MASTER

Fleet size decision under stochastic customer locations and forecasted demand at a transportation company

Pennartz, T.

Award date:
2020

Department of Industrial Engineering and Innovation Sciences

Fleet Size Decision under Stochastic Customer Locations and Forecasted Demand at a Transportation Company

Master's Thesis

T. (Thijs) Pennartz

BSc in Industrial Engineering - TU/e 2018

Student Identity Number: 0850936

In partial fulfillment of the requirements for the degree of Master of Science in Operations Management and Logistics

Supervisors:
dr. A.M. Florio (TU/e supervisor)
dr. V.J.C. Lurkin (TU/e supervisor)
dr. J. Kinable (TU/e supervisor)
A.P.N. Schuffelers (Company supervisor)

Eindhoven, 5th of May 2020

Eindhoven University of Technology
School of Industrial Engineering
Series Master Thesis Operations Management and Logistics

Keywords: fleet size decision, forecasting, time series analysis, scenario generation, resource requirements

Abstract

In order to prevent capacity shortages, many organizations rely on forecasting approaches. By providing an accurate forecast, these organizations can anticipate demand fluctuations and thus reduce overage and shortage costs. This master's thesis has been conducted at a transportation company to develop a quantitative framework to support the company in forecasting customer demand and in defining the number of required vehicles to serve its customers. The company's fleet size decision process is currently implemented in a qualitative and empirical fashion, where the number of required vehicles to serve customers is estimated based on the experience of the company's planners. Vehicle hiring costs start increasing two days before the delivery date and the timing of the fleet size decision has a major impact on the company's total routing costs. Five different time series forecasting methods based on daily order data have been implemented and compared. Besides, a neural network has been implemented to predict demand levels based on the order arrival process. The company's problem can be seen as a stochastic vehicle routing problem, where customer locations and demands are uncertain from day to day. Scenario generation has been applied to describe the uncertainty surrounding customers and to estimate the probability distribution of the number of required vehicles. The proposed framework provides a systematic method to determine the number of required vehicles from day to day.

Executive Summary

This master's thesis has been conducted at a transportation company located in Tilburg, the Netherlands. The company offers a wide range of logistics services to its customers in the Benelux such as multimodal transportation and warehousing. This research was aimed to determine the number of required vehicles to fulfill all customer demand per day at the DIS TIL planning department, which is one of the company's planning departments. Currently, the DIS TIL planning department determines the required fleet size based on the experience of the planners. Since vehicle hiring costs increase as the delivery day approaches, an accurate prediction of the fleet size is required to reduce operational expenses. Besides, orders that could not have been fulfilled due to vehicle shortages also lead to increased opportunity costs (lost sales) and customer dissatisfaction. The company is considering whether forecasting methods can help the planning department in making the fleet size decision with greater precision. Therefore, the following main research question has been proposed:

How to accurately predict the number of required vehicles, and, based on an estimate for the number of required vehicles, how many vehicles should be hired?

To provide an answer to this main research question, the research has been divided into several steps. First, a literature study has been conducted on forecasting methods and the fleet size decision. Thereafter, several factors that affect the fleet size decision have been analyzed such as the effect of customer time windows and customer locations on daily routing operations. Next, five time series methods have been implemented to generate a total demand forecast per day. Out of the implemented time series forecasting methods, the TBATS method showed the best forecast performance on the company's data set. Averaged over all test sets, the TBATS forecast had a Mean Absolute Error (MAE) of 74.56 loading meters and a Mean Absolute Percentage Error (MAPE) of 16.70%. Besides, the TBATS method is able to handle multiple seasonal cycles that may change over time, which makes this method robust and applicable for the long term in the context of the company.

The next step is to transform the demand forecast into the number of required vehicles to fulfill all customer demand. Solving a Vehicle Routing Problem (VRP) reveals the number of required vehicles. The company aims to establish the number of required vehicles for the short term while not all required data to make this decision is known beforehand. This problem is known as the Stochastic Vehicle Routing Problem (SVRP) and in the company's context, customer locations and demands are uncertain variables. Scenario generation can be seen as a means to provide an answer to the SVRP. Various different scenarios per weekday for one week (i.e. five weekdays) have been created and solved with the company's routing tool. The scenarios are based on forecasted demand and disaggregation rules. These rules are a way to generate customer locations and to allocate demand accordingly. Simulation has been

executed with the company's routing tool, considering five different demand levels and three unique sets of customer locations for each weekday. Since the majority of the forecast errors follow a normal distribution, a probability can be assigned to each scenario. These probabilities have been transformed into an empirical Cumulative Distribution Function (CDF). For each weekday under consideration, the number of required vehicles can be determined based on the empirical CDF function and the amount of risk the company wants to accept.

In addition, the company's routing tool has been utilized to solve the daily vehicle routing for October and November 2019. The goal was to gain insights on the interactions between the number of required vehicles, customer demand levels, number of customer locations, and the vehicle capacity utilization rate. For this analysis the routing results of the generated scenarios were also included. Based on these results, the vehicle utilization rate is stable and robust under various demand levels and customer location sets. Moreover, the average vehicle utilization rate for October and November 2019 are 92.28% and 93.01%, respectively. Since nearly all vehicles drive one route a day, the planning department can, as a rule of thumb, transform the demand forecast into the number of required vehicles by dividing the total demand forecast for a specific day by the capacity of one vehicle (corrected for the average vehicle utilization rate).

The majority of the customer orders arrive the day before the delivery date. To investigate whether an accurate demand prediction can be made based on the order arrival process, a neural network has been created. The neural network was able to predict the total demand level at 12pm based on the total demand level at 12am and calendar effects with a MAE of 34.59 and a MAPE of 8%. Although the vehicle hiring price increases significantly the day before the delivery date, the neural network can aid the planning department to signal a shortage or surplus of required vehicles at an earlier stage during the day. Currently, the planning department subjectively predicts the final demand level around 3pm.

To conclude, the company should adopt the proposed forecasting methods to make the forecasting procedure more systematic. Moreover, the fleet size decision can be made several days in advance and a surplus or shortage of vehicles can be signalled at an earlier stage. Determining the required fleet size at an earlier stage is expected to result in less vehicle shortages and reduced hiring costs. Besides, the planning department is less dependent on whether experienced planners are in office; the implemented forecasting methods are systematic and always applicable.

The recommendation to the company is to continue acquiring high quality order data. The implemented forecasting methods may show better performance with more representative daily order data that follows the trend from November 2018. Unfortunately, the accuracy of the DIS TIL planning department's vehicle predictions over time could not have been investigated since the frequency and magnitude of a surplus or shortage of vehicles is poorly recorded. Besides, a financial comparison between the number of vehicles that should have been deployed with the number of vehicles that the planning department actually deployed could not have been made. Whereas the vehicle underage costs are known, another recommendation to the company concerns establishing the vehicle's overage costs. These numbers are expected to reveal the financial impact of having a surplus or shortage of vehicles. The company is advised to acquire this data to complete this analysis. The last recommendation concerns the utilization of supply chain information. The company's current forecasting methods are based on local information. By using forecast information of parties in the supply chain, the company's forecasting accuracy is expected to improve.

Preface

This report marks the end of my master Operations Management and Logistics at Eindhoven University of Technology, and also the end of my life as a student. I would like to thank everyone who has been involved in my project, and several people in particular. First of all, I would like to express my gratitude to my first TU/e supervisor, Alexandre Florio. Alexandre, thank you for your help during the process of my master's thesis project. The conversations we had always helped me in understanding and solving the thesis' problem. You motivated me to think critically and I really appreciate the time you made for meetings. Besides, I would like to thank Virginie Lurkin, my second TU/e supervisor. I experienced your expertise and feedback as highly valuable for my master thesis project. Furthermore, I would like to thank Joris Kinable, my third TU/e supervisor. During my master thesis project, you moved to the United States of America. Despite the time zone difference, I would like to thank you for making time for me. I really appreciate you had a critical eye on my project and I learned a lot from you.

Moreover, I would like to thank several people at the company that supervised me during this project. First of all, I would like to thank my company supervisor, Armand Schuffelers. Armand, you gave me the freedom to execute my research while ensuring practical relevance. I really enjoyed the discussions we had about my project and the challenges logistics service providers currently face. Besides, I would like to thank Pim for his feedback and involvement in my project.

Last but not least, I would like to express my gratitude towards my family and friends. Mom and Dad, thank you for providing me the opportunity to study. I am happy to have experienced your unconditional support during my study. Furthermore, I would to thank the many friends I made during my time in Eindhoven.

Thijs Pennartz, April 2020

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Company Description

This master's thesis project has been conducted at a logistics service provider located in Tilburg, the Netherlands. The company is a multimodal logistics service provider that transports goods by road, water, and rail. Moreover, the company offers warehousing services to its customers. The company owns four distribution centers in the Netherlands and one distribution center in Belgium. Customers demanding a service are primarily located in the Benelux, known as the Netherlands, Belgium, and Luxembourg. The majority of the company's logistics activities take place in the Netherlands. The company's vehicle fleet consists of 375 pulling units and 725 pulled units. Each day, these vehicles visit around 5,000 customer locations.

1.2 Problem Statement

The problem of the company concerns the determination of the required resources in terms of required vehicles to perform deliveries for the next day. In 2019, around 500 orders, related to the DIS TIL planning department, have been recorded that could not be fulfilled due to a shortage in the number of vehicles. In practice, the number of unfulfilled orders is even larger since unfulfilled orders are not systematically recorded. Each order that cannot be fulfilled can be estimated having a financial cost of 50 Euros. The total financial costs of unfulfilled orders for the DIS TIL planning department due to a shortage in the number of vehicles is substantial. Moreover, a surplus of vehicles also leads to an increase in costs.

Each day, the planning department receives customer orders. The majority of the customer orders arrive the day before the delivery date. Order arrival times are highly variable during the day, even for existing customers that order frequently. Besides, customer demand quantities show high variation over time. The company deploys a fleet of company vehicles and a fleet of charter vehicles to serve its customers. Each day, extra vehicles can be hired or a surplus of vehicles can be sold in the market based on the required total vehicle capacity to fulfill customer orders. The planning department recognizes a shortage or surplus of vehicles around 3pm each day, based on the order arrival process and the planners' experience. The timing of the fleet size decision is of high importance, since hiring extra vehicles becomes more expensive as the day progresses. The opposite holds for selling surplus vehicles, where the revenue of selling shipments in the market decreases as the day progresses. Besides, a shortage of vehicles and thus unfulfilled demand may lead to customer dissatisfaction. A shortage or surplus of a few vehicles for the next day can be managed by contractual agreements with common carriers. A shortage or surplus of at least ten vehicles forces the planning department to hire or sell vehicles for high or low prices, respectively. Under vehicle shortage conditions, the planning department aims to fulfill all customer orders and consequently hires expensive extra vehicles.

The planning department currently determines the number of required vehicles based on the planners' experience, where the total amount of loading meters to be transported is a large determinant factor. A loading meter corresponds to one linear meter of loading space in a truck. A loading meter is a universal unit of measurement for freight that cannot be stacked, or when stacking on top of these goods is forbidden. Most of the order sizes range from zero up to four loading meters and order sizes of ten loading meters or more occur occasionally (when not considering full-truckload shipments) (Appendix A). At the end of the day, a shortage or surplus of vehicles may come to light after the planning department has created the final route planning.

Two days before the delivery date, vehicle hiring costs start increasing. The later the decision to hire extra vehicles is made, the more expensive these vehicles become. Therefore, the cheapest moment to hire extra vehicles is two days before the delivery date. However, the majority of the customer orders arrive one day before the delivery date. Since vehicle hiring costs increase significantly during the day before the delivery date, the company desires to establish the number of required vehicles two days before the order delivery date.

The company's problem shows similarities with the newsvendor problem, which is a mathematical model to determine the optimal inventory level (Axsäter, 2015). A decision maker, who faces uncertain demand, has to decide for a single period how many products to order to maximize its profits (Petruzzi & Dada, 1999). The order is placed before the period starts. At the end of the single period, the product perishes and can no longer be sold. There is a cost for ordering too few products and a cost for ordering too many products. Balancing the prices of underage costs (i.e. the cost per unit for unsatisfied demand) and overage costs (i.e. the cost per unit of excess inventory) is the basis to solving the newsvendor problem (Petruzzi & Dada, 1999). The company can reveal its underage costs by analyzing the vehicle hiring costs. In contrast, the overage costs are more difficult to estimate since surplus vehicles can either be sold in the market to common carriers or allocated to the company's other planning departments. In the latter case, the overage costs are even more difficult to estimate since vehicle re-allocations are poorly recorded.

The routing of vehicles impacts the fleet size decision at the company's planning departments. The following factors have the largest impact on vehicle routing activities and consequently the total number of required vehicles to perform customer deliveries. The first factor concerns geographical restrictions. Some customers cannot be served by certain truck types due to truck size or tail-lift restrictions. For example, truck-trailer combinations cannot enter the canals district in the city center of Amsterdam due to the trailer's size. Consequently, box trucks have to be deployed to serve these customers. Moreover, environmental zones may impede specific truck types to enter specific regions. The second factor concerns customer time windows. Customers prefer to be served during specific time windows, or are subject to given time windows due to geographical restrictions. Customers pay a fee to be served during a specific time window, where a small time window is more expensive compared to a large time window. Time windows can be either soft or hard, depending on agreements with the customer. Not complying to a hard time window may result in lost orders. The third factor concerns driver restrictions. The company has to comply to driver legislation. For example, the duration of any route is not allowed to exceed a work shift duration and truck drivers have to take breaks during their work shift. These breaks may complicate vehicle routing since trucks cannot be operated continuously. For large distance trips, the company overcomes this problem by assigning two truck drivers to one truck. Besides the factors that affect vehicle routing, the required number of vehicles to serve customer locations is also dependent on the timing of demand during the day. Most customers have a delivery time window. When demand is equally spread over the day, less vehicles are required to serve all customers compared to when most of the customer demand has to be delivered simultaneously.

1.3 Scope

The company provides logistics services to its customers in the Benelux. Since the thesis has been conducted at the main headquarters in Tilburg, this research focused on the distribution activities associated with this distribution center. For the Tilburg location, three main planning departments can be distinguished. The DIS TIL planning department transports products from the Tilburg distribution center to customers located in the Benelux Union. The order sizes range up to 12 loading meters and is mainly loaded onto pallets. Order sizes of at least 12 loading meters are considered as full-truckload (FTL) and are delivered by the FTL TIL planning department. The FTL TIL planning department manages the FTL shipments loaded in Tilburg. FTL shipments are dedicated to one customer only. Vehicles that ship FTL containers are able to drive multiple routes a day. FTL shipments are easier to sell or buy in the market since the transportation requires less handling costs. Also, FTL shipments are less encumbered by weight and size restrictions compared to DIS TIL shipments. Distribution of relatively small order sizes from the Tilburg distribution center to customer locations in the Benelux is handled by the DISK TIL planning department.

The DIS TIL planning department's delivery activities and its homogeneous vehicle fleet are the main focus of this thesis. The planning department's pick-up activities are out of scope. Each day, the DIS TIL planning department requires on average 50 tractor-trailer combinations and visits around 230 customer locations. Currently, nearly all trucks of the DIS TIL planning department complete one route per day.

1.4 Research Objective

In order to prevent capacity shortages, many organizations rely on forecasting approaches. By providing an adequate forecast, these organizations can anticipate demand fluctuations and thus reduce overage and shortage costs. The number of required vehicles to satisfy customer demand depends on many factors, such as the customer location, customer time windows, and customer demand quantities. Therefore, it is interesting to investigate the effect of these factors on the number of required vehicles to satisfy customer demand.

The number of required vehicles can be found by solving a Vehicle Routing Problem (VRP) (Toth & Vigo, 2002). An algorithm can generate an optimal routing plan that satisfies several constraints. A VRP requires data concerning customer locations and customer demand quantities. To estimate future customer demand, a demand forecast has to be generated. Besides, a set of customer locations needs to be generated that may be representative for the future. The objective of this research is to develop a quantitative framework to support the company in forecasting customer demand and in defining the required number of vehicles to serve its customers. This process is currently implemented in a qualitative and empirical fashion, where the number of required vehicles to serve these customers is estimated by the experience of the company's planners. Moreover, the objective of this research is not to build and solve a VRP to optimality, but to connect the domains of forecasting and vehicle routing to reveal the required resources in terms of vehicles to fulfill customer demand.

1.5 Research Questions

Since customer locations, order arrival times, and order sizes vary from day to day, the decision to hire or sell vehicles for the next day is of high importance for the planning department. The company would like to investigate whether forecasting methods can be helpful in the determination of the number of required vehicles for the short term. The fleet size decision is expected to become more efficient if the frequency and magnitude of a surplus or shortage of vehicles is reduced. In order to establish the number of required vehicles for the short term, the following main research question is proposed:

How to accurately predict the number of required vehicles, and, based on an estimate for the number of required vehicles, how many vehicles should be hired?

Currently, the planning department decides to hire or sell vehicles for the next day on a daily basis. Suppose the decision to hire extra vehicles can be established three days in advance, the vehicle hiring costs can be reduced. The following sub questions have been formulated to provide an answer to the main research question. The first sub research question relates to company's current performance on determining the total number of required vehicles for the short term. By analyzing the company's number of hired or sold vehicles from day to day, the company's current performance on its fleet size decision can be determined. Therefore, the following sub research question has been formulated:

RQ1: What is the company's current performance on their vehicle fleet size decision and what factors affect this decision?

A forecast of demand can provide input for the vehicle fleet size decision for the short term (Tsekeris & Tsekeris, 2011). Literature has shown that the performance of forecasting methods depends on the underlying data set (Shumway & Stoffer, 2017). Therefore, the prediction accuracy of multiple forecasting techniques has to be compared. The forecasting technique providing the highest prediction accuracy is of interest. Since the performance of several forecasting techniques has to be compared, the following sub research question aims to establish the best forecasting technique:

RQ2: Which demand forecasting method provides the highest prediction accuracy?

Lastly, a forecast of freight demand has to be transformed into the number of required vehicles. Therefore, the following sub research question has been proposed:

RQ3: How to transform the daily demand forecast into the total daily number of required vehicles that is able to fulfill customer demand?

The last sub research question is of high importance since it aims to transform a forecast of demand into the number of required vehicles and thus provides an answer to the main research question.

1.6 Thesis Outline

The outline of this report is as follows. A review on literature concerning the company's problem is presented in Chapter 2. In Chapter 3, an analysis of the company's data is given where the factors that affect the company's current forecasting process and vehicle routing policy are elaborated. The process of finding the forecasting method that is able to forecast freight demand with the highest prediction accuracy is given in Chapter 4. In Chapter 5, an approach to transform a demand forecast into the number of required vehicles to fulfill customer demand is presented. Finally, the thesis's conclusion, a discussion of the study's limitations, directions for further research, and recommendations to the company are provided in Chapter 6.

Chapter 2

Literature Review

The literature review in this chapter is a condensed version of the literature review written for the course 1ML05.

Forecasting of Demand

Time series analysis refers to the methods available to gain insights in data characteristics and statistics. It has been a widely used method to analyse freight demand and subsequently produce a demand forecast (Mrowczynska, Łachacz, Haniszewski & Sładkowski, 2012). Time series analysis methods require univariate or multivariate data. A univariate time series refers to an observation (e.g. total number of loading meters in the context of transportation) recorded sequentially in time increments of equal size. Multivariate time series consist of multiple time-dependent variables. Quantitative methods require historical data to reveal data patterns and to establish a forecast (Hart, Kubíková & Lukoszová, 2013). The company possesses time series data of customer demand (in loading meters) per day. Time series analysis methods can therefore be used to forecast future demand quantity per unit time for the company. In literature, several different time series methods have been studied and applied to forecast (freight) demand.

The paper that is most relevant to the company's problem setting is written by Zhou, Heimann and Clausen (2006). The authors conducted research on forecasting demand for the short term for a less-than-truckload (LTL) service provider. The study's company and problem environment shows many similarities with the company's case. Similar to the paper, the company's less-than-truckload activities are under consideration. The authors state that, due to LTL freight characteristics, accurate forecasting is important for effective resource allocation. Whereas the study uses monthly demand data, the company possesses daily demand data. An ARIMA (AutoRegressive Integrated Moving Average) model and neural network model have been utilized for short term demand forecasting. ARIMA has been chosen since it is a flexible method and widely used time series approach. The authors' motivation for a neural network is that it has proven to be useful for forecasting purposes due to the neural network model's learning capabilities and pattern recognition. The authors conclude the study by stating that the neural network provided the most satisfactory results concerning forecasting accuracy. Interestingly, the purpose of the study by Zhou et al. (2006) is not clear; the authors were able to forecast demand but did not utilize this demand forecast for short term or tactical decision making. Although the study examines monthly data, the study's applied forecasting techniques can be used for the company's daily data.

Time series data may contain multiple seasonal cycles (Montgomery, Jennings & Kulahci, 2015). For example, freight demand data may contain seasonality on both the week level and month level. An analysis of historical company data has revealed the existence of seasonal patterns in the data, where even multiple seasonal patterns may exist for specific customers. Therefore, methodologies on time series methods that can handle (multiple) seasonal patterns may be applicable to the company's case. Since the performance of each time series analysis method is dependent on the input data set, multiple time series methods that match the data characteristics should be tested. There is a wide range of literature available on the Holt-Winters exponential smoothing approach and the ARIMA approach, which are two of the most widely used approaches that are able to handle seasonal patterns (Gould et al., 2008). Both methods have been widely studied in general forecasting literature, but limited for freight demand forecasting in specific. Other methods that have been used for forecasting purposes are regression (Billings & Agthe, 1998), SARIMA (Xu, Chan & Zhang, 2019), TBATS (Brożyna, Mentel, Szetela & Strielkowski, 2018), and state-space models (Billings & Agthe, 1998).

Next to forecasting demand, the location of future demand has to be considered to determine the number of required vehicles (Gendreau, Laporte & Séguin, 1996). Some customers order every day, whereas other customer orders have a larger order interval. Since time series analysis methods require data of sequential equal time increments, any interruption in the sequence of data observations may cause problems when forecasting on the customer level (Zotteri, Kalchschmidt & Caniato, 2005). When establishing a forecast for each customer individually, the location of the demand is known. However, when considering a forecast on, for example, the postal code level the distribution of demand among the customers within the postal code region is unknown. Where forecasts based on time series data have been widely studied in literature, the usefulness and applicability of forecasts in practice for fleet size decisions for the short term has been limited. Articles on freight demand forecasting focus on forecasting practices and their performance instead of utilizing the forecast in operational decision making. To make a forecast suitable for operational decision making, a heuristic could be implemented to assign the total demand forecast for all customers combined to specific customer locations (Zotteri et al., 2005).

Time series analysis models require observations recorded sequentially in time increments of equal size. Since forecasting on the customer level has to cope with irregular observations, time series analysis methods are not suitable. A method of interest to tackle this problem is the utilization of a neural network. According to the study by Lee et al. (2018), demand can either be regular or irregular. Furthermore, irregular demand can be split into intermittent and non-intermittent demand, where periods of zero demand alternated with irregular demand are the main characteristics of irregular intermittent demand. For irregular intermittent demand, both the demand quantity and the timing of demand need to be forecasted. This is the main difference between irregular intermittent demand and irregular non-intermittent demand (Lee et al., 2018). Since the company is faced with irregular intermittent demand on the customer level, forecasting intermittent demand is a topic of interest. The study by Lee et al. (2018) aimed to forecast intermittent demand by means of an artificial neural network. The neural network was able to learn and investigate the length of intermittent demand periods as well as the demand quantity of positive demand observations.

Establishing the Number of Required Vehicles

The number of required vehicles to serve customers can be revealed by solving a Vehicle Routing Problem (VRP) as shown by Y. Y. Zhang and Li (2007). The classical VRP aims to find the optimal routing to visit all customers at minimum costs, while satisfying several constraints. It can be described by a directed graph $G(E, V)$ (Montoya-Torres, Franco, Isaza, Jiménez & Herazo-Padilla, 2015). The set of nodes V can be regarded as $V = \{0, 1, ..., n\}$ and E is the set of arcs between nodes. The depot (represented by node $j = 0$) is where all vehicles start and end their route. Nodes $j = 1, 2, ..., n$ represent customer locations. Each customer node has a positive demand, denoted by d_j (Montoya-Torres et al., 2015). Each arc, which is the connection from node i to node j, has an associated cost of c_{ij} (Toth & Vigo, 2002).

Researchers acknowledged practical contexts where several parameters are uncertain for the traditional VRP (Gendreau, Laporte & Seguin, 1996). These practical contexts gave rise to the introduction of the Stochastic Vehicle Routing Problem (SVRP), where, for instance, customer demand, customer locations, and travel times are uncertain (Berhan, Beshah, Kitaw & Abraham, 2014). Every day, the company is faced with variation in freight order sizes and customer locations and the SVRP might be applicable to provide an answer to the company's problem. The study by Gendreau, Laporte and Seguin (1996) modeled stochastic demand and stochastic customer presence by probability functions, where each customer is present with probability p_i and has stochastic demand with probability ξ_i. In practice, it is complex to determine the probability distribution functions. Therefore, the applicability of this paper to the company's problem setting may be limited. Scenario generation (i.e. simulation) could provide a solution to the SVRP (Kall & Wallace, 1994). Simulation is an attempt to model and solve real world problems. By altering the model's parameters, several (realistic) scenarios can be evaluated. Multiple papers have been published on the utilization of simulation in the domain of transportation and logistics. The study by Vonolfen et al. (2010) mentions simulation as a means to optimize VRP scenarios. By applying simulation, the authors tested and evaluated diverse problem environments. Examples of parameters that have been altered in this study are the customer order interval, the customer delivery strategy, the number of customers, and the number of vehicles (Vonolfen et al., 2010). The scenarios are generated by parameterizing a VRP model. Simulation has also been applied by Juan, Faulin, Pérez-Bernabeu and Domínguez (2013) to solve a stochastic VRP and by Fan, Xu and Xu (2009) to solve a VRP with time windows. Besides, Shyshou, Gribkovskaia and Barceló (2010) applied simulation to solve a fleet size problem for offshore mobile units that perform anchor handling operations. By carefully parameterizing the factors affecting the fleet size decision, the authors were able to model realistic scenarios and effectively evaluate the cost-optimal number of required vehicles. Since the fleet size decision at the company is stochastic in nature, simulation might be an adequate manner to determine the number of required vehicles.

Chapter 3

Data Analysis

To determine the number of required vehicles to visit customer locations from day to day, the company data and the factors that affect this decision are elaborated in this chapter.

3.1 Data Description

Historical order data is extracted from the company's data archive and transport management system. The historical order data comprises the following data classes: order number, order index number (indicates whether an order is an initial order or a backorder), customer name, delivery planning department, delivery date, delivery customer time windows, delivery address, and the order size in loading meters.

3.2 Data Preparation

Before the company's data set can be analyzed, data cleaning and an outlier analysis need to be executed. The company's data set includes order data from April 2015 up to December 2019. Only delivery orders are of interest. Also, only initial orders are used and backorders are removed from the data set. Moreover, the data set comprises only weekdays, since the total number of shipped loading meters on Saturdays and Sundays is negligible. Figure 3.1 depicts the total number of shipped loading meters per weekday from April 2015 up to December 2019 of the orders under consideration. As can be seen in this figure, multiple zero demand values are present in the data set. These zero demand values are caused by national holidays such as Easter, Christmas, Pentecost, Ascension Day, and Kingsday. The company does not perform any, or very little, transportation activities on these holidays.

In general, the presence of zero demand values in the data set is problematic for time series model fitting. Therefore, the zero demand values have to be transformed to facilitate time series model fitting. In total, only a few zero demand observations per year caused by holidays are present in the data set. Deterministic imputation is a method where the missing value is imputed by the mean of known values (Nordholt, 1998). According to Nordholt (1998), using the mean of similar observations is expected to provide more accurate results. Therefore, the zero demand values have been imputed by taking the average of the demand at the same weekday one week before and one week after. This approach is expected to have a minimal effect on the demand prediction, since the imputed value is based on the closest similar weekday day observation before and after the zero demand observation. Moreover,

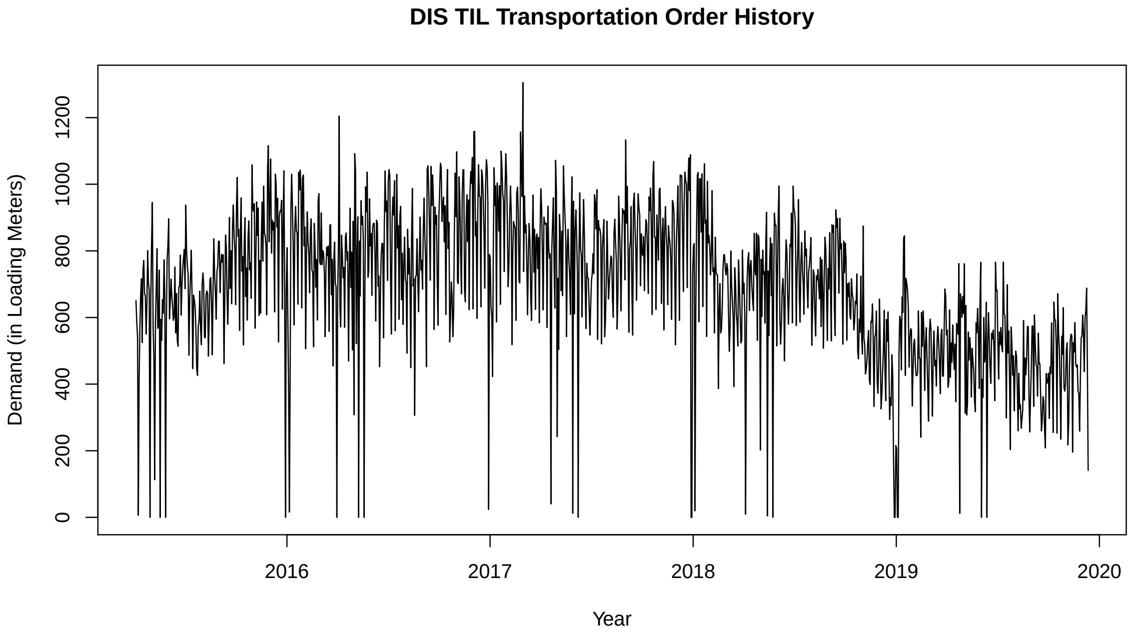

Figure 3.1: DIS TIL transportation order history

this approach ensures that the time series flow is not distorted. Besides, an outlier analysis has been executed to examine the presence of outliers in the data set. Non-systematic and extreme observations in time series data may impede time series model fitting. Therefore, outlier detection is important since outliers may have an effect on time series model selection, parameter setting, and the forecast (C. Chen & Liu, 1993). C. Chen and Liu (1993) describe an iterative procedure to detect outliers in time series data. The procedure starts by finding the most relevant anomaly and determining its effect on the time series' trend by means of multiple regression. Subsequently, this anomaly can be corrected by a replacement value and the model's parameters are estimated again. This process iterates until all anomalies are removed. The outlier analysis by C. Chen and Liu (1993) has been applied to the company's data set with imputed zero demand values. Based on this analysis, outliers are absent in the data set.

Figure 3.2 shows the DIS TIL order history from April 2015 up to December 2019 including imputed zero demand values. As can be seen in this figure, the average number of loading meters per day has decreased from November 2018 onwards, compared to the preceding years. Since November 2018, several DIS TIL logistics activities are handled by the FTL planning department. This shift causes the decrease in total number of shipped loading meters since November 2018 for the DIS TIL planning department. Besides, the DIS TIL planning department's demand is heavily influenced by the DIS TIL planning department's customer base and shipped product categories. Customers change over time, so do the customer demand quantities. Also, existing customers might decide to ship other product categories, which means that these shipments are handled by other planning departments. The DIS TIL order history from April 2015 up to December 2019 for each weekday is presented in Appendix B.

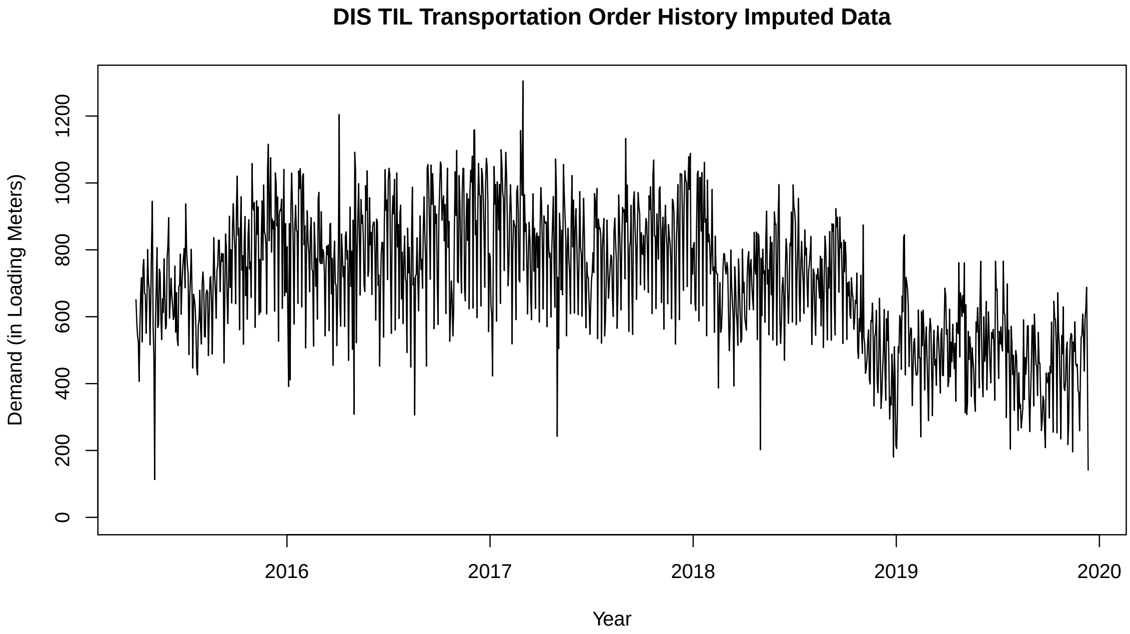

Figure 3.2: DIS TIL transportation order history including imputed data

3.3 Data Analysis

This chapter presents the results of the analysis of the company's data set. The demand per day, seasonal patterns, and the strength of the seasonal patterns have been analyzed. Insights in these factors can be helpful in deciding which forecasting methods to use.

3.3.1 Imbalance in Deliveries

Figure 3.3 shows the total number of shipped loading meters for each weekday of September, October, and November 2019. As can be seen in this figure, the total number of loading meters to be shipped differs for each weekday based on the most recent data. Monday is the weekday with the least customer demand, whereas the most customer demand is on Tuesday. The demand level decreases from Tuesday up to and including Friday. When considering each week individually, the ratio among the demand levels for each weekday is similar as depicted in Figure 3.3. The difference in total customer demand per day implies that the number of required vehicles also fluctuates from day to day. The demand level per weekday over time is highly volatile and is visually presented in Appendix B.

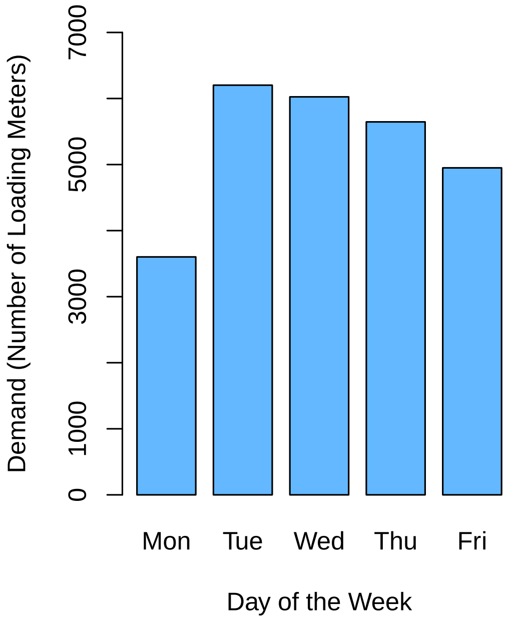

Figure 3.3: Total customer demand in loading meters per weekday for three months

3.3.2 Seasonality

The company's time series data set exhibits the following seasonal cycles: a weekly, monthly, and annual cycle. Since the data set exhibits multiple seasonal patterns, the data set can be decomposed to reveal each seasonal pattern. According to Hyndman and Athanasopoulos (2018), a time series data set can be decomposed into a trend component, a seasonal component, and a remainder component and can be written as

$$y_t = S_t + T_t + R_t,$$

where y_t represents the time series data, S_t denotes the seasonal component, T_t is the trend component, and R_t represents the remainder component. Cleveland, Cleveland, McRae and Terpenning (1990) have developed a method for time series decomposition using Loess, which is a method that applies a filtering procedure to reveal the trend component, remainder component, and seasonal components. Figure 3.4 depicts the original data set, trend component, seasonal components, and remainder component of the DIS TIL order history based on Seasonal and Trend decomposition using Loess (STL) (Hyndman & Athanasopoulos, 2018). In Figure 3.4, the graphs bottom left, top right, and center right show the weekly, monthly, and yearly seasonal cycle, respectively. The number in the y-axis of these graphs within Figure 3.4 denotes the length of the seasonal cycle in weekdays, corrected for leap years. The trend component (left center graph) shows the general behaviour of the time series data set. The seasonal components aim to model the variations due to calendar events. The error component (e.g. remainder component, right bottom graph) captures what the trend component and seasonal components cannot explain.

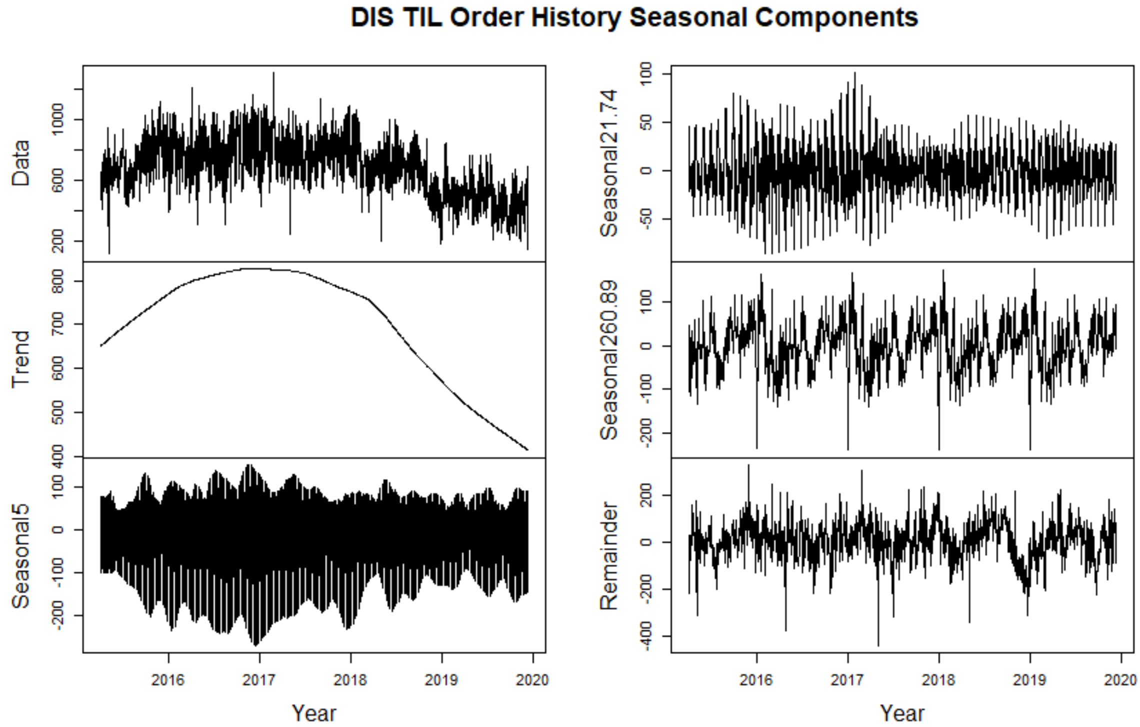

Figure 3.4: DIS TIL order history seasonal decomposition

When visually analyzing the decomposed components as depicted in Figure 3.4, the scale of the vertical axis of each decomposed component should be carefully considered. As can be seen in Figure 3.4, the trend component (left center graph) is fairly strong based on the range of its vertical scale compared to the other seasonal components' vertical scales. The weekly seasonal component is expected to be the strongest seasonal component based on the range of the graph's vertical scale and the graph's peak-to-peak amplitude. The monthly seasonal component seems to be the weakest of all three seasonal components. Besides, the yearly seasonal component seems the most stable of all seasonal components over time, whereas the weekly and monthly seasonal component show more variation over time. Furthermore, as can be seen in Figure 3.4, the magnitude of the weekly and monthly seasonal cycles varies over time. In contrast, the yearly seasonal cycle remains fairly constant.

3.3.3 Strength of the Seasonal Components

The strength of the seasonal components is relevant since the strongest seasonal cycle has to be precisely specified for specific time series forecasting methods. Also, it provides qualitative insights in the characteristics of the time series data. The time series decomposition notation by Hyndman and Athanasopoulos (2018) with multiple seasonal components is as follows:

$$y_t = \sum_{m=1}^{M} S_{t,m} + T_t + R_t,$$

where y_t represents the time series data, $\sum_{m=1}^{M} S_{t,m}$ denotes the M distinct seasonal components, T_t is the trend component, and R_t represents the remainder component. The strength of seasonal component M (F_{S_m}) is based on the variance of the remainder component and the variance of the respective seasonal component and can be defined as (Hyndman & Athanasopoulos, 2018):

$$F_{S_m} = max\ (0, 1 - \frac{Var(R_t)}{Var(S_{t,m}+R_t)})$$

The value variable F_{S_m} can take is between zero and one. Under conditions where the variance of the remainder component is larger compared to the variance of the respective seasonal component, the equation ensures that the value for F_{S_m} cannot become negative. In case a seasonal component exhibits a fairly weak seasonal effect, the value for its strength is close to zero. In contrast, seasonal components that exhibit a strong seasonal effect have a value close to one. The strength of the trend component (F_T) can be calculated in a similar manner, but requires the variance of the remainder component and the variance of the trend component (Hyndman & Athanasopoulos, 2018):

$$F_T = max\ (0, 1 - \frac{Var(R_t)}{Var(T_t+R_t)}),$$

The strength of the trend component and the seasonal components of the DIS TIL order data is presented in Table 3.1. As can be seen in this table, the weekly seasonal cycle is the strongest seasonal cycle, followed by the yearly seasonal cycle. The monthly seasonal cycle is the weakest of all seasonal cycles. The quantification of the strength of the components confirms the conclusions as stated in Section 3.3.2 based on Figure 3.4.

Table 3.1: Strength of the trend component and seasonal components

Time series component	Strength
Trend	0.70
Weekly seasonal cycle	0.56
Monthly seasonal cycle	0.13
Yearly seasonal cycle	0.34

3.4 Resource Requirements

This section discusses the factors that have an effect on daily vehicle routings and thus the number of required vehicles based on historical data. Customer time windows, customer location, and customer order sizes mainly determine the number of required vehicles.

3.4.1 Customer Time Windows

A customer time window is one of the factors that has a large influence on the daily routings of vehicles. To facilitate a quantitative analysis of the distribution of customer time windows during the day, an observation that represents the frequency of each customer demanding a service per minute has been created. More specifically, for a customer that prefers to be served between 8am and 5pm, an observation from 8am to 5pm per minute has been created. The frequency of each minute is presented in Figure 3.5 and shows the distribution of customer time windows during the day for September, October, and November 2019. The average customer time window length is four and a half to five hours. As can be seen in Figure 3.5, the peak of customers demanding a service is from 9am to 12am. The number of customers that prefer to be served decreases from 12am until 6pm. Since each vehicle deployed by the DIS TIL planning department performs one route per day, the peak during the morning is most determinant in establishing the number of required vehicles to fulfill all customer orders during the day.

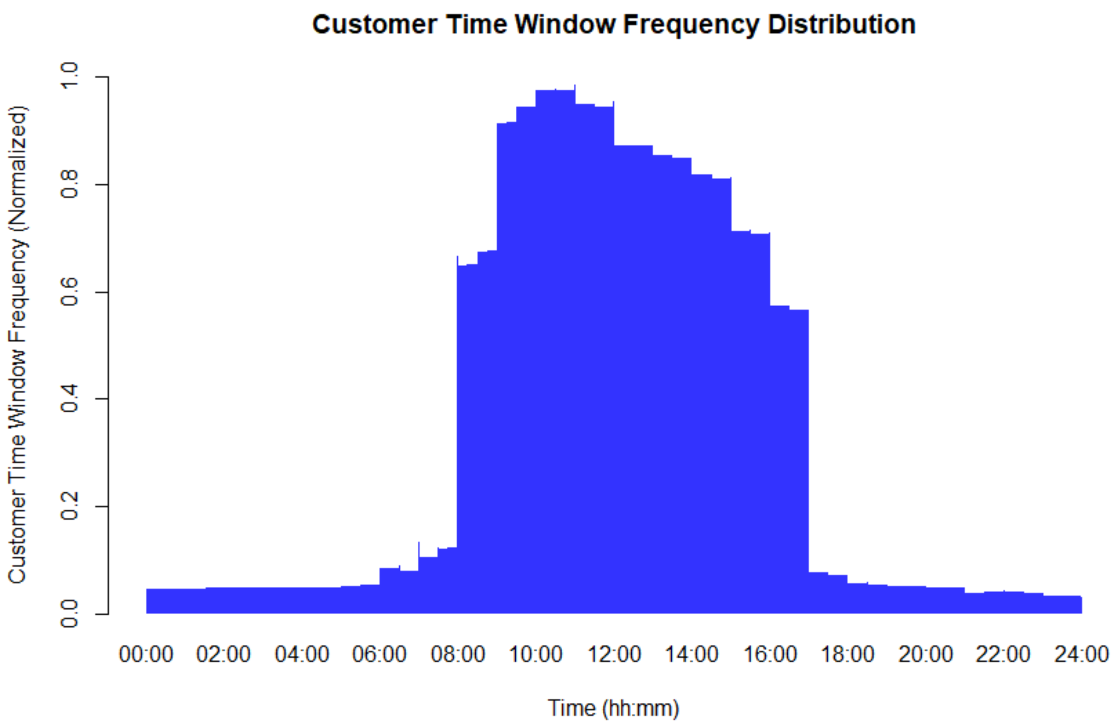

Figure 3.5: Customer time window distribution

The customer time window distribution differs per postal code region, of which an example is presented in Figure 3.6. Figure 3.6 depicts the time window distribution for September, October, and November 2019 for two geographical regions that are closely located to each other. The time window frequency distribution is normalized to compare both distributions based on the same scale. As can be seen in Figure 3.6, the time window distribution of one region has a peak before 12am and decreases thereafter. In contrast, the other region's customers prefer to be served more evenly throughout the day. Differences in time window distributions between regions can either be advantageous or disadvantageous for vehicle routing. When time window distribution peaks coincide, the objective to fulfill all customer orders in the customer's desired time window may not be achieved or forces the planning department to hire a large number of vehicles. In contrast, in case the peak of both time window distributions do not coincide, customer demand is more spread over the day and load sharing opportunities can be realized. By having customer demand more evenly spread over the day, the required number of vehicles to serve customers is expected to be reduced.

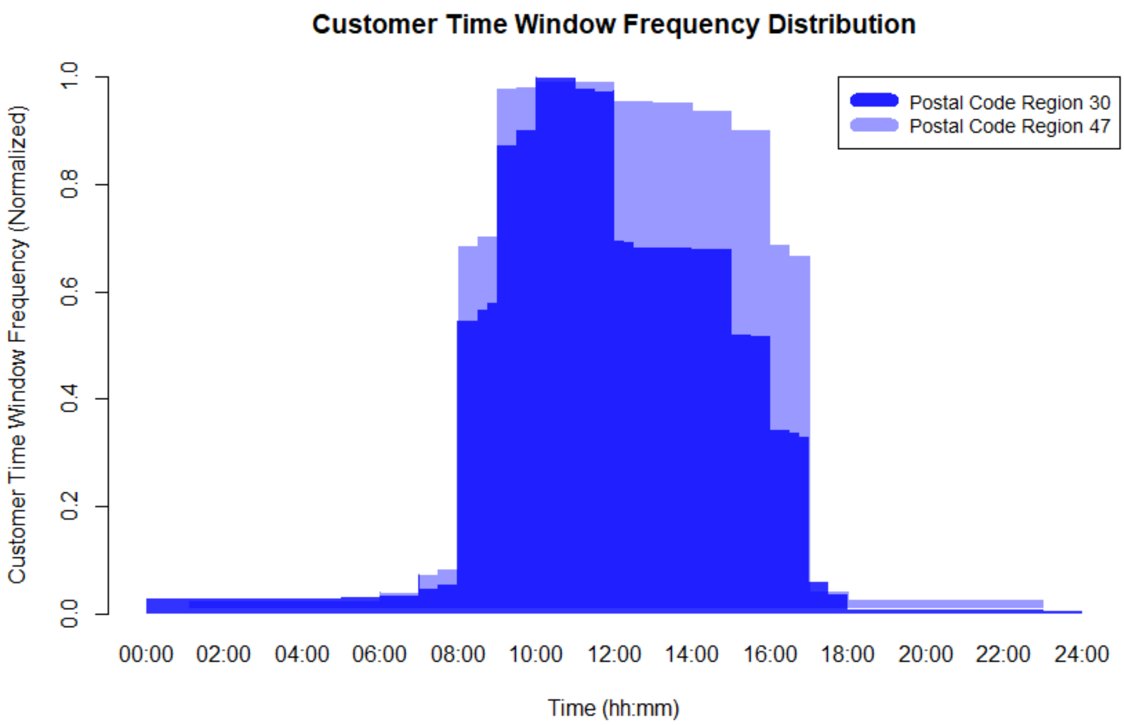

Figure 3.6: Customer time window distribution for two regions

3.4.2 Customer Locations

Customer locations also have a major impact on the required number of vehicles during the day. The impact of very narrow time windows of customers that are located far away from the depot have more impact on the vehicle routing than customer locations with narrow time windows close to the depot. Figure 3.7 depicts the postal code areas where customers are located for September, October, and November 2019. The presented postal code areas are based on the first two characters of the Dutch postal code system. The total number of shipped loading meters for September, October, and November 2019 is divided into classes of which its breakpoint is based on one half standard deviation. These classes determine the purple colour intensity in Figure 3.7. The intensity of the colour purple indicates the demand quantity in loading meters for each specific postal code region. This figure indicates that although the majority of demand is close to the depot in Tilburg, the company's vehicles also have to visit the outer regions of the Netherlands. The demand in Belgium is less compared to the demand in the Netherlands. The majority of the demand in Belgium is concentrated in the area between Brussels and Antwerp.

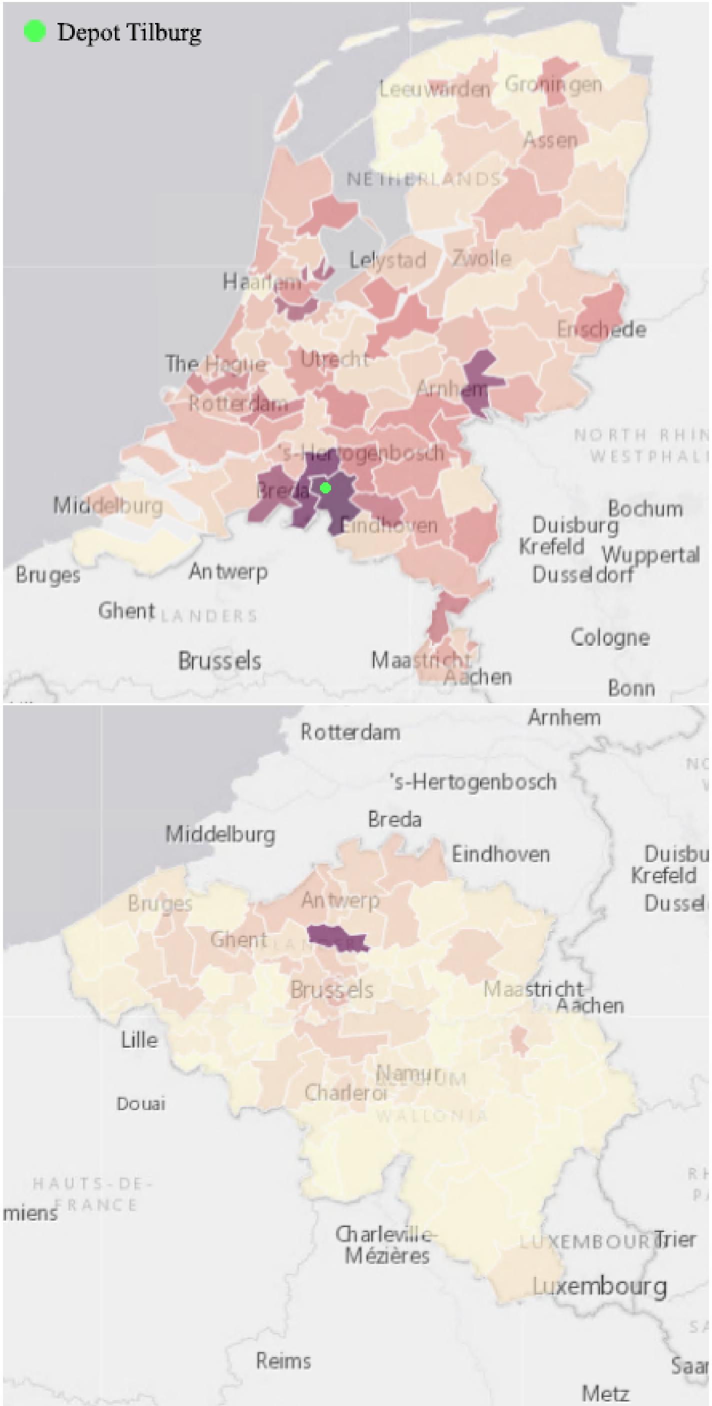

Figure 3.7: Geographical placement of customer demand in the Netherlands and Belgium

3.4.3 The Company's Fleet Size Decision

Based on company data, it is difficult to establish the accuracy of the DIS TIL planning department's vehicle predictions over time. First of all, the frequency and magnitude of a surplus or shortage of vehicles is poorly recorded. In case the DIS TIL planning department has a surplus of vehicles, these vehicles can be sold in the market. Data is available concerning the sale of vehicles in the market to other transportation companies. Preferably, the surplus of vehicles at the DIS TIL planning department is allocated to other planning departments since it may occur that other planning departments have a shortage of vehicles. Unfortunately, the allocation of vehicles to other planning departments is also poorly recorded which makes it difficult to assess the performance of the DIS TIL planning department's forecasting process.

The timing of the fleet size decision is also of high importance to assess the company's forecasting performance. Moreover, the later the decision to hire extra vehicles is made, the higher the vehicle hiring costs. Unfortunately, the timing of the fleet size decision is also poorly recorded. This makes it difficult to calculate the difference in vehicle hiring costs of the planning department's current method and the newly introduced forecasting method. If the planning department would wait until all customer orders for the next day have arrived, then the exact number of required vehicles would be known. In this case, the forecasting error is zero and no true forecast has been established. This latter example stresses the importance of high quality data concerning the timing of the fleet size decision.

Chapter 4

Forecasting of Freight Demand

To determine the number of required vehicles to fulfill customer demand for the short term, the design choice has been made to first generate a forecast of freight demand. Subsequently, this demand forecast is used to determine the number of required vehicles. This chapter explains the applied forecast methodology and presents the most accurate forecasting method based on the company's data set.

4.1 Forecasting Time Series Data

In early stages of the forecasting process, several decisions have to be made. For example, what should be forecasted, what should the forecast time horizon be, how frequently are forecasts required, what data is used to establish a forecast, how to treat outliers, and how to measure forecasting accuracy (Hyndman & Athanasopoulos, 2018). One critical issue is the forecast aggregation level (Zotteri et al., 2005). The appropriate forecast aggregation level depends on the decision making process that will utilize the forecast.

Research on forecast aggregation levels is often referred to as Hierarchical Forecasting and comprises two forecasting methods (Zotteri et al., 2005). The first method is called bottom-up forecasting, where forecasts on the individual (e.g. customer) level are accumulated to produce a forecast on a higher aggregation level (e.g. geographical region). The second method is called top-down forecasting, where the aggregate forecast is disaggregated for each segment (Zotteri et al., 2005). Disaggregation can, for example, be established by applying historical probability distributions to generate a forecast at a lower aggregation level (Weatherford, Kimes & Scott, 2001). Statistically, top-down forecasts should be more accurate than bottom-up forecasts (Weatherford et al., 2001). This is caused by the fact that the average of a number of observations is less variable than the average of individual observations (Ghiani, Laporte & Musmanno, 2013). Moreover, data aggregation might be important when intermittent demand with zero values are present in the data set. The presence of zero values, for example in the demand pattern of an individual customer, can be troublesome when fitting a time series forecasting method. An approach to overcome this problem is to group individual customers by, for instance, their similarity in product category or geographical region. Choosing the 'right' aggregation level is often a trade-off. The lower the aggregation level, the more meaningful and useful the forecast mostly is. However, the accuracy of a forecast at a lower aggregation level might be reduced due to a deficiency of data observations.

4.2 Forecasting Approach

The company's data set comprises the demand in loading meters per day for each customer location. Time series data on the individual customer level per day is irregular for the majority of customers. Moreover, the presence of zero demand values may impede adequate time series method fitting. The feasibility of applying time series methods on different data aggregation levels has been explored, such as creating a forecast per region (e.g. 2-digit postal code level). For several regions, sufficient data observations are available to fit a time series model. In contrast, time series methods cannot be fitted for regions having intermittent demand with zero values. Implementing a forecasting procedure that is valid for all cases is preferable. Therefore, the design choice has been made to forecast the total customer demand per day. A disaggregation rule can be applied to generate a forecast on the customer level. This way, the forecast can for instance be utilized by vehicle routing algorithms since the customer demand quantity and customer locations are generated.

Data Preparation, Forecast Horizon, and Cross-Validation

The company's data set comprises historical daily order data from April 2015 up to December 2019. The time series data has been separated into two parts: a training data set and a test data set. The training data set has been used to establish the forecasting model's parameters (model fitting) and the test data set has been used to evaluate the forecast accuracy (Hyndman & Athanasopoulos, 2018). Since the test data set has not been used to establish the forecast, the fitted model's forecast values can be evaluated with real data from the test data set.

The company's planning department currently establishes the fleet size decision the day before the order delivery date. By providing a forecast for the short term, the fleet size decision can be anticipated several days in advance. Although a forecast horizon of one week is sufficient to facilitate operational decision making, a forecast horizon of larger size can be helpful in determining whether the seasonal effects are captured correctly. A forecast for the very short term (one week) benefits from correct forecasting model parameter estimation. By testing the forecasting methods on a test set of small size, the seasonal effects on the forecast value cannot be tested. In addition, the generalization of the prediction accuracy of different time series methods cannot be assessed if the test set is of small size.

Usually, the size of the training set is about 80% and the size of the test set is about 20% of the total sample data (Hyndman & Athanasopoulos, 2019). The design choice has been made to generate a forecast for two months, which is mainly based on the fact that limited data is available that might be representative for the future (November 2018 up to December 2019). Still, the forecasting methods have been trained with data of prior years to establish the seasonal cycle parameters.

In addition, cross-validation has been a widely used technique to judge whether the forecasting performance generalizes to an independent data set (Fushiki, 2011). K-fold cross-validation randomly partitions the data into training and test sets and therefore the time component of time series data is distorted (Bergmeir & Benítez, 2012). Due to this distortion, k-fold cross-validation cannot be used for time series data. Instead, cross-validation on a rolling basis can be used (Hyndman & Athanasopoulos, 2018). The training data set has been used to set the time series method's parameters and subsequently make a prediction for two months. Thereafter, the forecasting accuracy has been evaluated by means of the test set. The previously used test set is included in the next training set and a new prediction for two months is made. A general example of this cross-validation on a rolling forecasting origin procedure is visually presented in Figure 4.1. By averaging the error metrics over the test sets, the time series method's forecasting accuracy can be computed (Hyndman & Athanasopoulos, 2018). The last half year of 2019 has been divided into three test sets of two months in size each. Note that the training sets differ in size, whereas the size of the test sets is equal but shifts in time. Only three different training and test sets have been used since the company has limited representative data from November 2018 up to December 2019.

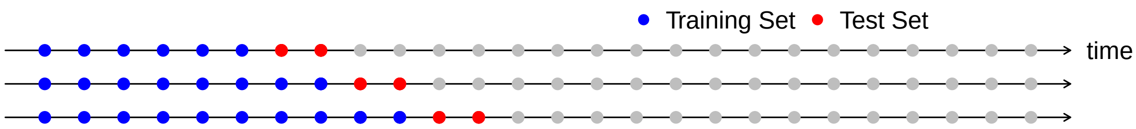

Figure 4.1: Example of time series cross-validation based on a rolling forecasting origin

4.3 Forecast Accuracy

The forecasting method's accuracy is based on several error metrics to determine which forecasting method is able to produce the most accurate forecast. Both in-sample performance and out-of-sample performance are examined, where in-sample performance evaluates the time series method's fitted values with the original values of the training set. Out-of-sample performance is evaluated by means of comparing the forecast values with the original observations from the test set. Both in-sample and out-of-sample performance is required to determine whether a fitted model overfits or underfits the data. If the forecasting model performs better on the training set in comparison to the test set, the forecasting model is likely overfitting the data. The opposite holds for underfitting. Cross-validation can be used as an approach to reduce the likelihood of overfitting.

Forecast Error Measures

Forecast residuals are calculated on the training data set and denote the part of the observation that could not be fitted to the training model (Hyndman & Athanasopoulos, 2018). The residuals can be defined as the error (e_t) between the observation value (y_t) and the fitted value ($\hat{y}_t$) (Montgomery et al., 2015):

$$e_t = y_t - \hat{y}_t$$

The discrepancy between an observed value and the forecast value is the forecast error and is calculated on the test data set (Hyndman & Athanasopoulos, 2018). The forecast error can be defined as

$$e_{T+h} = y_{T+h} - \hat{y}_{T+h|T},$$

where $\{y_1, ..., y_T\}$ denotes the training data set and $\{y_{T+1}, y_{T+2}, ...\}$ denotes the test data set (Hyndman & Athanasopoulos, 2018). Thus, e_{T+h} is the forecast error at time T plus forecast horizon h. Variable y_{T+h} represents the original observation at time T plus forecast horizon h. Lastly, variable $\hat{y}_{T+h|T}$ is the forecast value for time T plus forecast horizon h, based on the training data $y_1, ..., y_T$.

According to Wang and Chaovalitwongse (2010), the most widely used direct error measures are MSE (mean squared error), RMSE (root mean squared error), MAE (mean absolute error), and MAPE (mean absolute percentage error). These error measures are widely used due to their intuitive interpretation to assess the prediction accuracy of a forecasting model (Wang & Chaovalitwongse, 2010). The smaller the value for these error measures, the higher the model's prediction accuracy. Overfitting occurs when a model fits the training data set well, but does not adequately fit the test set data. Therefore, a good prediction model should perform well on both the training data set and test data set. Since cross-validation based on a rolling forecasting origin has been applied, the time series method's forecasting accuracy is computed by averaging over the error metrics of the test sets. The error measures that are used to evaluate forecasting performance are defined as follows (Wang & Chaovalitwongse, 2010):

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{4.1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \tag{4.2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \mid y_i - \hat{y}_i \mid \tag{4.3}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \mid \frac{y_i - \hat{y}_i}{y_i} \mid \tag{4.4}$$

Real observation data is indicated by set y_i of size n where $i = 1, 2, ..., n$ (Wang & Chaovalitwongse, 2010). Each real data observation has an associated forecast value indicated by $\hat{y}_i$. When the model's fitting accuracy needs to be assessed, variable $\hat{y}_i$ represents the value that has been fitted to the model. MSE (Equation 4.1), RMSE (Equation 4.2), and MAE (Equation 4.3) are scale-dependent error metrics, which means that comparing the performance of different forecasting methods is only applicable with a consistent data set. Since the implemented forecasting methods are based on a consistent data set, these scale-dependent measures are applicable to analyze forecasting performance.

The MAPE error measure (Equation 4.4) is based on percentage errors, where the percentage error is denoted as $p_t = 100(\frac{e_t}{y_t})$ (Hyndman & Athanasopoulos, 2018). The MAPE is frequently denoted as a percentage and can thus be rewritten as (Hyndman & Koehler, 2006):

$$MAPE = mean(\mid p_t \mid)$$

The MAPE error measure has several drawbacks. One of these drawbacks is that this error measure is not suitable when zero values are present in the data set (Hyndman & Koehler, 2006). Since the company data sets are based on aggregated demand and the zero demand observations have been imputed, the MAPE error measure is applicable since all data is positive and much larger than zero. Besides, the MAPE error measure penalizes negative errors more compared to positive errors and therefore requires careful interpretation.

The RMSE metric squares the errors prior to averaging the errors. The implication of this operation is that relative large errors have a relatively high weight in the computation of the error metric value. Hence, the RMSE might be more useful when large error values are undesirable. The MAE metric is easier to interpret since it measures the average error size where all errors have equal weight.

4.4 Analysis of Forecasting Methods

This chapter describes the implementation of five different time series models to generate a forecast of freight demand for the short term. A wide range of time series forecasting methods is available and discussed in literature (Brockwell & Davis, 2016). The selection of time series methods is mainly based on how these method handle (multiple) seasonal patterns and the methods' fit with the company's context. In total, five different forecasting methods have been implemented. More specifically, a triple exponential smoothing model, a model based on seasonal decomposition, an ARIMA model, a TBATS model, and a neural network model have been created and fitted to the company's data set. The forecasting performance is established by means of rolling forecasting horizon cross-validation as described in Section 4.2. Finally, a comparison of the time series models' forecasting performance is given in Section 4.5.

4.4.1 Triple Exponential Smoothing

Exponential smoothing methods generate a forecast by using weighted averages of historical observations (Goodwin, 2010). The more recent the observation, the higher the weight in the forecast. Holt-Winters triple exponential smoothing extends traditional exponential smoothing methods by considering three smoothing equations (Hyndman & Athanasopoulos, 2018). More precisely, one equation describes the time series level, one equation denotes the time series trend, and one equation outlines the seasonal component. In general, the Holt-Winters triple exponential smoothing method can solely handle one seasonal cycle. Therefore, the strongest seasonal cycle, which is the weekly seasonal cycle (Subsection 3.3.3), has been utilized for model fitting. The seasonal cycle can either be additive or multiplicative in nature (Hyndman & Athanasopoulos, 2018). The additive method can be described as

$$\hat{y}_{t+h|t} = l_t + hb_t + s_{t+h-m(k+1)},$$

where the triple exponential smoothing time series level is denoted by l_t, the trend is described by b_t, and the seasonal component is expressed by s_t (Hyndman & Athanasopoulos, 2018). The number of seasonal cycles within one year is m (here, m=52) and h is the forecast horizon. Variable k is is the integer part of $\frac{h-1}{m}$ and warrants that the latest year of the sample data set is used to determine the seasonal effects. Moreover, the multiplicative method is denoted as (Hyndman & Athanasopoulos, 2018):

$$\hat{y}_{t+h|t} = (l_t + hb_t)s_{t+h-m(k+1)}$$

Under conditions of a more or less constant seasonal effect, the additive method is able to describe this time series behaviour the best. In contrast, the seasonal effect may behave proportional to the level of the time series. In this case, the multiplicative method is the most suitable (Hyndman & Athanasopoulos, 2018). To establish whether the seasonal cycle of the company's time series data is additive or multiplicative in nature, both the additive and multiplicative methods have been implemented. The method that showed the best performance on fitting error metrics and forecasting error metrics is preferable.

Figure 4.2 depicts the fitted Holt-Winters' triple exponential smoothing method including a forecast for the best performing test set. The multiplicative method resulted the lowest error metric values. The green line, as can be seen in Figure 4.2, represents the fitted values that were used to establish the Holt-Winters' model parameters. The method does not have the capacity to use all training data; the most dated observations were omitted for model parameter estimation. The forecast values are indicated by blue observations and the test set values are represented by the red line. Moreover, Figure 4.2 shows the confidence level for the 80% and 95% prediction intervals. With a certain probability, the forecast value that has been generated by the forecasting method is within the prediction interval (Hyndman & Athanasopoulos, 2018). The uncertainty that is associated with the forecast value is therefore represented by the prediction interval. In general, the prediction interval becomes wider as the forecast horizon increases. Since the applied forecast horizon is relatively short, the prediction interval does not significantly increase with time.

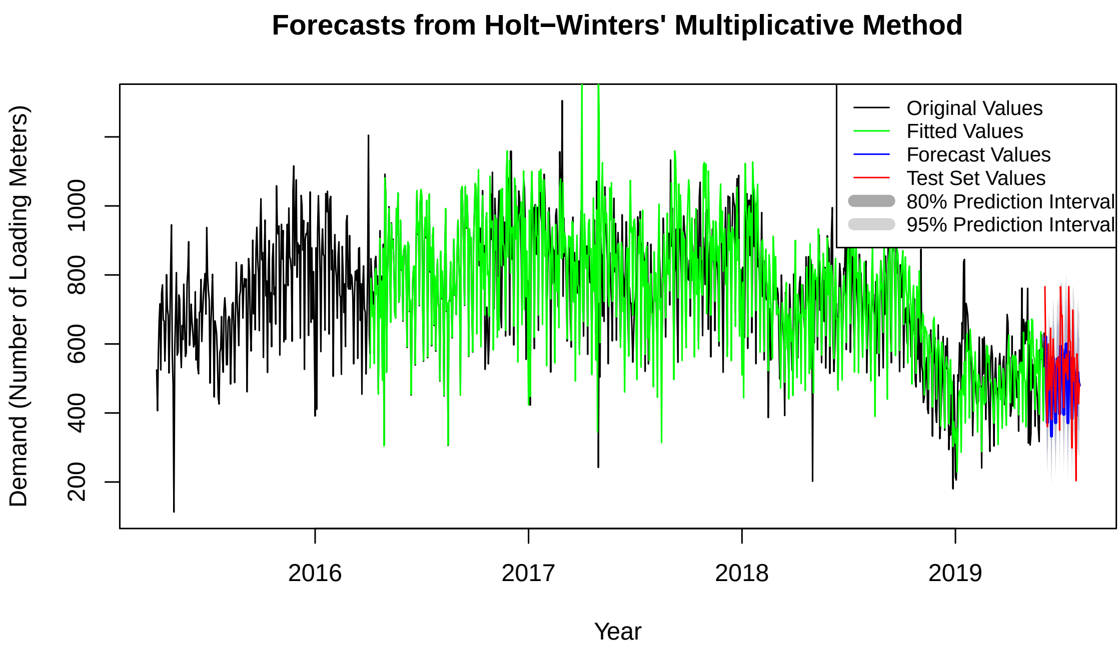

Figure 4.2: Triple exponential smoothing forecast

The error metrics for the training sets are presented in Table 4.1. For comparing the error metric values between the test sets and training sets, the focus is on the MAE and the MAPE due to their intuitive interpretation. Overall, all error metrics move in the same direction. As can be seen in this table, the error metrics of each training set show similar performance. Table 4.2 shows the error metrics for the different test sets. The first test sets shows the lowest error metrics value, with a MAE of 63.92 loading meters. Averaged for all different test sets, the MAE was 75.98 loading meters. A comparison of the average MAE of the training sets (78.02) and the average MAE (75.98) of the test sets shows that both error metrics perform similarly. Therefore, there is no indication of overfitting or underfitting the data. For the best performing test set, Figure 4.2 shows that the forecast is able to follow the red line reasonably well, except for the most extreme test set observations.

Table 4.1: Error metrics of triple exponential smoothing method (training sets)

Training Set	MSE	RMSE	MAE	MAPE
I	10,972.56	104.75	78.40	12.30
II	10,970.47	104.74	78.13	12.40
III	10,670.89	103.30	77.53	12.50
Average:	10,871.31	104.26	78.02	12.40

Table 4.2: Error metrics of triple exponential smoothing method (test sets)

Test Set	MSE	RMSE	MAE	MAPE
I	7,231.80	85.04	63.92	12.74
II	9,818.83	99.09	78.24	20.57
III	10,348.99	101.73	85.78	21.08
Average:	9,133.21	95.29	75.98	18.13

4.4.2 ETS Model with Seasonal Decomposition

Under specific circumstances, seasonal decomposition of time series data may result in better forecasting performance as mentioned by G. P. Zhang and Qi (2005). Therefore, a time series method that applies seasonal decomposition prior to model fitting has been implemented (Hyndman & Athanasopoulos, 2018). First, the time series data has been decomposed into a trend component, seasonal component, and remainder component in a similar manner as described in Section 3.3.2. So, the time series components have been established by means of Seasonal and Trend decomposition using Loess (STL) (Cleveland et al., 1990). Next, a non-seasonal forecasting method has been applied to fit the seasonally decomposed data. Lastly, the resulting forecast has been re-seasonalized by the last year of the seasonal component to correct for the applied seasonal decomposition.

The fitted model is referred to as a state space model (Hyndman & Athanasopoulos, 2018). A state space model is defined by a measurement equation and one or more state equations. More specifically, the observed data is defined by the measurement equation and state equations describe the behaviour of the level component, trend component, and seasonal component over time (Hyndman, Koehler, Ord & Snyder, 2008). The best fitted state space model is of the form $ETS(A, N, N)$, which implies that a simple non-seasonal exponential smoothing method with additive errors has been fitted to the seasonally adjusted data. ETS is an acronym for Error, Trend, and Seasonal, and thus refers to the decomposed seasonal components. The $ETS(A, N, N)$ state space model equations can be written as

$$y_t = l_{t-1} + \varepsilon_t, \; and$$

$$l_t = l_{t-1} + \alpha\varepsilon_t,$$

where y_t represents the measurement equation and l_t the state equation. Variable α is the smoothing parameter and controls the rate at which the weights of prior observations decrease. Variable ε_t captures the forecast error at time t.

Figure 4.3 depicts the forecast of a simple exponential smoothing method with additive errors for the test set that shows the best performance. The error metrics for the ETS method training sets are presented in Table 4.3. These metrics show similar results for all three different training sets. The error metrics for this method's test sets are presented in Table 4.4. The ETS model's performance on the test sets is poor, indicated by an average MAE of 148.03 loading meters and a MAPE of 33.37. An average MAPE value of 33.37 means that the forecast is on average off by 33.37%. Besides, the average MAE of the test sets (148.03) is much larger than the average MAE of the training sets (71.15). This is an indication that the ETS model overfits the data. In addition, Figure 4.3 shows that the forecast values follow a downward trend, whereas the test set values follow a slight upwards trend. Due to this downward trend, it is difficult to determine whether the ETS model is able to capture the seasonality well for the best performing test set.

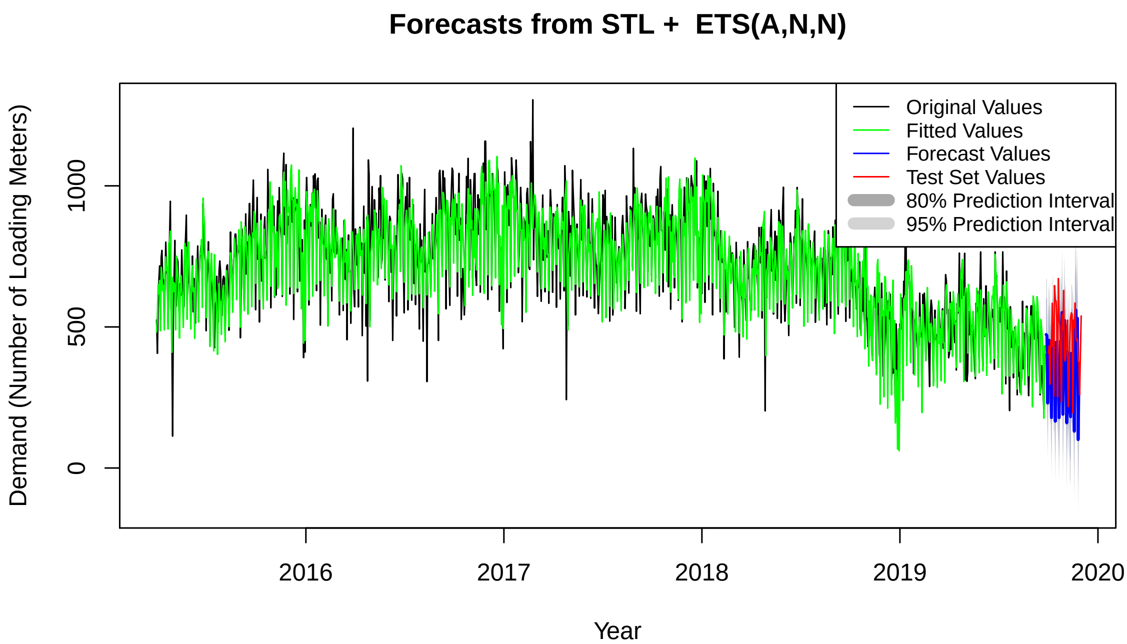

Figure 4.3: ETS model with seasonal decomposition forecast

Table 4.3: Error metrics of ETS model with seasonal decomposition (training sets)

Training Set	MSE	RMSE	MAE	MAPE
I	10,326.62	101.62	70.57	11.26
II	10,574.01	102.83	71.46	11.57
III	10,428.49	102.12	71.42	11.79
Average:	10,443.04	102.19	71.15	11.54

Table 4.4: Error metric of ETS model with seasonal decomposition (test sets)

Test Set	MSE	RMSE	MAE	MAPE
I	35,163.75	187.52	162.88	35.15
II	29,141.90	170.71	141.09	33.73
III	29,357.40	171.34	140.13	31.23
Average:	31,221.02	176.52	148.03	33.37

4.4.3 ARIMA

The forecasts generated by exponential smoothing methods are heavily influenced by the trend component and seasonal cycle in the data set. AutoRegressive Intregrated Moving Average (ARIMA) methods are based on a different concept and aim to define the autocorrelation of observations in the data set (Hyndman & Athanasopoulos, 2018). ARIMA methods are able to provide a forecast based on data that is either stationary or non-stationary, and data that has a trend and/or cyclic component. The notation of an ARIMA model is as follows:

$$ARIMA(p, d, q),$$

where p is the order of the autoregressive part, d represents the degree of data differencing, and q represents the order of the moving average part (Hyndman & Athanasopoulos, 2018). The method requires stationary data (i.e., the underlying stochastic time series process should not change over time) (Da Veiga, Da Veiga, Catapan, Tortato & Da Silva, 2014). Since time series data is mostly non-stationary, it is necessary to differentiate the data to make it stationary. Therefore, the degree of differencing has to be precisely stated in the ARIMA model formulation.

In case a strong seasonal cycle is present in the data set, the Seasonal ARIMA (SARIMA) model may show better forecasting performance since this model explicitly includes additional seasonal terms compared to traditional ARIMA models (K. Y. Chen & Wang, 2007). The notation of the SARIMA model is as follows (Hyndman & Athanasopoulos, 2018):

$$ARIMA(p, d, q)(P, D, Q)_m$$

The additional terms, compared to the aforementioned traditional ARIMA model formulation, are P (seasonal autoregressive order), D (seasonal difference order), Q (seasonal moving average order), and m (the number of observations for a single seasonal period) (Hyndman & Athanasopoulos, 2018). A disadvantage of the SARIMA method is that the seasonal component is periodic and thus cannot change over time.

Figure 4.4 depicts the fitted ARIMA model including a forecast for the best performing test set. The best fitting ARIMA model can be stated as $ARIMA(5,1,3)$. Even though the SARIMA method intuitively should fit the data better due to the presence of multiple seasonal cycles in the data set, the best performing method is the traditional ARIMA model. It could be that a SARIMA model could not be fitted due to an insufficient number of data observations that follow the trend since November 2018. Also, the seasonal cycles may be too capricious too fit a SARIMA model. As can be seen by the blue line in Figure 4.4, the fitted ARIMA model is able to provide a useful forecast for merely two weeks. The blue line stabilizes thereafter and the forecast shows poor performance.

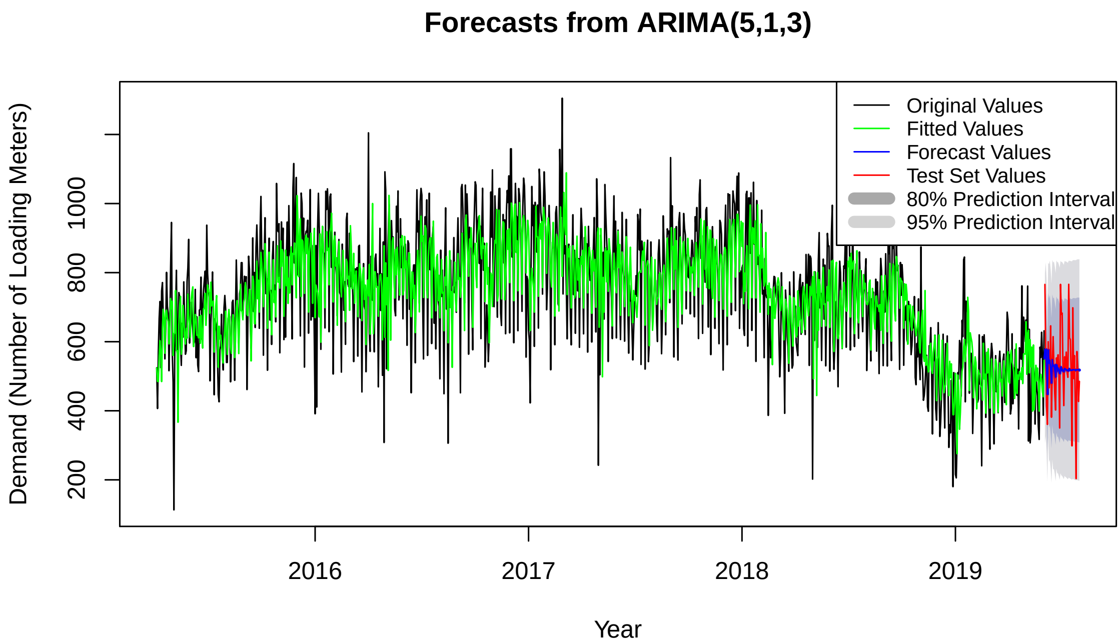

Figure 4.4: ARIMA forecast

The training set error metrics for the ARIMA model are presented in Table 4.5. The error metrics of the different training sets are nearly similar. The test sets' error metrics (Table 4.6) have an average MAE of 93.33 loading meters. Since the average MAE of the training sets (94.68) is nearly similar to the average MAE of the test sets (93.33), there is no indication of overfitting or overfitting. However, the ARIMA model that has been fitted to the best performing test set is not able to capture the seasonal effects, which is indicated by the blue line that shows damping behaviour over time (Figure 4.4).

Table 4.5: Error metrics of ARIMA method (training sets)

Training Set	MSE	RMSE	MAE	MAPE
I	15,560.07	124.74	94.68	14.69
II	15,557.57	124.73	94.94	14.98
III	15,366.08	123.96	94.41	15.21
Average:	15,494.57	124.48	94.68	14.96

Table 4.6: Error metrics of ARIMA method (test sets)

Test Set	MSE	RMSE	MAE	MAPE
I	12,802.92	113.15	83.65	18.42
II	16,154.41	127.10	99.82	23.30
III	14,061.22	118.58	96.52	25.69
Average:	14,339.52	119.61	93.33	22.47

4.4.4 TBATS

The TBATS method is able to handle multiple seasonal cycles and allows seasonal cycles to change over time (Hyndman & Athanasopoulos, 2018). Since the company data is characterized by multiple seasonal cycles that change over time, the TBATS method is expected to be capable of modelling the company's time series data well. TBATS stands for Trigonometric seasonality, Box-Cox transformation, ARIMA errors, Trend, and Seasonal components (de Livera, Hyndman & Snyder, 2011). The TBATS method fits seasonality cycles by means of Fourier terms; each seasonal cycle is thereby described by a unique sinusoids function. Moreover, the TBATS method may apply a Box-Cox transformation to transform non-normal data into a normal shape. In statistics, a normalized data set might be required to execute specific statistical analyses. The TBATS method aims to model error terms by ARMA terms. ARMA error terms imply that an error depends on errors of lagged observations. Moreover, ARMA error terms are characterized by fitting a moving average method to lagged observations. Figure 4.5 depicts the forecast of the TBATS model for the best performing test set.

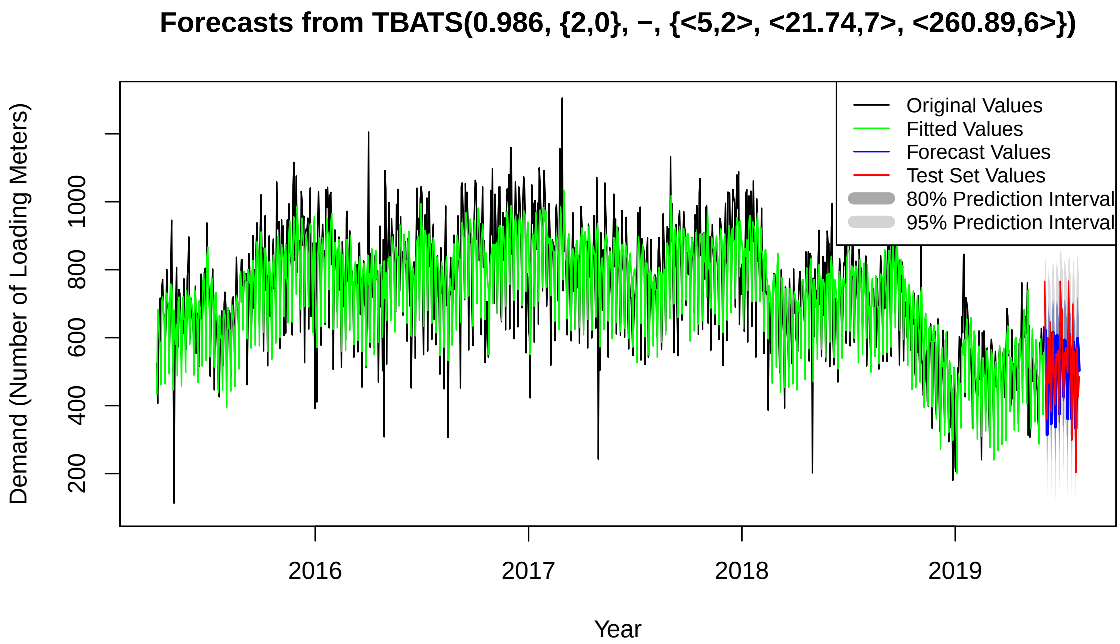

Figure 4.5: TBATS forecast

The fitted TBATS model applied a Box-Cox transformation with a Box-Cox parameter value of 0.986. The error term is modeled by ARMA(2,0) and no dampening parameters have been used. Besides, the weekly, monthly, and yearly seasonal cycles have a length of 5, 21.74, and 260.89 days respectively (corrected for leap years). These seasonal cycles are modeled by 2, 7, and 6 Fourier terms, respectively. The error metrics of the TBATS model's training sets are presented in Table 4.7. According to this table, the error metrics for all training sets show similar results. The test sets' error metrics, presented in Table 4.8, show an average MAE of 74.56 loading meters and a MAPE value of 16.70%. The difference in the average MAE of the training sets (78.02) and the average MAE of the test sets (74.56) is too small to conclude overfitting or underfitting. Based on Figure 4.5, the forecast of the best performing test set is able to follow the red line reasonably well, except for the most extreme test set values.

Table 4.7: Error metrics of TBATS method (training sets)

Training Set	MSE	RMSE	MAE	MAPE
I	10,972.56	104.75	78.40	12.30
II	10,970.47	104.74	78.13	12.40
III	10,670.89	103.30	77.53	12.50
Average:	10,871.31	104.26	78.02	12.40

Table 4.8: Error metrics of TBATS method (test sets)

Test Set	MSE	RMSE	MAE	MAPE
I	6,591.82	81.19	65.72	13.37
II	11,731.06	108.31	79.93	16.87
III	10,594.58	102.93	78.04	19.85
Average:	9,639.15	97.48	74.56	16.70

4.4.5 Neural Network

Neural networks are a domain of computational intelligence and have been widely studied in forecasting literature (Zhou et al., 2006). A neural network is an algorithm that mimics the capabilities of the human brain and is able to approximate functions that depend on a large number of inputs. The main characteristics of a neural network algorithm are its learning abilities, its ability to recognize patterns, and its ability to perform predictions on time series data with high accuracy (G. P. Zhang & Qi, 2005). Compared to other time series methods, a neural network does not require any assumptions being specified prior to model fitting since the neural network establishes the data set's underlying relationships by data mining (G. P. Zhang & Qi, 2005).

A neural network consists of an input layer, hidden layer(s), and an output layer. The input layer contains the raw input data. Each layer consists of nodes. Each node is connected with other nodes through weights (indicated by arcs). Nodes are given numerical input which is multiplied by the weights of the arcs. The neural network can process input, perform weight adjustments, and subsequently produce output. The hidden layers process the input nodes' information. The resulting forecast is represented by the output layer. According to G. P. Zhang (2003), the single hidden layer feedforward network is one of the most widely used models for neural network time series forecasting. Figure 4.6 provides a visual example of a single hidden layer neural network.

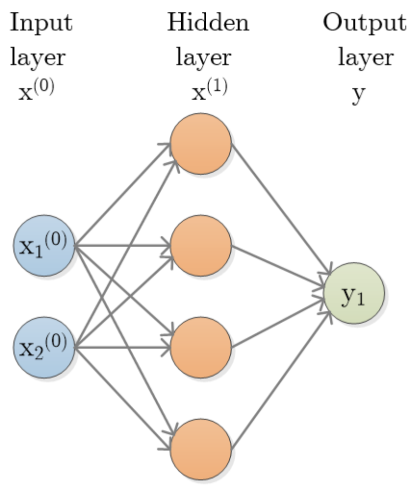

Figure 4.6: Example of a single hidden layer neural network

Figure 4.7 depicts the forecast of the neural network for the test set having the highest prediction accuracy. The resulting model is a feedforward neural network consisting of a single hidden layer and lagged inputs. The neural network's model can be described as

$$NNAR(p, P, k)_m,$$

where p represents the number of lagged observations, P depicts the number of seasonal lags used as input, and k is the optimal number of neurons (where $k = \frac{p+P+1}{2}$) (Hyndman & Athanasopoulos, 2018). Variable m indicates the length of the seasonal cycle. The fitted neural network has parameters p=30, P=1, k=16, and m=260. In other words, the neural network model has 31 input nodes ($y_{t-1}, y_{t-2}, ..., y_{t-30}$ and y_{t-260}), 16 nodes in the hidden layer, and one node in the output layer.

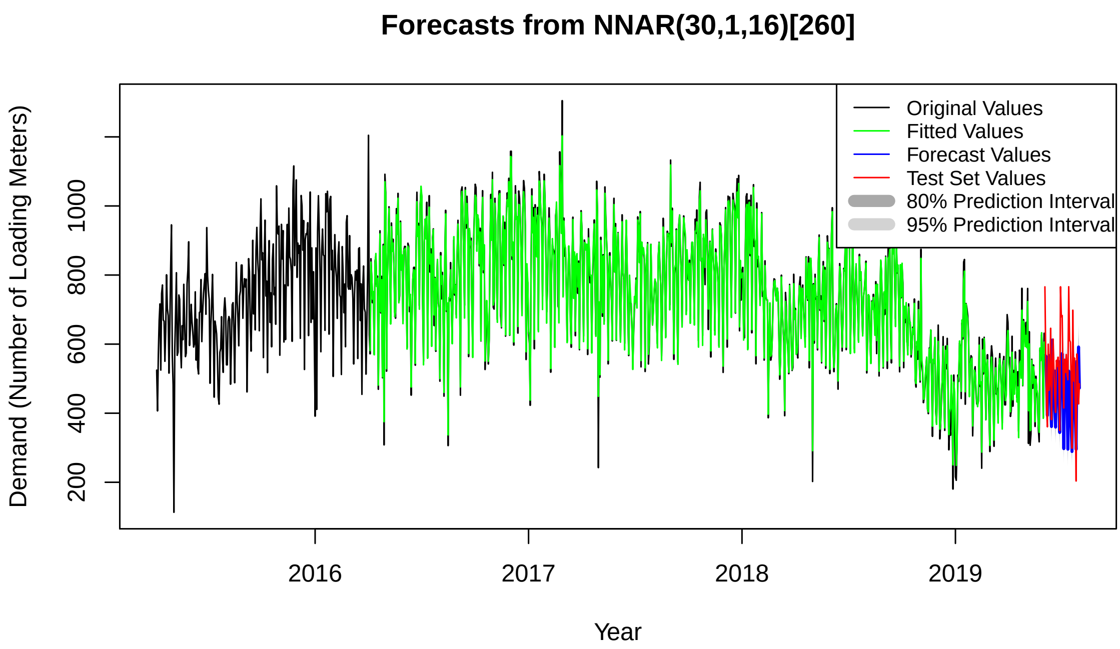

Figure 4.7: Neural network forecast

The error metrics of the neural network's training sets are presented in Table 4.9. Based on this table, the average MAE is 14.43 loading meters and the MAPE is 2.44. The MAE and MAPE error metric values indicate that the neural network is able to adequately fit the company's data. The neural network's error metric results for the test sets are presented in Table 4.10. The average MAE of the three test sets is 83.12 loading meters and the MAPE has a value of 17.35%. The neural network overfits the data since the error metric values of the training sets are much lower than the error metric values of the test sets. Also, based on Figure 4.7, the neural network of the best performing test set does not capture seasonality very well due to the deviation between the blue and red line, especially for the extreme test set values.

Table 4.9: Error metrics of neural network (training sets)

Training Set	MSE	RMSE	MAE	MAPE
I	365.57	19.12	12.44	2.05
II	445.63	21.11	13.94	2.33
III	665.12	25.79	16.91	2.95
Average:	492.11	22.01	14.43	2.44

Table 4.10: Error metrics of neural network (test sets)

Test Set	MSE	RMSE	MAE	MAPE
I	8,582.17	92.64	76.43	18.74
II	10,759.91	103.73	82.07	15.88
III	14,332.88	119.72	90.85	17.42
Average:	11,224.99	105.36	83.12	17.35

4.5 Review of Forecasting Methods

In order to establish the best time series forecasting method for the company, this section is dedicated to the analysis and comparison of the five different implemented forecasting methods. The smaller the error metric values, the higher the method's prediction accuracy. Table 4.11 shows the average of the training sets' error metrics for all five forecasting methods. As can be seen in this table, the neural network outperforms the other forecasting methods when considering the fit to the training data as indicated by a MAE value of 14.43 loading meters. Moreover, the neural network's MAPE and MSE values are the lowest of all forecasting methods. The neural network's outstanding fit can be explained by the neural network's learning and pattern recognition abilities. Of all forecasting methods, the ARIMA model performed the worst on the training sets and has an average MAE of 94.68 loading meters. The MAE of the triple exponential smoothing model, ETS model with seasonal decomposition, and the TBATS model are nearly equal with a value of 78.02, 71.15, and 78.02, respectively.

Table 4.11: Overview of average forecasting method results on training sets

Forecasting model	MSE	RMSE	MAE	MAPE
Triple exponential smoothing	10,871.31	104.26	78.02	12.40
ETS model with seasonal decomposition	10,443.04	102.19	71.15	11.54
ARIMA	15,494.57	124.48	94.68	14.96
TBATS	10,871.31	104.26	78.02	12.40
Neural Network	492.11	22.01	14.43	2.44

Table 4.12 depicts the average of the test sets' error metrics for all five forecasting methods. As can be seen in this table, the TBATS has the lowest MAE value (74.56). The triple exponential smoothing method performs slightly worse compared to the TBATS model with a MAE value of 75.98 loading meters. Despite its fit on the training data sets, the neural network does not show the best forecasting accuracy based on the test sets compared to the

other forecasting models. Furthermore, the triple exponential smoothing model and ETS with seasonal decomposition model show conceptually many similarities. As can be seen in Table 4.11 and Table 4.12, the triple exponential smoothing model showed a higher MAE on the training sets compared to the ETS with seasonal decomposition model, whereas the opposite holds for the test sets.

Table 4.12: Overview of average forecasting method results on test sets

Forecasting model	MSE	RMSE	MAE	MAPE
Triple exponential smoothing	9,133.21	95.29	75.98	18.13
ETS model with seasonal decomposition	31,221.02	176.52	148.03	33.37
ARIMA	14,339.52	119.61	93.33	22.47
TBATS	9,639.15	97.48	74.56	16.70
Neural network	11,224.99	105.36	83.12	17.35

Due to variations in the underlying data set, some time series models do not perform as expected in reference to the model's theoretical capabilities. For example, the SARIMA model seemed promising due to the way it handles seasonal cycles. Unfortunately, this model could not be fitted to the training data sets. Moreover, even though the neural network outperformed the other time series methods on its fit to the training data sets, the resulting forecast values did not fit the test data sets well. The neural network thus overfits the data. All implemented time series forecasting methods are affected by the shift in the average level of the planning department's loading meters since November 2018. The time series models could have performed better when the training data set consisted of more observations that follow the recent average demand levels.

When the end of a month coincides with the beginning of a week, the fitted time series models showed poor performance. In the context of the company, freight demand increases towards the end of the month, whereas the demand at the beginning of the week is the lowest. This distortion in the weekly seasonal cycle impedes time series model fitting. Having more training observations from November 2018 onwards may increase time series model fitting when seasonal cycles show conflicting behaviors. Moreover, especially neural network forecasting performance may increase by having more training observations. A characteristic of neural networks is that these models require many data observations and the number of observations since November 2018 might be of insufficient size to produce an accurate forecast.

Based on the evaluation of the implemented forecasting methods, the TBATS method is the most applicable in the context of the company. The TBATS model does not seem to overfit or underfit the data. Besides, the TBATS model outperforms the other forecasting models and has a MAPE value of 16.70: averaged over the three test sets, the TBATS model's forecast is off by 16.70%. Although the triple exponential smoothing model has a lower MSE and RMSE value compared to the TBATS model, the TBATS model's characteristics make it the most suitable to forecast the company's demand. One of the advantages over the other methods is the TBATS method's ability to handle multiple seasonal cycles that are able to change over time. Since the company data set is subject to many variations in customer types, customer order intervals, and customer order sizes, the TBATS method is the most applicable and robust for the long term.

Chapter 5

Determining the Required Fleet Size

In Chapter 4, the best performing time series model based on company data has been established. The design choice has been made to base the number of required vehicles on a demand forecast. One method to determine the number of required vehicles is to establish the routes the vehicles have to drive to perform daily customer order deliveries. The company has a routing tool at its disposal that can be utilized to optimize the daily vehicle routes. The routing tool implements an algorithm for the VRP. The VRP aims to create routes for a fleet of vehicles to visit a set of customers at minimum costs (Toth & Vigo, 2002). The decision has been made to use the company's routing tool. This tool is able to generate a feasible solution considering 6,000 customer locations in a short time frame using heuristics. In addition, the utilization of the routing tool is representative for the company's daily operations. Besides, the routing tool considers all customer specific constraints such as customer time windows.

5.1 Vehicle Routing Tool

The company's routing tool requires data of customer locations and customer demand. The routing tool considers the following factors for route optimization. First, each vehicle drives a single route per day and each vehicle departs and returns at the company's depot in Tilburg. Next, all customer demand (expressed in loading meters) needs to be satisfied. The routing tool ensures that each customer is visited within its specified time window. A time window between 9am and 5pm has been assigned to orders without a specified time window. Besides, each vehicle has a capacity of 13.6 loading meters. To comply to driver legislation, each trip is subject to a maximum duration of 15 hours (sum of driving and break time). Also, each trip is subject to a maximum driving duration of 10 hours. Within a driving time of 4.5 hours, the driver has to take a total break time of 45 minutes. Within a work shift of 6 hours, the driver has to take a total break time of 45 minutes. A break can have a duration of 15, 30, or 45 minutes. The distance matrix comprises the distances between locations which is based on real distances, instead of Euclidean distances as considered by many VRPs. In addition, the routing tool considers each vehicle having an average speed of 60 kilometers per hour. The vehicle unloading time at a customer location consists of two parts. First, it takes 10 minutes to start unloading the vehicle. Second, unloading one loading meter takes 5 minutes and the total unloading time at a customer location is calculated based on this rate.

The main objective of the vehicle routing tool is to fulfill all customer orders. After all customer orders are planned, the number of required vehicles is minimized. A visualization of the routing for an arbitrary day in November 2019 based on real order data is presented in Appendix C.

5.2 Scenario Generation

In practice, the company aims to establish the number of required vehicles for the short term while not all required data to make this decision is known beforehand. This problem is known as the Stochastic Vehicle Routing Problem (SVRP), where several components of the SVRP are random (Oyola, Arntzen & Woodruff, 2018). For the company, the customer locations and customer order quantities can be considered as random variables that change from day to day.

Several approaches that provide an answer to the SVRP are discussed in literature. For instance, Gendreau, Laporte and Seguin (1996) assigned a probability function to the presence of a customer location (p_i) and the demand of this customer location (ξ_i). In practice, it is complex and time consuming to determine these probability distribution functions. Another approach to solve a SVRP is to perform a scenario analysis based on simulation (Kall & Wallace, 1994). This approach has been chosen due to its applicability in the context of the company since the vehicle routing tool is suitable to perform simulation.

The experimental design to provide an answer to the main research question is to start by generating a total demand forecast for one week. Next, disaggregation rules have been applied to create a set of customer locations based on historical order data and to apply a procedure to disaggregate the total demand forecast to each customer for each weekday. Thus, each weekday is characterized by a unique set (i.e. a scenario) of customer demand and customer locations. Next, each scenario has been solved as a static VRP to reveal the required resources in terms of vehicles (Hvattum, Løkketangen & Laporte, 2006). Subsequently, a probability has been assigned to each scenario for each weekday. Based on these probabilities, the company can determine the number of required vehicles for each day by the amount of risk they want to cover. The scenarios are used to represent uncertainty; they are a means to provide an answer to the main research question. Lastly, the number of required vehicles revealed by scenario generation can be compared with the actual number of required vehicles based on historical data.

To present the aforementioned process more formally, given customer locations l and customer demands d, the number of required vehicles can be found by means of the routing tool that is based on the principles of a VRP. The VRP algorithm takes customer locations and demands as input variables and results the number of required vehicles k (Equation 5.1).

$$k = \textsc{vrp}(l, d) \tag{5.1}$$

Since customer locations L and customer demands D are stochastic variables, the number of required vehicles K is a random variable and a function of L and D (Equation 5.2).

$$K = \textsc{vrp}(L, D) \tag{5.2}$$

The goal is to generate sets of customer demands and locations by applying scenario generation. Let ω be a scenario that is created by means of disaggregation rule f (which is a scenario generation function) that takes input variable Z (the total demand forecast) and Q (historical customer order data) (Equation 5.3). Each scenario ω is not a random variable but a particular realization of two random variables. The result is two vectors: customer locations l_f and customer demands d_f. Both vectors have the same dimension. So, each scenario based on a particular disaggregation rule can be described as follows:

$$\omega_f = f(Z, Q) = l_f, d_f \tag{5.3}$$

Each scenario can be regarded as a result of ω_f and is thus represented by a unique customer location (l_f) and customer demand (d_f) vector combination. To reveal the number of required vehicles k for each scenario ω_f, the variables l_f and d_f have to be used as input variables in Equation 5.1. In addition, a probability has been assigned to each scenario since not all scenarios have an equal probability of occurring, which is explained in Section 5.2.1. Moreover, the demand levels of each scenario are also presented in Section 5.2.1. Furthermore, an elaboration on the disaggregation rules is presented in Section 5.2.2.

5.2.1 Scenario Demand Generation

The customer demands for each scenario are based on a forecast generated by the TBATS model. In Chapter 4, the TBATS model showed the highest prediction accuracy on the test data sets (Table 4.12). The TBATS model has been used to generate a demand forecast since this model is the most appropriate to forecast demand in the context of the company. The design choice has been made to create scenarios for one week (October 14, 2019 up to and including October 18, 2019). This period is referred to as Monday, Tuesday, Wednesday, Thursday, and Friday, hereafter. Since the price of hiring extra vehicles starts increasing two days before the delivery date, the demand prediction is based on a rolling forecast with a forecast horizon of two days.

According to Hyndman and Athanasopoulos (2019), the range of values the demand forecast can take, including its probabilities, is the forecast distribution where the point forecast can be regarded as the mean of this distribution. Each forecast is subject to uncertainty, which is modeled by the point forecast's prediction interval (Hyndman & Athanasopoulos, 2018). For instance, the 95% prediction interval corresponds to the 97.5% and 2.5% quantiles of the forecast distribution under the assumption of a normal distributed forecast (Hyndman, 2014). The demand levels for scenario generation are derived from the point forecast and the forecast values at one and two standard deviations. The demand forecasts at one and two standard deviations have a lower probability of occurring compared to the point forecast itself. In total, five demand values are derived from the forecast distribution to create scenarios for each day of the five weekdays under consideration.

The standard deviation of the forecast distribution of a one-step ahead forecast corresponds to the standard deviation of the errors of this forecast (Hyndman & Athanasopoulos, 2018). In general, the prediction interval increases as the forecast horizon increases. So, the standard deviation increases with the forecast horizon for multi-step ahead forecasts. The forecast horizon that has been used for scenario generation is relatively short (two days) in comparison to the size of the training data set. Therefore, the standard deviation of a multi-step ahead forecast in this context is nearly similar to the standard deviation of the forecast's errors.

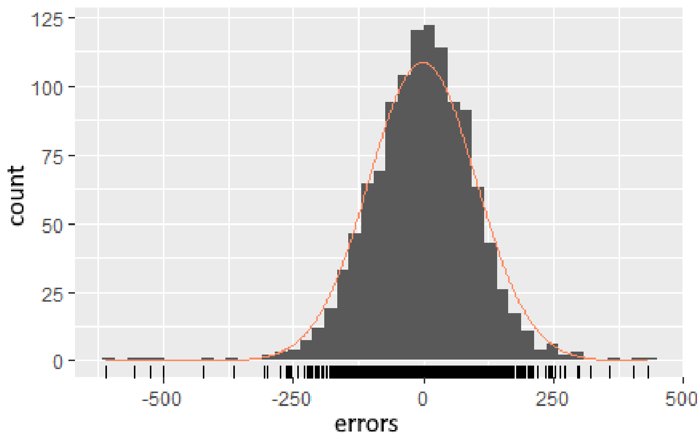

Figure 5.1: Histogram of TBATS forecast errors including normal curve

Figure 5.1 depicts the distribution of the forecast errors of the fitted TBATS model. This figure presents a histogram of the forecast errors, where the errors can be positive (overpredicting) or negative (underpredicting). Moreover, the forecast error distribution follows the shape of a normal distribution which is indicated by the orange line in Figure 5.1. Under the condition of normal distributed forecast errors, the prediction interval of the forecast can be notated as

$$\hat{y}_{T+h|T} \pm c\hat{\sigma}_h,$$

where $\hat{y}_{T+h|T}$ is the estimate of y_{T+h} based on previous observations $y_1, ..., y_T$ (Hyndman & Athanasopoulos, 2018). Variable c represents the multiplication factor derived from the cumulative distribution function of the normal distribution. Moreover, the estimate of the standard deviation that corresponds to the h-step ahead forecast distribution is represented by variable $\hat{\sigma}_h$ (Hyndman & Athanasopoulos, 2019). For instance, a 95% prediction interval corresponds to a value of 1.96 for variable c and can be stated as:

$$\hat{y}_{T+h|T} \pm 1.96\hat{\sigma}_h$$

Figure 5.2 shows the forecast errors' Q-Q plot. The Q-Q plot can be used to assess whether data follows a normal distribution (Shumway & Stoffer, 2017). The Q-Q plot is a scatterplot that plots two sets of quantiles; the theoretical quantiles on the x-axis and the quantiles of the errors (sample quantiles) on the y-axis. Both quantiles follow the same distribution if the observations fall on a straight line. As can be seen in Figure 5.2, the majority of the errors seem normally distributed. The outer values of the forecast errors show departure from normality. This might be an indication that the error data has more extreme values in comparison to what would be assumed from a normal distribution. Based on the finding that the majority of the forecast errors follow a normal distribution, a probability can be assigned to various demand levels of the corresponding forecast distribution. These probabilities can be used to assign a probability to each scenario which is explained in Section 5.3.1.

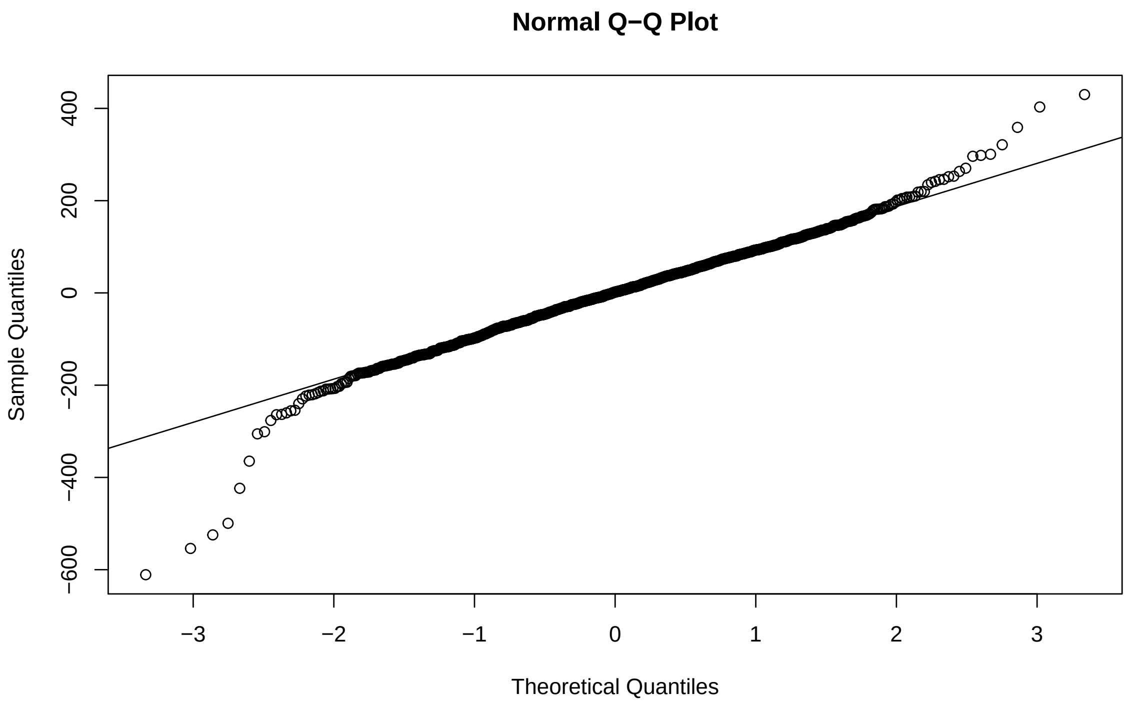

Figure 5.2: Q-Q plot of TBATS forecast errors

Table 5.1 depicts the forecast demand values for one week that have been used to generate the scenarios. The demand forecast is based on a rolling forecast with a forecast horizon of two days. The training data for the TBATS forecasting method ranges from April 2015 up to two days before the weekday that requires a demand forecast. For each weekday, the generated demand levels in loading meters at the point forecast and at one and two standard deviation from the point forecast are presented in Table 5.1. The disaggregation rules that have been used to allocate the total demand forecast to the customer level are presented in Section 5.2.2.

Table 5.1: Demand levels (point forecast and standard deviations) in loading meters

Weekday	Point forecast	$-\sigma$	$+\sigma$	-2σ	$+2\sigma$
Monday	306.38	204.10	408.66	104.80	507.96
Tuesday	550.17	446.63	653.70	346.11	754.22
Wednesday	521.78	418.91	624.64	319.04	724.51
Thursday	515.26	411.04	619.49	309.84	720.68
Friday	485.59	380.69	590.48	278.85	692.33

5.2.2 Disaggregation Rules

The disaggregation rules determine the customer locations and the disaggregation of the total demand forecast to the customer level. The purpose of the disaggregation rules is to create scenarios so that the fleet size decision is more robust against variations in the distribution of demand. Because of practical limitations, only three disaggregation rules have been applied since the preparation of data is time consuming.

Customer locations are derived from historical customer orders. It may occur that multiple customer orders are related to a single customer location. Several high-volume customers order frequently. Therefore, customer locations derived from order data of the previous week are expected to be representative future customer locations. Below, the three disaggregation rules to generate unique scenarios are explained:

I The customer orders of a specific weekday are based on the customer orders of the same weekday of the previous week. The demand forecast is disaggregated to the customer level by applying the demand ratios among customer locations of the previous weekday to the demand forecast value.

II The customer orders of a specific weekday are based on the customer orders of the same weekday of the previous week. The demand forecast is disaggregated to the customer level by allocating the demand randomly, where each customer order is allowed to have a minimum demand of 0.3 loading meters and a maximum demand of 12 loading meters. This range in order sizes is one of the characteristics of the customer orders that are associated to the DIS TIL planning department (Appendix A). Based on historical data, each day there are more orders of up to four loading meters than orders from four to 12 loading meters (Appendix A). Therefore, the allocated orders of the disaggregated forecast of size up to four loading meters have a higher frequency in the data set compared to orders above four loading meters in size.

III The customer orders of a specific weekday are randomly chosen from all customer orders of the same weekday of the last four weeks. The total number of randomly chosen customer orders is the average number of customer orders per weekday of the four last weeks. The demand forecast is disaggregated to the customer level by allocating the demand randomly, where each customer order is allowed to have a minimum demand of 0.3 loading meters and a maximum demand of 12 loading meters. Similar to disaggregation rule II, the allocated orders of the disaggregated forecast of size up to four loading meters have a higher frequency in the data set compared to orders above four loading meters in size.

5.3 Computational Results

This section discusses the computational results generated by the routing tool. The results based on scenario generation are discussed in Section 5.3.1, whereas the results based on real historical data are discussed in Section 5.3.2. Finally, a review on all computational results generated by the routing tool is provided in Section 5.3.3.

5.3.1 Scenario Generation Results

Three disaggregation rules (Section 5.2.2) and five different demand levels (Table 5.1) have been applied for each weekday. This makes a total number of 15 generated scenarios for each weekday. In total, the routing tool has solved 75 scenarios for the five weekdays under consideration. The vehicle routing tool results for Monday, Tuesday, Wednesday, Thursday, and Friday are presented in Table 5.2, Table 5.3, Table 5.4, Table 5.5, Table 5.6, respectively. A scenario is characterized by a unique demand value and disaggregation rule combination and corresponds to the rows of these tables. These tables depict for each scenario the demand level in loading meters, the applied disaggregation rule, the number of customer locations (number of vehicle stops), the number of deployed vehicles, the total travelled distance in kilometers, the average used vehicle capacity in loading meters, and the average vehicle utilization rate. The demand levels as presented in Table 5.2, Table 5.3, Table 5.4, Table 5.5, Table 5.6 correspond to the demand levels as depicted in Table 5.1.

For most of the scenarios, the routing tool is able to construct an efficient routing policy as indicated by an average vehicle capacity utilization rate of at least 90%. Several scenarios consider relatively low customer demand, such as the two lowest demand levels of the Monday scenarios (Table 5.2). For these scenarios, the total customer demand level is insufficient to achieve a high vehicle capacity utilization rate while complying to the routing tool's constraints. Moreover, the standard deviation of the vehicle's used capacity is high for the scenarios with a very low utilization rate. The opposite holds for scenarios having a utilization rate of at least 90%, where the standard deviation of the vehicle's used capacity is low.

The number of required vehicles is relatively similar when considering the three disaggregation rules that have identical demand levels. Disaggregation rules I and II differ in the way the total demand is disaggregated among customers. For rule I, the demand is disaggregated based on historical demand ratios, whereas the demand is disaggregated randomly for rule II. For these two disaggregation rules, the way the demand is allocated among customers does not result in a significant difference in the number of required vehicles. In addition, the number of required vehicles for rule III deviates from rules I and II by one or two vehicles for identical demand levels. For a multitude of scenarios, the total distance travelled differs significantly between rule I and II. For these cases, the placement of demand within the network does not necessarily impact the number of required vehicles.

Table 5.2: Monday scenarios

Demand	Disagg. rule	Stops	Vehicles	Distance (km)	Capacity	Utilization rate
104.80	I	200	19	6,654.96	5.50	40.4%
104.80	II	200	19	6,621,52	5.50	40.4%
104.80	III	186	17	6,641,32	6.16	45.3%
204.10	I	200	19	6,661.23	10.72	78.8%
204.10	II	200	20	7,055.86	10.18	74.9%
204.10	III	186	22	8,102.94	9.28	68.2%
306.38	I	200	25	7,259.50	12.25	90.1%
306.38	II	200	26	8,059.63	11.78	86.6%
306.38	III	186	25	7,818.21	12.25	90.1%
408.66	I	200	32	8,053.18	12.78	94.0%
408.66	II	200	32	9,000.62	12.78	94.0%
408.66	III	186	32	9,909.41	12.71	93.5%
507.96	I	200	39	9,078.23	12.98	95.4%
507.96	II	200	39	10,149.75	13.06	96.0%
507.96	III	186	39	10,589.99	13.03	95.8%

Table 5.3: Tuesday scenarios

Demand	Disagg. rule	Stops	Vehicles	Distance (km)	Capacity	Utilization rate
346.11	I	291	31	9,858.88	11.15	82.0%
346.11	II	291	31	10,009.39	11.15	82.0%
346.11	III	276	29	9,256.60	11.95	87.9%
446.63	I	291	37	10,479.40	12.07	88.8%
446.63	II	291	35	10,449.16	12.76	93.8%
446.63	III	276	36	10,612.70	12.41	91.3%
550.17	I	291	44	11,493.17	12.51	92.0%
550.17	II	291	42	11,676.45	13.11	96.4%
550.17	III	276	43	11,921.25	12.80	94.1%
653.70	I	291	51	12,435.78	12.82	94.3%
653.70	II	291	51	13,108.04	12.82	94.3%
653.70	III	276	50	12,819.91	13.07	96.1%
754.22	I	291	59	13,576.63	12.93	95.1%
754.22	II	291	58	15,245.30	13.01	95.7%
754.22	III	276	59	15,984.32	12.78	94.0%

Table 5.4: Wednesday scenarios

Demand	Disagg. rule	Stops	Vehicles	Distance (km)	Capacity	Utilization rate
319.04	I	253	27	9,103.91	11.81	86.8%
319.04	II	253	27	9,019.87	11.81	86.8%
319.04	III	250	26	9,676.09	12.28	90.3%
418.91	I	253	34	10,130.31	12.31	90.5%
418.91	II	253	34	10,341.48	12.31	90.5%
418.91	III	250	33	11,359.26	12.70	93.4%
521.78	I	253	40	11,489.44	13.04	95.9%
521.78	II	253	41	11,877.14	12.72	93.5%
521.78	III	250	41	10,966.29	12.72	93.5%
624.64	I	253	48	13,088.58	13.01	95.7%
624.64	II	253	48	12,639.09	13.01	95.7%
624.64	III	250	48	12,635.12	13.02	95.7%
724.51	I	253	56	14,836.36	12.94	95.1%
724.51	II	253	56	14,356.95	12.94	95.1%
724.51	III	250	57	14,588.56	12.72	93.5%

Table 5.5: Thursday scenarios

Demand	Disagg. rule	Stops	Vehicles	Distance (km)	Capacity	Utilization rate
309.84	I	314	27	9,802.95	11.47	84.3%
309.84	II	314	29	9,428.72	10.67	78.5%
309.84	III	293	29	10,263.27	10.69	78.6%
411.04	I	314	34	10,252.83	12.09	88.9%
411.04	II	314	34	10,219.91	12.09	88.9%
411.04	III	293	34	10,899.92	12.08	88.8%
515.26	I	314	40	11,402.86	12.87	94.6%
515.26	II	314	39	11,640.42	13.20	97.1%
515.26	III	293	40	12,103.11	12.88	94.7%
619.49	I	314	48	12,619.60	12.91	94.9%
619.49	II	314	49	12,374.30	12.65	93.0%
619.49	III	293	48	13,610.73	12.90	94.9%
720.68	I	314	57	13,734.55	12.63	92.9%
720.68	II	314	55	13,954.35	13.11	96.4%
720.68	III	293	55	15,914.53	13.08	96.2%

Table 5.6: Friday scenarios

Demand	Disagg. rule	Stops	Vehicles	Distance (km)	Capacity	Utilization rate
278.85	I	236	25	8,498.87	11.16	82.1%
278.85	II	236	25	8,331.07	11.16	82.1%
278.85	III	245	24	8,997.70	11.61	85.4%
380.69	I	236	31	9,387.82	12.29	90.4%
380.69	II	236	31	9,706.46	12.29	90.4%
380.69	III	245	31	10,009.56	12.27	90.2%
485.59	I	236	38	10,090.18	12.78	94.0%
485.59	II	236	37	11,120.40	13.12	96.5%
485.59	III	245	38	11,447.99	12.78	94.0%
590.48	I	236	45	11,771.84	13.13	96.5%
590.48	II	236	45	13,518.13	13.13	96.5%
590.48	III	245	45	12,882.42	13.12	96.5%
692.33	I	236	54	12,756.17	12.84	94.4%
692.33	II	236	53	13,378.32	13.07	96.1%
692.33	III	245	53	14,569.61	13.06	96.0%

The number of required vehicles k is a function of customer demand and customer locations (Equation 5.1). Since the majority of the forecast errors follow a normal distribution, a probability can be assigned to each of the fifteen generated scenarios for each specific weekday. The probabilities are based on the normal probability density function of mean zero and standard deviation one. The point forecast corresponds to a probability of 0.3989, the 68% prediction interval (one standard deviation) corresponds to a probability of 0.2419, and the 95% prediction interval (two standard deviation) corresponds to a probability of 0.0539. An empirical Cumulative Distribution Function (CDF) for the number of required vehicles has been generated based on the scenarios and their probabilities. The empirical CDF provides the fraction of sample observations less than or equal to a particular value of k. The empirical CDFs for Monday and Tuesday are plotted in Figure 5.3. Reading this figure, starting from this figure's x-axis, if the company would deploy 26 vehicles for Monday, the company would cover 70% of the scenarios for this day. Similarly, if the company would deploy 37 vehicles for Tuesday, the company would cover 30% of the scenarios for this day. The company can establish the number of required vehicles to fulfill all customer demand based on the empirical CDFs and the amount of risk they want to incorporate. The more scenarios the company wants to cover, the more vehicles are required. The empirical CDFs for Wednesday, Thursday, and Friday are presented in Figure D.1, Figure D.2, and Figure D.3 in Appendix D, respectively. The number of required vehicles for Wednesday, Thursday, and Friday can be established in a similar manner as has been shown for Figure 5.3.

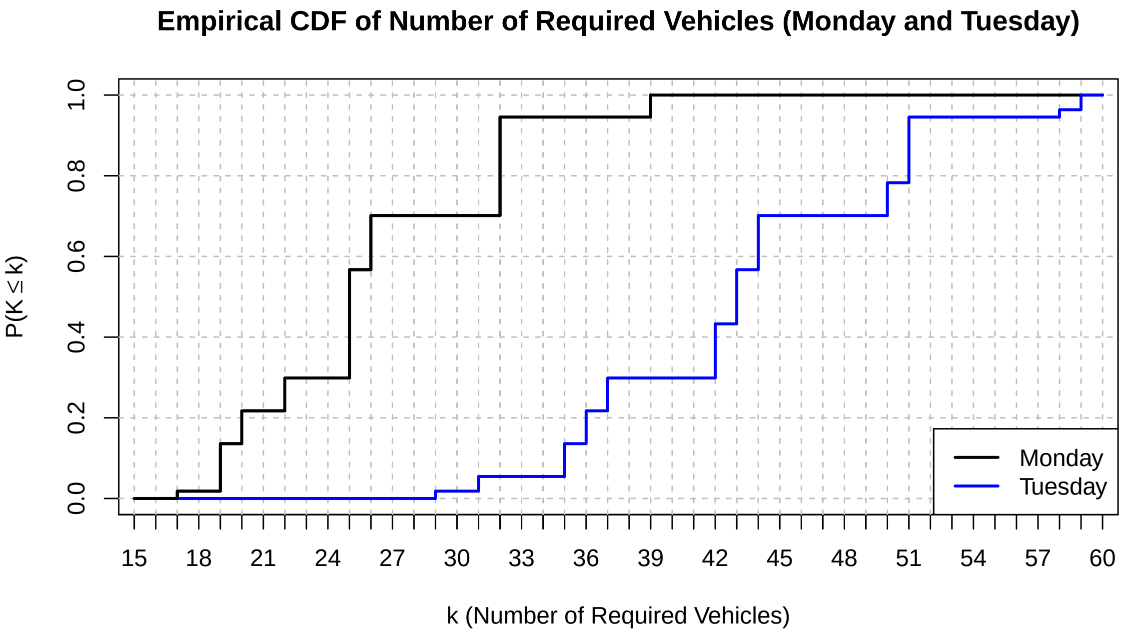

Figure 5.3: Number of required vehicles to cover Monday and Tuesday scenarios

The number of required vehicles ranges from 17 to 39 and 29 to 59 for Monday and Tuesday, respectively (Figure 5.3). Moreover, the step graphs as depicted in this figure show significant jumps. For example, with a probability of 70%, the company does not need more than 26 vehicles on Monday (Figure 5.3). A probability of 72% corresponds to 32 vehicles. Thus, a small increase in the empirical CDF may result in a relatively large increase in the number of required vehicles. This is caused by the number of required vehicles for each disaggregation rule; there are no major differences in the number of required vehicles for each disaggregation rule for identical demand levels (Table 5.2, Table 5.3, Table 5.4, Table 5.5, and Table 5.6). Moreover, the number of required vehicles for each demand level shows significant jumps because of the magnitude of deviation between each demand level. Jumps of smaller size in the number of required vehicles might have been realized by including demand levels at 0.5 and 1.5 standard deviation from the point forecast.

The number of required vehicles based on scenario generation can be compared with the real number of required vehicles. The real number of required vehicles is based on historical data comprising real customer demand and customer locations of the same period as the generated forecast. The routing tool has been used to reveal the number of required vehicles in a similar manner as for scenario generation. The number of vehicles the company would actually need in an optimal situation to serve its customers for each weekday under consideration is presented in Table 5.7. This table shows the real demand level, real number of customer locations (stops), number of required vehicles to serve all customers, total travelled distance in kilometers, average used vehicle capacity, and average utilization rate for each weekday under consideration. The routing tool is able to achieve a high average vehicle utilization rate for both the generated scenarios and the real demand and customer locations.

Table 5.7: Number of required vehicles based on historical data

Day	Demand	Stops	Vehicles	Distance (km)	Capacity	Utilization rate
Mon	259.75	199	24	7,019.57	10.82	80.0%
Tue	650.33	284	50	12,433.89	13.00	95.6%
Wed	547.18	285	43	11,650.78	12.73	93.6%
Thu	504.87	310	40	12,620.76	12.60	92.6%
Fri	393.63	232	33	9,984.06	11.93	87.7%

Uncertainties are inherent to the predictions by forecasting methods. Comparing the point forecast with the actual demand, the point forecast values for each weekday (Table 5.1) are off by 46.63, -100.16, -25.40, 10.39, and 91.96 loading meters for Monday, Tuesday, Wednesday, Thursday, and Friday, respectively (Table 5.7). By considering different demand levels, as derived from the forecasting distribution, and considering different customer locations, several scenarios and their corresponding probabilities have been generated. To conclude, the number of required vehicles can be found by establishing a coverage level that covers a desired number of generated scenarios. The company has to define this coverage level which reflects the amount of risk the company wants to protect against.

5.3.2 Vehicle Routing Results based on Historical Data

In Section 5.3.1, the company's routing tool has been used to establish the number of required vehicles based on generated scenarios. This section is dedicated to finding the minimum number of required vehicles to serve customers based on real historical customer locations and customer demand. More specifically, the routings per weekday for two months have been solved by the routing tool. By analyzing these cases, the interaction between various variables such as the number of stops, customer demand levels, total travelled distance, and the vehicle's utilization rate can be investigated.

For October and November 2019, the routing tool established the minimal number of required vehicles for each weekday. The results for these months are presented in Table 5.8. The period comprises two months, eight weeks, and two holidays in Belgium. For each weekday, these tables depict the date, the weekday, the customer demand level in loading meters, the number of customer locations (stops), the number of required vehicles, the total travelled distance in kilometers, the average used vehicle capacity in loading meters, and the average vehicle utilization rate. Both months cover four weeks each. Besides, no customer orders were placed in Belgium due to Belgian holidays on November 1 and November 11, which is reflected by a decrease in the number of stops. For October, the average vehicle capacity utilization rate is 92.28%, whereas the average vehicle capacity utilization rate for November is 93.01%. Although the demand level and the number of stops on November 1 and November 11 decreased significantly, the routing tool was still able to achieve a high vehicle capacity utilization rate. This is likely caused by the centralization of customer locations in the network.

Table 5.8: Vehicle routing results October and November per weekday

Date	Day	Demand	Stops	Vehicles	Distance	Cap.	Utilization rate
07-10-2019	Mon	280.01	203	22	7,549.01	12.73	93.6%
08-10-2019	Tue	551.31	244	43	11,434.78	12.82	94.3%
09-10-2019	Wed	541.03	290	42	11,298.26	12.88	94.7%
10-10-2019	Thu	523.11	319	41	11,582.39	12.76	93.8%
11-10-2019	Fri	493.22	235	38	10,486.78	12.98	95.4%
14-10-2019	Mon	259.75	199	24	7,019.57	10.82	80.0%
15-10-2019	Tue	650.33	284	50	12,433.89	13.00	95.6%
16-10-2019	Wed	547.18	285	43	11,650.78	12.73	93.6%
17-10-2019	Thu	504.87	310	40	12,620.76	12.60	92.6%
18-10-2019	Fri	393.63	232	33	9,984.06	11.93	87.7%
21-10-2019	Mon	227.17	180	19	5,550.65	11.96	87.9%
22-10-2019	Tue	536.92	219	42	11,041.11	12.78	94.0%
23-10-2019	Wed	481.51	249	38	10,821.57	12.67	93.2%
24-10-2019	Thu	627.42	305	48	12,116.59	13.07	96.1%
25-10-2019	Fri	379.55	208	30	7,777.72	12.65	93.0%
28-10-2019	Mon	357.88	194	29	9,119.95	12.34	90.7%
29-10-2019	Tue	429.61	262	34	10,041.02	12.64	92.9%
30-10-2019	Wed	509.47	291	41	10,711.22	12.43	91.4%
31-10-2019	Thu	520.32	380	41	12,406.47	12.69	93.3%
01-11-2019	Fri	212.10	137	17	4,049.65	12.48	91.8%
04-11-2019	Mon	264.87	202	22	6,853.09	12.04	88.5%
05-11-2019	Tue	441.82	237	34	9,874.81	12.99	95.5%
06-11-2019	Wed	510.69	279	40	10,893.77	12.77	93.9%
07-11-2019	Thu	522.63	277	40	11,599.43	13.07	96.1%
08-11-2019	Fri	538.60	267	42	11,475.53	12.82	94.3%
11-11-2019	Mon	179.99	104	14	3,415.85	12.86	94.6%
12-11-2019	Tue	503.89	340	40	13,690.37	12.60	92.6%
13-11-2019	Wed	543.19	322	43	13,119.62	12.63	92.9%
14-11-2019	Thu	615.02	283	47	12,467.27	13.09	96.3%
15-11-2019	Fri	503.85	290	40	12,132.19	12.59	92.6%
18-11-2019	Mon	439.87	215	34	8,876.86	12.94	95.1%
19-11-2019	Tue	484.75	247	39	11,199.06	12.43	91.4%
20-11-2019	Wed	407.84	237	32	10,906.52	12.74	93.7%
21-11-2019	Thu	363.76	245	30	9,227.35	12.13	89.2%
22-11-2019	Fri	260.22	215	23	8,436.82	11.31	83.2%
25-11-2019	Mon	397.83	234	32	8,335.65	12.43	91.4%
26-11-2019	Tue	538.82	254	42	10,896.04	12.83	94.3%
27-11-2019	Wed	547.05	250	42	12,291.21	13.02	95.7%
28-11-2019	Thu	602.02	280	47	13,567.63	12.81	94.2%
29-11-2019	Fri	522.56	242	40	11,376.40	13.06	96.0%

5.3.3 Review of Computational Results

Scenario generation revealed that the number of required vehicles can be established by the amount of risk the company wants to protect against. For both the generated scenarios and the routings based on real historical demand, the routing tool is generally able to achieve an average and stable vehicle utilization rate of at least 90%. Since each vehicle drives one route a day, the vehicle capacity utilization rate can be considered as the link between customer demand and the number of required vehicles. Therefore, as a rule of thumb, the planning department can transform the total demand forecast for a specific day into the number of required vehicles by dividing the demand forecast by the capacity of one vehicle (corrected by the average vehicle utilization rate). The main difference in the number of required vehicles revealed by the scenarios and the required number of vehicles based on real historical data is the error in the point forecast value. If the total demand forecast per day would have been more accurate, the number of required vehicles based on scenario generation at the point forecast would be closer to the number of vehicles the company would actually need.

5.4 Forecasting Demand based on Order Arrival Process

As stated in Section 1.2, the most desirable moment to hire extra vehicles is two days before the delivery date. Vehicle hiring costs are lowest two days before the delivery date and start increasing thereafter. The majority of customer orders arrive the day before the delivery date. At the end of the day before the delivery date, a shortage or surplus of vehicles may come to light when all orders have been received. This section describes a method to forecast the total customer demand level at the end of the day based on the order arrival process by using a neural network.

The forecasting methods that have been implemented in Chapter 4 require univariate data. Neural networks are able to handle multivariate data and are therefore more suitable to create a forecast based on an order arrival process. Although it is undesirable to hire extra vehicles at the day preceding the delivery date for a high price, this section can be seen as a way to forecast the final demand level by the order arrival process during the day. By implementing a neural network, a shortage or surplus of vehicles might be signalled at an earlier stage during the day compared to the subjective approach the planning department currently applies. Other methods, such as linear regression, could also have been implemented to forecast the final demand level based on the order arrival process. Due to time constraints, such methods have not been implemented.

An example of the order arrival process of an arbitrary day in November 2019 is presented in Figure 5.4. The observations of the order level at different moments during the day have been recorded and connected by a smooth line. As can be seen in this figure, the demand up to the start of the day preceding the delivery date (at 00:00) is presented. Furthermore, the total demand to be delivered on the next day is an increasing function in time. The demand level at different moments during the day can be monitored and a prediction of the demand level at the end of the day can be made accordingly. The design choice has been made to predict the demand level at 12pm based on the demand level at 12am. At 12am, a significant amount of orders have already arrived. Besides, signalling a shortage or surplus of vehicles at 12am provides the planning department sufficient time to anticipate the fleet size decision.

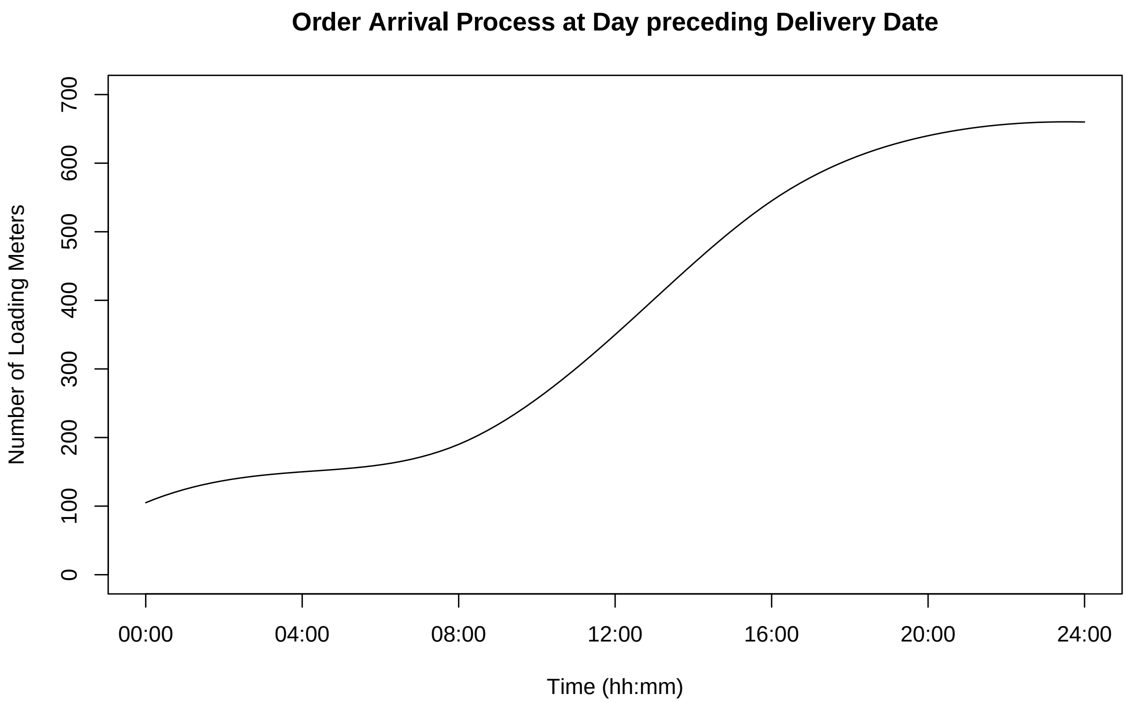

Figure 5.4: Order arrival process of an arbitrary day in November 2019

The data set that has been used consists of daily demand observations from April 2015 up to December 2019. A neural network has been trained with the demand level at 12am and the final demand level at 12pm. In addition, the day number of the month, the month number of the year, the year number, and whether the respective day is a holiday or not are factors that have been used to train the neural network. The use of calendar effects is motivated by the fact that the company's data is subject to seasonal cycles. Cross-validation has been used to test the neural network's performance on new data that has not been used for neural network training purposes. Besides, by applying cross-validation, the goal is to gain insights into how the neural network will generalize to an independent data set. More specifically, stratified cross-validation has been applied to ensure that the frequency of holidays is approximately preserved in both the training and test data sets. The company's data set comprises data of 1,240 days. The training and tests sets are 85% and 15% of the sample size, respectively. In total, six different training and test set combinations have been created to assess the neural network's prediction accuracy.

According to Stathakis (2009), it is difficult to say how many hidden layers and hidden nodes an optimal neural network requires. Therefore, several neural network configurations have been tested and the configuration that generated satisfactory results has been used. The fitted neural network consists of six nodes in the input layer, one hidden layer with ten nodes, and one node in the output layer. The neural network uses feed-forward backpropagation to update its arc weights. The error metric values of the training data sets are presented in Table 5.9. The error metrics as introduced in Section 4.3 have been used to establish the neural network's forecasting performance. As can be seen in Table 5.9, the average MAE and

MAPE of all training sets have a value of 46.71 loading meters and 9%, respectively. The error metrics of each training set show similar results.

Table 5.9: Error metrics of neural network on order arrival process (training sets)

Training set	MSE	RMSE	MAE	MAPE
I	4,084.01	63.91	46.47	0.09
II	3,879.08	62.28	46.30	0.09
III	4,152.61	64.44	47.38	0.10
IV	4,355.15	65.99	46.09	0.08
V	4,242.37	65.13	46.96	0.09
VI	4,197.07	64.78	47.05	0.10
Average:	4,151.72	64.42	46.71	0.09

The error metrics of the test data sets are presented in Table 5.10. As can be seen in Table 5.10, the average MAE is 34.59 loading meters per day. The average MAPE is 8%, which means that on average the forecast is off by 8%. The error metrics of each test set show more variation compared to the error metrics of each training set (Table 5.9 and Table 5.10). Since the average error metric values of the training sets are higher than the average error metric values of the test sets, this can be an indication of underfitting. Based on the relatively low average MAE value of the test sets, the fitted neural network is fairly accurate in predicting the demand level of a specific day at 12pm by considering the demand level at 12am and calendar effects.

Table 5.10: Error metrics of neural network on order arrival process (test sets)

Test set	MSE	RMSE	MAE	MAPE
I	2,091.26	45.73	36.35	0.08
II	1,986.40	44.57	36.05	0.10
III	1,436.21	37.90	29.53	0.07
IV	2,193.08	46.83	36.72	0.07
V	819.52	28.63	22.65	0.06
VI	2,823.08	53.13	46.22	0.10
Average:	1,891.59	42.80	34.59	0.08

The rule of thumb as mentioned in Section 5.3.3 can be used to transform the total demand forecast at 12pm into the number of required vehicles in the context of the company. To conclude, the implemented neural network can aid the planning department in accurately signalling a shortage or surplus of vehicles at an earlier stage during the day. In addition, the neural network is more systematic compared to the planning department's current subjective forecasting procedure.

Chapter 6

Conclusion, Limitations, Future Research, and Recommendations

6.1 Conclusion

This research has been conducted to determine the number of required vehicles to fulfill all customer demand per day at a transportation company. Since vehicle hiring costs increase as the delivery day approaches, an accurate required fleet size prediction is required to reduce operational costs. Besides, orders that could not have been fulfilled due to vehicle shortages also lead to increased lost sales and customer dissatisfaction. Therefore, the following main research question has been proposed in Section 1.5:

How to accurately predict the number of required vehicles, and, based on an estimate for the number of required vehicles, how many vehicles should be hired?

To provide an answer to this main research question, first a literature study has been conducted on forecasting methods and the fleet size decision. Thereafter, the effect of customer time windows and customer locations on daily routing operations has been assessed. Next, five different time series methods have been implemented that generate a demand forecast. The TBATS model showed the highest prediction accuracy with a MAE of 74.56 loading meters and a MAPE of 16.70%. Since the TBATS method is able to handle multiple seasonal cycles that may change over time, this method is robust and applicable for the long term in the context of the company. Subsequently, the design choice has been made to transform the demand forecast into the number of required vehicles. Simulation has been executed with five different demand levels and three unique sets of customer locations for each weekday under consideration to provide an answer to the SVRP. Each scenario can occur with a probability. These probabilities have been transformed into an empirical Cumulative Distribution Function (CDF). For each weekday under consideration, the number of required vehicles can be determined based on the empirical CDF functions and the amount of risk the company wants to accept.

In addition, the company's routing tool has been utilized to solve the vehicle routing for real historical order data. The goal was to gain insights on the interactions between the number of required vehicles, customer demand levels, number of customer locations, and the vehicle capacity utilization rate. Based on the results, the vehicle utilization rate is stable and robust under various demand levels and customer location sets. Since each vehicle drives one route a day, the planning department can, as a rule of thumb, transform the demand forecast into the number of required vehicles by dividing the total demand forecast for a specific day by the capacity of one vehicle. The capacity of one vehicle must be corrected for the vehicle utilization rate.

Lastly, a neural network has been created to predict the demand level at 12pm based on the order arrival process. The input data consists the demand level at 12am and calendar effects. The neural network has a MAE of 34.59 and a MAPE of 8%. The neural network can aid the planning department to signal a shortage or surplus of required vehicles at an earlier stage during the day.

The main scientific contribution of this thesis is the scenario-based method to determine the fleet size. Moreover, the generated empirical CDF functions provide insights in the distribution of the number of vehicles and the associated risk for each specific weekday. The applied quantitative framework connects the domains of demand forecasting and vehicle routing to ultimately determine the required fleet size.

Based on the executed analyses, the company is recommended to apply the implemented forecasting procedures. By implementing the TBATS time series method and the neural network, the forecasting procedure becomes more systematic. Moreover, the fleet size decision can be made several days in advance and a surplus or shortage of vehicles is signalled at an earlier stage. Since the implemented forecasting methods are systematic and always applicable, the planning department is less dependent on whether experienced planners are on duty. Unfortunately, the forecasting accuracy of the DIS TIL planning department's current vehicle prediction could not be assessed due to lack of data. In addition, the frequency and magnitude of a surplus or shortage of vehicles is poorly recorded, so are the vehicle overage costs. The company should acquire this data to reveal the financial differences between the fleet size prediction of the proposed forecasting procedure and the planning department's current process.

6.2 Limitations

The first limitation of this study concerns the quality of the data set. The performance of the forecasting methods is dependent on the underlying data set. The company's data set consists of relatively few observations that are representative for the company's current operations. Since November 2018, the average demand per day of the DIS TIL planning department decreased due to a shift in the company's planning policy. Representative training observations are of limited size and thus have an effect on forecasting performance.

The second limitation of this study concerns the absence of company data on the frequency and magnitude of a surplus or shortage of vehicles over time. Re-allocations of vehicles among different planning departments make it difficult to establish the prediction performance of the DIS TIL planning department since re-allocation data is poorly recorded. Another important factor to assess the DIS TIL planning department's forecasting performance is the timing of the fleet size decision. Data concerning the timing is necessary to know when and how many vehicles were considered to hire. This data is poorly recorded and is also necessary to make a financial comparison with the applied forecasting procedures. The later the decision to hire extra vehicles is made, the higher the vehicle hiring costs. The vehicle hiring costs (underage costs of vehicles) are known and systematically recorded. In contrast, the vehicles overage costs are poorly recorded and often not established. This is caused by the re-allocation of vehicles among planning departments, where vehicle costs and re-allocations are not systematically recorded. The limited data availability on the frequency, magnitude, and overage costs make it difficult to compare the applied forecasting procedures with the company's forecasting procedure financially and in terms of required vehicles.

6.3 Future Research

Since data preparation is time consuming, only three unique disaggregation rules have been used to create a set of customer locations and customer demand per location. Future research could be conducted on assessing more distinctive disaggregation rules that may be more representative for the company's future operations. By including more disaggregation rules, it is likely that more knowledge will be acquired on the relationship between these rules and the number of required vehicles.

Another direction for future research is to approach the fleet size decision in a dynamic manner. Suppose the forecasting method reveals a shortage of ten vehicles five days from now. 80% of the required vehicles could be hired at a low rate now, and the fleet size decision should be evaluated as several days elapsed and demand information is more accurate. In practice, this staged approach is undesirable when initially hired vehicles have to be rejected as more advanced demand information becomes available. Rejecting vehicles that were initially hired is harmful for the company's relationship with other transportation companies that rent out these vehicles. This staged approach might reduce vehicle hiring costs and is therefore a direction for future research.

6.4 Recommendations

The data from the end of 2018 up to and including the end of 2019 is the most representative for the current DIS TIL activities. The company is recommended to continue acquiring high quality order data, so the forecasting performance could be improved. Besides, it is difficult to investigate how accurate the company's vehicle predictions are; how often they need to rent extra vehicles last minute, or how often do they have too many vehicles. The proposition is made to systematically record vehicle surpluses and shortages. This information can be used to gain better insights in the company's vehicle prediction process.

The forecasting methods that have been used for this master's thesis are based on local information. By considering forecasts generated by partners in the supply chain, the applied forecasting methods are expected to improve in forecasting performance. Currently, the DIS TIL planning department receives a demand forecast of a few large customers. These customers send their forecast a few days in advance to the company, where the orders are finalized the day before the delivery date. The validity of these customer forecasts has never been investigated. In addition, customer forecasts have never been compared to the final customer demand levels. The company should use forecast information of parties in the supply chain to improve the company's forecasting accuracy.

Whereas the vehicle underage costs are known, the last recommendation concerns establishing the vehicle's overage costs. These numbers are expected to reveal the financial impact of having a surplus or shortage of vehicles.

References

Axsäter, S. (2015). *Inventory Control*. Springer.

Bergmeir, C. & Benítez, J. M. (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences*, *191*, 192–213.

Berhan, E., Beshah, B., Kitaw, D. & Abraham, A. (2014). Stochastic Vehicle Routing Problem: A Literature Survey. *Journal of Information & Knowledge Management*, *13*(3).

Billings, R. B. & Agthe, D. E. (1998). State-Space versus Multiple Regression for Forecasting Urban Water Demand. *Journal of Water Resources Planning and Management*, *124*(2), 113–117.

Brockwell, P. J. & Davis, R. A. (2016). *Introduction to Time Series and Forecasting*. Springer.

Brożyna, J., Mentel, G., Szetela, B. & Strielkowski, W. (2018). Multi-seasonality in the tbats model using demand for electric energy as a case study. *Economic Computation and Economic Cybernetics Studies and Research*, *52*(1), 229–246.

Chen, C. & Liu, L. M. (1993). Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association*, *88*(421), 284–297.

Chen, K. Y. & Wang, C. H. (2007). A hybrid SARIMA and support vector machines in forecasting the production values of the machinery industry in Taiwan. *Expert Systems with Applications*, *32*(1), 254–264.

Cleveland, R., Cleveland, W., McRae, J. & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics*, *6*(1), 3–73.

Da Veiga, C. P., Da Veiga, C. R. P., Catapan, A., Tortato, U. & Da Silva, W. V. (2014). Demand forecasting in food retail: A comparison between the Holt-Winters and ARIMA models. *WSEAS Transactions on Business and Economics*, *11*(1), 608–614.

de Livera, A. M., Hyndman, R. J. & Snyder, R. D. (2011). Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, *106*(496), 1513–1527.

Fan, W., Xu, H. & Xu, X. (2009). Simulation on vehicle routing problems in logistics distribution. *COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering*, *28*(6), 1516–1531.

Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing*, *21*(2), 137–146.

Gendreau, M., Laporte, G. & Séguin, R. (1996). Stochastic vehicle routing. *European Journal of Operational Research*, *88*(1), 3–12.

Gendreau, M., Laporte, G. & Seguin, R. (1996). A Tabu Search Heuristic for the Vehicle Routing Problem with Stochastic Demands and Customers. *Operations Research*, *44*(3), 469–477.

Ghiani, G., Laporte, G. & Musmanno, R. (2013). *Introduction to Logistics Systems Management.* John Wiley & Sons.

Goodwin, P. (2010). The Holt-Winters Approach to Exponential Smoothing: 50 Years Old and Going Strong. *Foresight*, *19*, 30–33.

Gould, P. G., Koehler, A. B., Ord, J. K., Snyder, R. D., Hyndman, R. J. & Vahid-Araghi, F. (2008). Forecasting time series with multiple seasonal patterns. *European Journal of Operational Research*, *191*(1), 207–222.

Hart, M., Kubíková, J. & Lukoszová, X. (2013). Logistics management based on demand forecasting. *Research in Logistics & Production*, *3*(1), 71–80.

Hvattum, L. M., Løkketangen, A. & Laporte, G. (2006). Solving a Dynamic and Stochastic Vehicle Routing Problem with a Sample Scenario Hedging Heuristic. *Transportation Science*, *40*(4), 421–438.

Hyndman, R. J. (2014). *Generating quantile forecasts in R.* Retrieved from `robjhyndman.com/hyndsight/quantile-forecasts-in-r`

Hyndman, R. J. & Athanasopoulos, G. (2018). *Forecasting: principles and practice, 2nd edition.* Retrieved from `OTexts.com/fpp2`

Hyndman, R. J. & Athanasopoulos, G. (2019). *Forecasting: principles and practice, 3rd edition.* Retrieved from `OTexts.com/fpp3`

Hyndman, R. J. & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, *22*(4), 679–688.

Hyndman, R. J., Koehler, A. B., Ord, J. K. & Snyder, R. D. (2008). *Forecasting with Exponential Smoothing: The State Space Approach.* Springer Science & Business Media.

Juan, A. A., Faulin, J., Pérez-Bernabeu, E. & Domínguez, O. (2013). Simulation-optimization methods in vehicle routing problems: A literature review and an example. *Lecture Notes in Business Information Processing*, *145*, 115–124.

Kall, P. & Wallace, S. W. (1994). *Stochastic Programming.* Chichester: John Wiley & Sons.

Lee, K., Kang, D. Y., Choi, H. R., Park, B. K., Cho, M. J. & Kim, D. H. (2018). Intermittent demand forecasting with a recurrent neural network model using IoT data. *International Journal of Control and Automation*, *11*(3), 153–168.

Montgomery, D. C., Jennings, C. L. & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting.* John Wiley & Sons.

Montoya-Torres, J. R., Franco, J. L., Isaza, S. N., Jiménez, H. F. & Herazo-Padilla, N. (2015). A literature review on the vehicle routing problem with multiple depots. *Computers and Industrial Engineering*, *79*, 115–129.

Mrowczynska, B., Łachacz, K., Haniszewski, T. & Sładkowski, A. (2012). A comparison of forecasting the results of road transportation needs. *Transport*, *27*(1), 73–78.

Nordholt, E. S. (1998). Imputation: Methods, Simulation Experiments and Practical Examples. *International Statistical Review*, *66*(2), 157–180.

Oyola, J., Arntzen, H. & Woodruff, D. L. (2018). The stochastic vehicle routing problem, a literature review, part I: models. *EURO Journal on Transportation and Logistics*, *7*(3), 193–221.

Petruzzi, N. C. & Dada, M. (1999). Pricing and the newsvendor problem: A review with extensions. *Operations Research*, *47*(2), 183–194.

Shumway, R. & Stoffer, D. (2017). *Time series analysis and its applications: with R examples.* Springer.

Shyshou, A., Gribkovskaia, I. & Barceló, J. (2010). A simulation study of the fleet sizing problem arising in offshore anchor handling operations. *European Journal of Operational*

Research, *203*(1), 230–240.

Stathakis, D. (2009). How many hidden layers and nodes? *International Journal of Remote Sensing*, *30*(8), 2133–2147.

Toth, P. & Vigo, D. (2002). *The vehicle routing problem.* Society for Industrial and Applied Mathematics.

Tsekeris, T. & Tsekeris, C. (2011). Demand Forecasting in Transport: Overview and Modeling Advances. *Economic Research-Ekonomska Istraživanja*, *24*(1), 82–94.

Vonolfen, S., Wagner, S., Beham, A., Kofler, M., Affenzeller, M., Lengauer, E. & Scheucher, M. (2010). A simulation-based approach to the vehicle routing problem. *22th European Modeling and Simulation Symposium, EMSS 2010*(October), 363–368.

Wang, S. & Chaovalitwongse, W. A. (2010). Evaluating and Comparing Forecasting Models. *Wiley Encyclopedia of Operations Research and Management Science*.

Weatherford, L. R., Kimes, S. E. & Scott, D. A. (2001). Forecasting for hotel revenue management: Testing aggregation against disaggregation. *Cornell Hotel and Restaurant Administration Quarterly*, *42*(4), 53–64.

Xu, S., Chan, H. K. & Zhang, T. (2019). Forecasting the demand of the aviation industry using hybrid time series SARIMA-SVR approach. *Transportation Research Part E: Logistics and Transportation Review*, *122*(October 2018), 169–180.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175.

Zhang, G. P. & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, *160*(2), 501–514.

Zhang, Y. Y. & Li, J. B. (2007). Dynamic optimal model of vehicle fleet size and exact algorithm. *System Engineering - Theory and Practice*, *27*(2), 83–91.

Zhou, L., Heimann, B. & Clausen, U. (2006). *Short Term Demand Forecasting For A Typical Logistics Service Provider.*

Zotteri, G., Kalchschmidt, M. & Caniato, F. (2005). The impact of aggregation level on forecasting performance. *International Journal of Production Economics*, *93*, 479–491.

Appendix A

Histogram Order Sizes

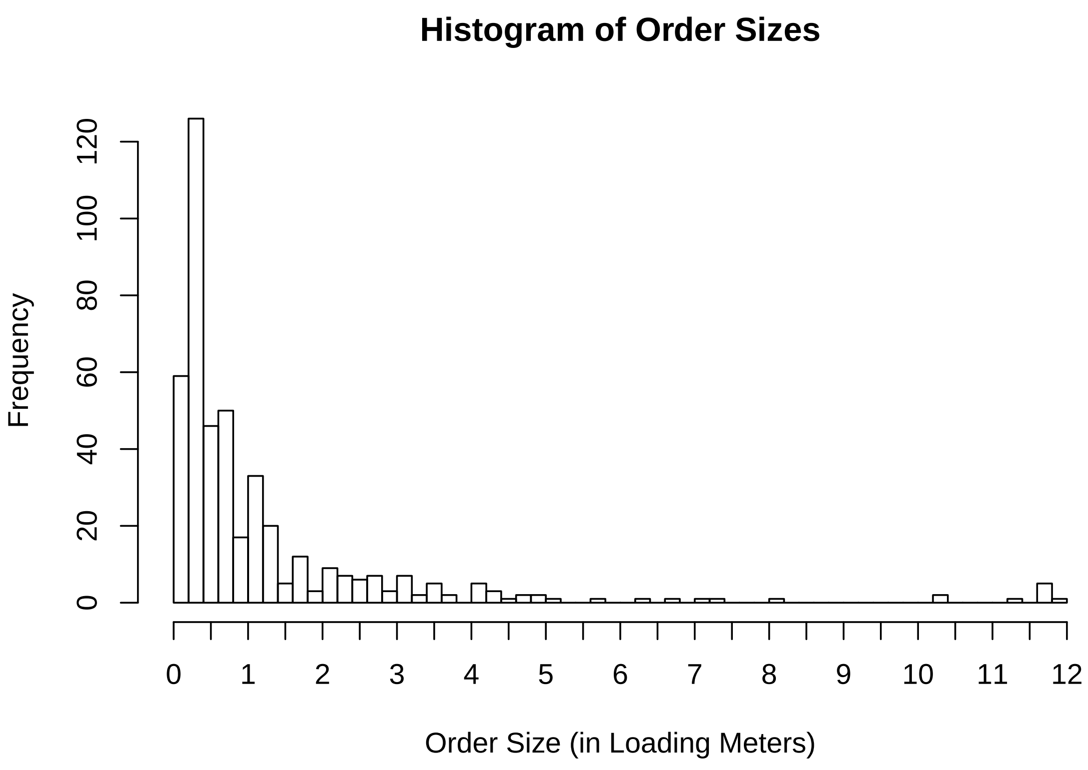

Figure A.1: Histogram of order sizes of an arbitrary day in November 2019

Note: for nearly all weekdays, the order distribution of customers of the DIS TIL planning department follows the same distribution as depicted in Figure A.1.

Appendix B

Weekday Time Series

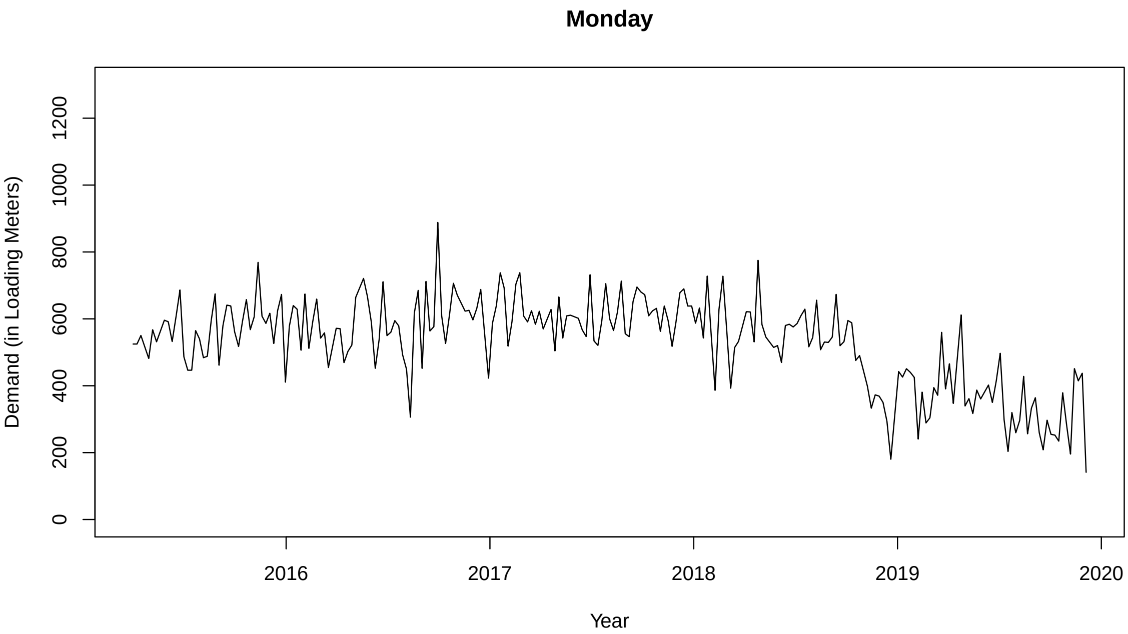

Figure B.1: DIS TIL transportation order history Monday

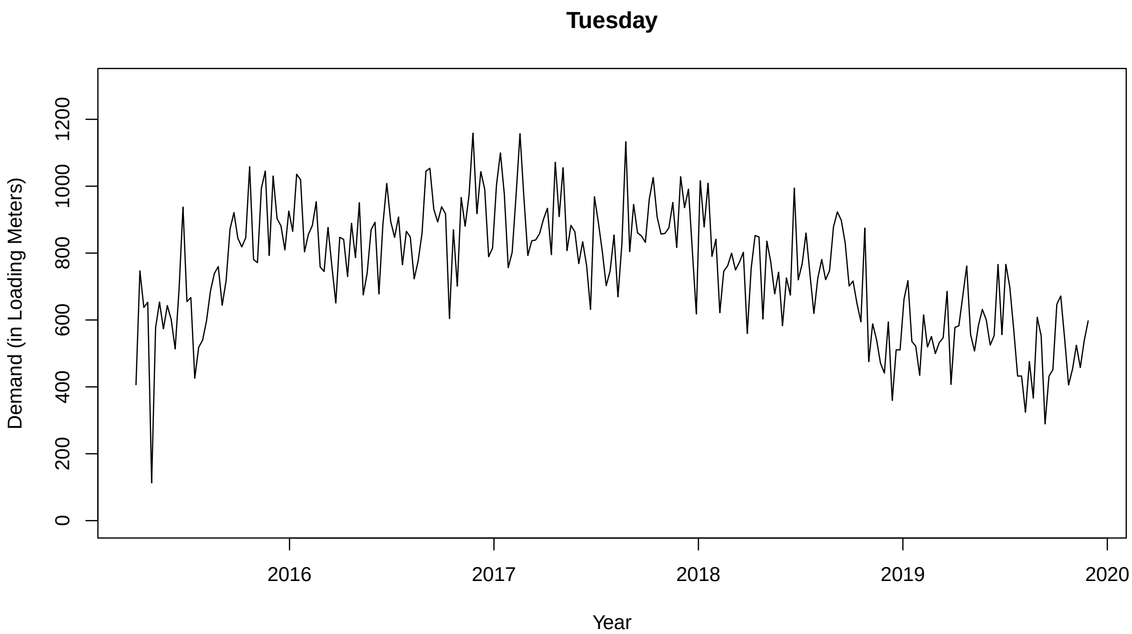

Figure B.2: DIS TIL transportation order history Tuesday

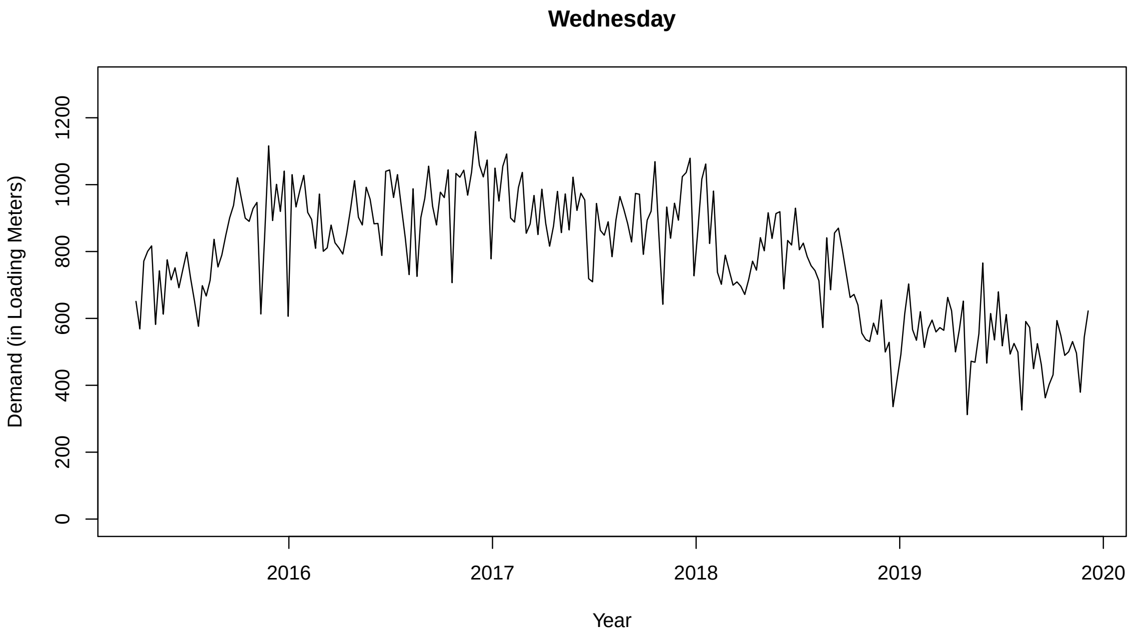

Figure B.3: DIS TIL transportation order history Wednesday

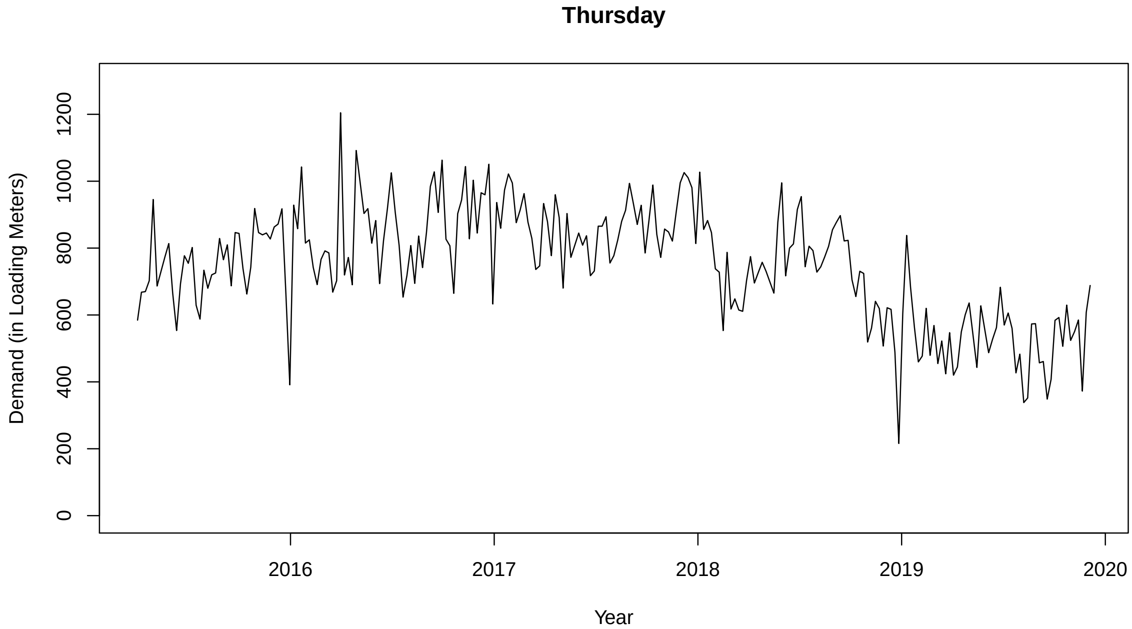

Figure B.4: DIS TIL transportation order history Thursday

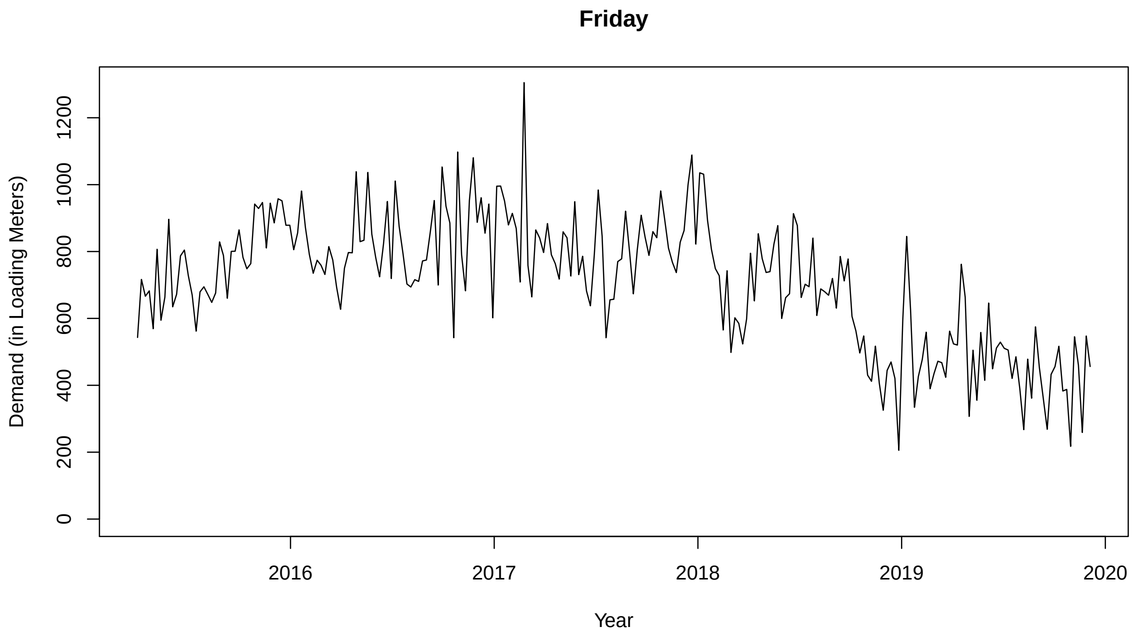

Figure B.5: DIS TIL transportation order history Friday

Appendix C

Vehicle Routing Map

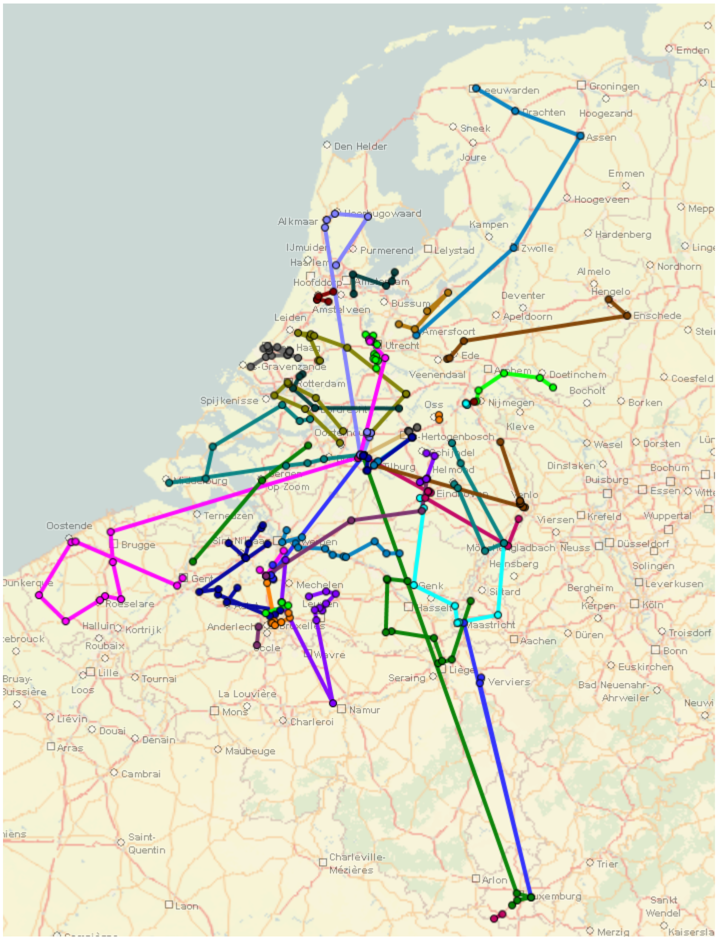

Figure C.1: Vehicle routing map

Note: the lines connecting the route to the depot have been eliminated to increase the figure's readability.

Appendix D

Empirical CDFs for Wednesday, Thursday, and Friday

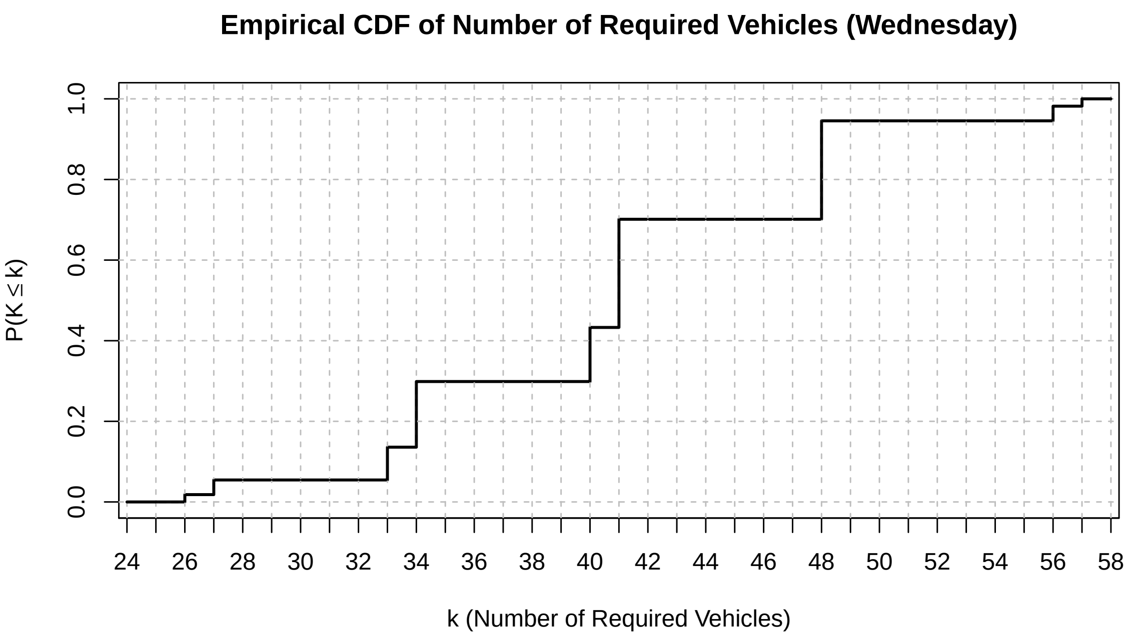

Figure D.1: Number of required vehicles to cover Wednesday scenarios

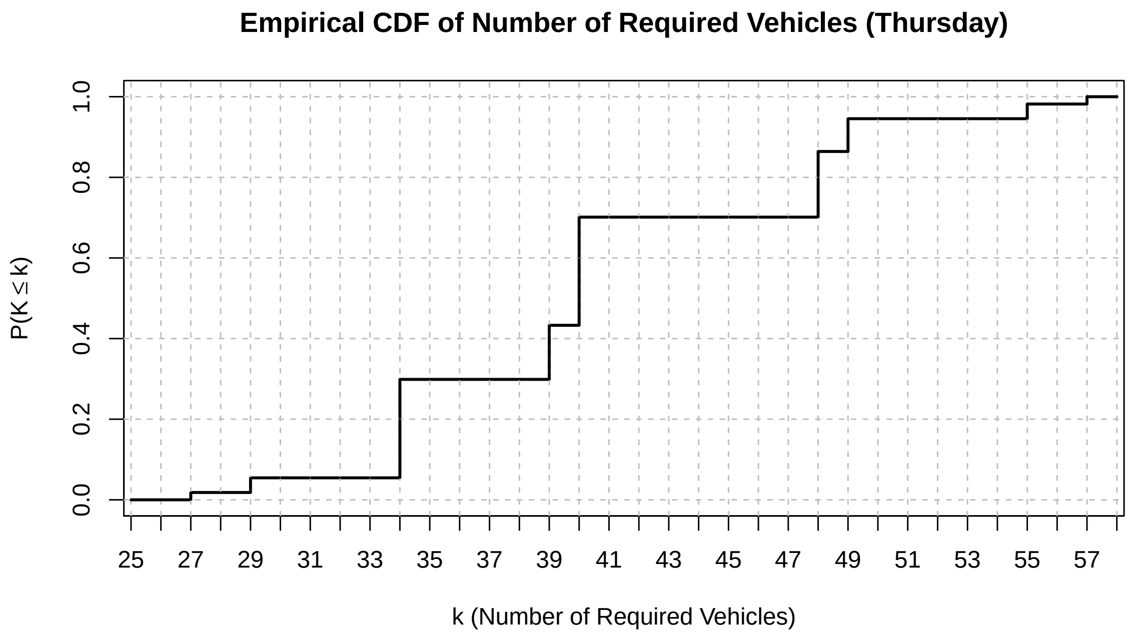

Figure D.2: Number of required vehicles to cover Thursday scenarios

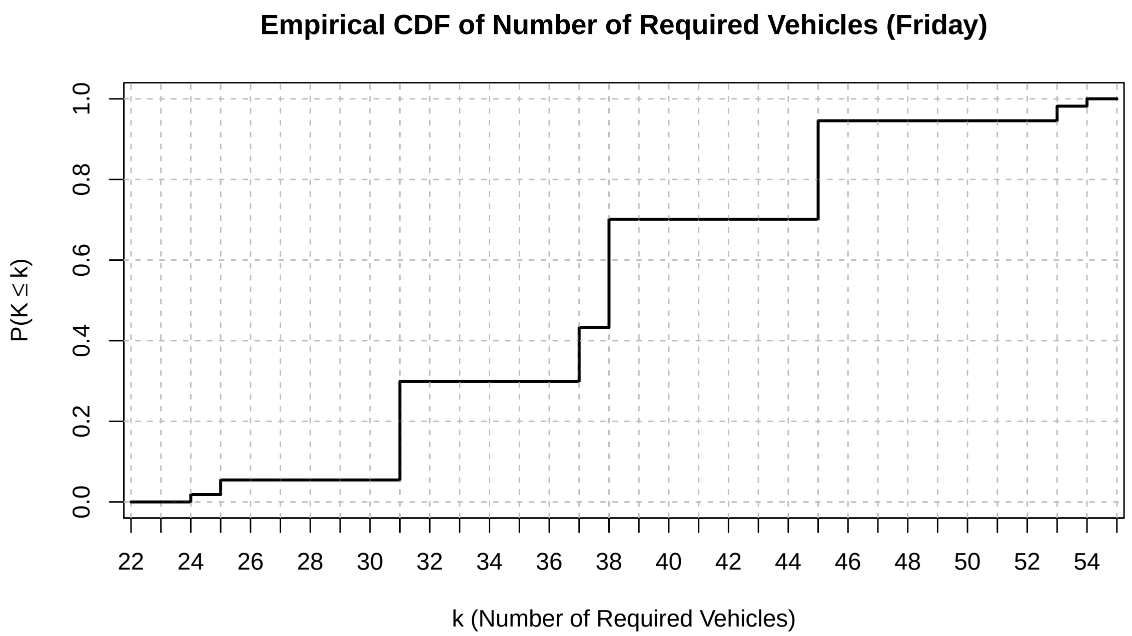

Figure D.3: Number of required vehicles to cover Friday scenarios