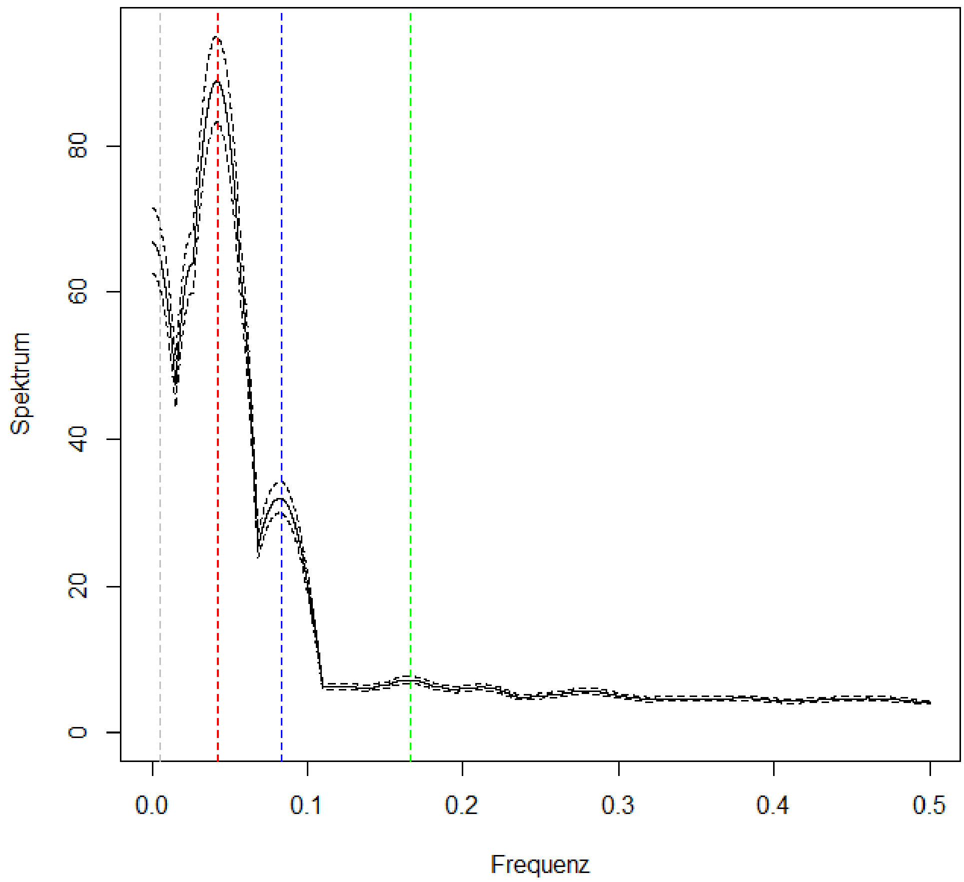


Figure 34: Spectrogram exemplarily shown by postal code area 10115 (Berlin).

sors of the 168 h-cycle: the frequencies of a week ($\lambda = 1/168 \approx 0.005$), a day ($\lambda = 1/24 \approx 0.042$), 12 h ($\lambda = 1/12 \approx 0.083$) and 6 h ($\lambda = 1/6 \approx 0.166$). The graph of the spectral estimation can be regarded as an overlay of cycles and different harmonics. The spectrum is therefore an indicator of the intensity of the different frequencies. Examples of spectrograms are shown in Fig. 47 and 48 in appendix B.4.

When focusing on the amplitude of the spectrum estimation at these four frequencies and comparing them with the number of bookings a strong correlation is obvious. That might be not surprising but it excludes situations where zip code areas are highly frequented for a certain time only. Such a hot spot might be a stadium that is an attractive spot for visitors only on weekends.

Zone A	highly frequented areas	week frq. > 50	daily frq. > 50
Zone B	normally frequented areas	week frq. > 10	daily frq. > 10
Zone C	lowly frequented areas	week frq. > 1	daily frq. > 1
Zone D	peripheral areas	week frq. > 0	daily frq. > 0

Table 19: Clustering for Berlin on the basis of the spectrum estimation

Zone A	highly frequented areas	week frq. > 10	daily frq. > 10
Zone B	normally frequented areas	week frq. > 5	daily frq. > 5
Zone C	lowly frequented areas	week frq. > 1	daily frq. > 1
Zone D	peripheral areas	week frq. > 0	daily frq. > 0

Table 20: Clustering for Munich on the basis of the spectrum estimation

To see this relationship between the spectra and the booking frequencies the following clusters for Berlin (Table 19) and Munich (Table 20) are proposed.

The relationship between the clusters and the number of bookings is easy to see when one looks at the maps colored by clusters on the one side and by the quartiles of the booking frequencies on the other side. The direct comparison is shown in Fig. 35. A difference plot is not done, since the accordance is obvious. The proposed clusters are hence the evidence for the fact that highly frequented areas are highly demanded at all time, lowly frequented areas have less bookings at all time of the day. This result complies with previous analyses from Schmöller et al. in [123] and also with the spatio-temporal analyses in section 4.3. Hot spot analyses for different time zones have already shown that the intensity of the hot spots in the districts may vary but they do not change significantly. This corresponds to the results of the second part of this thesis. Carsharing works optimally in areas with a mixed use of residential and business areas because it is then used during all times of the day.
After this clustering the focus first rests on the description of the spectrum estimation performed with `specest`. As explained in the theory of the function only a comparison of the postal code areas makes sense. The values of the y-scale cannot be interpreted. The focus is thus on the amplitude of the particular frequencies. The spectra of both cities have a quite similar characterization that differs mostly in Zone C and D.

- Zone A has its highest peak at the 24 h frequency. And also the weekly harmonic is strongly pronounced.
- Zone B's spectrum also has its biggest harmonic at the 24 h-frequency, a weaker weekly frequency and almost no special manifestations at 12 h and 6 h.
- A typical Zone C in Berlin has generally a low spectrum level as it can be seen at the y-scale. Here the week frequency has the biggest impact followed by the 24 h and the 12 h harmonics. In Munich, however, the 24 h-frequency has the most influence.
- The spectral density of Zone D has no typical pattern and is generally low scaled in Berlin. This is also valid for Munich but there are some areas with a peak at the 24 h harmonics.

The interpretation of the frequencies is quite easy. A high spectrum at the weekly frequency means that the number of bookings in the according zip area is similar for every week. In other words, the same days of the week are strongly correlated. Analogous interpretation can be done for the other frequencies.
In the following, four postal code areas are the representatives for their zone. The four zip code areas for Berlin are 10115 (Mitte) for Zone A, 10178 (Mitte) for Zone B, 10317 (Rummelsburg) for Zone C and 12165 (Steglitz) for Zone D. Munich's representative postal code areas are 80333 (Altstadt-Lehel), 80802 (Maxvorstadt), 80992 (Moosach) and 81549 (Obergiesing) for Zone A to D, respectively. The map of clusters in Fig. 35 shows the position of these postal code areas. All analyses, models and forecasts are done for every zip code area. Nevertheless, the graphical results in appendix B.4 are just shown for these districts and all descriptions of the four zones in each cities are oriented towards these postal code areas. As conclusion of the spectral density, one get as a result that the decision to take the 168 h as cycle for the time series was an acceptable decision although the daily recurrent harmonics are very strong. It should be take into account that an extension of the model with an additional 24 h cycle may improve the results. But especially for areas of Zone C which make up the biggest part of both cities the 168 h frequency has the biggest impact on the time series.

This result can also be concluded by a look at the autocorrelation functions. A perfect model would have an acf where all correlations are statistically zero.

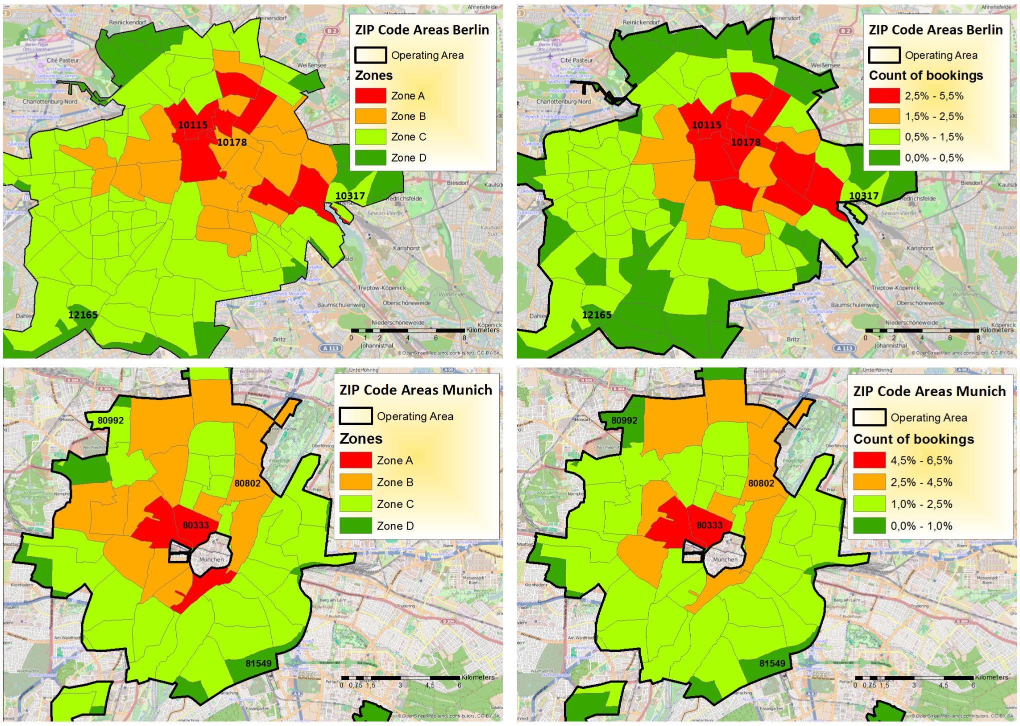


Figure 35: Postal code areas colored by the number of bookings and by the defined clusters

That means when looking at the graph in Fig. 36 (or Fig. 45 and 46 in the appendix) that the bars are all within the confidence interval determined by the Box-Ljung test that is marked by the dashed lines. That is obviously not the case.

The height of the bars in the diagrams represent the correlation of the lags and the current data. So the first bar indicates the average correlation between the current data and the 0 th lag. This is 1 in every case. The y value of the second bar at $x = \frac{1}{168} \approx 0.006$ shows the average correlation between current data and those of one time step (1 hour) ago and so on. Interestingly there is in almost every of the Zones a notable correlation at the 24 h lag. In very busy areas there is additionally a negative correlation at the 12 h lag visible.

The interpretation of these phenomena is simple. As seen in the first chapters of the dissertation there are two peaks in the daily profile of the bookings. The morning peak is however not as high as the evening one that is almost exactly 12 h later. So for most of the time steps the correlation between the step 12 h ago was negative. The other peak that appears in nearly every Zone is the 24 h frequency. This positive correlation is expected since the number of bookings

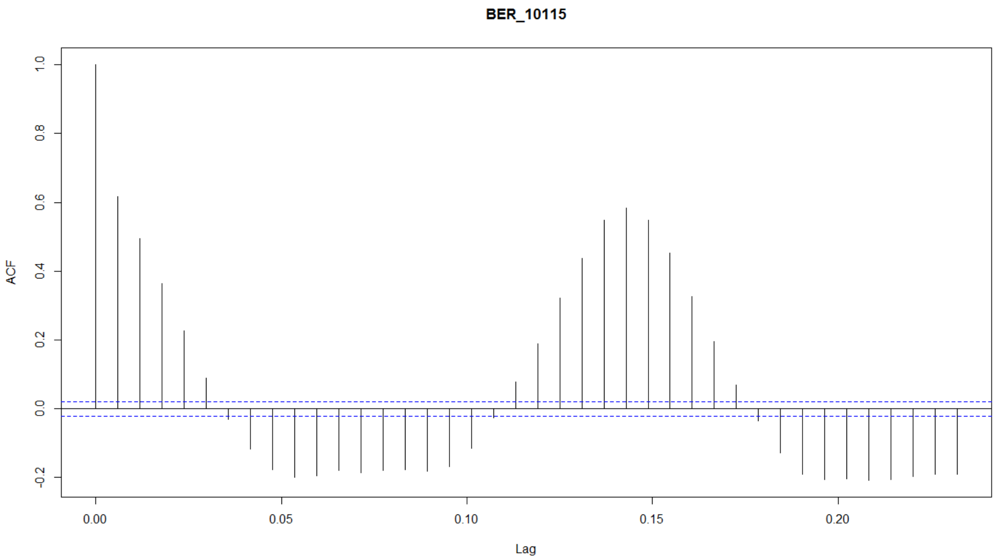


Figure 36: The autocorrelation function exemplarily shown by postal code area 10115 (Berlin).

on a certain time of the day is strongly dependent on the booking frequency at the same time one day ago. The 6 h frequency does not play a role at all: The autocorrelation at that time is in every case close to zero. The fact that areas of Zone D are nearly not autocorrelated has to do with the very low number of bookings in these districts. A booking is - now also statistically proved - coincidental.

The results of the autocorrelation analysis correspond to those of the spectral analysis. The booking frequency does not only depend on the number of bookings of the previous week but also of the frequency one day before. Nonetheless, it is only the 168th cycle that will be taken into consideration. The other fact is that the negative 12 h autocorrelation appears usually in areas of Zone A or B. As explained they are a sign for a difference in the number of booking in one part of the day. From the results of the booking data analysis it is obvious that there is a higher booking frequency in the afternoon or evening time. Combined with the results of the acf analysis it is now visible that the increase of bookings compared to other areas is mainly done in these time of the day.

As last part of the Stage A of the BJA the decomposition is pending. This was originally the first performed analysis but due to the cluster proceeding the decomposition is explained only now. The spectral estimation as well as the acf are methods to analyze the seasonal recurrent trends in a time series. The

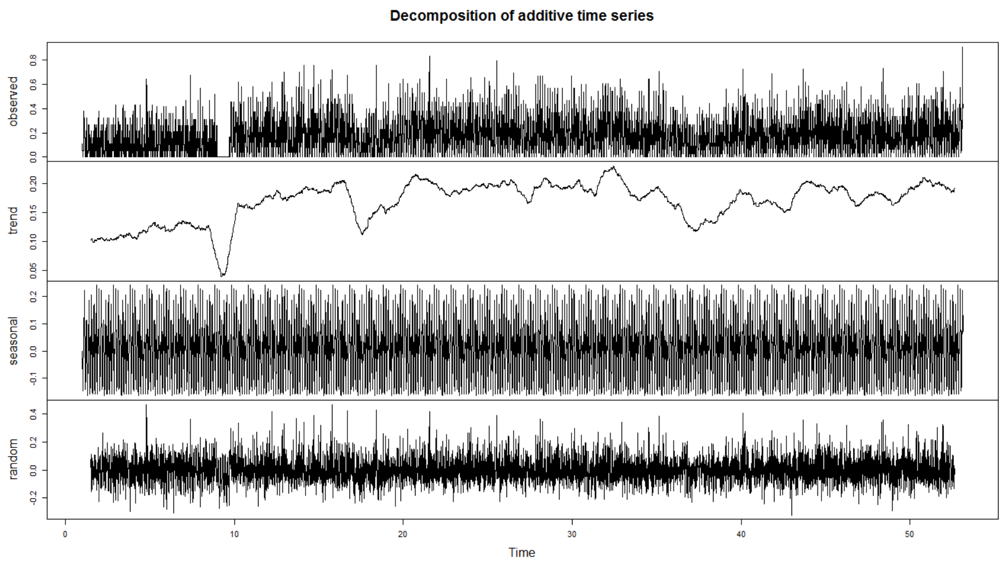


Figure 37: Decomposition exemplarily shown by postal code area 10115 (Berlin).

decomposition is in contrary a good instrument to detect general trends in the time series.
The results of the decomposition can be found in the appended Fig. 43 for Berlin and Fig. 44 for Munich or exemplarily in Fig. 37. Each figure consists of four graphs. The first shows the profile of the observed bookings, the second the calculated trend, the third the seasonal trend and the last the random error. R offers two methods of decomposition: an additive and a multiplicative solution. Since all values of the time series have to be positive for the multiplicative method, the additive decomposition is chosen. The focus is on the second graph to prove the stationarity of the time series. The graphs do not show the original data but an already standardized booking profile. There was an increase in the number of bookings visible in the original data. Since the data are from the period of September 2012 to August 2013 – so just around one year after launching the carsharing system in the cities – it was obvious that an increasing number of vehicles and a growing number of active users is responsible for this trend. To normalize the data only the relevant customers are considered. The relevant customers are the active ones, i. e. the registered users who did at least one booking per month. The number of cars measured by used vehicles with a different license number actually raised from around 550 in September 2012 to around 850 in August 2013. The number of active customers is confidential and thus cannot be published at this place.

For normalization the booking numbers are multiplied with the rate of monthly users and current vehicle fleet size.

$$\text{bookings}_{\text{standardized}} = \text{bookings} \times \frac{\text{available cars}}{\text{active users}}$$

When looking to the number of bookings in Berlin, one conspicuity is around week 10 (November 2012). In each area there is a break in the booking history that was caused by a technical disruption of the system for the period of 2 days. Regarding the trend one observes almost a stationary tendency which is a good sign for the standardization method.
The trends of the decomposition in Munich show a similar result. The trend function in Zone A and C even makes the Christmas holidays visible (at week 18). In some postal code areas there is even a slight negative trend from week 25 (March 2013) noticeable. But due to the negligible decrease, the standardization is taken for all further analysis. The spectrum as well as the acf analyses are already performed with this normalization of the booking data.

8.2 BOX-JENKINS-APPROACH: STAGE B

This section explains the search for the optimal parameters of the seasonal ARIMA model and the HWF model. The calibration is performed with the `R` functions `auto.arima` and `HoltWinters` which are explained in theory in section 7.2. The computation times for the algorithms refers to a ThinkPad Intel(R) Core(TM) i5-3320M CPU 2.60GHz with 8 GB RAM.

8.2.1 *ARIMA*

`auto.arima` calculates the optimal values of the parameters of the seasonal ARIMA model by comparing the AIC of each model. The parameters of the model with the lowest AIC are stored to use them for the forecasting stage.
As one can imagine it may take a long time to calculate the AIC of all thinkable models of just one postal code area if the parameters are not limited. Usually, the AIC converges to a particular value or increases strictly from a particular value of the parameters. The optimal model is hence that model with the lowest parameters and where the AIC does not increase significantly any longer.
The problem during the performance was the fact that there are several models that do not show this convergence of the AIC. To prevent this, one can change the default settings of the `auto.arima` function. One possibility is to limit the

parameters p, d, q, P, D and Q. Another option is to limit the number of iterations of the algorithm. Especially the parameters d and D which are responsible for the number of differentiations are usually 0 or 1. They also need the highest computation time for modeling.

Problems with the convergence of the models and the computation time appear especially when calculating the models for data set I of Berlin. All limitation of the parameters were no proper solution for the run time and convergence issues. The long cycle of 168 h and the long data period seemed to be the main problem. Hence, the data set I was read in again, but with a cycle of 24 h. The computation time still lasts at around 48h but at least almost every model could be calculated.

The cycle for all other data sets were set to 168 h. For models of postal code areas that did not converge the parameters d and D were set to 0; additionally p and q were limited to 4 whereas the limit of P and Q were set to 1. By this, the time series may not be modeled by a global optimal model but with an acceptable local optimum. This kind of avoiding computational errors was only performed for Berlin. In Munich, non converging models were not analyzed any longer.

The list of parameters for the representatives of the Zones is shown in Fig. 21. It is conspicuous that in the models for the short data sets III and IV the optimal parameters of the seasonal ARIMA part are all 0. That reveals that this period of training data set is too short for an exact modeling or the limitation of the parameters is too strict. But a relaxation would easily lead to run time or convergence error. The optimal value for D is in every area of the representatives 0 which means that there is no remarkable seasonal trend in all data sets.

The computation time became less the shorter the data period of the set was. But it took still around 12 h for the shortest data set IV. The data sets with a short period of historic data tend not to converge especially for lowly frequented areas and even when the parameters were limited in the preset of the function. The reason for that is the too low number of bookings that may let the booking profile look arbitrary and thus difficult to model.

There are a lot of other settings that were used to shorten the computation time of the algorithm. It is e. g. possible to choose the value for `approximation`. If the AIC of the next calculated model is less than this value, the algorithm stops. Another option to accelerate the algorithm is to allow a stepwise search for an optimal model. Next to these settings it is possible to consider stationary models only. That means in consequence the parameters d and D are set to 0. That definitely makes sense since the decomposition already showed no remarkable trend in most time series and a first check of the optimal parameters often leads exactly to this value for the optimal d and D.

8.2.2 *HWF*

The key parameters in `HoltWinters` are the parameters α, β and γ. They correspond to μ, b and ω, respectively, from the equations (28), (26) and (29). The values for them can be preset, set to `FALSE` or undetermined. With the third option – that is chosen for the present data – the values for the parameters are optimized automatically by minimizing the squared one-step prediction error. The start values for μ, b and ω are taken from the `decompose` function (see [99]). The results have, similar to the values of the parameters of the `auto.arima` function, no special pattern. It is just conspicuous that the $1-\beta$ values are mostly 0 or close to 0 and the $1-\alpha$ values are all almost 0 as it can be seen in Fig. 21. The high values of β have a strong smoothing for the fitted and predicted values as consequence. That also means that sudden changes and breaks in the booking frequencies only happen occasionally. Since standardized booking numbers are regarded the results can be interpreted as a general stationarity in the capacity of the vehicles.

The computation time for finding the optimal parameters for HWF is negligible. The longest run time took about 20 seconds for data set I. There was no correction or limitation of parameters necessary for any time series. The much shorter computation time is caused by the lower number of parameters and the smaller range of the values. α, β and γ are all out of the $[0, 1]$ interval.

8.3 BOX-JENKINS-APPROACH: STAGE C

Two main options exist to produce a forecast with `R`. The first one is the function `predict` that one can use for many purposes and not only for time series. `forecast` however is a comprehensive `R` package with more options. One reason to use this package is the confidence intervals which are one part of the output. They are set to 0.95 and 0.999 and appear in figures in light gray buffers around the dotted line which marks the forecasted data. The disadvantage of the function is that the forecast is calculated at the utmost for two periods ahead.
In this section, there will be a short description of the graphical output of the forecast function `Arima.forecast` and `HoltWinter.forecast`. The values of the

			p	d	q	P	D	Q	$1-\alpha$	$1-\beta$	$1-\gamma$
Berlin	I	A	1	0	3	2	0	2	0.044	0	0.145
		B	3	0	1	2	0	2	0.028	0	0.142
		C	3	0	1	2	0	2	0.009	0.001	0.137
		D	1	0	1	1	0	1	0.004	0.001	0.092
	II	A	3	1	3	1	0	1	0.041	0.001	0.213
		B*	1	1	4	1	0	1	0.014	0.001	0.194
		C*	3	1	2	1	0	1	0.010	0.001	0.169
		D	4	0	3	1	0	0	0.005	0.002	0.187
	III	A*	0	1	1	0	0	0	0.039	0	0.301
		B	0	0	1	0	0	0	0.016	0	0.302
		C*	4	0	4	0	0	0	0.011	0.002	0.264
		D	3	1	3	0	0	0	0.003	0	0.280
	IV	A*	0	1	2	0	0	0	0.132	0	0.415
		B*	0	1	1	0	0	0	0.036	0	0.372
		C*	3	0	3	0	0	0	0.006	0	0.509
		D*	2	0	3	0	0	0	0.007	0.010	0.479
Munich	I	A	2	1	2	1	0	1	0.023	0	0.176
		B	1	1	4	1	0	1	0.008	0.001	0.163
		C	1	1	2	1	0	1	0.004	0.001	0.148
		D	3	0	2	1	0	0	0.004	0	0.142
	II	A	5	1	5	1	0	0	0.032	0.001	0.199
		B	2	1	3	1	0	1	0.005	0.002	0.236
		C	5	0	2	1	0	0	0.004	0.001	0.166
		D	1	1	3	1	0	1	0.007	0	0.199
	III	A	4	1	3	0	0	1	0.038	0	0.265
		B	2	1	3	1	0	1	0.005	0.003	0.270
		C	4	0	2	1	0	0	0.004	0.003	0.260
		D	2	1	4	1	0	0	0.003	0.002	0.278
	IV	A	2	0	1	1	0	1	0.059	0	0.500
		B	4	0	3	1	0	1	0	0	0.365
		C	4	0	4	1	0	1	0	0	0.375
		D	4	0	0	0	0	1	0	0	0.479

Table 21: Optimal parameters of the ARIMA and HWF model for the representatives of the Zones for each data set. Parameters of areas marked with $*$ are optimal for limited parameter settings.

x-axis tell the week of the data set plus 1. For the forecast of e. g. data set I the observed training data set is 52.14 weeks long. From the point $x = 53.14$ on the figure shows the forecasted values and the observed data of the independent data set. In every figure the forecasted values are drawn with a dashed line, the observed data with a continuous line. The confidence intervals always start at the time of the forecast. The forecast is generally shown for one week ahead.

The first focus is on the validation of the forecasts of the ARIMA models. Because of the difficult performance especially for the yearly data set of Berlin it is necessary to detect potential errors of the forecasts.
The forecast for the ARIMA model of data set I for Berlin has the particularity that it is just 2 days long (Fig. 49). That has to do with the necessary change of the cycle of the time series as explained in the section before. Since the `forecast` function produces just forecasts for up to two cycles, the predicted values are consequently calculated for the next two days. The results seem quite good. The same is valid for the city of Munich where the cycle was again set to a week (Fig. 50). Only the forecast for the representative area of Zone D shows an unsatisfactory prediction. The reason is probably a general low booking frequency in these areas.
A similar output can be seen for the Zone D of Berlin (Fig. 53) and Zone C for Munich (Fig. 54) for data set II. The cause is supposed to be the same. The figure of Berlin shows also some examples of forecasts where the model's parameter were limited. As one can see this has not necessarily a bad prediction as consequence. The labels of the x-axis of these changed models can be ignored.
The forecast starts struggling when the duration of the training data set becomes too short. Berlin's representative postal code area for Zone C for instance does not include a forecast for data set III even if limited parameters were applied for the model (Fig. 57). This phenomenon appears also in these models because the locally optimal parameter can be so small so that they actually cannot be regarded as the best choice for modeling. A forecast with that little parameters cannot provide a graphical output. But also models with usual parameter may fail because of wrong start values or a general bad fit. Munich also has got some forecasts of that type at data set III but the representatives still work acceptable except for the areas of Zone D (Fig. 58).
The duration of one month training data set seems much too short when looking to the results in Berlin (Fig. 61). With exception of the highly frequented

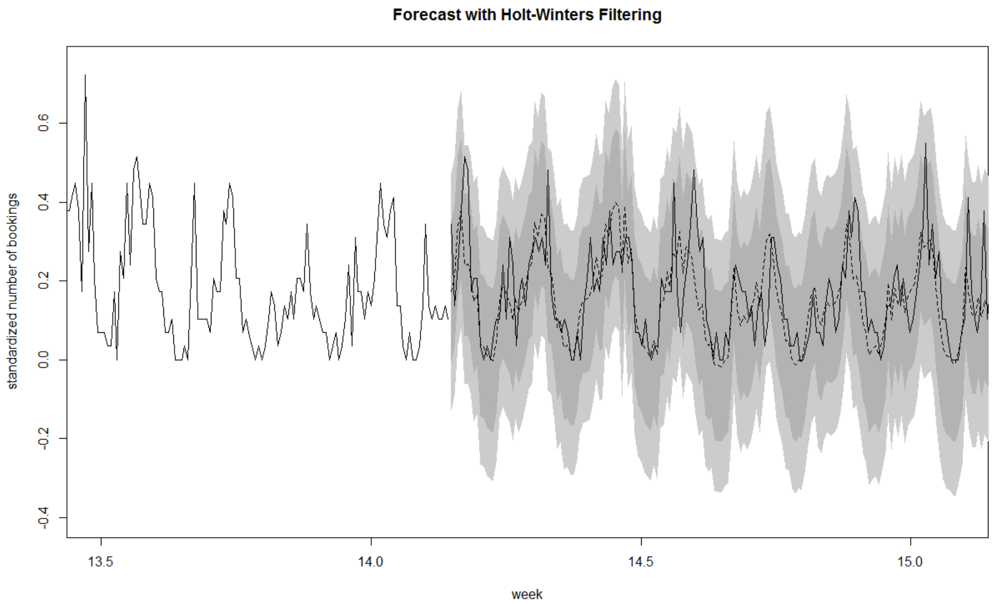


Figure 38: Forecast output exemplarily shown by postal code area 10115 (Berlin) with a three month training data set based on the HWF model.

areas only some postal code areas provide a satisfactory forecast. The representatives of Zone C and D totally failed. An even worse situation is presented in Munich (Fig. 62). In all representative zip code areas, the optimal parameters does not lead to a good fit and the forecast could not be performed.
One positive point of the ARIMA modeling has to be mentioned. In case of a good forecast the quality usually keeps on the same level for the whole week and not just for the first time of the prediction period.

A much smoother performance was the prediction of the HWF model. Results for a period of a year, a half year, a quarter of a year and a month for Berlin and Munich can be found in Fig. 51 and 52, Fig. 55 and 56, Fig. 59 and 60 and Fig. 63 and 64, respectively. Due to the problem-free calculation of the optimal parameters for the model it is not very surprising that the forecasts provide very good results that can already be observed by graphical comparison between the predicted and observed data.
Even if just one month of historic data is taken to model the time series the extrapolation works fine. The problematic areas of Zone D also show acceptable results. There does not seem to be a big difference between predicted values some hours or some days ahead. The precision keeps constantly.

		Berlin				Munich			
		I	II	III	IV	I	II	III	IV
ARIMA		**0.96**	0.99	0.97	1.03	**0.77**	0.82	0.87	0.88
HWF		0.93	0.93	**0.84**	0.90	**0.76**	0.77	0.78	0.83
ARIMA	A	2.14	**2.05**	2.06	2.37	1.53	**1.40**	1.73	1.77
	B	1.41	**1.48**	1.49	1.72	**1.23**	1.51	1.35	1.35
	C	0.86	0.93	0.83	**0.80**	**0.82**	**0.82**	0.91	0.92
	D	0.48	**0.39**	0.48	0.41	0.27	**0.24**	0.30	0.30
HWF	A	2.25	2.29	**1.94**	2.22	**1.41**	1.42	1.59	1.68
	B	1.45	1.45	**1.31**	1.42	1.16	1.17	**1.09**	1.17
	C	0.86	0.86	**0.77**	0.84	**0.77**	0.78	0.84	0.89
	D	0.40	0.40	**0.27**	0.27	**0.28**	**0.28**	**0.28**	0.30

Table 22: Absolute errors in veh/h in average for the particular Zone. The best results in the row for every city are written in bold.

8.4 BOX-JENKINS-APPROACH: STAGE D

In the last stage the observations of stage C are quantified. The failed ARIMA models are not considered in this evaluation.
There are a lot of popular methods to measure the difference between a forecasted and an observed data point. The most usual way is to calculate the squared error. This approach penalizes high deviations more than small ones. This is not considered necessary so that the sum of absolute errors was chosen as an appropriate measurement for the differences.
The absolute errors are not only summed over the week but also aggregated by all postal code areas of each Zone. The sum is subsequently divided by the number of areas per zone and by 168 to get an average error per area and hour. The results are afterwards transformed back from standardized to absolute values. The numbers listed in Table 22 can thus be interpreted as the average error of the particular Zone in vehicles per hour. Berlin's forecast of data set I was averaged not by the seven but by the two days of its prediction.

It is easy to explain the higher errors in the areas of Zone A. A high number of bookings implicates directly a higher deviation of errors. The focus of the error evaluation is hence on a comparison of the different data set, i.e. a comparison column by column, and a comparison of the two methods of modeling. The numbers that are written bold mark the best result of the city for that data set. The HWF in Berlin provides the best forecast with a training data set of three months. The ARIMA modeling works in the best way with a data set of a half year. But the differences differ not that much. Regarded solely from the quantification of the errors, HWF wins the battle with ARIMA. Although there are some Zones that reach better forecasts with ARIMA the HWF has lower errors in total. A comparison of the first two rows shows a clear preference for HWF in every data set. The average error of 0.84 vehicles per hour in Berlin and 0.78 in Munich is satisfying. Although it is hard to compare these results with other studies, the in Table 3 mentioned studies for a prediction approach are recalled. Kaltenbrunner et al. presented their results in [77] and came to an average prediction error of 1.39 bicycles per hour and station. Regue and Recker obtained in [119] an average error of 1.68 bikes per station with a maximum capacity of 15 bicycles. The results of this dissertation are thus satisfying.

Before coming to the conclusion of the chapter it is mentioned briefly that in almost every case more than 95 % of the forecasted data are in the 95 % confidence interval. A detailed analysis of the confidence intervals has been waived since the graphical outputs show the sometimes large size of the intervals.

8.5 CONCLUSION

The motivation for this chapter was to find an appropriate method for providing exact forecasts and to find an adequate duration of the training data set which is used for modeling. The answers for both questions are found.

Without doubt, there exist much more methods than the described two forecast instruments. But the two analyzed ways of modeling are also a comparison of stochastic modeling and numeric modeling. A clear result of the evaluation is the failure of the ARIMA. This has to do both with the too short duration of the training data set and the length of the cycle of the time series. Especially for the three and one month data set the problem of the shortness of the data set occurs so that parameter optimization failed at most time. When modeling very long and a lot of data sets like that data set I for Berlin, a run time error sometimes occurs. It is thus not surprising that the best results for ARIMA are from data set I and II.

Comparing the computation time it is also not very advisable to use a stochastic method for modeling. The Holt-Winters-Filtering provides similar or even better solutions in less than one minute. In practice, this or an equivalent method is consequently the preferred way to forecast the data. Three months of historic data is necessary to gain an optimal forecasting result for even every postal code area. The reason that the first data set does not provide a better forecast is that yearly seasonal variations are more considered than in shorter data sets. The quality of the forecast does not change over the week as the figures show. That also means that the provider does not have to refresh the input very often. A proposed time for a routine update of the data is a week or more. The graphical comparison of the forecasts from data set I/II and III/IV also shows that it does not play a role whether the time of the forecasts ends in a highly demanded time or not.

Nevertheless, there are some issues that have not been solved yet. The table of absolute errors may be a first hint for the operator which is the best duration for a historic training data set. But the error just measures the average over the week. The carsharing operator may be interested in a special time slot of the forecast (e. g. monday mornings) for which he plans his relocation. And it is also possible that he just wants to have a forecast for the highly demanded areas. These special analyses are an option for future research.

Another point is the range of the forecast. The present methods map the values to $\mathbb{R}$, necessary would be $\mathbb{N} \cup \{0\}$. The easiest solution would be the Gaussian floor function. It could also be used with a shift of 0.5 such that all predicted values in $\mathbb{R}^+$ are rounded off or up in a regular way.

A byproduct of the spectral analysis was the clustering that can be regarded as an activity clustering. It means that highly frequented areas show a high number of bookings all the time, and lowly demanded areas are unattractive for the user over the whole day. A similar result was found by Froehlich et al. in [61] for the already mentioned bikesharing system in Barcelona. They created a Bayesian network for the prediction and yielded a quite well-working model with the biggest error during peak hours. They also distinguished the different city areas in activity clusters, but not on the basis of the spectrum.

It is not clear if these activity clusters are caused by the (theoretical) demand of the user or simply by the nature of the system. The demand might be higher in other regions but since trips end in particular areas at a particular time of the day the vehicles cannot be located to other spots in the city for another time.

9 CONCLUSIONS AND FUTURE RESEARCH

In this final chapter all results of this dissertation are reflected and scrutinized regarding their validity. It is moreover detected which future research has to be done to get satisfying answers to still open questions. Eventually, the author makes some proposals concerning the improvement of the environmental effects of FFCS systems.

The first part of the dissertation intended to come to a deeper understanding of the FFCS system. The new mobility service is mainly used after the evening rush hour and at the end of the week. Preferred destinations are central districts of the city so that a map of bookings resembles a map that marks public life of a town.

The regression analyses is retrospectively seen as an appropriate approach to explain the spatial distribution of carsharing demand. The defined clusters consolidate the results of the regression models with land-use and election results. The latter data set describes the local milieu and is therefore in most times already represented by the socio-economic variables. The clusters *open-mindedness*, *type of car user* and *financial situation* describe attributes of FFCS users: They are above average open to new technologies, financially well-off and not the classic car owners but prefer business or leasing cars. It is pleasant to see that the results correspond to findings of other researchers who used in most cases surveys to draw a picture of customers.

The other clusters *centrality*, *parking availability* and *number of companies* are factors indicating the city structure. Although the variables do not give information about the concrete destination they describe the conditions which are necessary for a successful carsharing utilization rate. To summarize, FFCS works best in areas with a vibrant urban life. A mix of business, recreational and residential use of a district guarantees a high demand over the whole day. Hot spots for carsharing are therefore places in a city with a general attractiveness. These are also typical locations with a high parking pressure. Nonetheless, a direct relation between the kind of parking license area and FFCS booking frequencies cannot be observed. There is just a slight trend for short-term parking

zones observable which is assumed to be caused by a high fluctuation of parking cars and the consequent easier availability of parking lots for non-residents. The influence of weather seems to be more complex, too. Heavy users tend to use the system more intensively during bad weather conditions while there is no significant difference of booking frequencies observable in general. Options for improving this approach are already mentioned in section 6.6.
Modeling carsharing demand with time series models is better performed with HWF than with ARIMA. The main problems for ARIMA are the number of parameters and the recursive algorithms. A period of three months of historical data are sufficient to get an acceptable forecast result. Nevertheless, it has to be regarded that the error of 0.84 and 0.76 vehicles per hour in Berlin and Munich, respectively, is an average over all postal zip codes and hours. The distinction between the defined zones shows higher errors for highly demanded areas and lower ones for peripheral areas but it would also be of interest how the forecast errors differ over the day. A qualitative result is already given in the diagrams of the predicted booking profile.
The by-products of the time series analysis are insightful. Clustering zip code areas by their spectra results in a similar distinction than the classification on the basis of the number of bookings. The demand is thus more or less evenly distributed on a spatial level over the whole day although there are some specific spots demanded only temporarily (e. g. Munich Airport). This also clarifies that a free-floating carsharing system is not as flexible as its name pretends. The vehicles have to be moved by other users who have their own destinations and do not think about other users' need for a vehicle. Trips by the operator's staff and incentives for users are currently the only chance to locate the fleet in an optimal manner.
A second by-product of the time series analysis is the fact that the standardization by the number of active customers per month divided by the fleet size leads to a quite stationary time series. Consequently, an increase of number of vehicles will also mean more active users. This implies that there is still a potential for the FFCS market in Germany.

Next to these research results some questions remain open.

1. A disadvantage of the regression analysis approach is that there is no distinction between user types possible. An option to characterize customers in a better way is to *match user data with booking data*. This combination of actual booking behavior and statements in surveys would be of high

value but is difficult to perform in practice because of strict data privacy policy.

2. Parking management zones – even if they are determined in a well-elaborated way like in Munich – are not *a practical measure for parking pressure*. This pressure occurs when driving cars find in an area around their destination too less parking lots. This is a spatial phenomenon but also dependent on time. There need to be detailed information available about the parking situation and also about the number of cars seeking a parking lot. It would be for instance interesting how to define a parking lot seeking car just by a track of its GPS position.

3. Someone has to be aware that booking data is used for all methods to model the demand of FFCS. But this quantity just illustrates the observed demand. Future research should also concern about the *theoretical or hypothetical demand of carsharing*. The potential does not seem to have been exhausted yet.

4. Future research should also regard how an *implementation of the provided forecast in the fleet management* could be realized. The predicted number of bookings per hour is nice to have but how to determine a lack or overflow of vehicles is another question. The information that is presented to the customer must be easily comprehensible. An indication of availability probability would demand an underlying statistical model.

5. *Relocations with autonomous vehicles* are at present future visions and interesting topics for researchers. Next to the question of the practical implementation, it is interesting to think about how the FFCS system would change. The demand of vehicles could be much better satisfied, but it remains open if the system would still be profitable and sustainable with many idle trips.

6. The segmentation of zip code areas may also not be appropriate for the operator's purpose. Another tendentially simpler grid is thinkable. The findings of the regression analysis can also help to take the spatial differences into account. The NB model for Berlin shows a sufficient transferability. *The union of the HWF and NB model* is not easy to handle but would get along without any segmentation.

7. An interesting research topic is how the carsharing and especially *the FFCS market* will develop *in future times*. This does not only concern the

fleet size and the number of bookings but also the characterization of customers. It is revealed in past studies that the customer structure of station-based carsharing systems has changed from the typical young, well-educated early adopter to an average citizen. This trend is also thinkable for FFCS. It is of interest if the vehicles will then be used in the same way or if the spatial and temporal demand will vary from the status quo.

As mentioned in the introduction of this dissertation, sharing economy may make parts of the economy more effective and sustainable. All efforts and research of this dissertation aimed to understand FFCS systems and demonstrate tools for an optimization. Under the assumption that carsharing and especially FFCS is a good alternative to private car-ownership, these works help to make traffic in a city more environmentally friendly.
This final paragraph wants to scrutinize this hypothesis. In the following, the positive environmental effects of FFCS are illuminated critically. The author aims to make the reader think about the system in a different way. Additionally, he likes to give an outlook and recommendations for the FFCS system.
Eichhorst and Spermann analyzed in 2015 ([51]) sharing economy systems and discovered rebound effects. Rebound effects in economics are in general consequences of inventions and launched products whose original purpose reverses into its opposite. A typical example is the introduction of more energy-efficient products which should safe energy and money for the customers but often have bigger, more qualitative and entirely considered more energy consumptive products on the market in consequence. One analyzed example of the study is *airbnb*. The positive effect for the environment is the lower necessity of hotels and the better efficiency of dispensable private living space. The observed rebound effect results from a "misuse" of the system. The original positive purpose of earning money with private room availabilities changes to a lucrative business model for tenants. Apartments which offer space for e. g. three people were used as two-person households. It is more profitable to let the available room for *airbnb* customers than for regular tenants. This makes housing space scarcer for permanent residents and lets the rents increase.
Transferring this scenario to the carsharing market would mean that people do not only use a vehicle in a smarter way by sharing it with others but drive more often by car than without the FFCS availability. The effect of more parking space due to less FFCS caused private car-ownership can also be destroyed if the theoretically unnecessary parking lots would not directly be converted. Other non-owners can realize the lower parking pressure and tempt to pur-

chase a car.
This rebound effect will unfortunately be visible in Munich. The city council decided in December 2015 ([26]) in consequence of the project results of EVA-CS and WiMobil to promote FFCS by repealing the limit of 500 vehicles per carsharing operator and reducing the parking license fees for the fleet. It was not resolved to reduce and convert 1500 parking lots as the study also proposed to see positive effects. This case shows impressively the missing courage of especially conservative governments to restrict comfort for car drivers.

The potential positive impacts of FFCS to the environment are without tackling its rebound effects repressed.

It is assumed that the decision of the Munich city council satiesfies the carsharing operators. Especially BMW as partner of DriveNow would have welcomed this resolution because it makes more people use cars and especially their car models. The involvement of BMW in the project WiMobil which final results serve as basis for the decision of the city council is therefore considered critically. To the author's impression BMW used research results for their own interests by omitting facts about the rebound effects of FFCS consciously. It is obvious that FFCS operators primarily do not care about environmental sustainability but need this label to get a political and societal legitimation for offering their products.
But it is not only the non-reduced parking space, it is the way FFCS is used in general. 60 % of all trips are less than 5 km ([94]). The average speed of trips with a beeline of more than 800 m between start and end point is less than 25 km/h. It is expected to be even lower for shorter trips. All these rides could have been easily replaced by bikes or public transport. Another critical fact is that most of FFCS trips take place in the direct city center where just a few people need a car. The public transport system and bike lanes offer a much better and much more sustainable infrastructure for inner city trips. Only those who can afford the high prize of the FFCS system can use its convenience – at the expense of all other citizens' health.

Nevertheless, there are undeniable positive effects for urban traffic and the environment.

- FFCS has the opportunity to support public transport by filling the gap in timetables especially during nights. This is particularly noticeable for trips to Munich airport.
- FFCS serves as a good and practical solution for companies. Instead of offering some employees business cars it can make much more sense to launch an own carsharing system. The observed BMW effect is an example of such a system.

In consideration of the assumed rebound effects of carsharing the author wants to make some proposals to counteract negative influences on the environment by FFCS.

- *Change the prize system!* The current prize system forces users to drive as fast as they can. And it can be assumed that most customers follow the philosophy „Don't be gentle, it's a rental.". Combined with the fact that almost all trips are in town, fuel consumption is supposed to be much higher than average. An additional inclusion of the trip distance to the prize could obviate non-eco-friendly driving behavior. This is already common practice at most station-based carsharing operators. Another option is to change the linear prize system and introduce exponential falling rates per minutes. If the customer will have to pay much more for e. g. his first ten minutes those unnecessary short-term trips will soon disappear.
- *Foster multimodality!* A car key was in previous times the key to individual freedom. To offer a competitive product with an equal convenience FFCS has to be conceived as one part of many mobility offers. To register for each mobility service is for many people simply too inconvenient. A mobility card enabling access to public transport, station-based carsharing, free-floating carsharing etc. is necessary to facilitate a life without an own private car. Moreover, it makes sense to launch carsharing options for longer trips, e. g. over the weekend. Mobility stations including parking facilities for station-based and free-floating carsharing, a bikesharing station and a well-connected public transport stop are also an attractive future scenario.
- *Include the periphery!* The carsharing operator runs its business following the rules of free enterprise economy. This has a FFCS system in consequence as it is nowadays. The environmental effects are not necessarily

positive. The relinquishment of car-ownership is the clue. FFCS has therefore especially be attractive for car owners. This goal can only be reached by extending existing operating areas to the city's periphery or to smaller cities which do not have a comparable high density as towns with over a million inhabitants. Even though the capacity and idle times of vehicles will be high initially, the positive effects of the system are assumed to be better. Since no operator will take this risk of loss of profit the municipalities should offer financial help. It is also thinkable to launch carsharing systems in a tendering procedure. The operators would only be allowed to run their system if they integrate tendentially unattractive areas at the city border into their operating area.

Without regulations for FFCS operators the system will worsen parking pressure, traffic and life quality of a city. A good cooperation between providers and municipalities is necessary to strengthen the positive environmental effects of free-floating carsharing and make it to an attractive and affordable mobility service for large parts of citizenry.

Part IV

APPENDIX

A DATA

A.1 BOOKING DATA OF THE FFCS OPERATOR

Field name	Type	Description	Example	further use
`NAM1`	Text	city where the trip takes place	Berlin, Germany	no
`C.MODL`	Text	car model	MINI	no[1]
`AMT`	Text	license number	M-DX 4981	no[2]
`RES_TIME`	Date, Time	Date and exact time of the trip reservation	09.09.2013 08:00:00	no
`START_TIME`	Date, Time	Date and exact time of the trip start	09.09.2013 08:08:08	yes
`END_TIME`	Date, Time	Date and exact time of the trip end	09.09.2013 08:24:20	yes
`START_KM`	Long	mileage [in km] at the begin of the trip	9158	no
`END_KM`	Long	mileage [in km] at the end of the trip	9167	no
`DISTANCE`	Long	driven kilometers during the trip	9	yes
`START_ZIP`	Long	zip code of the origin of the trip	10119	yes
`END_ZIP`	Long	zip code of the end of the trip	10117	yes
`JORO`	Text	address of the origin of the trip based on the GPS coordinates [3]	Christinenstraße 22, 10119 Berlin	no
`JORI`	Text	address of the end of the trip based on the GPS coordinates	Unter den Linden 38, 10117 Berlin	no
`LATO`	Double	latitude of the GPS coordinates of the origin	52.506039	yes
`LOTO`	Double	longitude of the GPS coordinates of the origin	13.375334	yes

1 only useful to distinguish between BEV and conventional cars
2 only used to sort the data by vehicle
3 used for customers to show them the position of the vehicle

Field name	Type	Description	Example	further use
LATI	Double	latitude of the GPS coordinates of the end	52.517158	yes
LOTI	Double	longitude of the GPS coordinates of the end	13.386676	yes
MINF	Long	vehicle in driving mode [in min]	8	yes
MINH	Long	vehicle in parking mode [in min]	8	yes
PIDN	Long	unique customer identification number	987654	yes[4]
TRIP_TYPE	Text	distinction between private, business and service trips	Private	yes
RES_TYPE	Text	distinction between the different reservation types: Mobile (i.e. via Smartphone-App), Spontan and Online	Mobile	no

Table 23: List of all variables of the booking data provided by DriveNow

4 used e.g. to identify heavy user

A.2 LAND-USE DATA

Variable group	Type	Description	variables
`Bebauung_Ortslage`	Long, Double	number/rate of households classified by their sites	city center, city, outskirts, outside of the city
`Bebauung_Wohnlage`	Long, Double	number/rate of property environments	very good, good, satisfactory, sufficient, poor, deficient
`Beschaeftigung_Ausbildung`	Long	number of students	students in school, students at colleges, apprentices
`Beschaeftigung_Beschaeftigte`	Long	number of employees	—
`Beschaeftigung_Erwerbstaetige`	Long	number of working people (incl. freelances etc.)	—
`Demogr_17_AK`	Long, Double	number/rate of people classified by their age	0-2, 3-5, 6-9, 10-14, 15-17, 18-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-74, ⩾75
`Demogr_17_AK_M_W`	Long	see above, but also distinguished by sex and just absolute numbers	see above
`Demogr_5_AK`	Long, Double	number/rate of people classified by their age	0-14, 25-24, 25-49, 50-64, ⩾65
`Demogr_5_AK_M_W`	Long, Double	see above, but also distinguished by sex	see above
`Demogr_Auslaender`	Long, Double	number/rate of foreigners	—
`Demogr_Auslaender_Ethnisch`	Long	number of foreigners classified by their origin	Western Europe/North America/Australia, Turkey, Greece, Asia/Africa/Middle East, Russia/Belarus/Ukraine, Balkan, East Europe, others

Variable group	Type	Description	variables
Demogr_Einwohner	Long, Double	population figure/rate, also distinguished by sex	—
EinzelhandelZentralitaet_EK	Double	retail purchasing power	per inhabitant, in Mio €, ppm Germany, as index (Germany=100)
EinzelhandelZentralitaet_ZK	Double	purchasing budget	per inhabitant, in Mio €, as index (Germany=100)
Firmenzaehler15	Long, Double	number of companies distinguished by their business	companies per km^2, businesses (each additionally classified in big (more than 100 employees), medium (10 to 100 employees) and small (1 to 10 employees)): total, government agencies, doctors, car dealers and car repair shops, banks, services, retail, wholesale, craftman's businesses, producers, hotels/gastronomy, agriculture, advisors/consultants, insurances, other businesses, unknown businesses
Gebaeude_Bauweise	Long, double	number/rate of houses classified by the quality of the architecture	exclusive, good, satisfactory, sufficient, simple, very simple
HH_Einkommen6	Long	number of households classified by their income [in €]	0-900, 900-1500, 1500-2600, 2600-3600, 3600-5000, >5000
HH_Groesse	Long, Double	number of households and their average household size	—
HH_Groesse3	Long, Double	number/rate of households classified by the number of persons	1,2, $\geqslant 3$

Variable group	Type	Description	variables
`HH_Kinder`	Long	number of households classified by the number of children	total, 1 child, 2 children $\geqslant$ 3 children
`HH_Kinder_4AK`	Long	number of households classified by the age of children	0-5, 6-14, 15-17, $\geqslant$ 18
`HH_Lebensform`	Long	number of households classified by special forms of the household	DINKS[5], high-consumption singles, yuppie-households[6]
`HH_SozialeSchicht`	Long	number of households classified by the social class	upper class, upper middle class, middle class, lower middle class, lower class
`High_Tech_Idx`	Double	affinity to high-tech products	
`KK`	Double	purchasing power	per person [in €], Mio €, ppm Germany, as index (Germany=100)
`KK_DISP`	Double	liquid purchasing power	see above
`KK_KON`	Double	purchasing power for consumption	see above
`KK_TB`	Double	purchasing power for diurnal goods	see above
`Konsum_Idx`	Double	importance of factors for purchase decisions	ecology, price, new product, brand
`Lifestyle`	Long	number of households classified by their lifestyle	active center, down-to-earth people, innovative newcomers, first-time consumers, material orientated people, established cosmopolitans
`Mieten`	Double	rents not including bills	per sqm, as index (Germany=100)
`PKW_Dichte`	Double	density of automobiles [per 1000 inhabitants]	—

5 double income no kids
6 DINKS and high-consumption singles $\leqslant$ 35 years

Variable group	Type	Description	variables
PKW_KFZ	Long	density of private automobiles [as index (Germany=100)]	total, share of used cars, car performance, frequent drivers, company car drivers, affinity to leasing cars, affinity to natural gas passenger cars
PKW_Diverse	Double	number of automobiles	cars, motorcyles
PKW_Konzern	Text	name of the most frequent brand	—
PKW_Nutzung	Long, Double	number/rate of automobiles	total, commercial, private
Profit_Idx	Double	rentability as index (Germany=100)	—
Telkotyp_Idx	Double	type of communication as index (Germany=100)	multimedia professionals, open communication type, classical fixed network user
Typologie_Wohnumfeld	Text	type of the residential quarter	
UMTS_Idx	Double	affinity to UMTS [as index (Germany=100)]	—
Umweltaffinitaet_Idx	Double	environmental affinity	—
Verkehrsanbindung	Long	distance to the national traffic network [in m]	nearest airport, nearest long-distance train station, nearest autobahn entrance
Zentralörtliche_Einstufung	Text	local classification	—
ZM_Typ	Text	market classification	—

Table 24: List of all variables of the land-use data provided by infas for 2012

B RESULTS

B.1 BOOKING DATA ANALYSIS - OD MATRIX ANALYSIS

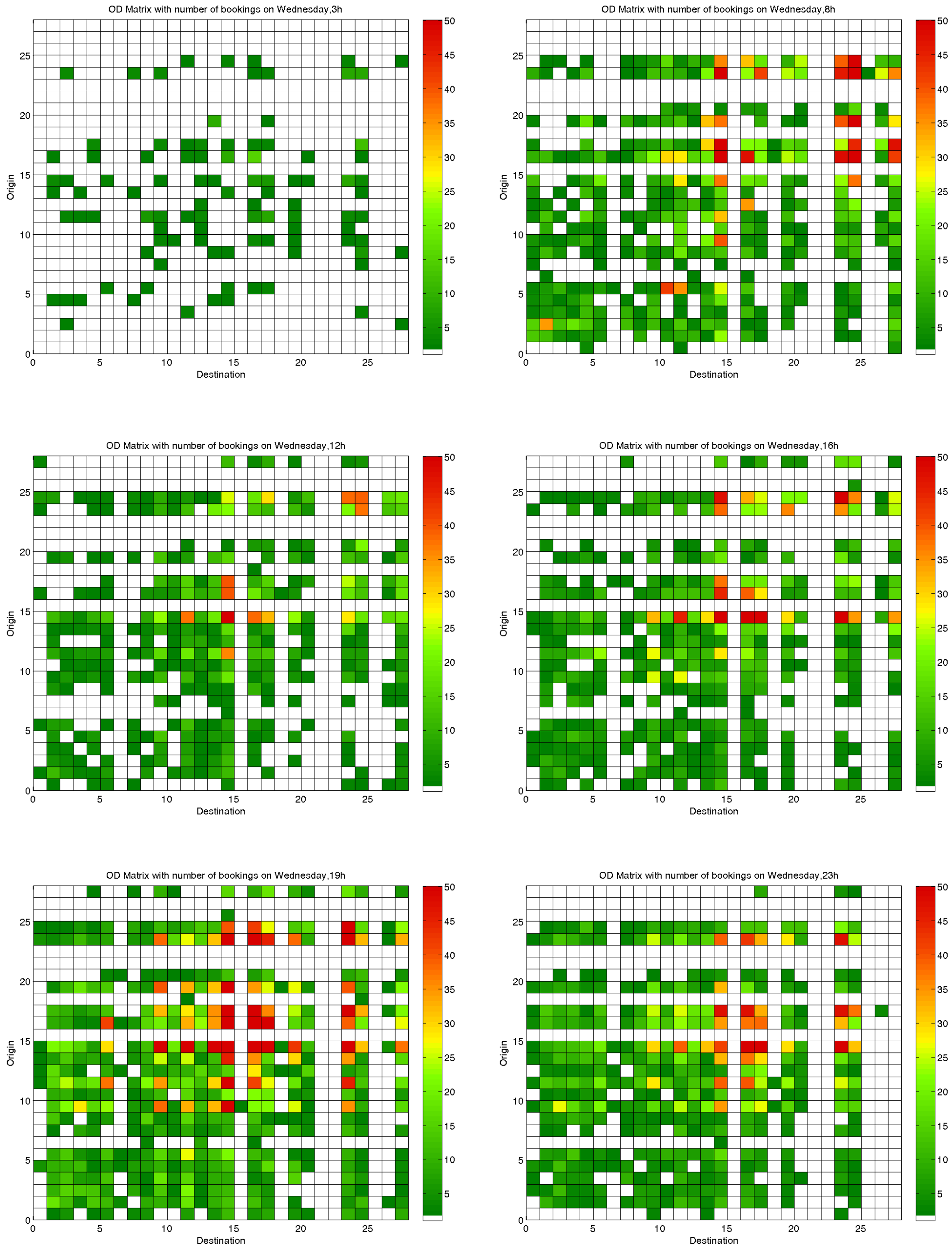


Figure 39: OD matrices for the number of bookings for particular time slots. The district numeration is based on the map in Fig. 15. The chosen time periods are from 3am-4am, 8am-9am, 12pm-1pm, 4pm-5pm, 7pm-8pm and 11pm-12am (from left to right and above to below) averaged over all Wednesdays. White fields mean that no or just one booking took place.

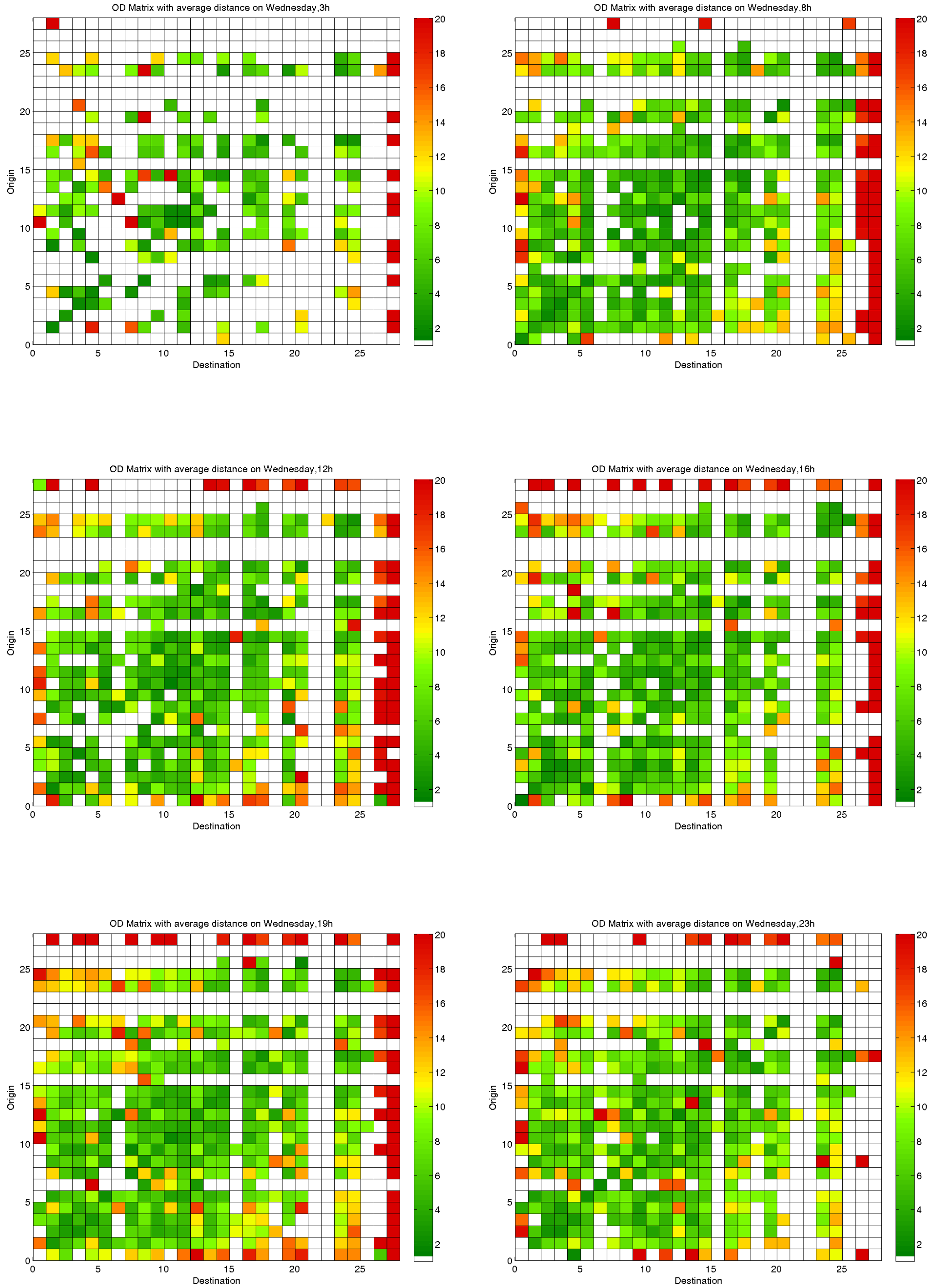


Figure 40: OD matrices showing the average trip distance for particular time slots. The district numeration is based on the map in Fig. 15. The chosen time periods are from 3am-4am, 8am-9am, 12pm-1pm, 4pm-5pm, 7pm-8pm and 11pm-12am (from left to right and above to below) averaged over all Wednesdays. White fields indicate that no booking took place.

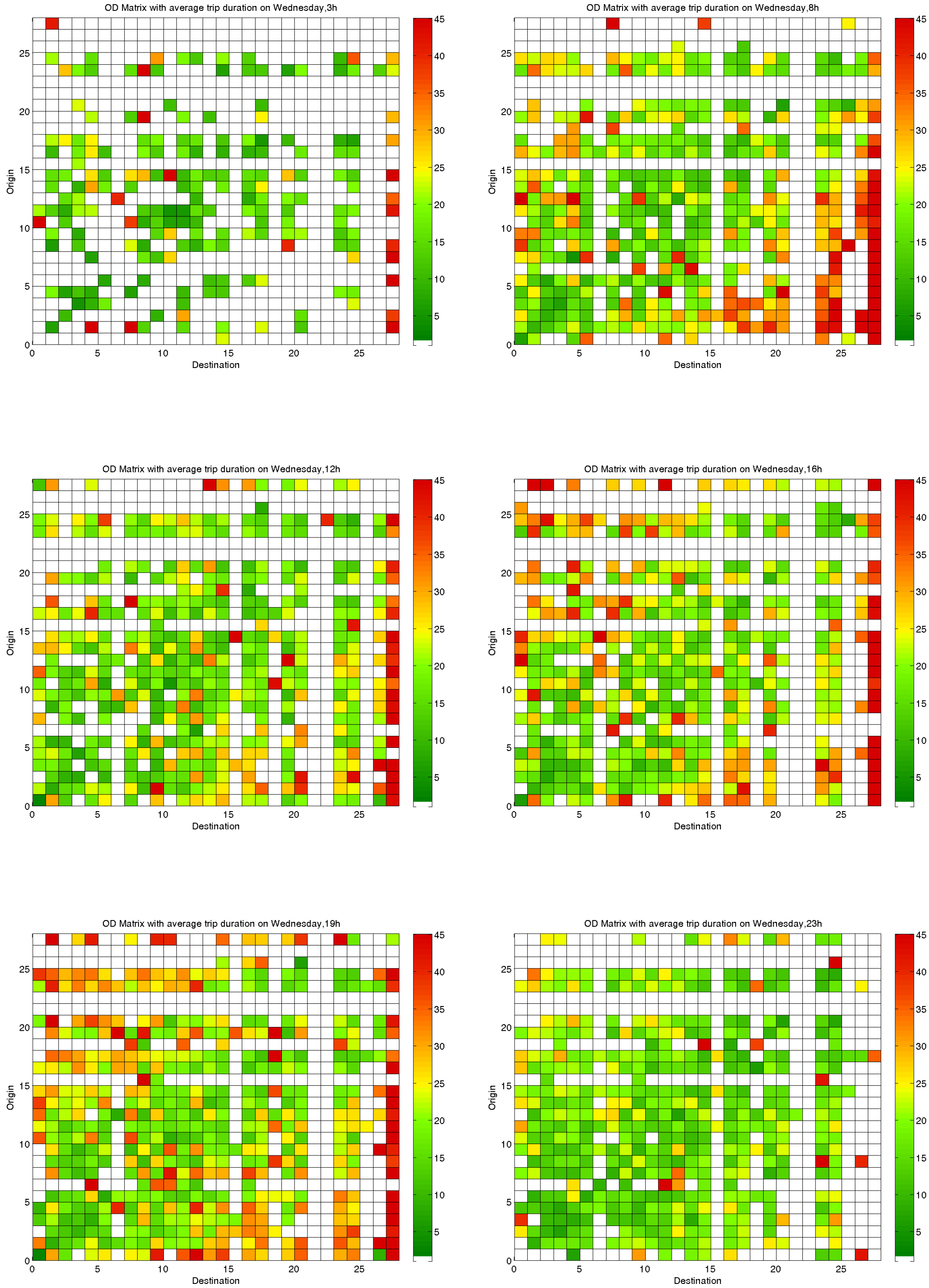


Figure 41: OD matrices showing the average trip duration for particular time slots. The district numeration is based on the map in Fig. 15. The chosen time periods are from 3am-4am, 8am-9am, 12pm-1pm, 4pm-5pm, 7pm-8pm and 11pm-12am (from left to right and above to below) averaged over all Wednesdays. White fields indicate that no booking took place.

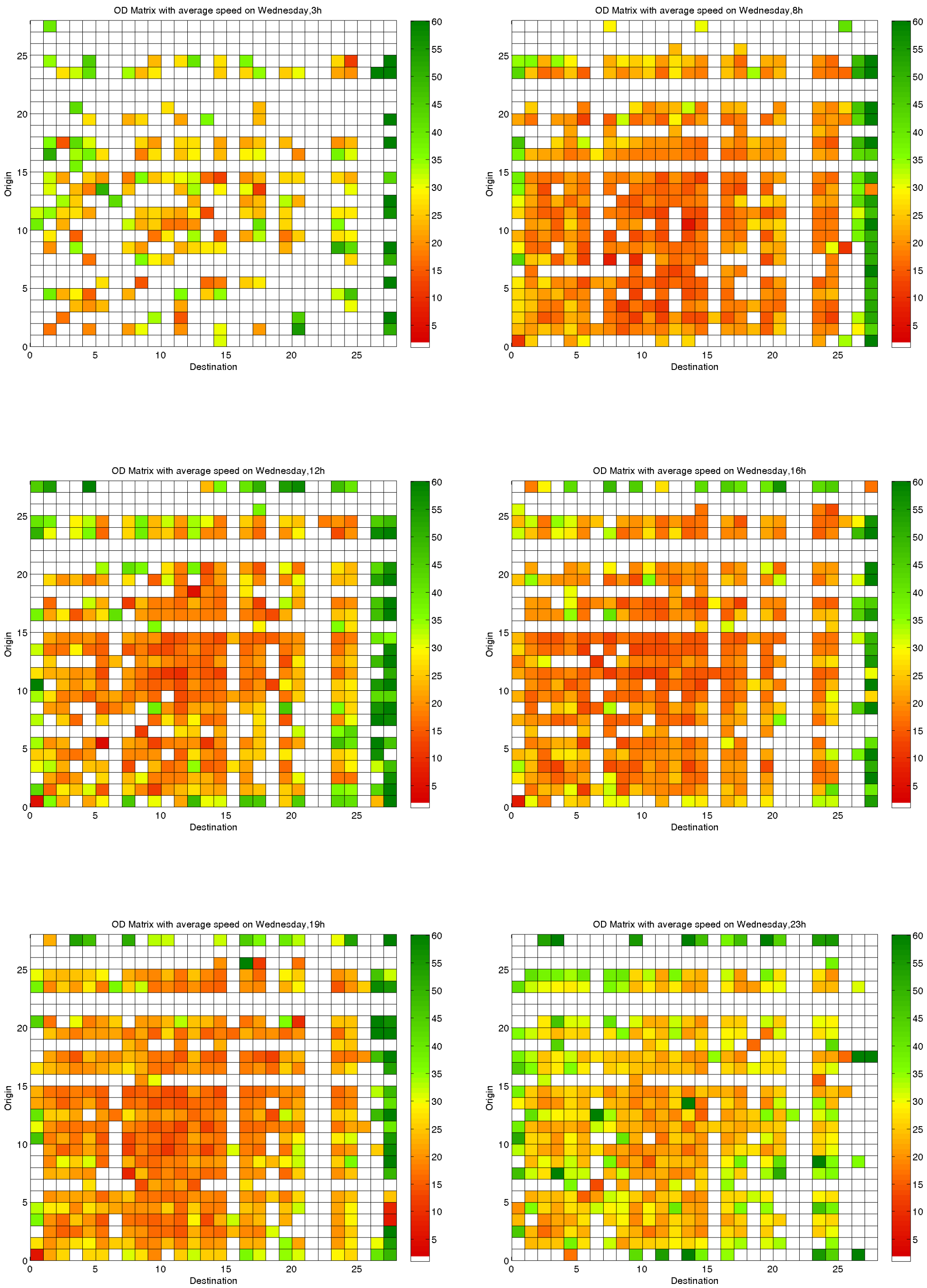


Figure 42: OD matrices for the average speed for particular time slots. The district numeration is based on the map in Fig. 15. The chosen time periods are from 3am-4am, 8am-9am, 12pm-1pm, 4pm-5pm, 7pm-8pm and 11pm-12am (from left to right and above to below) averaged over all Wednesdays. The scale shows the speed in km/h. White fields indicate that no booking took place.

B.2 RESULTS OF THE LINEAR REGRESSION ANALYSIS

	Berlin		Munich	
	t	adj.R^2	t	adj.R^2
citizen data				
# citizens per sqkm	-12.05	0.08		
type of telecommunication: classical fixed network user (index)	-10.69	0.07		
company car driver (index)	18.42	0.18	10.34	0.10
affinity to leased private cars (index)	13.80	0.11		
frequent drivers (index)	16.89	0.15		
environmental affinity (index)	12.02	0.09		
household data				
# buildings	19.05	0.18		
% buildings (best quality)	14.12	0.11		
rent	22.02	0.23		
rent (index)	22.02	0.23		
% private cars (business)	12.78	0.09	12.62	0.14
% private cars (private)	-12.78	0.09	-12.88	0.14
number of companies				
# government agencies and administrative offices	38.56	0.49	18.89	0.27
big	28.55	0.30	14.05	0.17
medium	35.75	0.43	20.04	0.29
small	34.40	0.30		
# medical surgeries	12.94	0.45		
small	12.96	0.10		
# car dealers and car repair shops	10.55	0.07		
medium	12.37	0.09		
# banks	35.53	0.45	11.60	0.12
big	15.04	0.13		
medium	35.17	0.44	11.16	0.11
small	30.83	0.38	11.46	0.12
# services	39.18	0.49	18.36	0.26
big	28.54	0.34	11.65	0.12
medium	40.46	0.51	18.92	0.27
small	35.36	0.44	16.48	0.22
# retail	23.40	0.26		

medium	22.47	0.24		
small	21.60	0.23		
# hotels	34.65	0.43	10.92	0.11
big	23.06	0.25		
medium	28.73	0.34		
small	32.75	0.40	10.68	0.10
# mechanics	20.67	0.21		
medium	19.74	0.20		
small	15.92	0.14		
# manufacturers	26.99	0.32	10.59	0.10
big	13.18	0.10		
medium	25.14	0.29		
small	18.67	0.18		
# wholesale markets	28.95	0.35	12.46	0.14
big	16.56	0.15		
medium	26.91	0.31	13.27	0.15
small	26.00	0.30	10.02	0.09
# other type of commerce	19.18	0.19		
medium	16.72	0.15		
small	19.49	0.15		
# consulting for legal, business and investment	26.61	0.31	12.29	0.13
big	15.53	0.13		
medium	24.40	0.27		
small	24.57	0.28	12.61	0.14
# insurance companies	17.84	0.17		
medium	16.14	0.14		
small	14.73	0.12		
# unknown business	38.01	0.48	16.49	0.22
medium	29.45	0.35	14.34	0.17
small	37.96	0.48	15.80	0.20
# companies (big)	35.64	0.45	16.05	0.21
# companies (medium)	39.31	0.49	16.97	0.23
# companies (small)	35.64	0.45	14.90	0.18
# companies (total)	38.47	0.48	16.25	0.21
miscellaneous				
# cars (total)	16.27	0.14		
street length	26.69	0.31	13.39	0.15
purchasing power in retail per citizen	24.47	0.27		

purchasing power in retail in Mio €	33.84	0.28	10.87	0.11
purchasing power in retail (index)	24.66	0.28		

Table 25: Results of the linear regression with land-use data for Berlin and Munich. The listed variables have an absolute t-value of at least 10.

	Poisson regression				Quasipoisson regression				Negative binomial regression			
Variable name	Estimate	Std. Error	z value	Sign	Estimate	Std. Error	t value	Sign	Estimate	Std. Error	t value	Sign
Intercept	14.18	0.16	86.25	***	14.18	2.63	5.40	***	8.92	2.70	3.31	***
citizen data												
rate of foreigners	-0.01	1.01e-04	-57.74	***	-0.01	1.62e-03	-3.61	***	-4.21e-03	1.57e-03	-2.68	**
average number of persons per household	0.24	0.01	26.46	***	0.24	0.15	1.66	.	0.50	0.16	3.09	**
citizens per sqkm	-3.13e-05	1.32e-07	-236.05	***	-3.13e-05	2.12e-06	-14.77	***	-2.98e-05	2.12e-06	-14.06	***
% citizens classified by their age: 3 - 5 yrs old	-0.10	4.25e-03	-23.68	***	-0.10	0.07	-1.48		-0.23	0.07	-3.37	***
% citizens classified by their age: 10 - 14 yrs old	0.16	2.74e-03	57.64	***	0.16	0.04	3.61	***	0.12	0.04	2.70	**
% citizens classified by their age: 18 - 19 yrs old	-0.24	4.47e-03	-54.13	***	-0.24	0.07	-3.39	***	-0.13	0.08	-1.69	.
% citizens classified by their age: 30 - 34 yrs old	0.01	1.15e-03	4.60	***	0.01	0.02	0.29		0.02	0.02	1.10	
% citizens classified by their age: 50 - 64 yrs old, female	0.03	1.86e-03	17.72	***	0.03	0.03	1.11		0.03	0.03	1.03	
% female citizens	-0.02	1.91	-12.76	***	-0.02	0.03	-0.80		2.36e-03	0.03	0.08	
% quality of buildings: good	-0.01	6.50e-05	-83.85	***	-0.01	1.04e-03	-5.25	***	-4.60e-03	11.01e-03	-4.57	***

% quality of buildings: sufficient	-9.89e-04	5.24e-05	-18.86	***	-9.89e-04	8.38e-04	-1.18		-4.60e-04	8.39e-04	-0.55	
% quality of buildings: simple	-3.64e-03	5.98e-05	-60.86	***	-3.64e-03	9.55e-04	-3.81	***	-2.88e-03	8.70e-04	-3.31	***
% quality of buildings: very simple	-7.65e-03	2.37e-04	-32.27	***	-0.01	3.79e-03	-2.02	*	-1.41e-03	3.44e-03	-0.41	
density of private cars (index)	-3.55e-03	5.56e-05	-63.90	***	-3.55e-03	8.89e-04	-4.00	***	-4.78e-03	8.99e-04	-5.32	***
used cars (index)	-3.39e-03	6.60e-05	-51.41	***	-3.39e-03	1.05e-03	-3.22	**	-1.56e-03	1.08e-03	-1.44	
annual mileage (index)	2.48e-03	5.44e-05	45.57	***	2.48e-03	8.70e-04	2.85	**	1.59e-03	8.88e-04	1.79	.
purchase decision: innovation is crucial	-7.01e-05	1.01e-04	-0.69		-7.01e-05	1.61e-03	-0.04		-3.17e-03	1.70e-03	-1.86	.
purchase decision: environment is crucial	6.63e-03	1.14e-04	58.03	***	0.01	1.83e-03	3.63	***	3.57e-03	1.86e-03	1.923	
frequent drivers (index)	-2.25e-03	1.27e-04	-17.73	***	-2.25e-03	2.03e-03	-1.11		-4.24e-03	2.19e-03	-1.94	.
profitability (index)	1.20e-04	1.60e-04	0.75		1.20e-04	2.56e-03	0.05		-0.01	2.67e-03	-2.26	
type of telecommunication: multimedia professionals (index)	2.67e-05	1.30e-03	0.21		2.67e-05	2.07e-03	0.01		-1.77e-03	2.04e-03	-0.87	*
type of telecommunication: open communication type (index)	5.29e-03	2.36e-04	22.37	***	5.29e-03	3.78e-03	1.40		0.01	3.94e-03	1.93	.

type of telecommunication: classical fixed network user (index)	-9.79e-03	2.82e-04	-34.76	***	-9.79e-0	4.50e-03	-2.17	*	-0.01	4.89e-03	-2.06	*
environmental affinity (index)	2.77e-03	1.33e-04	20.87	***	2.77e-03	2.12e-03	1.31		3.35e-03	2.21e-03	1.52	
density of private cars per citizen	-0.02	3.82e-04	-41.76	***	-0.02	6.11e-03	-2.61	**	-0.01	0.01	-0.86	
% private cars	-4.40e-03	9.40e-05	-46.84	***	-4.40e-03	1.50e-03	-2.93	**	-3.37e-03	1.72e-03	-1.96	.
purchasing power in retail (index)	-3.95e-04	2.91e-06	-135.72	***	-3.95e-04	4.65e-05	-8.49	***	-1.02e-04	6.33e-05	-1.61	
household data												
# buildings in total	-2.12e-04	3.57e-05	-5.93	***	-2.12e-04	5.71e-04	-0.37		-2.06e-03	7.44e-04	-2.77	**
% households with 2 persons	-0.01	4.67e-04	-25.86	***	-0.01	0.01	-1.62		1.53e-03	0.01	0.20	
% households with 3 or more persons	-0.02	4.44e-04	-50.06	***	-0.02	0.01	-3.13	**	-0.03	0.01	-4.09	***
# households with net income 900-1500€	4.93e-03	3.73e-05	132.08	***	4.93e-03	5.97e-04	8.26	***	5.91e-03	6.56e-04	9.02	***
# households with net income 3600-5000€	1.51e-03	9.55e-05	15.78	***	1.51e-03	1.53e-03	0.99		2.77e-03	1.71e-03	1.62	
# households with characteristic "active center"	-5.84e-05	4.07e-05	-1.44		-5.84e-05	6.50e-04	-0.09		-3.84e-04	6.44e-04	-0.60	
# households with characteristic "down-to-earth"	-1.02e-03	1.50e-05	-67.95	***	-1.02e-03	2.39e-04	-4.25	***	-7.63e-04	2.50e-04	-3.05	**

# households of the upper class	-3.81e-03	6.38e-05	-59.71	***	-3.81e-03	1.02e-03	-3.74	***	-2.87e-04	9.67e-04	-2.96	**
# households of the upper middle class	2.44e-05	4.59e-05	0.53		2.44e-05	7.34e-04	0.03		-1.37e-04	7.69e-04	-0.18	
# households of the lower class	-1.91e-04	1.43e-05	-13.32	***	-1.91e-04	2.29e-03	-0.83		-2.57e-04	2.49e-04	-1.03	
rent	0.06	1.17e-03	47.86	***	0.06	0.01	2.99	**	0.14	0.02	6.38	***
number of companies												
medical surgeries: small	2.34e-03	1.58e-04	14.88	***	2.34e-03	2.52e-03	0.93		4.30e-03	3.10e-03	1.39	
medical surgeries: medium	0.01	1.96e-03	3.04	**	0.01	0.03	0.19		-0.03	0.04	-0.80	
car dealers and car repair shops: big	-0.22	0.01	-23.61	***	-0.22	0.15	-1.48		0.05	0.21	0.23	
car dealers and car repair shops: medium	0.08	1.54e-03	48.79	***	0.08	0.02	3.05	**	0.05	0.03	1.64	
car dealers and car repair shops: small	-0.01	8.31e-04	-15.46	***	-0.01	0.01	-0.97		-0.01	0.02	-0.58	
# banks: big	-0.13	0.01	-25.37	***	-0.13	0.08	-1.59		-0.04	0.13	-0.33	
# banks: small	0.02	4.49e-04	43.74	***	0.02	0.01	2.74	**	0.02	0.01	1.72	.
# services: big	0.09	1.28e-03	66.37	***	0.09	0.02	4.15	***	0.07	0.04	1.95	.
# retail: big	0.21	0.01	35.02	***	0.21	0.10	2.19	*	0.04	0.12	0.37	
# retail: medium	0.05	6.26e-04	85.48	***	0.05	0.01	5.35	***	0.03	0.01	2.24	*
# retail: small	4.17e-03	1.34e-04	31.00	***	4.17e-03	2.15e-03	1.94	.	3.77e-03	3.05e-03	1.23	
# wholesale markets: big	-0.07	4.35e-03	-16.43	***	-0.07	0.07	-1.03		-0.06	0.12	-0.49	

# wholesale markets: medium	6.31e-03	4.19e-04	15.07	***	6.31e-03	0.01	0.94		-3.02e-03	0.01	-0.35	
# wholesale markets: small	1.29e-03	5.09e-04	2.53	*	1.29e-03	0.01	0.16		2.26e-03	0.0	0.21	
# mechanics: big	0.22	4.75e-03	46.69	***	0.22	0.08	2.92		0.09	0.11	0.81	
# mechanics: medium	-2.48e-03	5.19e-04	-4.78	***	-2.48e-03	8.30e-03	-0.30		1.55e-03	0.01	0.14	
# mechanics: small	4.04e-03	4.85e-04	8.33	***	4.04e-03	0.01	0.52		7.30e-04	0.01	0.08	
# manufacturers: big	0.05	3.62e-04	13.99	***	0.05	0.06	0.88	**	0.06	0.08	0.81	
# manufacturers: medium	0.02	5.39e-04	40.11	***	0.02	0.01	2.51		-0.01	0.01	-0.69	
# manufacturers: small	0.01	8.97e-04	12.10	***	0.01	0.01	0.76		0.03	0.02	1.45	
# hotels: big	-0.02	0.01	-3.16	**	-0.02	0.11	-0.20		-0.37	0.18	-2.10	*
# hotels: medium	-0.03	5.33e-04	-56.52	***	-0.03	0.01	-3.54	***	-0.02	0.01	-1.31	
# hotels: small	0.01	3.18e-04	28.25	***	0.01	0.01	1.77		0.01	0.01	0.97	
# agriculture business: big	-0.03	0.03	-0.96		-0.03	0.49	-0.06		-0.29	0.58	-0.51	
# agriculture business: medium	0.01	2.38e-03	4.77	***	0.01	0.03	0.30		0.09	0.06	1.56	
# agriculture business: small	-0.05	1.94e-03	-27.70	***	-0.05	0.03	-1.73	.	-0.05	0.04	-1.38	
# companies per sqkm	3.03e-05	1.47e-06	20.65	***	3.03e-05	2.35e-05	1.29		5.68e-05	2.82e-05	2.01	*
# consulting for legal, business and investment: big	0.02	0.01	3.82	***	0.02	0.10	0.24		-0.15	0.15	-1.01	
# consulting for legal, business and investment: medium	-1.83e-03	1.18e-04	-15.51	****	-1.83e-03	1.89e-03	-0.97		-3.37e-03	3.05e-03	-1.11	

# other type of commerce: big	-6.42e-03	0.02	-0.35		-6.42e-03	0.29	-0.02		-0.45	0.43	-1.05	
# other type of commerce: medium	0.02	1.49e-03	11.99	***	0.02	0.02	0.75		2.48e-03	0.03	0.07	
# other type of commerce: small	-0.01	6.26e-04	-13.17	***	-0.01	0.01	-0.82		1.79e-03	0.01	0.15	
# unknown business: big	0.25	0.03	7.48	***	0.25	0.54	0.47		0.05	0.99	0.05	
# unknown business: medium	-0.02	5.74e-04	-31.88	***	-0.02	0.01	-1.99	*	-0.02	0.01	-1.53	
# insurance companies: big	0.14	4.93e-03	28.55	***	0.14	0.08	1.79	.	0.31	0.14	2.29	*
# insurance companies: medium	-0.03	1.80e-03	-16.42	***	-0.03	0.03	-1.03		-0.01	0.04	-0.31	
# insurance companies: small	2.37e-03	8.26e-04	2.87	**	2.37e-03	0.01	0.18		-0.01	0.02	-0.54	
miscellaneous												
distance to city center	-1.04e-04	1.01e-06	-102.88	***	-1.04e-04	1.61e-05	-6.44	***	-1.04e-04	1.56e-05	-6.63	***
distance to district center	3.38e-05	2.03e-06	16.65	***	3.38e-05	3.24e-05	1.04		-2.33e-05	3.39e-05	-0.69	
distance to the nearest airport	-4.03e-07	4.25e-07	-0.95		-4.03e-07	6.80e-06	-0.06		4.20e-06	6.23e-06	0.67	
distance to the nearest long distance train station	-3.33e-05	9.58e-07	-34.80	***	-3.33e-05	1.53e-05	-2.18	*	-2.76e-05	1.49e-05	-1.85	.
distance to the nearest freeway entrance	-1.97e-05	8.04e-07	-24.54	***	-1.97e-05	1.29e-05	-1.54		-2.19e-05	1.27e-05	-1.72	.
CDU, first vote	-4.36e-03	.92e-04	-14.92	***	-4.36e-03	4.67e-03	-0.93		-2.44e-03	4.47e-03	-0.54	

SPD, first vote	1.10e-03	4.46e-04	2.47	*	1.10e-03	0.01	0.15		0.01	0.01	0.94	
SPD, second vote	-0.04	6.07e-04	-58.79	***	-0.04	0.01	-3.68	***	-0.03	0.01	-2.83	**
FDP, first vote	0.12	2.77e-03	42.66	***	0.12	0.04	2.67	**	0.09	0.04	1.95	.
Die Linke, first vote	0.01	2.68e-04	48.31	***	0.01	4.28e-03	3.02	***	4.13e-03	4.33e-03	0.95	
Piraten, first vote	-0.10	1.41e-03	-68.68	***	-0.10	0.02	-4.30	***	-0.06	0.02	-2.54	*
Piraten, second vote	0.06	1.17e-03	50.79	***	0.06	0.01	3.18	**	0.05	0.02	2.73	**
AfD, first vote	0.04	2.49e-03	14.560	***	0.04	0.04	0.91		0.05	0.04	1.25	
AfD, second vote	-0.01	2.21e-03	-2.83	**	-0.01	0.04	-0.18		-0.01	0.03	-0.43	
NPD, first vote	-0.18	2.60e-03	-70.21	***	-0.18	0.04	-4.39	***	-0.20	0.04	-5.30	***
density of bars	0.03	5.22e-04	65.94	***	0.03	0.01	4.13	***	0.03	0.01	3.08	**
density of cinemas	0.07	2.06e-03	33.64	***	0.07	0.03	2.10	*	0.01	0.04	0.30	
density of attractions	0.04	3.06e-03	12.56	***	0.04	0.05	0.79		-0.04	0.05	-0.77	
density of hotels	0.08	2.97e-03	27.46	***	0.08	0.05	1.72	.	0.08	0.07	1.22	
density of banks	0.01	1.30e-03	5.87	***	0.01	0.02	0.37		-3.95e-03	0.03	-0.15	
density of taxi stands	-0.01	2.65e-03	-1.92	.	-0.01	0.04	-0.12		-0.01	0.05	-0.24	
density of bus stations	-0.04	9.68e-04	-43.46	***	-0.04	0.02	-2.72	**	-0.03	0.02	-1.86	.
density of subway and sub-urban stations	-0.07	1.60e-03	-42.27	***	-0.07	0.03	-2.64	**	-0.06	0.03	-1.69	.
street length	1.67e-04	8.41e-07	198.57	***	1.67e-04	1.34e-05	12.42	***	2.40e-04	1.94e-05	12.35	***
# bars	1.17e-03	1.83e-03	6.40	***	1.17e-03	2.92e-03	0.40		0.01	4.34e-03	1.34	
# cinemas	0.02	1.74e-03	12.74	***	0.02	0.03	0.80		0.02	0.04	0.55	
# attractions	4.97e-03	1.27e-03	3.90	***	4.97e-03	0.02	0.24		0.03	0.03	0.91	

# hotels	0.01	1.09e-03	7.21	***	0.01	0.02	0.45		-4.23e-03	0.03	-0.15	
# banks	4.39e-03	1.09e-03	4.03	***	4.39e-03	0.02	0.25		-4.89e-04	0.02	-0.02	
# taxi stands	-0.16	2.30e-03	-67.66	***	-0.16	0.04	-4.23	***	-0.08	0.05	-1.61	
# bus stations	-0.01	5.24e-04	-16.98	***	-0.01	0.01	-1.06		6.97e-04	0.01	0.07	
# subway and suburban stations	0.08	1.96e-03	42.03	***	0.08	0.03	2.63	**	-0.04	0.04	-0.85	
area size	5.75e-09	5.61e-09	1.02		5.75e-09	8.97e-08	0.06		2.98e-07	1.27e-07	2.35	*

Table 26: Estimate, Standard error, z-/t-value and significance of all non-redundant variables of land-use data and election results for the Poisson, Quasipoisson and NB model

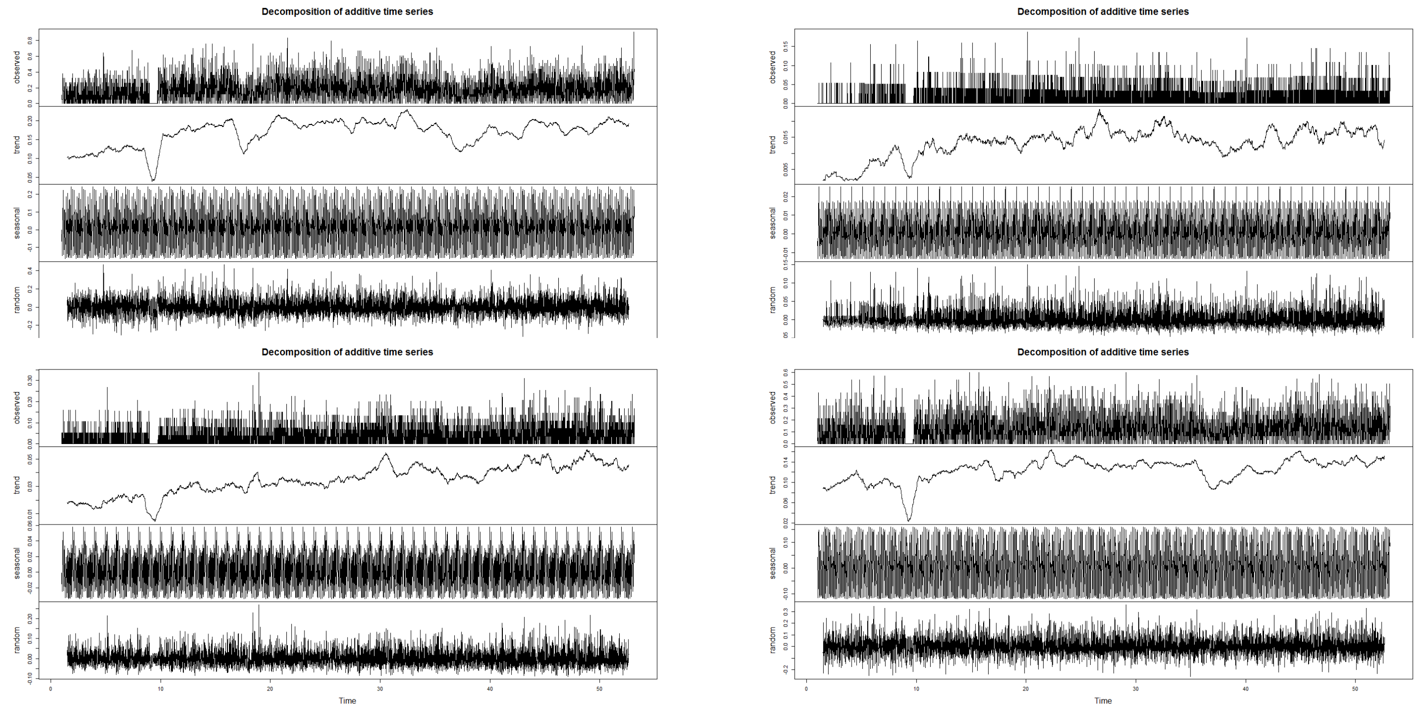


Figure 43: Decomposition for the representatives of Zone A-D in Berlin

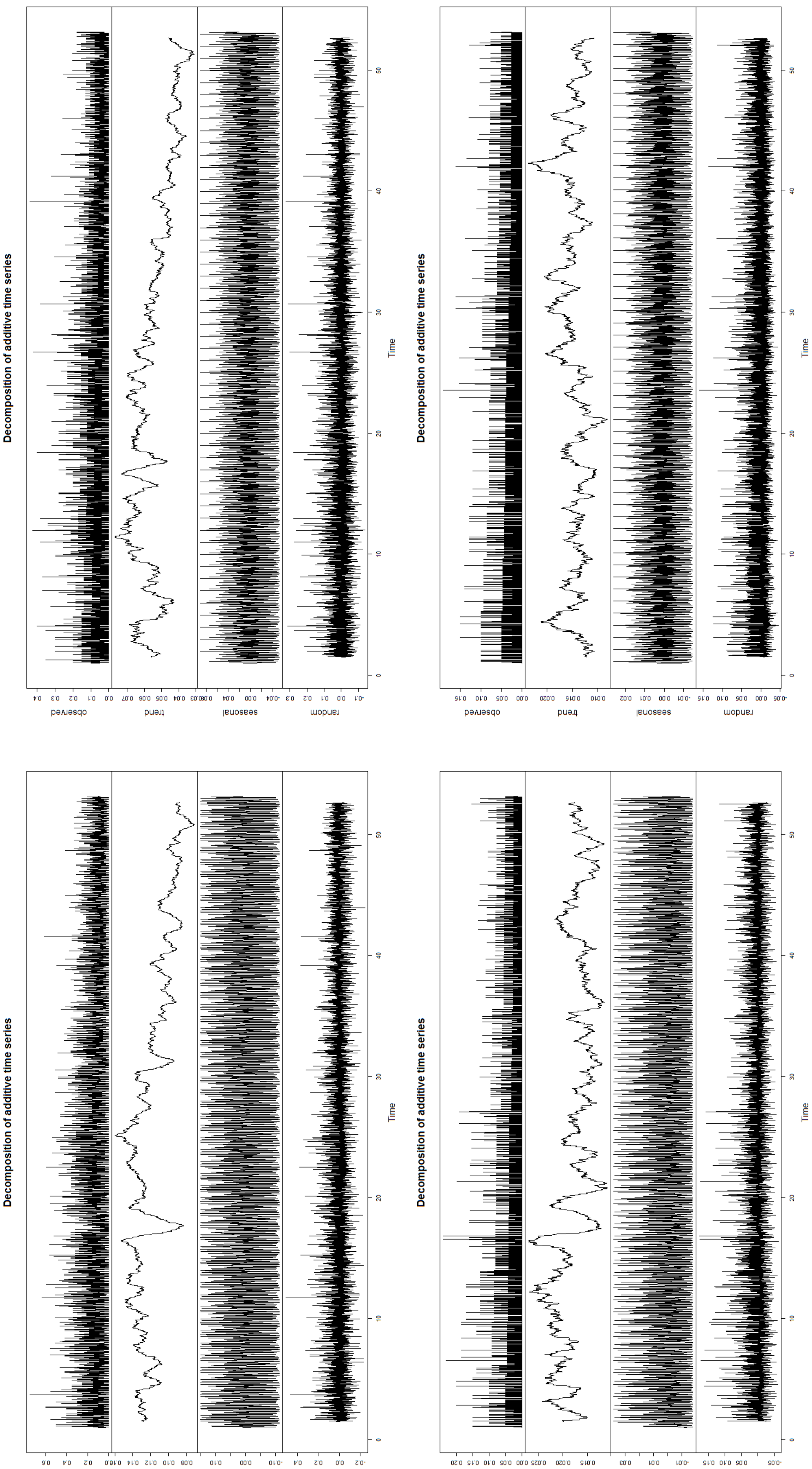


Figure 44: Decomposition for the representatives of Zone A-D in Munich

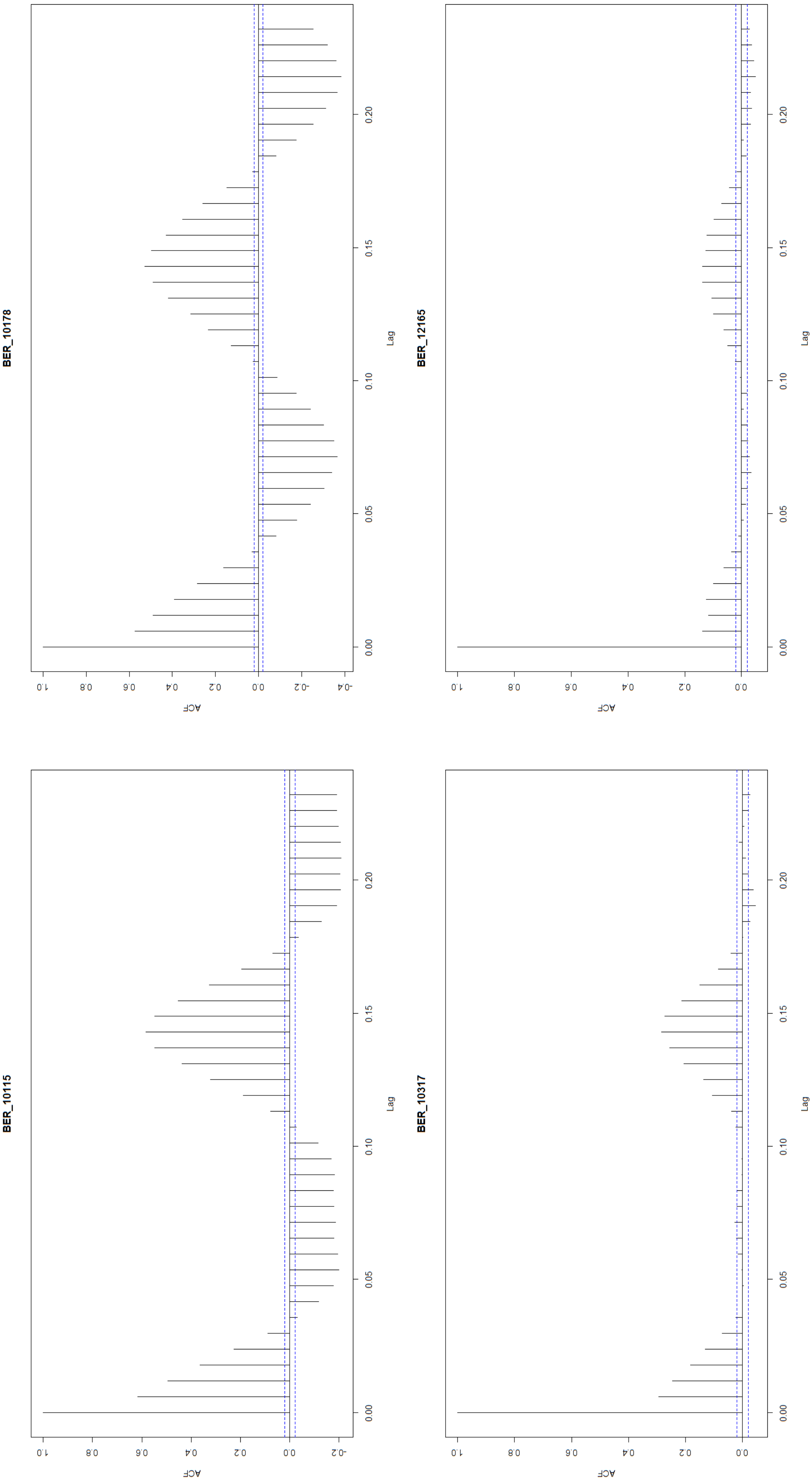


Figure 45: Autocorrelation function for the representatives of Zone A-D in Berlin

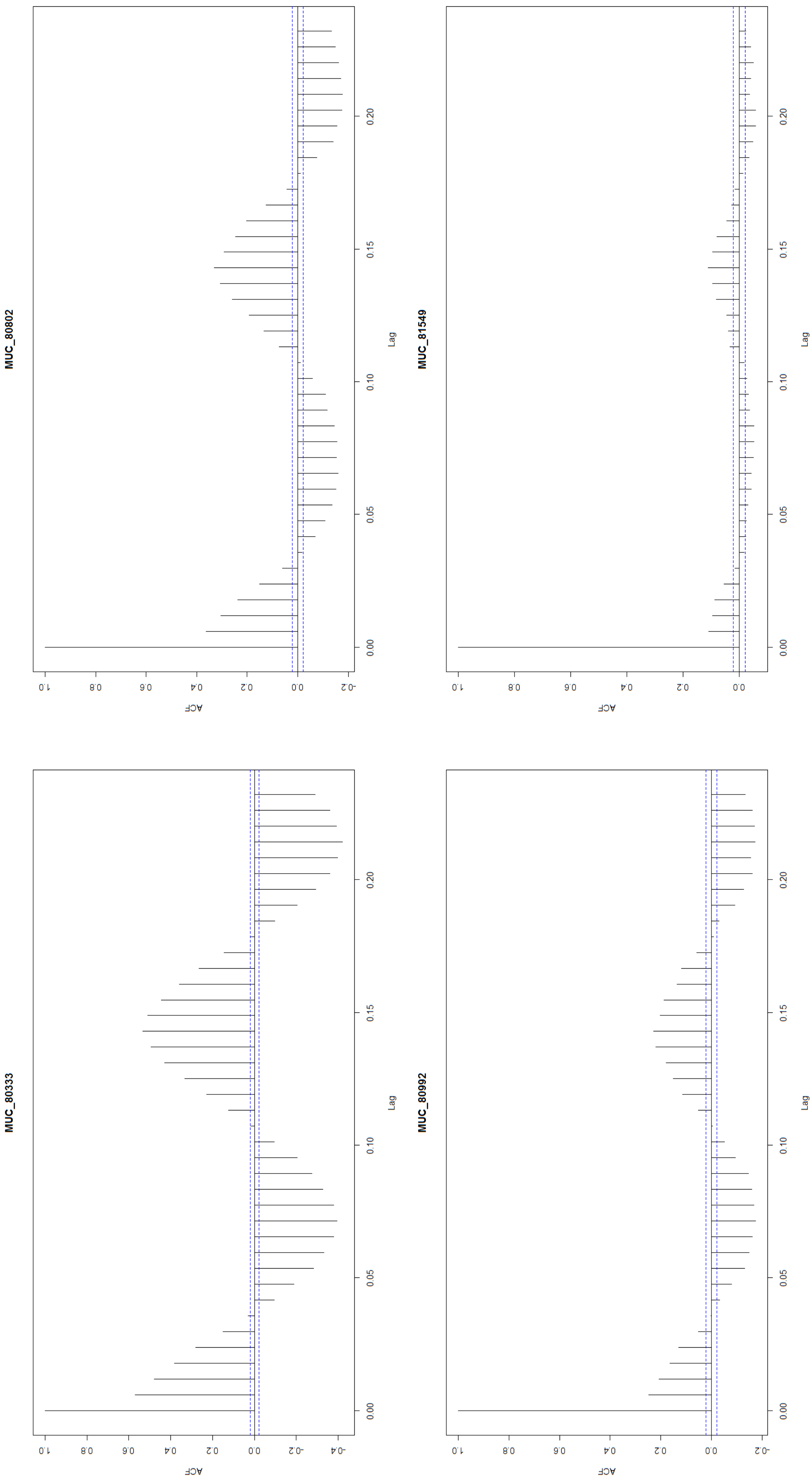


Figure 46: Autocorrelation function for the representatives of Zone A-D in Munich

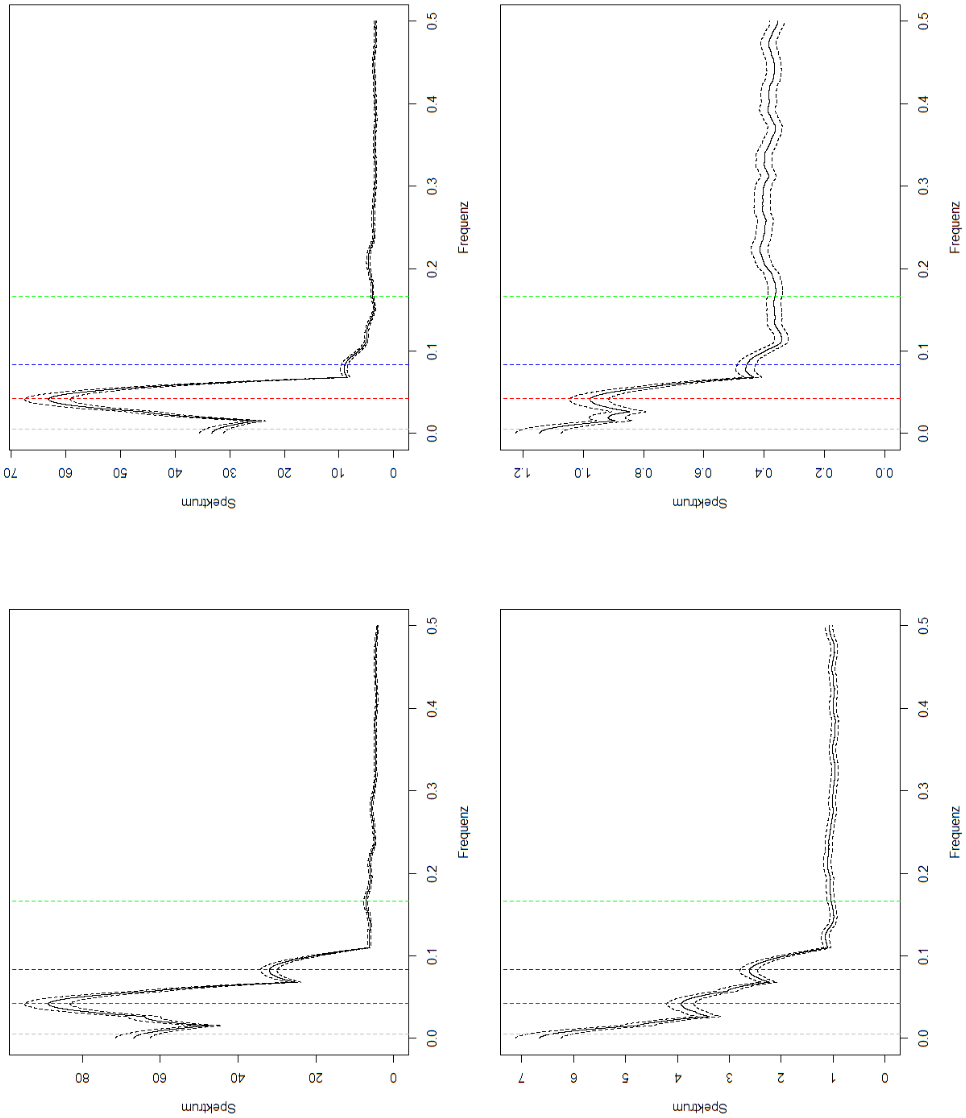


Figure 47: Spectrum for the representatives of Zone A-D in Berlin

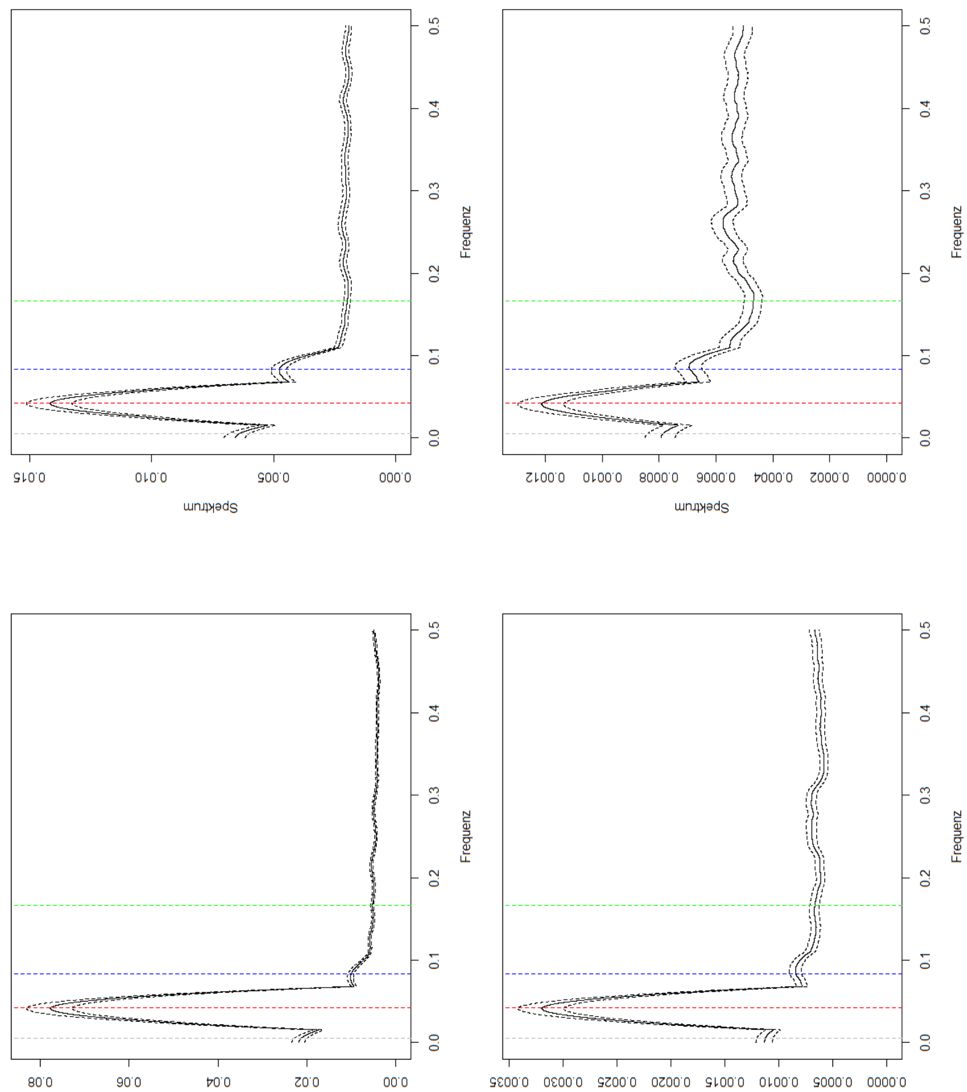


Figure 48: Spectrum for the representatives of Zone A-D in Munich

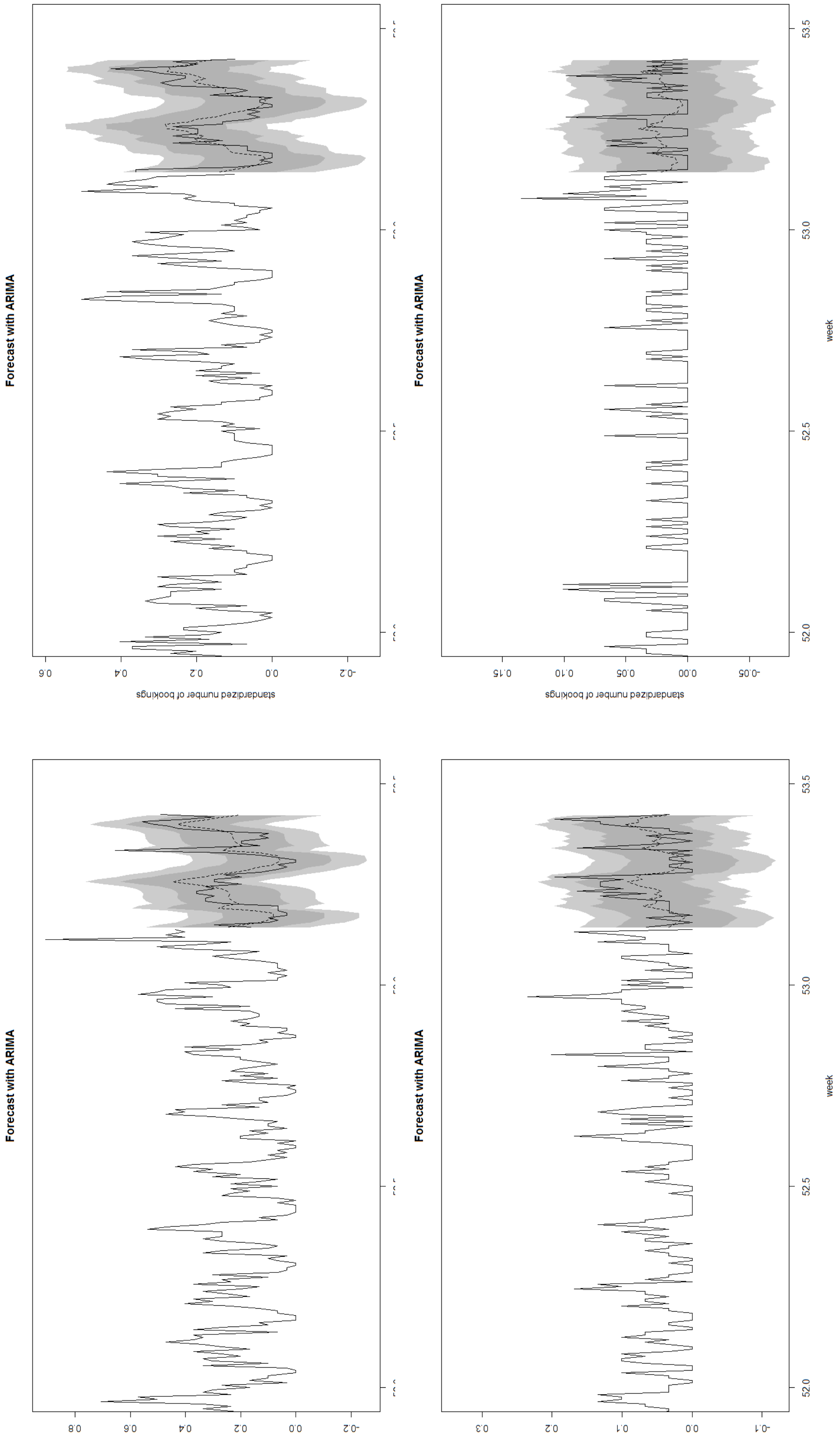


Figure 49: Forecast with ARIMA for the representatives of Zone A-D in Berlin based on data set I (year)

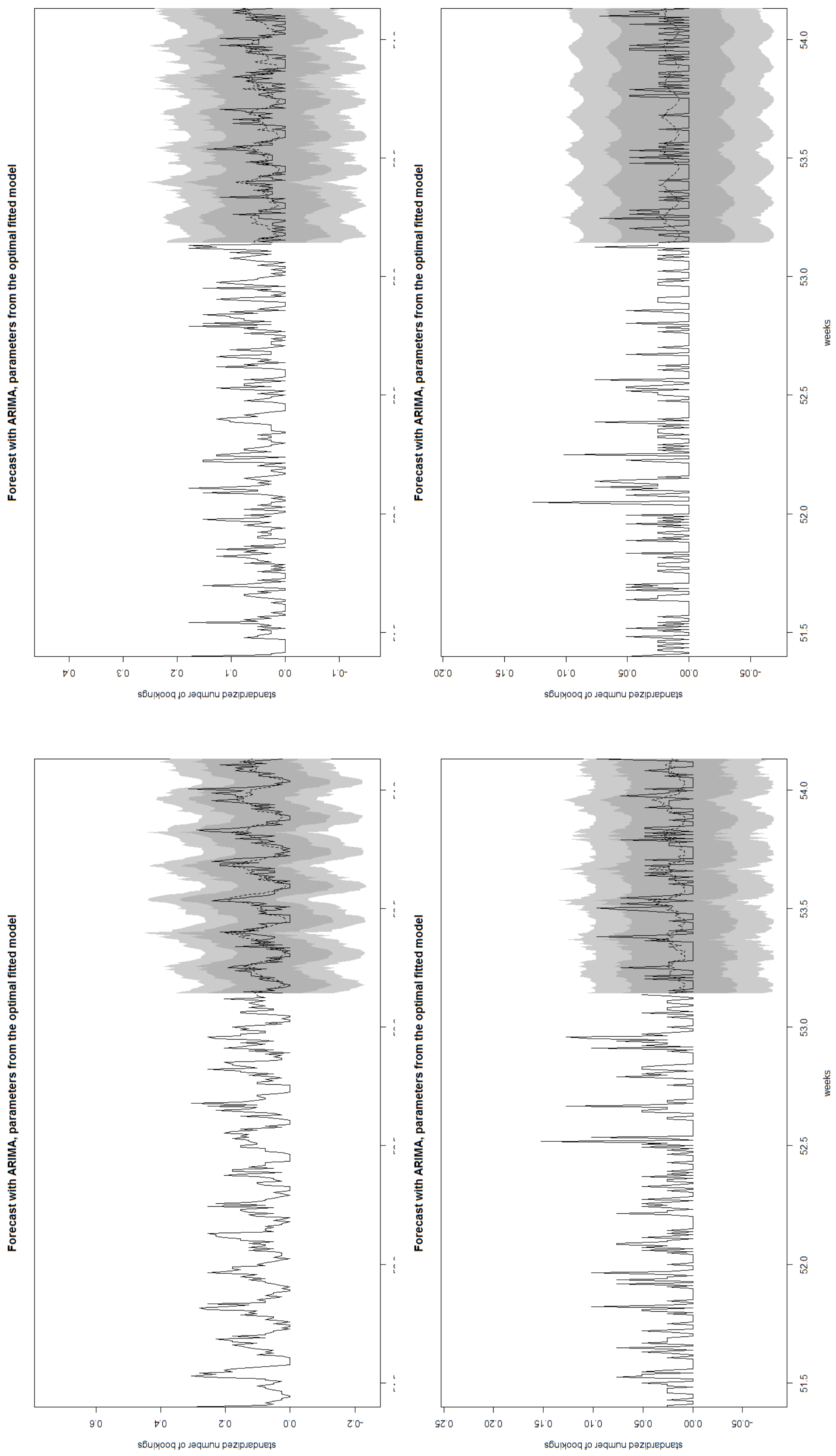


Figure 50: Forecast with ARIMA for the representatives of Zone A-D in Munich based on data set I (year)

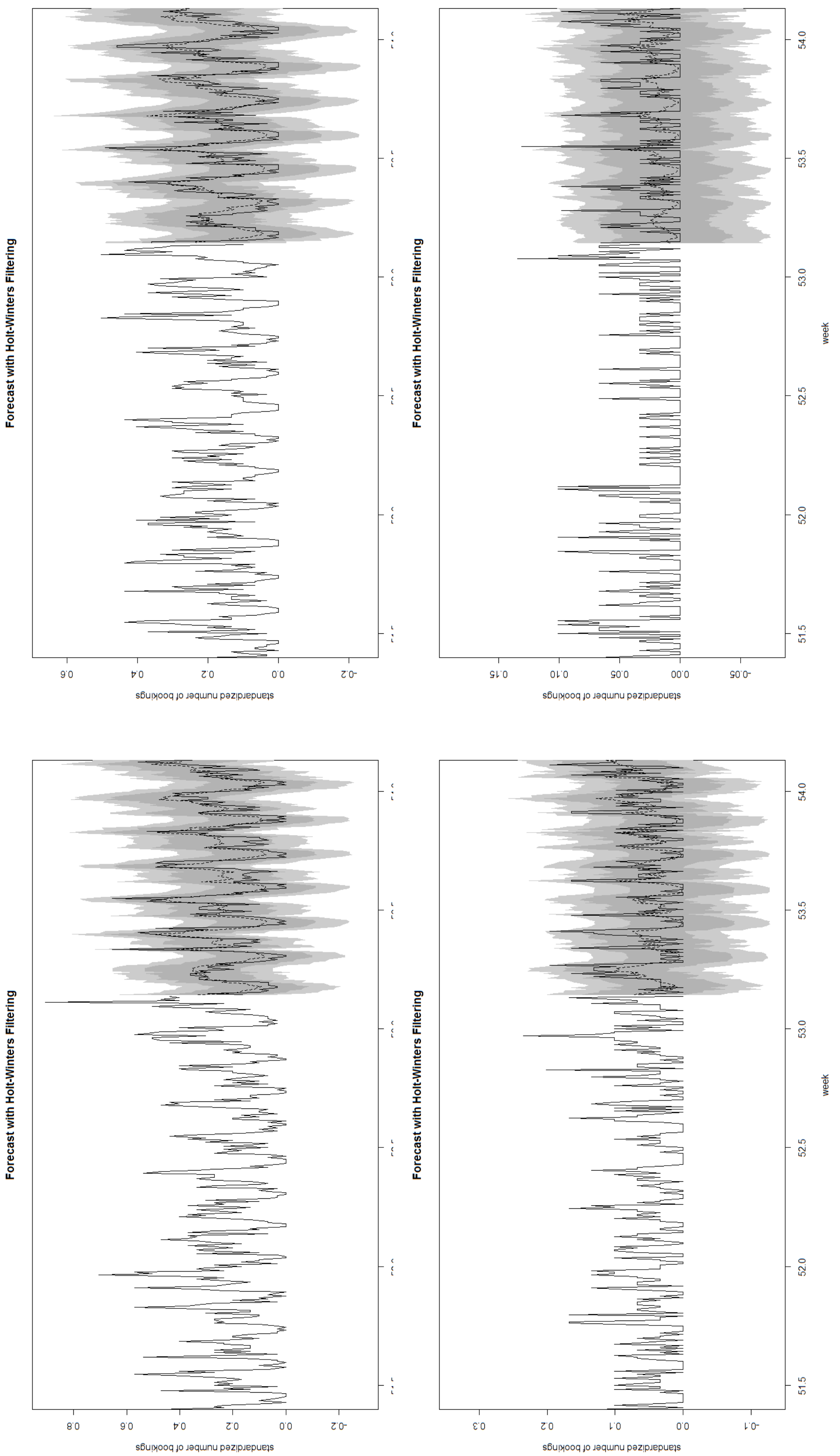


Figure 51: Forecast with HWF for the representatives of Zone A-D in Berlin based on data set I (year)

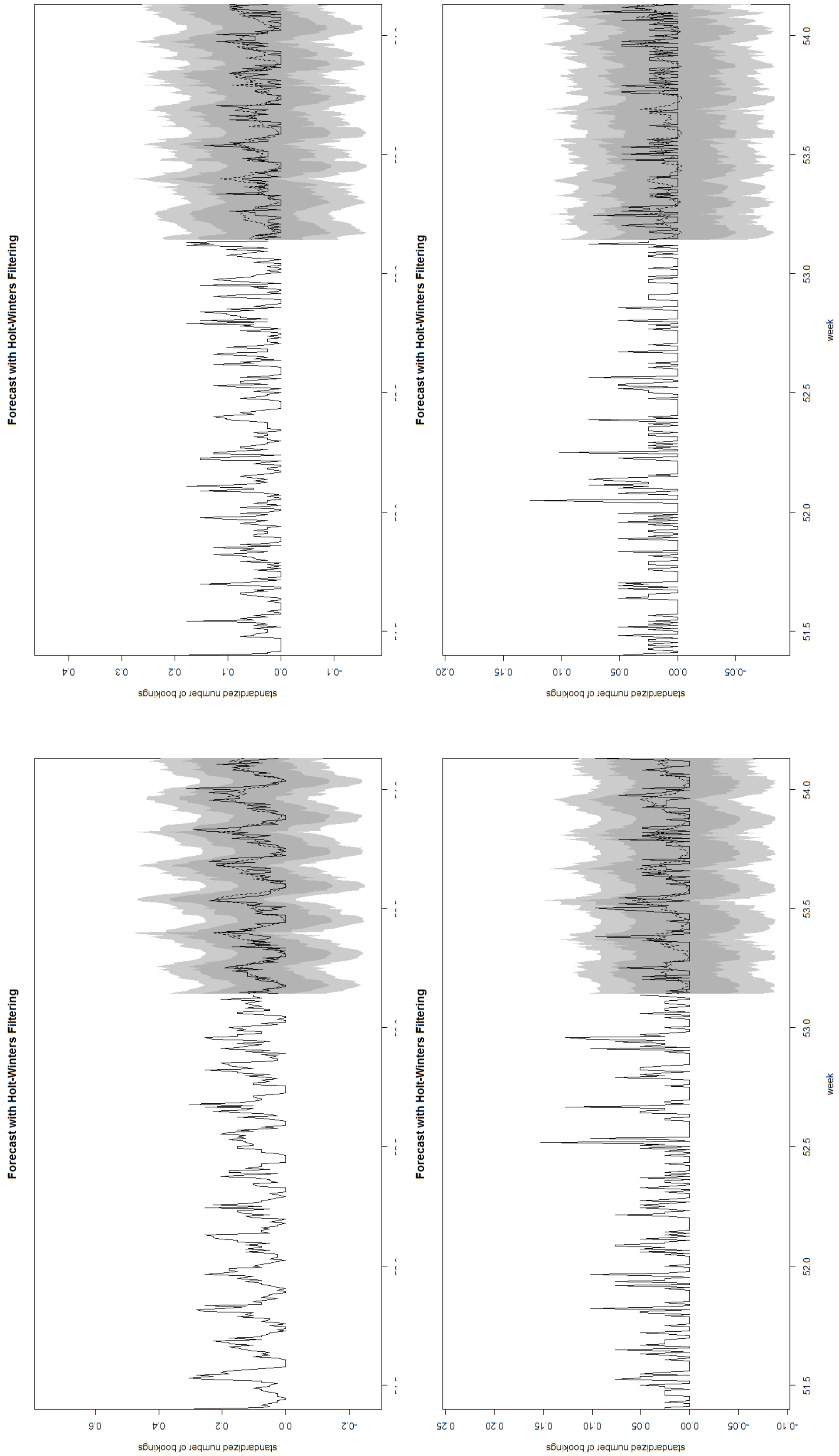


Figure 52: Forecast with HWF for the representatives of Zone A-D in Munich based on data set I (year)

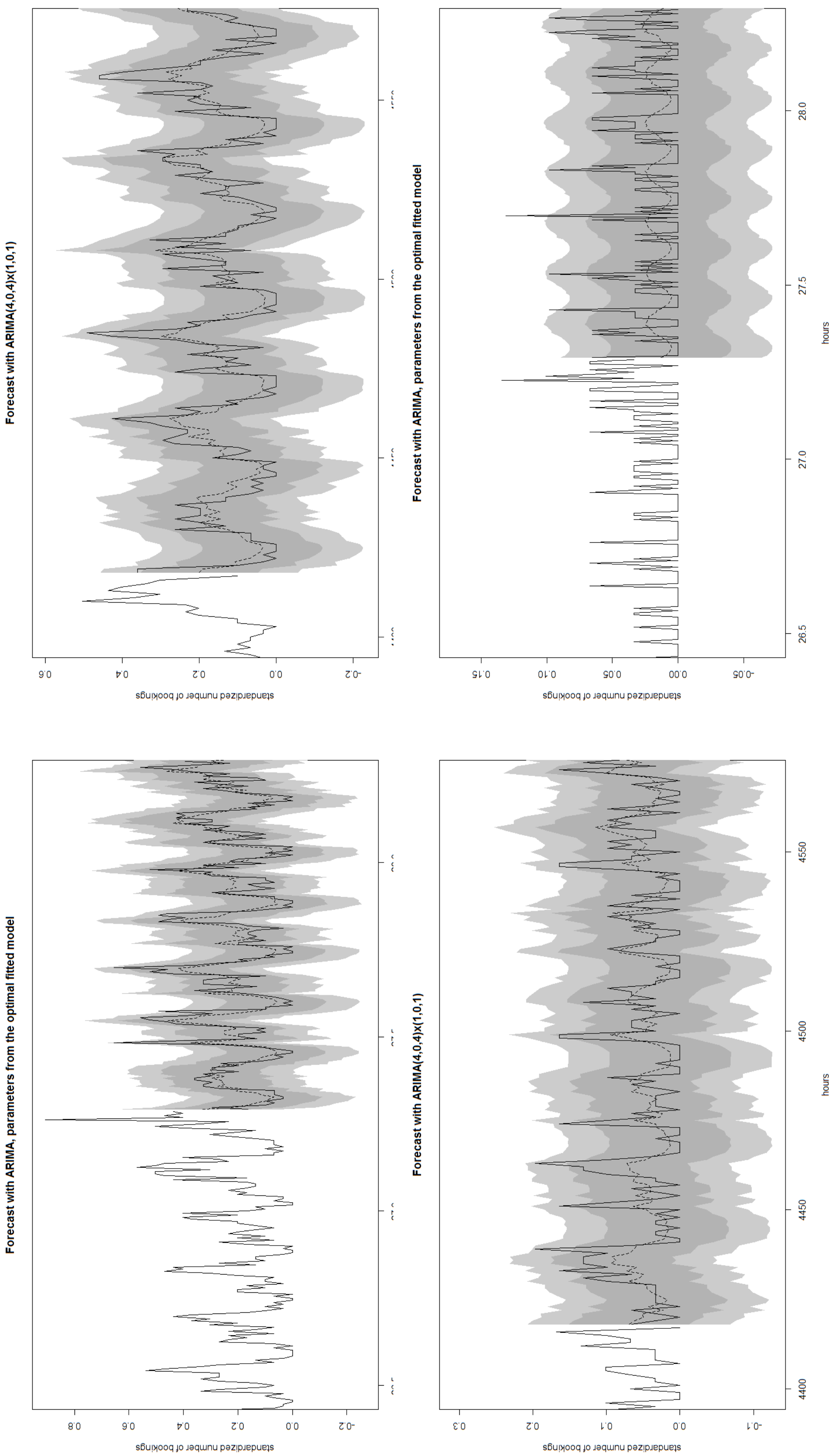


Figure 53: Forecast with ARIMA for the representatives of Zone A-D in Berlin based on data set II (half year)

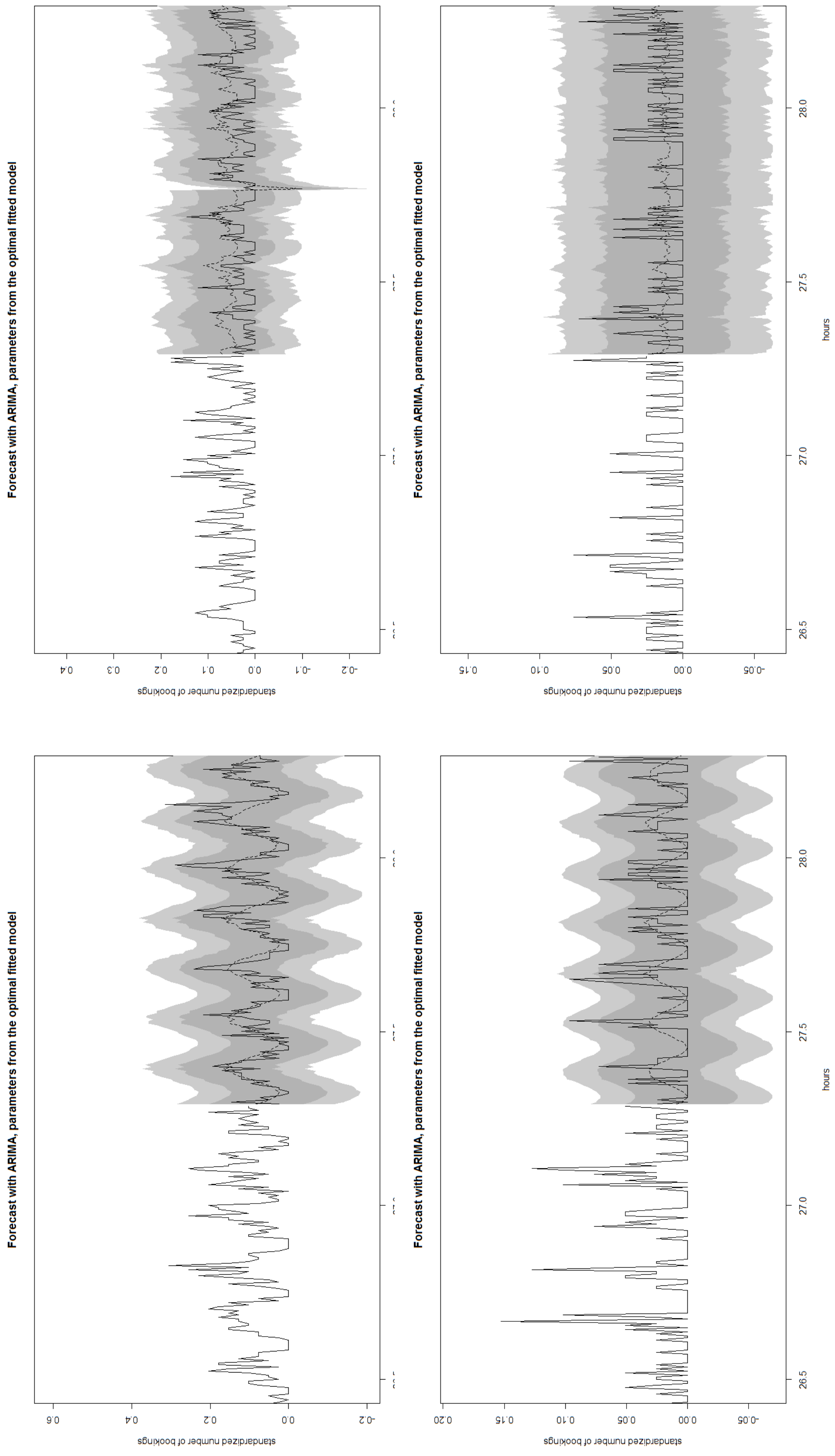


Figure 54: Forecast with ARIMA for the representatives of Zone A-D in Munich based on data set II (half year)

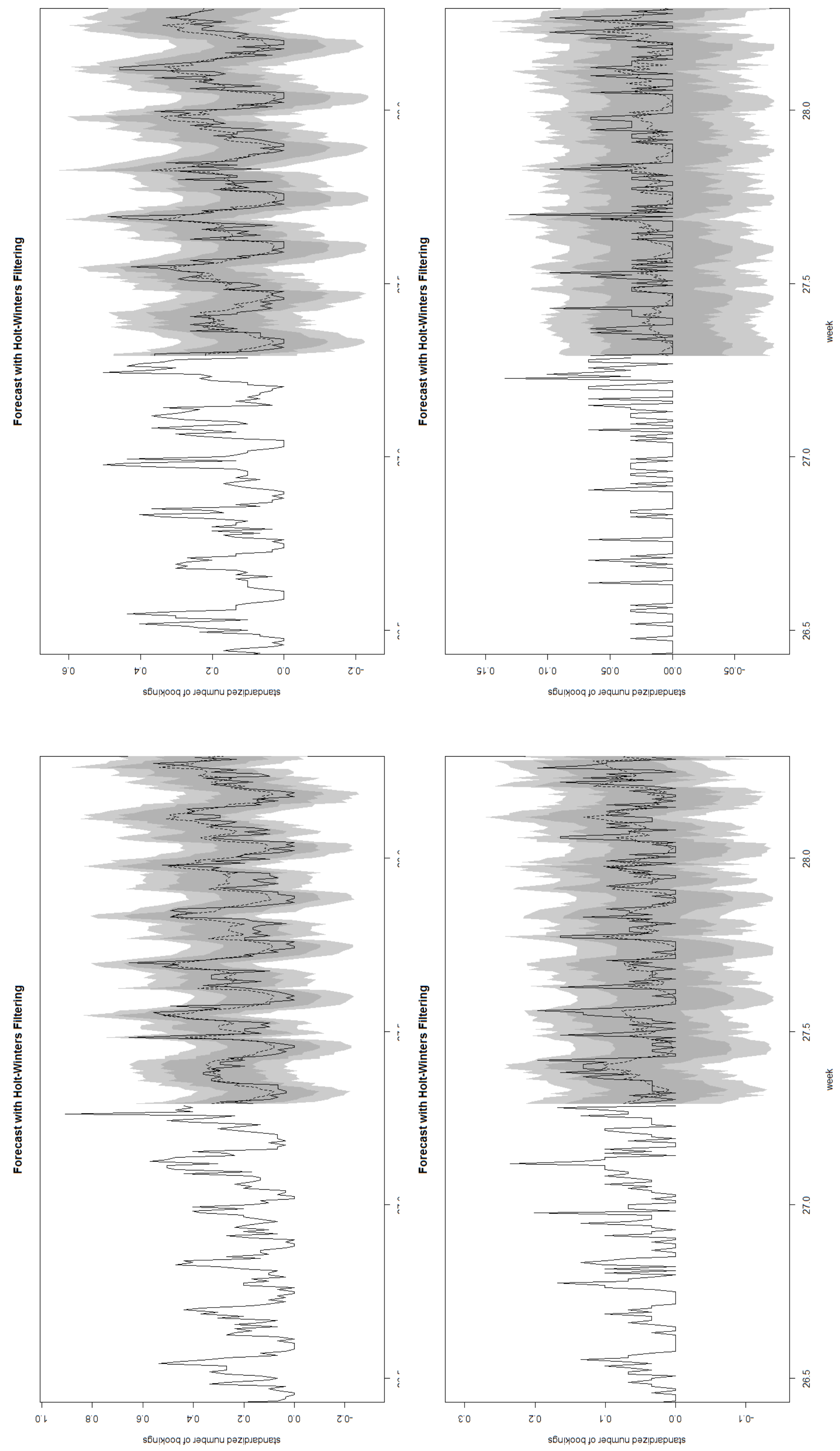


Figure 55: Forecast with HWF for the representatives of Zone A-D in Berlin based on data set II (half year)

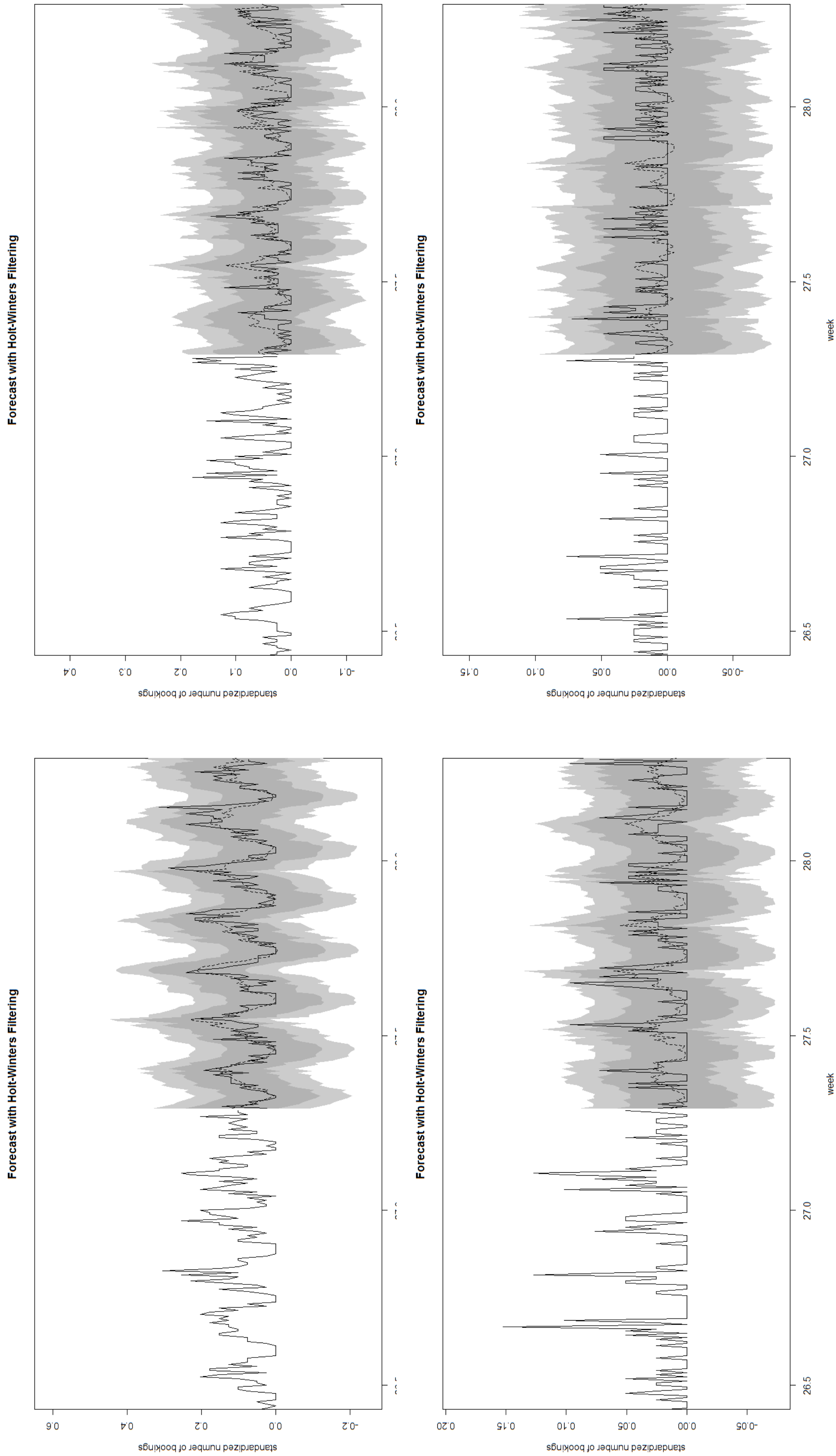


Figure 56: Forecast with HWF for the representatives of Zone A-D in Munich based on data set II (half year)

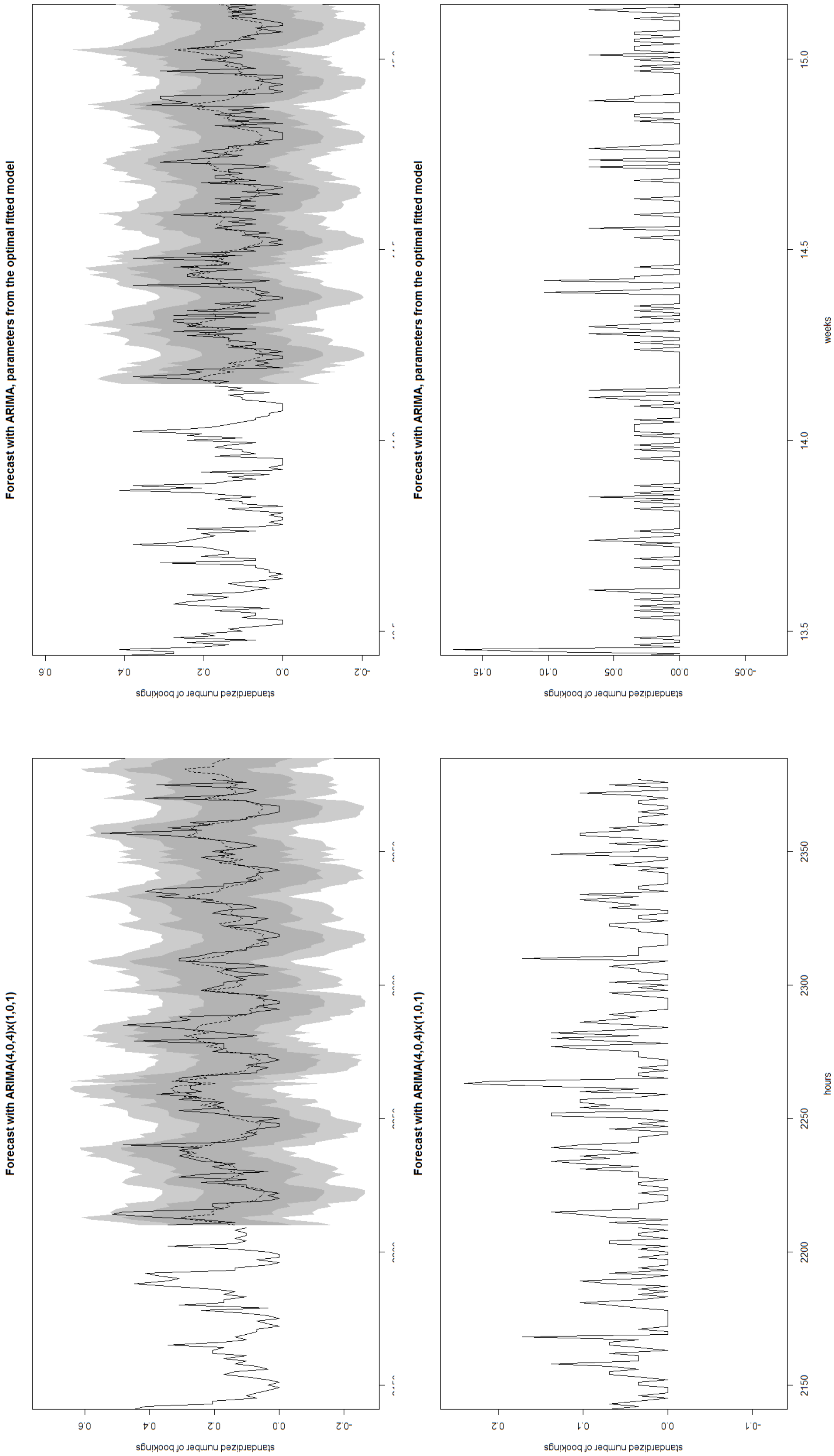


Figure 57: Forecast with ARIMA for the representatives of Zone A-D in Berlin based on data set III (quarter of a year)

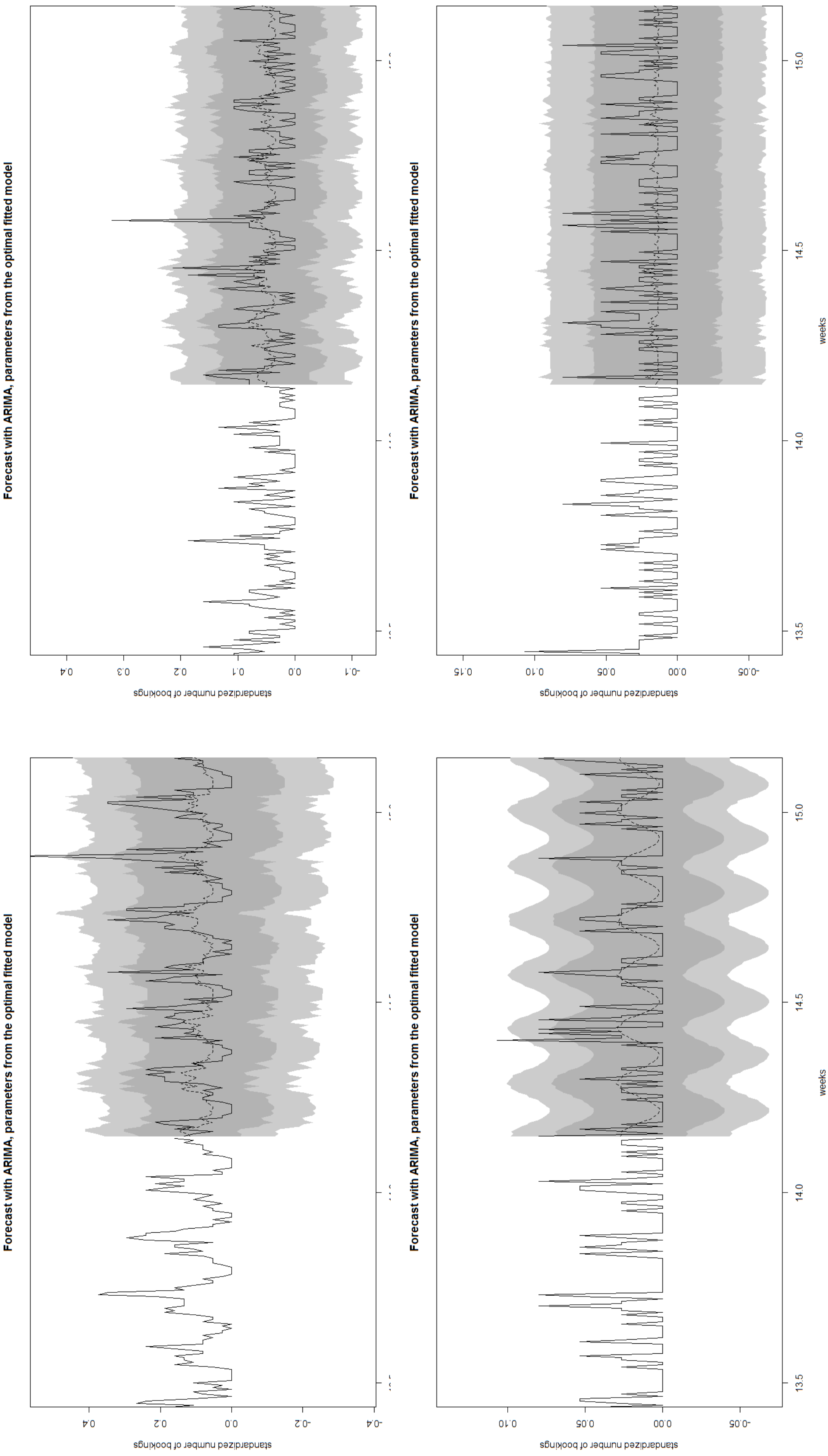


Figure 58: Forecast with ARIMA for the representatives of Zone A-D in Munich based on data set III (quarter of a year)

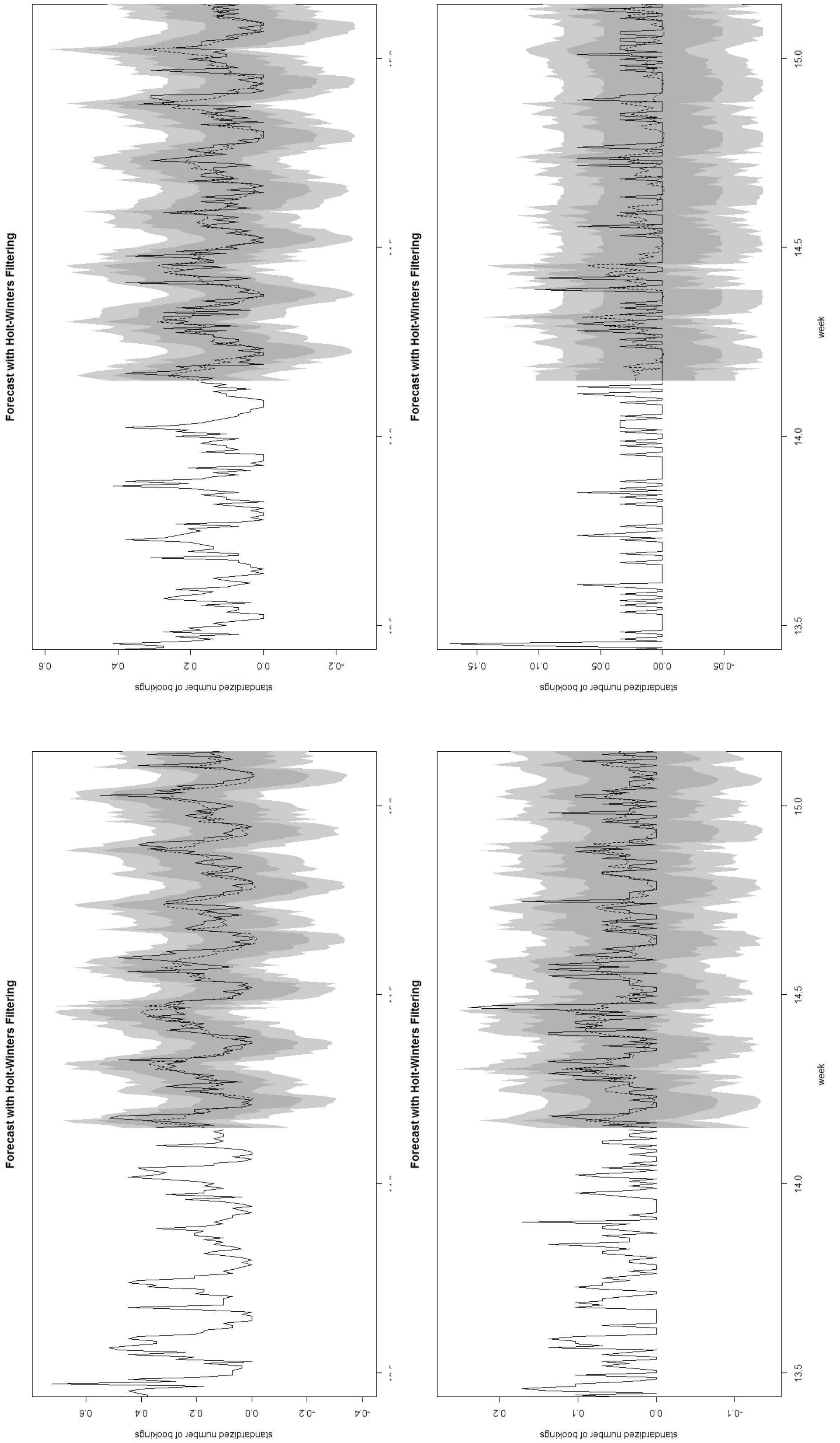


Figure 59: Forecast with HWF for the representatives of Zone A-D in Berlin based on data set III (quarter of a year)

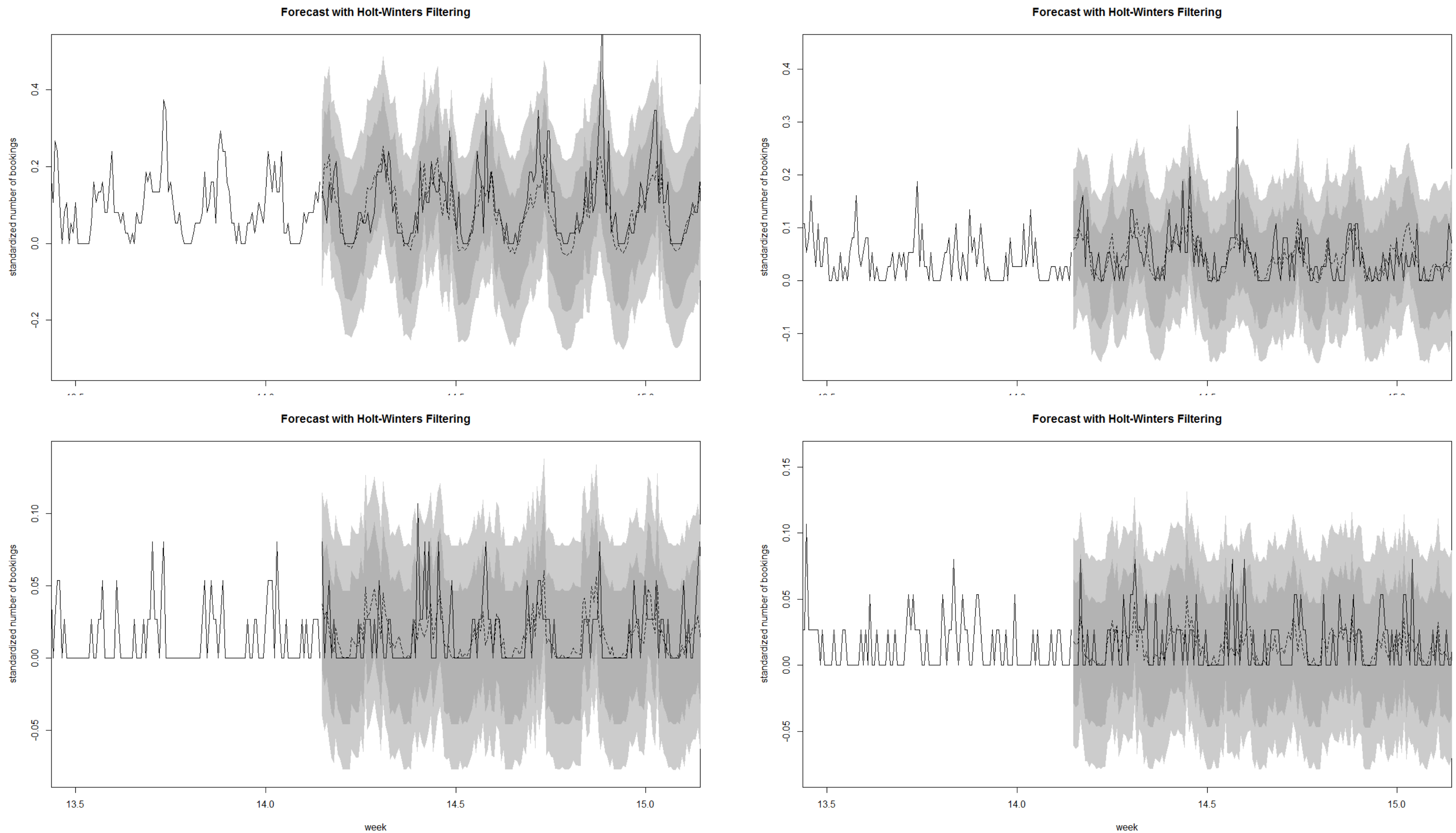


Figure 60: Forecast with HWF for the representatives of Zone A-D in Munich based on data set III (quarter of a year)

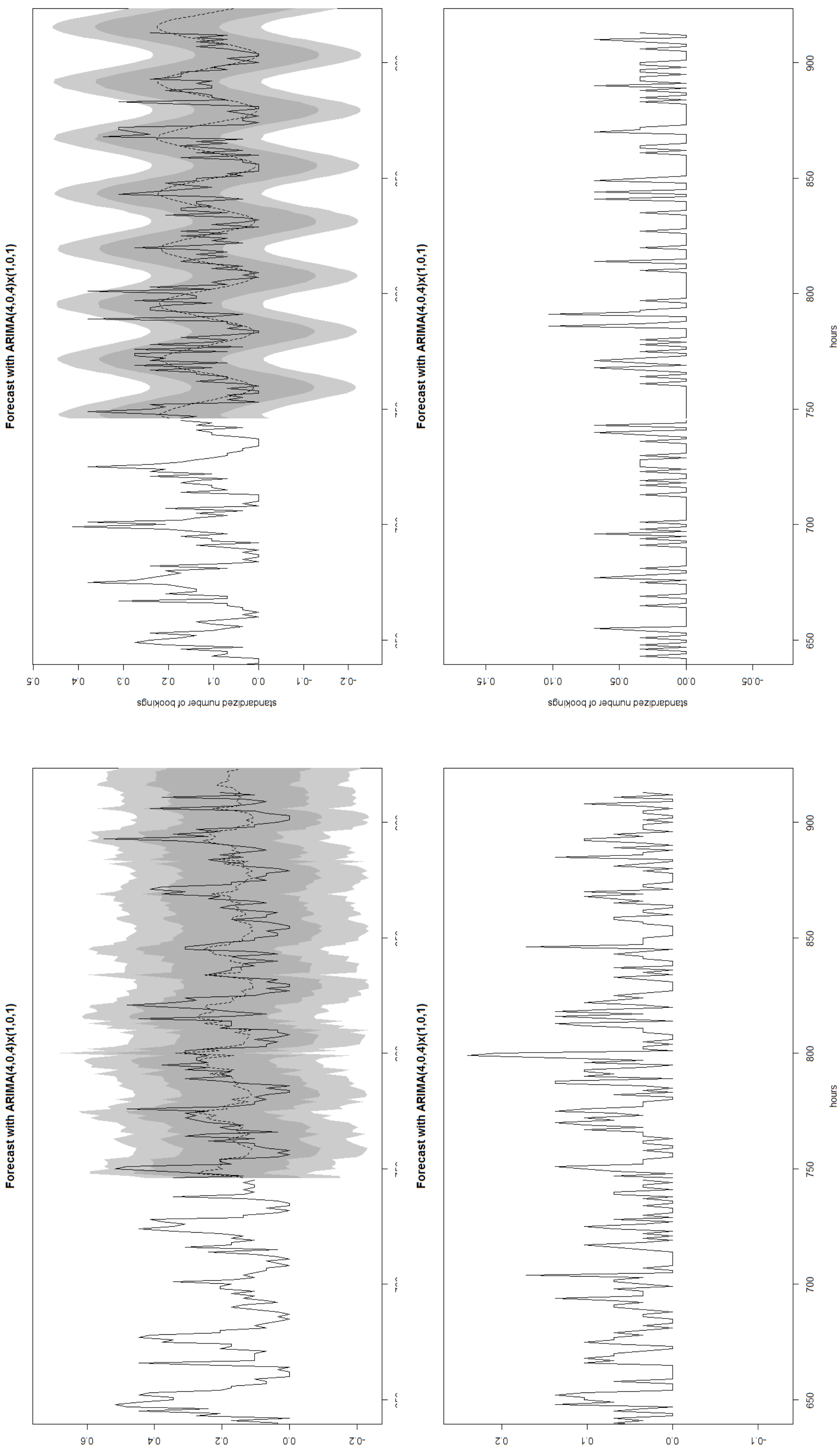


Figure 61: Forecast with ARIMA for the representatives of Zone A-D in Berlin based on data set IV (month)

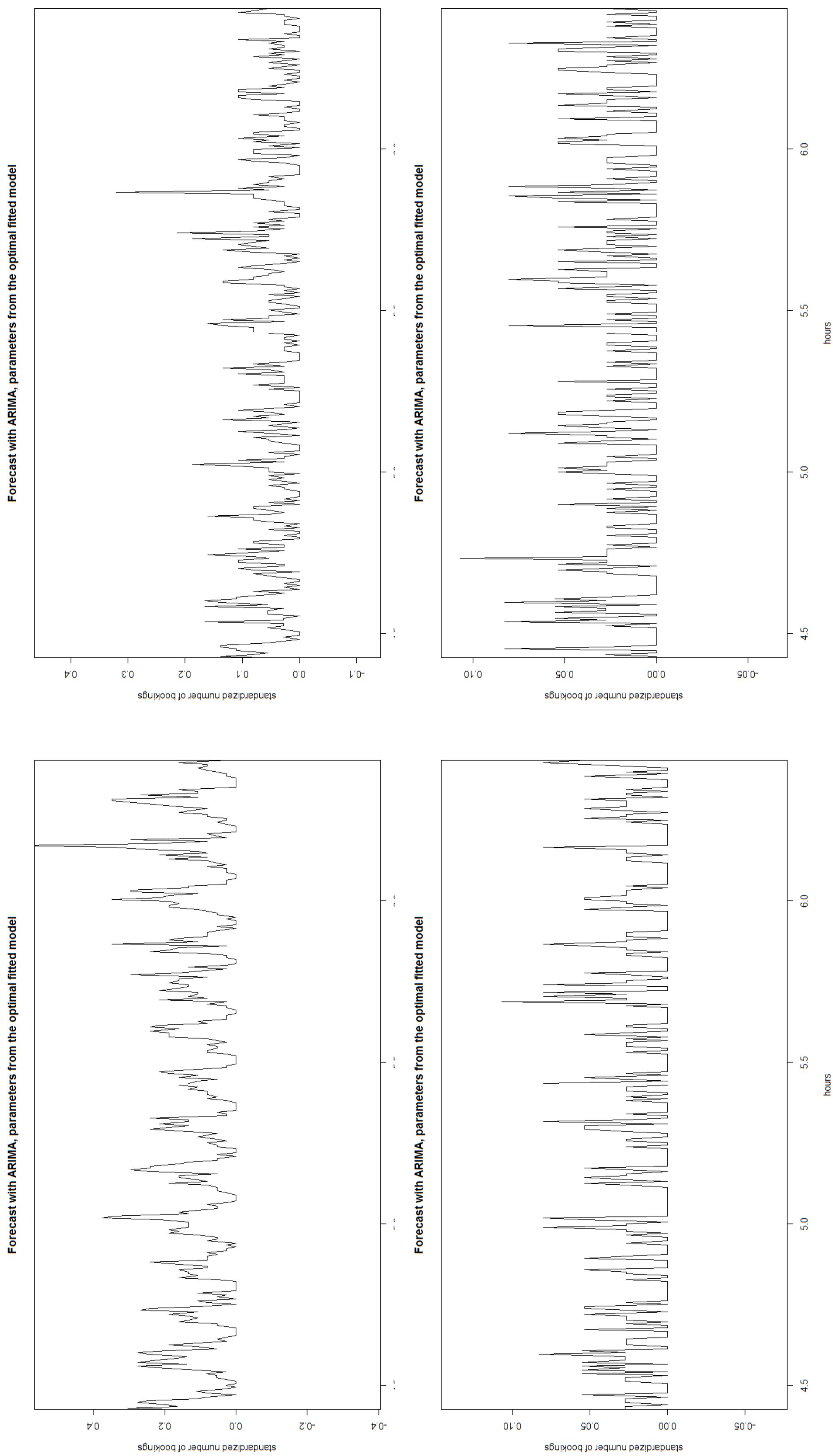


Figure 62: Forecast with ARIMA for the representatives of Zone A-D in Munich based on data set IV (month)

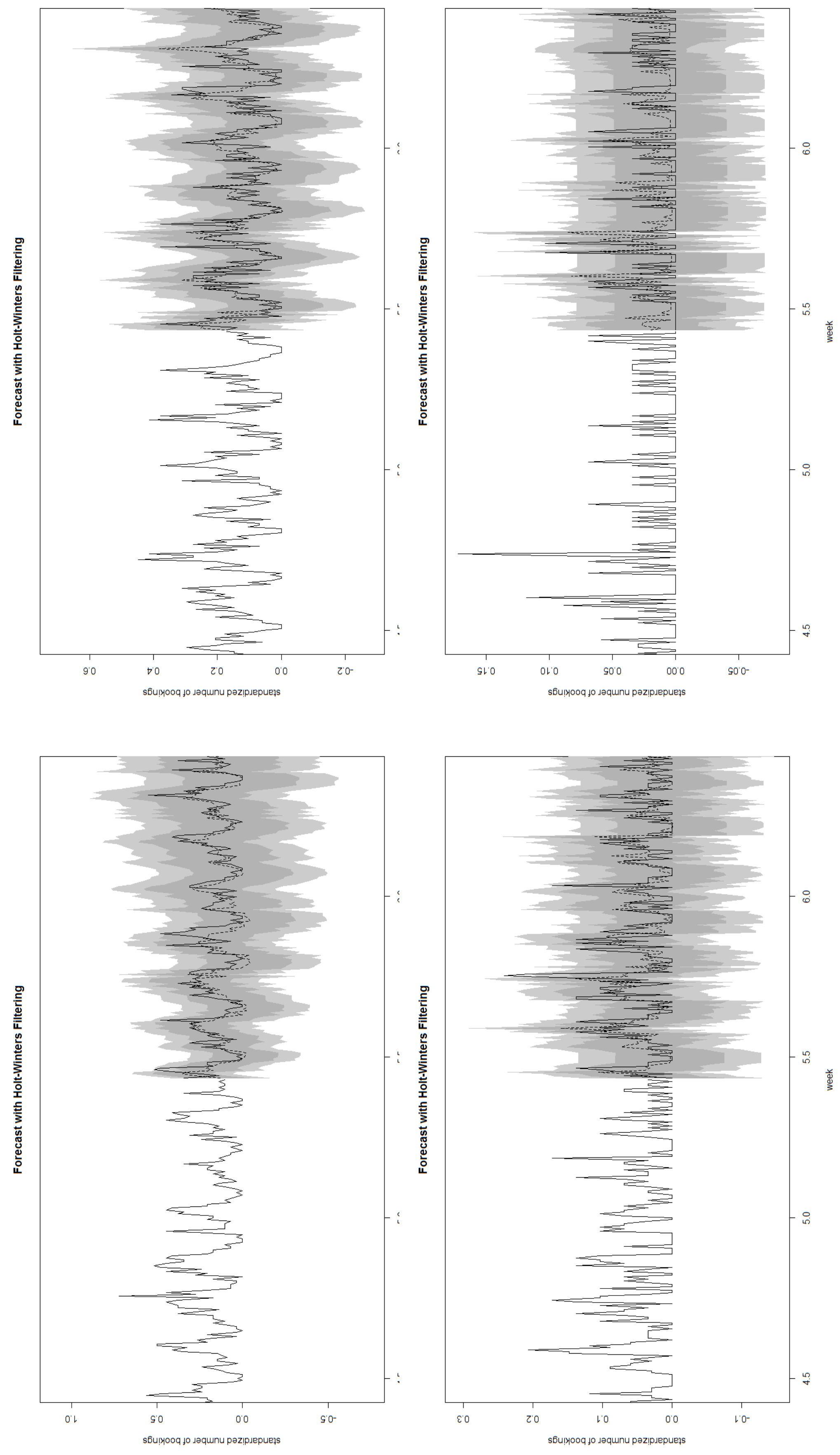


Figure 63: Forecast with HWF for the representatives of Zone A-D in Berlin based on data set IV (month)

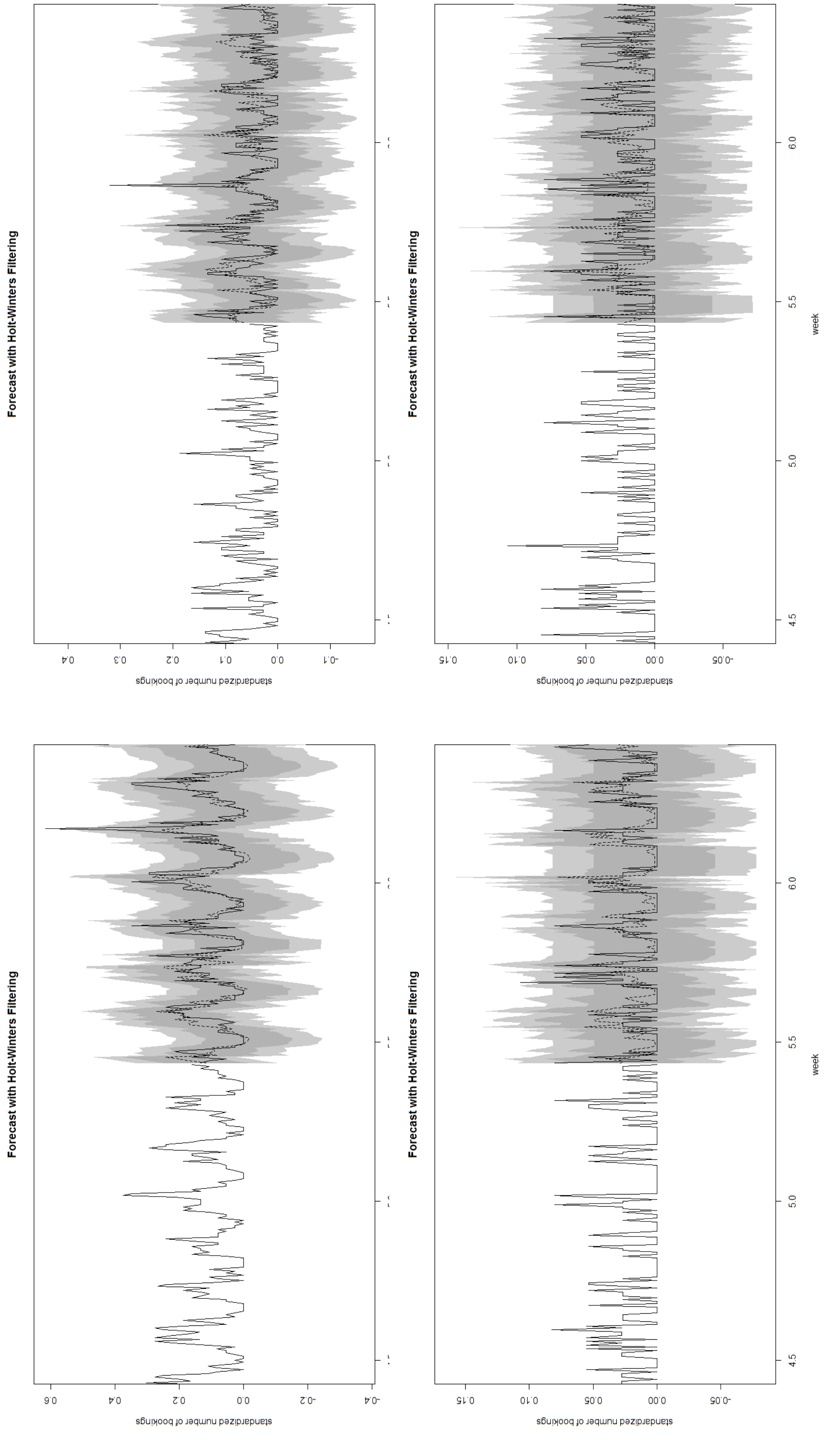


Figure 64: Forecast with HWF for the representatives of Zone A-D in Munich based on data set IV (month)

LIST OF FIGURES

LIST OF TABLES

BIBLIOGRAPHY

[1] Amt für Statistik Berlin-Brandenburg: *Bevölkerungsstand in Berlin am 31. August 2013 nach Bezirken*. https://www.statistik-berlin-brandenburg.de/Publikationen/OTab/2014/OT_A01-10-00_124_201308_BE.pdf. Aug 2013. – retrieved on 12/06/2015

[2] Amt für Statistik Berlin-Brandenburg: *Bevölkerungsstand in Berlin und im Land Brandenburg am 30. November 2014*. https://www.statistik-berlin-brandenburg.de/publikationen/otab/2015/OT_A01-14-00_124_201411_BBB.pdf. 2015. – retrieved on 12/06/2015

[3] Amt für Statistik Berlin-Brandenburg: *Katasterflächen nach Art der tatsächlichen Nutzung am 31.12.2014*. https://www.statistik-berlin-brandenburg.de/BasisZeitreiheGrafik/Bas-Flaechennutzung.asp?Ptyp=300&Sageb=33000&creg=BBB&anzwer=5. 2015. – retrieved on 12/06/2015

[4] ArcGIS Help 10.1: *What they don't tell you about regression analysis*. http://resources.arcgis.com/en/help/main/10.1/index.html#//005p00000053000000. 2011. – retrieved on 12/08/14

[5] Bach, S. ; Grabka, M. M.: Parteianhänger: Wohlhabende neigen zu Union und FDP - und zu den Grünen. In: *DIW-Wochenbericht* 80 (2013), 11-18. http://hdl.handle.net/10419/81553

[6] Balac, M. ; Ciari, F. ; Axhausen, K. W.: Carsharing Demand Estimation: Zurich, Switzerland, Area Case Study. In: *Transportation Research Record: Journal of the Transportation Research Board* 2536 (2015), p. 10–18. – DOI 10.3141/2536–02

[7] Barth, M. ; Shaheen, S. : Shared-Use Vehicle Systems: Framework for Classifying Carsharing, Station Cars and Combined Approaches. In: *Transportation Research Board: Journal of the Transportation Research Board* 1791 (2002), p. 105–112

[8] Barth, M. ; Todd, M. : Simulation model performance analysis of a multiple station shared vehicle system. In: *Transportation Research Board: Journal of the Transportation Research Board* 7 (1999), p. 237–259

[9] Baureferat der Landeshauptstadt München: *Münchner U-Bahn-Bau im Wandel der Zeit*. http://www.muenchen.de/rathaus/Stadtverwaltung/baureferat/ubahn-bau.html. 2014. – retrieved on 12/06/2015

[10] Bayerisches Landesamt für Statistik: *Bevölkerung: Gemeinden*. https://www.statistikdaten.bayern.de/genesis/online?sequenz=tabelleErgebnis&selectionname=12411-001. 2013. – retrieved on 12/06/2015

[11] Beck, M. : *Collinearity and stepwise VIF selection*. https://beckmw.wordpress.com/2013/02/05/collinearity-and-stepwise-vif-selection/. 2013. – retrieved on 12/06/2015

[12] Berlin cambio: *cambioJournal, Ausgabe 31*. http://www.cambio-carsharing.de/cms/downloads/31c8b394-ce98-49b7-8b3b-3cc587d277fd/cambioJournal_31_Nov15_A4.pdf. Nov 2015

[13] Berliner Verkehrsbetriebe: *Zahlenspiegel 2014*. http://unternehmen.bvg.de/images/content/unternehmen/medien/1801663BroschuerenZahlenspiegel2014derBVG.pdf. 2014. – retrieved on 12/06/2015

[14] BMUB ; Umweltbundesamt: *Fact Sheet zur Umweltbewusstseinsstudie 2014*. http://www.bmub.bund.de/fileadmin/Daten_BMU/Download_PDF/Forschung/umweltbewusstsein_2014_factsheet_bf.pdf. 2015. – retrieved on 12/08/2015

[15] BMW AG: *Abschlusspräsentation zum Forschungsprojekt WiMobil*. http://www.erneuerbar-mobil.de/de/projekte/foerderung-von-vorhaben-im-bereich-der-elektromobilitaet-ab-2012/ermittlung-der-umwelt-und-klimafaktoren-der-elektromobilitaet/20151014_WiMobil_Abschluss.pdf. 2015. – retrieved on 09/11/2014

[16] Borgnat, P. ; Abry, P. ; Flandrin, P. ; Robardet, C. ; Rouquier, J. B. ; Fleury, E. : Shared bicycles in a city: a signal processing and data analysis perspective. In: *Advances in Complex Systems* 14 (2011), p. 415–438

[17] Box, G. E. P. ; Jenkins, G. M. ; Reinsel, G. C.: *Time Series Analysis - Forecasting and Control*. 4th Edition. John Wiley & Sons, Inc., 2008

[18] Breusch, T. S. ; Pagan, A. R.: A simple test for heteroscedasticity and random coefficient variation. In: *Econometrica: Journal of the Econometric Society* 47 (1979), p. 1287–1294

[19] Brockmeyer, F. ; Frohwerk, S. ; Weigele, S. : *Free-Floating Carsharing: Urbane Mobilität im Umbruch*. http://matters.civity.de/. 2014. – retrieved on 31/03/15

[20] Bundesagentur für Arbeit: *Statistik der Bundesagentur für Arbeit - Statistik nach Regionen*. https://statistik.arbeitsagentur.de/Navigation/Statistik/Statistik-nach-Regionen/Statistik-nach-Regionen-Nav.html. Oct 2015. – retrieved on 11/04/2015

[21] Bundesanzeiger: *Bekanntmachung der Technischen Richtlinie zum Gesetz über die geodätischen Referenzsysteme, -netze und geotopographischen Referenzdaten des Bundes (Technische Richtlnie Bundesgeoreferenzdatengesetz - TR BGeoRG)*. https://www.bmi.bund.de/SharedDocs/Downloads/DE/Gesetzestexte/tr_bgeorg.pdf?__blob=publicationFile. 2012. – retrieved on 11/11/2015

[22] Bundesministerium für Umwelt, Naturschutz, Bau und Reaktorsicherheit: *WiMobil - Projekthomepage*. http://www.erneuerbar-mobil.de/de/projekte/foerderung-von-vorhaben-im-bereich-der-elektromobilitaet-ab-2012/ermittlung-der-umwelt-und-klimafaktoren-der-elektromobilitaet/wimobil. 2015. – retrieved on 12/

[23] Bundesministerium für Verkehr und digitale Infrastruktur: *Straßenverkehrsordnung*. http://www.stvo.de/strassenverkehrsordnung. Mar 2013. – retrieved on 11/15/2015

[24] CAR2GO: *Meine Stadt, meine Freiheit*. www.car2go.com. 2015. – retrieved on 12/11/2015

[25] CAR2GO BLACK: *car2go black - Die neue Art Ihr Ziel zu erreichen*. https://www.car2goblack.com/de/DE/. 2015. – retrieved on 12/06/2015

[26] CARSHARING-NEWS.DE: *Stadtrat München*. http://www.carsharing-news.de/stadtrat-muenchen-carsharing/. Dec 2015. – retrieved on 01/11/2016

[27] CELSOR, C. ; MILLARD-BALL, A. : Where Does Carsharing Work? Using GIS to Assess Market Potential. In: *Transportation Research Record: Journal of the Transportation Research Board* 1992 (2007), p. 61–69

[28] CERVERO, R. : San Francisco City CarShare: First-Year Travel Demand Impacts. In: *Transportation Research Board 82nd Annual Meeting* (2003)

[29] CERVERO, R. : San Francisco City CarShare: Second-Year Travel Demand and Car Ownership Impacts. In: *Transportation Research Board 83rd Annual Meeting* (2004)

[30] CIARI, F. ; BOCK, B. ; BALMER, M. : Modeling station-based and free-floating carsharing demand: A test case study for Berlin, Germany. In: *Transportation Research Record: Journal of the Transportation Research Board* 2416 (2014), 37-47. http://www.innoz.de/fileadmin/INNOZ/pdf/In_den_Medien/TRBBENNO.pdf

[31] CIARI, F. ; SCHUESSLER, N. ; AXHAUSEN, K. W.: Estimation of Carsharing Demand Using an Activity-Based Microsimulation Approach: Model Discussion and Some Results. In: *International Journal of Sustainable Transportation* 7 (2013), p. 70–84. – DOI 10.1080/15568318.2012.660113

[32] CITEECAR BLOG: *Wichtige Mitteilung der Insolvenzverwaltung an alle Kunden*. http://blog.citeecar.com/insolvenzantrag/. 2015. – retrieved on 01/06/2016

[33] CITROËN DEUTSCHLAND GMBH: *Citroën Multicity Carsharing Berlin*. https://www.multicity-carsharing.de/. 2015. – retrieved on 12/11/2015

[34] CORREIA, G. ; ANTUNES, A. P.: Optimization approach to depot location and trip selection in one-way carsharing systems. In: *Transportation Research Part E: Logistics and Transportation Review* 48 (2012), Nr. 1, 233-247. – DOI 10.1016/j.tre.2011.06.003. – ISSN 1366–5545. – Select Papers from the 19th International Symposium on Transportation and Traffic Theory

[35] CORREIA, G. ; JORGE, D. R. ; ANTUNES, D. M.: The Added Value of Accounting For Users' Flexibility and Information on the Potential of a Station-Based One-Way Car-Sharing System: An Application in Lisbon, Portugal. In: *Journal of Intelligent Transportation Systems* 18 (2014), Nr. 3, p. 299–308

[36] CÍZEK, P. ; HÄRDLE, W. ; WERON, R. : *Statistical Tools for Finance and Insurance*. http://sfb649.wiwi.hu-berlin.de/fedc_homepage/xplore/tutorials/. 2005. – retrieved on 12/08/2015

[37] DB RENT GMBH: *Flinkster - Mein Carsharing*. https://www.flinkster.de/index.php. 2015. – retrieved on 12/11/2015

[38] Deutsche Bahn AG: *Presseinformation: 10 Jahre Carsharing der Bahn - eine Erfolgsgeschichte*. https://www.flinkster.de/fileadmin/www.flinkster.de/redaktion/images/PDF/Presseinfo_Berlin_10_Jahre_Carsharing_01.pdf. 2011. – retrieved on 12/11/2015

[39] Deutsche Bahn AG: *250000 Kunden bei "Flinkster - Mein Carsharing"*. http://www.deutschebahn.com/de/konzern/im_blickpunkt/4869360/flinkster.html. Oct 2013. – retrieved on 11/06/2015

[40] Deutsche Bahn AG: *Die S-Bahn Berlin*. http://www.s-bahn-berlin.de/unternehmen/firmenprofil/kurzfassung.htm. Nov 2015. – retieved on 12/06/2015

[41] Deutsche Bahn AG: *Zahlen, Daten und Fakten über die S-Bahn München*. http://www.s-bahn-muenchen.de/s_muenchen/view/wir/daten_fakten.shtml. 2015. – retrieved on 12/06/2015

[42] Deutscher Wetterdienst: *Erläuterungen zur Beaufort-Skala*. http://www.dwd.de/bvbw/appmanager/bvbw/dwdwwwDesktop?_nfpb=true&_pageLabel=dwdwww_menu2_wetterlexikon&_nfls=false. 2009. – retrieved on 07/08/2014

[43] Deutscher Wetterdienst: *Formulierung zu den Wetterelementen "Temperatur", "Niederschlag", "Bewölkung" in den Wetterberichten*. http://www.deutscher-wetterdienst.de/lexikon/download.php?file=Wetterelemente_Formulierungen.pdf. 2009. – retrieved on 07/08/2014

[44] DriveNow: *DriveNow International*. https://de.drive-now.com/#!/carsharing/international. – retrieved on 12/11/2015

[45] DriveNow: *Blog von DriveNow*. http://blog.drive-now.de/. 2011. – retrieved on 08/20/14

[46] DriveNow: *DriveNow Carsharing Berlin*. https://de.drive-now.com/#!/carsharing/berlin. Dec 2015. – retrieved on 12/03/2015

[47] DriveNow: *DriveNow Carsharing München*. https://de.drive-now.com/#!/carsharing/muenchen. Dec 2015. – retrieved on 12/03/2015

[48] DriveNow: *DriveNow Tarife und Pakete*. https://de.drive-now.com/#!/tarife. Nov 2015. – retrieved on 11/05/2015

[49] Düwel, J. et al.: *1945, Krieg - Zerstörung - Aufbau, Architektur und Stadtplanung 1940-1960*. Henschel Verlag, Berlin, 1995

[50] Ehrenpfordt, R. : *Systematische Fehler von Wahlprognosen*. Seminary paper. http://www.grin.com/de/e-book/84278/systematische-fehler-von-wahlprognosen. 2007. – retrieved on 12/22/2015

[51] Eichhorst, W. ; Spermann, A. : Sharing Economy - Chancen, Risiken und Gestaltungsoptionen für den Arbeitsmarkt. In: *IZA Research Report* 69 (2015). http://www.iza.org/en/webcontent/publications/reports/report_pdfs/iza_report_69.pdf. – retrieved on 01/11/2016

[52] Eugster, J. : *Einfluss des Wetters auf das Verkehrsverhalten*, ETH Zürich, Master thesis, 2001

[53] Fahrmeir, L. ; Kneib, T. ; Lang, S. ; H. Dette, W. Härdle (ed.): *Regression*. Springer Verlag Berlin, 2009

[54] FairVote2K: *Gerry's Gallery*. http://www.westmiller.com/fairvote2k/IN_GERRY.HTM. 2006. – retrieved on 12/18/2015

[55] Fehndrich, M. : *Wahlkreisgeometrie - Gerrymandering*. http://www.wahlrecht.de/lexikon/gerrymander.html. Jul 2006. – retrieved on 12/18/2015

[56] Firnkorn, J. ; Müller, M. : What will be the environmental effects of new free-floating car-sharing systems? The case of car2go in Ulm. In: *Ecological Economics* 70 (2011), p. 1519–1528

[57] Fliegner, S. ; Werner, M. (ed.): *Car Sharing als Alternative? Mobilitätsbasierte Potenziale zur Autoabschaffung*. MetaGIS Fachbuch, 2002

[58] Florax, R. ; Nijkamp, P. : Misspecification in Linear Spatial Regression Models. In: *Encyclopedia of Social Measurement* (2004)

[59] Flughafen München GmbH: *Parking at Munich Airport: for a relaxed journey*. http://www.munich-airport.com/media/download/general/publikationen/en/parken_en.pdf. Feb 2015. – retrieved on 11/19/2015

[60] Forbes Magazine: *The World's Biggest Public Companies*. http://www.forbes.com/global2000/list/. 2011. – retrieved on 11/19/2015

[61] Froehlich, J. ; Neumann, J. ; Oliver, N. : Sensing and Predicting the Pulse of the City through Shared Bicycling. In: Boutilier, C. (ed.): *IJCAI*, 1420-1426

[62] Gallop, C. ; Tse, C. ; Zhao, J. : Seasonal Autoregressive Model of Vancouver Bicycle Traffic Using Weather Variables. In: *Transportation Research Board 91st Annual Meeting* (2012)

[63] Getis, A. ; Ord, J. K.: The analysis of spatial association by use of distance statistics. In: *Geogr. Anal.* 24 (1992)

[64] GfK Geomarketing GmbH: *Pressemitteilung 15.12.2014: Kaufkraft der Deutschen steigt 2015 moderat*. http://www.gfk-geomarketing.de/fileadmin/newsletter/pressemitteilung/PM_kaufkraft-DE-2015.html. Dec 2014. – retrieved on 11/04/2015

[65] Glotz-Richter, M. : *Car-Sharing in Bremen*. http://www.cities-for-mobility.net/documents/wc11/michael_glotz_richter.pdf. 2011. – retrieved on 12/04/2015

[66] Goodwin, T. : *The Battle Is For The Customer Interface*. http://techcrunch.com/2015/03/03/in-the-age-of-disintermediation-the-battle-is-all-for-the-customer-interface/#.qmud3w:0sCd. Mar 2015. – retrieved on 11/11/2015

[67] Hanitzsch, A. ; Hauenstein, A. ; Korda, J. ; Grötsch, M. : Parkraummanagement in München - Planung und Umsetzung. In: *Straßenverkehrstechnik* 8 (2012), p. 481–488

[68] Herminghaus, H. : *CO2-Emissionen eines PKW*. http://www.co2-emissionen-vergleichen.de/verkehr/PKW/CO2-Emissionen-PKW.html. Oct 2011. – retrieved on 12/06/2015

[69] Hinkeldein, D. ; Schoenduwe, R. ; Graff, A. ; Hoffmann, C. : Who Would Use Integrated Sustainable Mobility Services - And Why? In: *Sustainable Urban Transport* 7 (2015), p. 177–203

[70] Holt, C. C.: *Forecasting seasonals and trends by exponentially weighted moving averages*. Carnegie Institute of Technology, Pittsburgh, Pennsylvania, 1957

[71] Hyndman, R. J. ; Athanasopoulos, G. : *Forecasting principles and practice*. OTexts https://www.otexts.org/fpp/

[72] infas Institut für angewandte Sozialwissenschaften GmbH: *Mobilität in Deutschland (MiD) - Alltagsverkehr in München, im Münchner Umland und im MVV Verbundraum*. http://www.mvv-muenchen.de/fileadmin/media/download/downloadbereich/Publikationen/documents/INFAS_Kurzbericht_MiD_20101124.pdf. 2008. – retrieved on 12/06/2015

[73] infas Institut für angewandte Sozialwissenschaft GmbH: *Mobilität im Wandel - Potenziale des Carsharing*. Luxembourg, 2012

[74] Jones, A. : *Lecture Notes - The ARIMA Procedure*. http://www.okstate.edu/sas/v8/saspdf/ets/chap7.pdf. 2012. – retrieved on 07/29/14

[75] Jorge, D. ; Correia, G. : Carsharing systems demand estimation and defined operations: a literature review. In: *European Journal of Transport and Infrastructure Research* Issue 13(3) (2013), p. 201–220

[76] Jorge, D. ; Correia, G. ; Barnhart, C. : Testing the Validity of the MIP Approach for Locating Carsharing Stations in One-way Systems. In: *Procedia - Social and Behavioral Sciences* 54 (2012), 138-148. – DOI 10.1016/j.sbspro.2012.09.733. – ISSN 1877–0428. – Proceedings of EWGT 2012 - 15th Meeting of the EURO Working Group on Transportation, September 2012, Paris

[77] Kaltenbrunner, A. ; Meza, R. ; Grivolla, J. ; Codina, J. ; Banchs, R. : Urban cycles and mobility patterns Exploring and predicting trends in a bicycle-based public transport system. In: *Pervasive and Mobile Computing* Vol. 6 (2009), p. 455–466

[78] Kawgan-Kagan, I. : Early adopters of carsharing with and without BEVs with respects to gender preferences. In: *European Transport Research Review* 7(33) (2015), p. 1–11. – DOI 10.1007/s12544–015–0183–3

[79] Kek, A. G. H. ; Cheu, R. L. ; Chor, M. L.: Relocation Simulation Model for Multiple-Station Shared-Use Vehicle Systems. In: *Transportation Research Board 85th Annual Meeting* (2005)

[80] Kek, A. G. H. ; Cheu, R. L. ; Xu, J. : Trip Forecasting Models for Intelligent Community Vehicle Systems. In: *Transportation Research Board 84th Annual Meeting* (2005)

[81] Kek, A. G. H. ; Cheu, R. L. C. ; Meng, Q. ; Fung, C. H.: A decision support system for vehicle relocation operations in carsharing systems. In: *Transportation Research Part E: Logistics and Transportation Review* 45 (2009), p. 149–158

[82] Kempen, E. E. M. M. ; Kruize, H. ; Boshuizen, H. C. ; Ameling, C. B. ; Staatsen, B. A. M. ; Hollander, A. E. M.: The association between noise exposure and blood pressure and ischemic heart disease. In: *Environ Health Perspect.* 110(3) (2002), p. 307–317

[83] Knoflacher, H. : *Virus Auto - Die Geschichte einer Zerstörung*. Verlag Carl Ueberreuter, Wien, 2009

[84] Knoflacher, H. : *Zurück zur Mobilität - Anstöße zum Umdenken*. Verlag Carl Ueberreuter, Wien, 2013

[85] Kopp, J. : *Carsharing - Ein neuer Element multimodaler Verkehrsangebote*. http://www.vt.bgu.tum.de/fileadmin/w00bnf/www/VKA/2013_14/140113_VerkehrAktuell_Multimodalitaet_Kopp.pdf. 2013. – retrieved on 12/10/2015

[86] Kopp, J. P.: *GPS-gestützte Evaluation des Mobilitätsverhaltens von free-floating CarSharing-Nutzern*, ETH Zürich, PhD thesis, 2015. – DOI 10.3929/ethz–a–010453113

[87] Kortum, K. ; Machemehl, R. : *Free-floating Carsharing Systems: Innovations in membership prediction, mode share, and vehicle allocation optimization methodologies*, Center for Transportation Research, University of Texas at Austin, PhD thesis, 2012

[88] Kraftfahrtbundesamt: *Erteilungen von Fahrerlaubnissen nach Führerscheinstellen*. www.kba.de/SharedDocs/Publikationen/DE/Statistik/Kraftfahrer/FE/2013/fe5_2013_xls.xls?__blob=publicationFile&v=5. 2013. – retrieved on 04/14/15

[89] Kraftfahrtbundesamt: *Bestand an Kraftfahrzeugen und Kraftfahrzeuganhängern nach Gemeinden*. http://www.kba.de/SharedDocs/Publikationen/DE/Statistik/Fahrzeuge/FZ/2014/fz1_2014_pdf.pdf?__blob=publicationFile&v=8. Jan 2014. – retrieved on 11/09/2015

[90] Kumar, V. ; Bierlaire, M. : Optimizing Locations for a Vehicle Sharing System. (2012)

[91] Kwiatkowski, D. ; Phillips, P. ; Schmidt, P. ; Shin, Y. : Testing the null hypothesis of stationarity against the alternative of a unit root. In: *Journal of Econometrics* 54 (1992), p. 159–178

[92] Landeshauptstadt München: *München in Zahlen*. http://www.muenchen.de/sehenswuerdigkeiten/muenchen-in-zahlen.html. 2015. – retrieved on 12/06/2015

[93] Landeshauptstadt München: *Allgemeine Informationen des Kreisverwaltungsreferats zum Thema Parken*. http://www.muenchen.de/rathaus/Stadtverwaltung/Kreisverwaltungsreferat/Verkehr/Parkraummanagement/Allgemeine-Informationen.html. 2013. – retrieved on 08/19/14

[94] Lenz, B. ; Bogenberger, K. : *Wirkungen von E-Carsharing-Systemen auf Mobilität und Umwelt in urbanen Räumen - Halbzeitkonferenz*. http://www.erneuerbar-mobil.de/de/events/pdf-dateien/folien-wimobil.pdf. 2014. – retrieved on 09/11/2015

[95] Loose, W. ; Mohr, M. ; Nobis, C. ; Holm, B. ; Bake, D. : *Bestandsaufnahme und Möglichkeiten der Weiterentwicklung von Carsharing*. http://www.carsharing.de/images/stories/pdf_dateien/wp2_endbericht_deutsch_final_4.pdf. 2004. – retrieved on 04/15/15

[96] Markenartikel Magazin ; Focus: *Durchschnittsalter von Neuwagenkäufern in Deutschland von 1995 bis zum Jahr 2015 (in Jahren)*. Statista - Das Statistik-Portal. http://de.statista.com/statistik/daten/studie/215576/umfrage/durchschnittsalter-von-neuwagenkaeufern/. 2015. – retrieved on 11/05/2015

[97] Martin, E. ; Shaheen, S. : The Impact of Carsharing on Public Transport and Non-Motorized Travel: An Exploration of North American Carsharing Survey Data. In: *Energies* 4 (2011), p. 2094–2114

[98] Mayer, D. : *R Documentation: decompose - Classical Seasonal Decomposition by Moving Averages*. http://stat.ethz.ch/R-manual/R-devel/library/stats/html/decompose.html. 2009. – retrieved on 07/31/14

[99] Mayer, D. : *R Documentation: Holt-Winters Filtering*. http://stat.ethz.ch/R-manual/R-patched/library/stats/html/HoltWinters.html. 2012. – retrieved on 07/14/14

[100] McFadden, D. : Conditional Logit Analysis of Qualitative Choice Behavior. In: *Frontiers of Econometrics* (1973), p. 105–142

[101] Mendes Lopes, M. ; Martínez, L. M. ; A. Correia, G. H. ; Moura, F. : Insights into Carsharing Demand Dynamics: Outputs of an Agent-Based Model Application to Lisbon, Portugal. In: *Transportation Research Board 94th Annual Meeting* (2015)

[102] Millard-Ball, A. ; Murray, G. ; Schure, J. ter ; Fox, C. : *TRCP Report 108 - Carsharing: Where and How It Succeeds*. 2005

[103] Müller, J. ; Bogenberger, K. : Time Series Analysis of Booking Data of a Free-Floating Carsharing System in Berlin. In: *Transportion Research Procedia* 10 (2015), 345-354. – DOI 10.1016/j.trpro.2015.09.084

[104] Müller, J. ; Schmöller, S. ; Bogenberger, K. : Empirische Datenanalyse von Free-Floating Car-Sharing-Systemen. In: *Straßenverkehrstechnik* 2 (2015), p. 75–80

[105] Müller, J. ; Schmöller, S. ; Giesel, F. : Identifying Users and Use of (Electric) Free-Floating Carsharing in Berlin and Munich. In: *Proceedings to the IEEE ITSC Conference 2015* (2015)

[106] Morency, C. ; Habib, K. ; Grasset, V. ; Islam, M. : Understanding members' carsharing (activity) persistency by using economic model. In: *Journal of advanced Transportation* 46 (2012), p. 26–38

[107] Morgenstern, V. ; Zutavern, A. ; Cyrys, J. ; Brockow, I. ; Koletzko, S. ; Krämer, U. ; Behrendt, H. ; Herbarth, O. ; Berg, A. von ; Bauer, C. P. ; Wichmann, H.-E. ; Heinrich, J. : Atopic Diseases, Allergic Sensitization and Exposure to Traffic-related Air Pollution in Children. In: *American Journal of Respiratory and Critical Care Medicine* 177(12) (2008), 1331-1337. – DOI 10.1164/rccm.200701–036OC

[108] Morrison, G. ; Pike, D. H.: Kalman Filtering to Statistical Forecasting. In: *Management Science* 23 (1977), p. 768–774

[109] nexiga GmbH: *Datenkatalog Marktinformationen - Geomarketing nach Maß*. http://www.nexiga.com/cms/wp-content/uploads/2014/12/Nexiga_Marktdatenkatalog_web.pdf. 2014. – retrieved on 07/28/2015

[110] Niedermayer, O. et al. ; Niedermayer, O. (ed.): *Handbuch Parteienforschung*. Springer VS http://link.springer.com/book/10.1007/978-3-531-18932-1

[111] Niehues, M. : *Analyse von Free-Floating Car-Sharing in Bereichen außerhalb der Kerngeschäftsgebiete*, University of Federal Armed Forces Munich, Master thesis, 2015

[112] O'Brien, R. : A Caution Regarding Rules of Thumb for Variance Inflation Factors. In: *Quality & Quantity* 41(5) (2007), p. 673–690

[113] Ord, J. K. ; Getis, A. : Local spatial autocorrelation statistics: distributional issues and an application. In: *Geogr. Anal.* 4 (1995)

[114] Parikh, P. ; Ukkusuri, S. V.: Estimation of Optimal Inventory Levels at Stations of a Bicycle Sharing System. In: *Transportation Research Board 94th Annual Meeting* (2015)

[115] Petersen, M. ; Petersen, M. (ed.): *Ökonomische Analyse des Car-Sharing*. Springer, 1995

[116] Prettenthaler, F. E. ; Steininger, K. W.: From ownership to service use lifestyle: the potential of carsharing. In: *Ecological Economics* 28 (1999), p. 443–453

[117] Randeloff, M. : *Europäische Städte mit der niedrigsten Durchschnittsgeschwindigkeit im Straßenverkehr (2008)*. http://www.zukunft-mobilitaet.net/9995/analyse/durchschnittsgeschwindigkeit-europa-2008-berlin-deutschland/. Aug 2012. – retrieved on 12/06/2015

[118] Rat für nachhaltige Entwicklung: *Nutzt Carsharing? Warum zwei Studien zu unterschiedlichen Ergebnissen kommen*. http://www.nachhaltigkeitsrat.de/news-nachhaltigkeit/2015/2015-11-05/nutzt-carsharing-warum-zwei-studien-zu-unterschiedlichen-ergebnissen-kommen/?blstr=0. Nov 2015. – retrieved on 12/06/2015

[119] Regue, R. ; Recker, W. : Using Gradient Boosting Machines to Predict Bikesharing Station States. In: *Transportation Research Board 93rd Annual Meeting* (2014)

[120] Schlesinger, C. : DriveNow vs. Car2Go - BMW hat bei Carsharing die Nase vorne. In: *Wirtschaftswoche* (2014). http://www.wiwo.de/unternehmen/auto/drivenow-vs-car2go-bmw-hat-bei-carsharing-die-nase-vorne/10335938.html. – retrieved on 08/24/2014

[121] Schlittgen, R. : *Angewandte Zeitreihenanalyse mit R*. 2nd. Oldenbourg Verlag, München, 2012

[122] Schlittgen, R. ; Streitberg, B. H. J.: *Zeitreihenanalyse*. 9th. Oldenbourg Wissenschaftsverlag GmbH, 2001

[123] Schmöller, S. ; Weikl, S. ; Müller, J. ; Bogenberger, K. : Empirical analysis of free-floating carsharing usage: The Munich and Berlin case. In: *Transportation Research Part C: Emerging Technologies* 56 (2015), p. 34–51. – DOI 10.1016/j.trc.2015.03.008

[124] Seign, R. : *Model-based design of free-floating carsharing systems*, University of Federal Armed Forces, PhD thesis, 2015

[125] Senatsverwaltung für Stadtentwicklung und Umwelt Berlin: *"Mobilität in Städten - SrV 2013" - Neue Mobilitätsdaten für Berlin und seine Bezirke*. http://www.stadtentwicklung.berlin.de/verkehr/politik_planung/zahlen_fakten/mobilitaet_2013/. 2013

[126] Senatsverwaltung für Stadtentwicklung und Umwelt Berlin: *Parkraumbewirtschaftung*. http://www.stadtentwicklung.berlin.de/verkehr/politik_planung/strassen_kfz/parkraum/. 2014. – retrieved on 08/20/14

[127] Shaheen, S. ; Sperling, D. ; Wagner, C. : Carsharing in Europe and North America: Past, Present, and Future. (2001)

[128] Stadtmobil Berlin GmbH: *CarSharing mit stadtmobil*. http://berlin.stadtmobil.de/privatkunden/fuhrpark/. 2015. – retrieved on 12/11/2015

[129] Statista GmbH: *Bestand an Personenkraftwagen in Berlin von 2008 bis 2015*. http://de.statista.com/statistik/daten/studie/255179/umfrage/bestand-an-pkw-in-berlin/. 2015. – retrieved on 12/06/2015

[130] Statistisches Amt München: *Statistisches Taschenbuch 2013*. www.muenchen.de/rathaus/dms/Home/Stadtinfos/Statistik/Taschenbuch/Stat_Taschenbuch_2013.pdf. 2013. – retrieved on 11/20/2015

[131] Statistisches Amt München: *Statistisches Taschenbuch 2014*. www.muenchen.de/rathaus/dms/Home/Stadtinfos/Statistik/Taschenbuch/Stat_Taschenbuch_2013.pdf. 2014. – retrieved on 12/06/2015

[132] Stattauto München: *Umweltschonende Mobilität - CarSharing für München seit 1992*. https://www.stattauto-muenchen.de/. 2015. – retrieved on 12/11/2015

[133] Stillwater, T. ; Mokhtarian, P. ; Shaheen, S. : Carsharing and the built environment: A GIS-based study of one U.S. operator. In: *Transportation Research Board 88th Annual Meeting* (2009)

[134] team red: *Endbericht Evaluation Carsharing (EVA-CS)*. http://www.ris-muenchen.de/RII/RII/DOK/SITZUNGSVORLAGE/3885730.pdf. Dec 2015. – retrieved on 01/11/2016

[135] Thomsen, L. : *Video of Lars Thomsen's talk on the 26th international conference "Motor und Umwelt" about the future of electromobility*. https://www.youtube.com/watch?v=JHUzfw24oCk. Sep 2012. – retrieved on 11/04/2015

[136] Umweltbundesamt: *Feinstaub (PM10) im Jahr 2014*. https://www.umweltbundesamt.de/sites/default/files/medien/378/dokumente/pm10_2014_0.pdf. Jan 2015. – retrieved on 11/05/2015

[137] Umweltbundesamt: *Presseinfo Nr. 14 vom 31.03.2015*. http://www.umweltbundesamt.de/sites/default/files/medien/381/dokumente/pi_2015_31_03_uba-emissionsdaten_2014_zeigen_trendwende_beim_klimaschutz.pdf. Mar 2015. – retrieved on 11/05/2015

[138] Wagner, M. : *Pressemitteilung Bundesverband Carsharing: Jahresbilanz 2012: So viel Carsharing wie noch nie*. http://www.carsharing.de/presse/pressemitteilungen/bundesverband-carsharing-jahresbilanz-2012-so-viel-carsharing-zuwachs-wie. Feb 2013. – retrieved on 11/09/2015

[139] Wagner, S. ; Brandt, T. ; Neumann, D. : Data Analytics in Free-Floating Carsharing: Evidence from the City of Berlin. In: *Proceedings of the 48th Annual Hawaii International Conference on System Sciences* (2015), p. 897–907

[140] Weikl, S. ; Bogenberger, K. : A practice-ready relocation model for free-floating carsharing systems with electric vehicles - Mesoscopic approach and field trial results. In: *Transportation Research Part C: Emerging Technologies* 57 (2015), p. 206–223

[141] Weikl, S. ; Bogenberger, K. : An Integrated Relocation Model for Free-Floating Carsharing Systems: Field Trial Results. In: *Transportation Research Record: Journal of the Transportation Research Board* 2536 (2015), p. 19–27

[142] Weikl, S. ; Bogenberger, K. : Nachfragephänomene von Free Floating Carsharing-Systemen - Räumlich-Zeitliche Angebots-Nachfrage-Asymmetrie. In: *Straßenverkehrstechnik* 59/3 (2015), p. 156–163

[143] Winters, P. R.: Forecasting Sales by Exponentially Weighted Moving Averages. In: *Graduate School of Industrial Administration* (1960), p. 324–342

[144] Zeileis, A. ; Kleiber, C. ; Jackman, S. : *Regression Models for Count Data in R*. Apr 2007. http://epub.wu.ac.at/1168/ retrieved on 12/06/2015

DECLARATION

This dissertation has been submitted for the degree of *Doktor der Ingenieurwissenschaften*. I hereby declare that I am the sole author of this thesis.
I have fully acknowledged and referenced the ideas and work of others, whether published or unpublished, in my thesis.
I have prepared my dissertation specifically for the degree of *Doktor der Ingenieurwissenschaften*, while under supervision at the Universität der Bundeswehr München.
My Ph.D. thesis does not contain work extracted from a thesis, dissertation or research paper previously presented for another degree or diploma at this or any other university.

Munich, February 14, 2016

Johannes Müller

COLOPHON

This dissertation is written in LaTeX using `classicthesis`. The author likes to thank André Miede for developing this appealing template and making it openly available.

This dissertation was submitted on February 27, 2016 and electronically published in September 2016.

Final Version as of September 18, 2016 (`classicthesis`).